NEW ZEALAND

N Mortimer, Institute of Geological and Nuclear Sciences, Dunedin, New Zealand

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The south-west Pacific Ocean is a region of isolated islands and submerged plateaus and ridges (Figure 1). The three main islands of New Zealand (North, South, and Stewart) make up the largest landmass group in the region. Schists, greywackes, and granitoids are exposed in islands on the Challenger Plateau and Chatham Rise and have been sampled in dredges on the Campbell Plateau and Norfolk Ridge, thus demonstrating their continental geological character. Abyssal Pacific oceanic-crustal floor typically lies at water depths of about 5000 m, and the boundary between continental crust and oceanic crust is marked by a generally pronounced slope break at about 2500 m water depth. The wider area of continental crust in the New Zealand region (Figure 1) is about one-third the area of on-land Australia and is commonly referred to as Zealandia.

On-land New Zealand contains a wide variety of Phanerozoic rocks (Figure 2), which preserve a detailed record of the Cambrian to early Early Cretaceous convergent margin of southern Gondwana, late Early Cretaceous rifting, a Late Cretaceous–Palaeogene passive margin, and the Neogene–Holocene active convergent and strike-slip margin. So much of continental Zealandia is submerged because of the widespread Cretaceous extension and rifting. It was only with the development of the Neogene–Holocene convergent plate boundary that about 10% of Zealandia emerged above sea-level.

A distinction is generally drawn in New Zealand between pre-late Early Cretaceous (more than 105 Ma) 'basement' rocks, which are commonly metamorphosed and generally highly deformed, and 'cover' rocks, which are younger than 105 Ma, poorly indurated, well stratified, and less deformed.

Palaeogeographical Reconstructions

The shape of the continental crust of Zealandia has changed throughout the Phanerozoic. From the Cambrian to the Early Cretaceous, the New Zealand part of the Gondwanan margin grew by the magmatic and tectonic addition of batholiths and terranes. In the last 100 Ma this continental crust has been thinned, rotated, and translated in response to multiple tectonic events. Reconstructions (Figure 3) involve subtracting the 480 km Neogene dextral strike-slip movement on the Alpine Fault, the 45° Oligocene–Miocene rotation between the Pacific and Australian plates, and 4000 km of northwards drift. Small crustal blocks within 100 km of the Alpine Fault (i.e. most of on-land New Zealand) have undergone strong Cenozoic deformation. In the pre-rift (100 Ma) palaeogeography (Figure 3A) Zealandia is in a near-polar position and contiguous with Tasmania and Antarctica. By 10 Ma, some movement on the Alpine Fault had taken place and modern-day New Zealand had been isolated by seafloor spreading.

Geological Basement

At a regional scale, the volcanic, sedimentary, plutonic, and metamorphic basement rocks of New Zealand can be described in terms of a number of western and eastern tectonostratigraphic terranes, composite regional batholiths intruding these terranes, and schist, gneiss, and mélange overprints on

Figure 1 The outline of the area of continental crust in the New Zealand region (Zealandia). Land and major islands are pale brown; water less than 2500 m deep is pale blue. Deep ocean floor is dark blue and Hikurangi plateau large igneous province is intermediate blue. The present-day Pacific–Australian plate boundary is shown by the thick black line, with teeth on the overriding plate. Only about 10% of Zealandia is emergent above sea-level as the North and South Islands. NR, Norfolk Ridge.

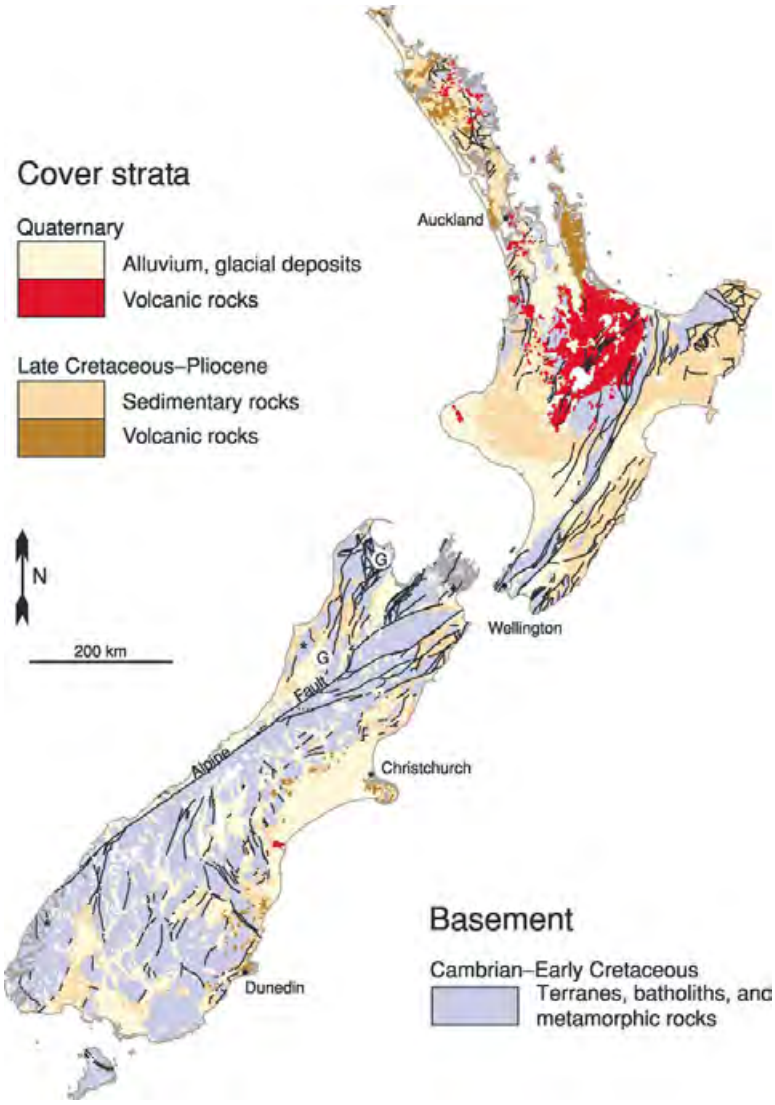


Figure 2 Simplified present-day on-land geology of New Zealand. The black lines are faults. Many of the lakes and valleys of the South Island are the result of Quaternary glacial erosion. G, locations of Devonian to Jurassic Gondwanan sequences; ∗, locations of Cretaceous metamorphic core complexes.

the terranes and batholiths (Figure 4). Figure 5 shows the names, age ranges, and mutual geometric relationships of the constituent basement units on an Early Cretaceous reconstruction (Figure 3A). No Precambrian rocks are exposed; New Zealand has been near a continent–ocean margin throughout the Phanerozoic.

Western Province Terranes

The Western Province terranes lie west of the Median Batholith and comprise the Early Palaeozoic Buller and Takaka terranes. The Buller Terrane consists of variably metamorphosed siliciclastic sandstones and mudstones, of continental Gondwanan provenance, and is the westernmost recognized terrane in New Zealand (i.e. the terrane closest to the Gondwanan cratonic core). Rare fossils are of Ordovician age, but a Buller Terrane paragneiss contains detrital zircons as old as 3400 Ma (Archean; New Zealand's oldest known geological material). Intercalated volcanics are absent. The Takaka Terrane consists of siliciclastic, carbonate, and volcanic rocks. Middle Cambrian trilobites in the Takaka Terrane are New Zealand's oldest known fossils. The Takaka Terrane is generally well stratified and lithologically diverse, and includes Cambrian ultramafics and boninites, Ordovician limestones, and Silurian orthoquartzites. The Buller and Takaka Terranes were accreted to Gondwana by the Devonian.

Eastern Province Terranes

The Eastern Province terranes lie east of the Median Batholith and comprise the Brook Street, Murihiku,

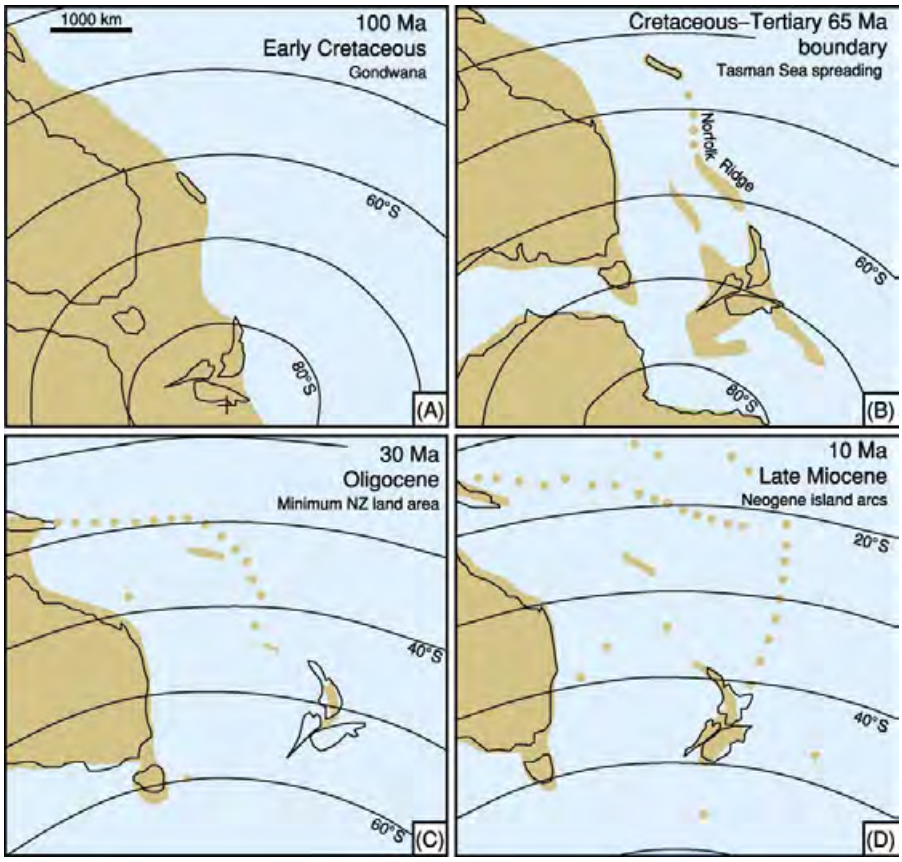


Figure 3 Palaeogeographical reconstructions of New Zealand at approximately (A) 100 Ma, (B) 65 Ma, (C) 30 Ma, and (D) 10 Ma. Brown, inferred land including schematic island-arc chains; black lines, present-day coastlines of North and South Islands and the Alpine Fault. On these reconstructions, the Alpine fault divides the South Island into eastern and western parts. (Reproduced with permission from Lee DE, Lee WG, and Mortimer N (2001) Where and why have all the flowers gone? Depletion and turnover in the New Zealand Cenozoic angiosperm flora in relation to palaeogeography and climate. *Australian Journal of Botany* 49: 341–356.)

Maitai, Caples, Rakaia, and Pahau terranes. Carboniferous conodonts are known from a limestone in the Rakaia Terrane, but the age range of clastic rocks in Eastern Province terranes is from Permian to Early Cretaceous (Figure 5). The Eastern Province terranes are thus entirely younger than the Western Province terranes and represent accretion of material to Gondwana in the Mesozoic.

The Brook Street Terrane is a Permian subduction-related isotopically primitive pyroxene-rich basalt-dominated volcanic pile and volcaniclastic apron, in places up to 14 km thick, which is intruded by Permian layered gabbros and trondhjemite plutons that are now part of the Median Batholith. New Zealand's only known *Glossopteris*, a Gondwanan leaf fossil, occurs in the Brook Street Terrane. The Murihiku Terrane comprises a 9–13 km Late Permian to Late Jurassic volcaniclastic marine succession of sandstone with lesser conglomerates, mudstones, and tuffs. It has the simplest internal structure of all the Mesozoic New Zealand terranes, a broad synclinorium that is traceable for 450 km through the North and South Islands.

The Maitai Terrane consists of the eastern Early Permian (285–275 Ma, according to uranium–lead dating of zircon) Dun Mountain Ophiolite Belt, which is unconformably overlain by 6 km of well-stratified Late Permian to Middle Triassic volcaniclastic sedimentary rocks. The ophiolite originated in a near-arc setting. The Brook Street, Murihiku and Maitai terranes are adjacent to each other as a Permian–Triassic arc, fore-arc, and exhumed near-arc ophiolite, respectively.

The Caples, Bay of Islands, and Rakaia terranes contrast with the aforementioned Eastern Province terranes in that their Permian–Jurassic clastic sequences are tectonically imbricated with ocean-floor basalt, chert, and limestone associations; all three terranes grade into the pumpellyite–actinolite to amphibolite facies Haast Schist. Deposition occurred as submarine-fan deposits in lower trench-slope basins, before juxtaposition in an accretionary

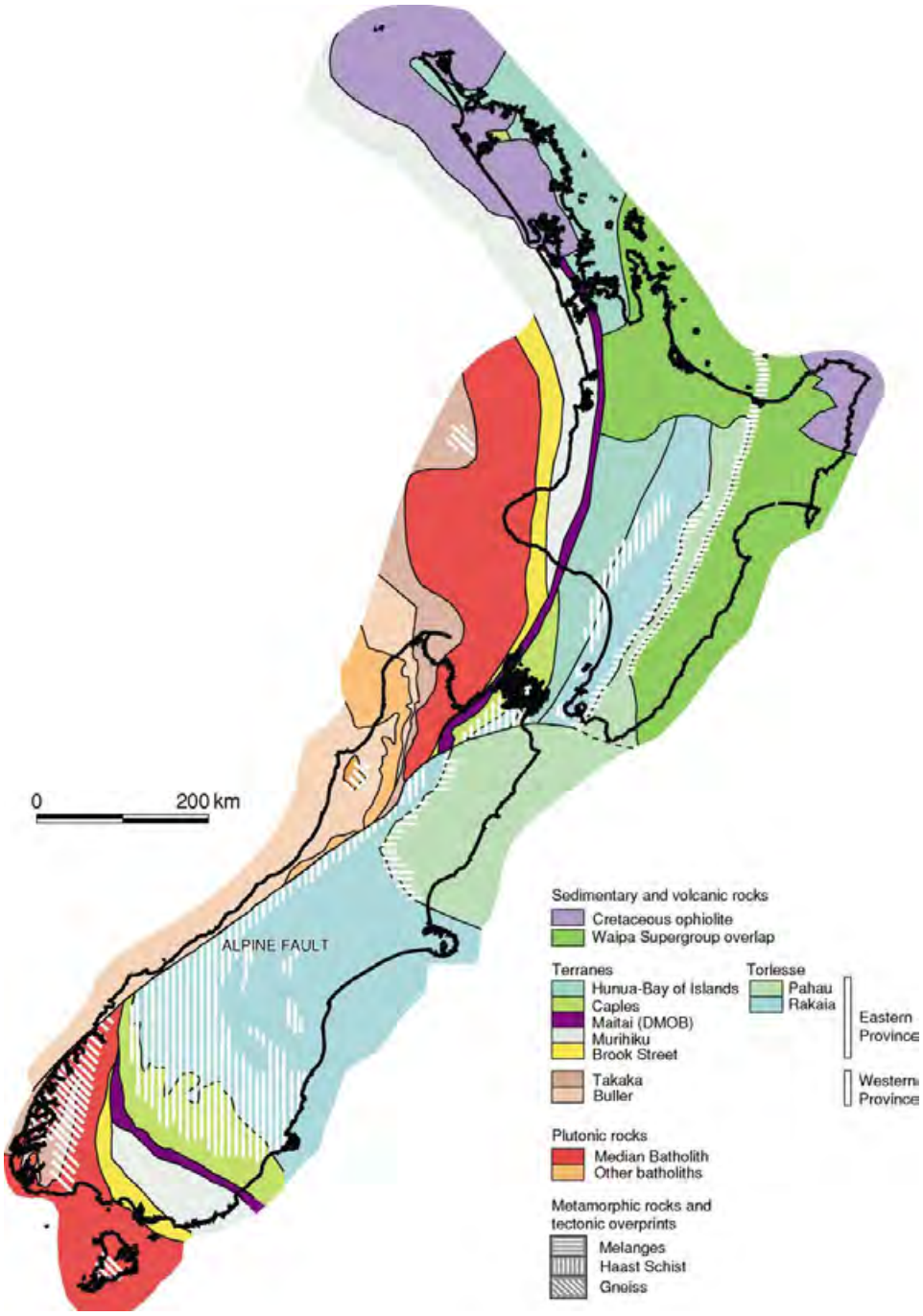


Figure 4 Basement geological subdivisions of New Zealand. Minor outliers of Permian Gondwanan sequences rest on Takaka Terrane, and Devonian, Triassic and Jurassic Gondwanan sequences on Buller Terrane. DMOB, Dun Mountain Ophiolite Belt. (Reproduced with permission from Mortimer N (2004) New Zealand's geological foundations. *Gondwana Research* 7: 261–272. © International Association for Gondwana Research.)

prism. Compositional and provenance differences are used to discriminate the three terranes: Rakaia sandstones are quartz rich, plutoniclastic, and of average rhyodacitic composition and are thus compositionally distinct from the more dacitic to andesitic volcaniclastic-dominated sandstones in Caples and the Bay of Islands. The Pahau Terrane has a similar lithology and structure to the Rakaia Terrane, but its depositional ages extend into the Late Jurassic and Early Cretaceous and it contains tuffs. Much of the Pahau clastic detritus is probably recycled from Rakaia rocks, but a volcanic input, probably from the Median Batholith, is also required. The Pahau Terrane probably represents trench deposits that

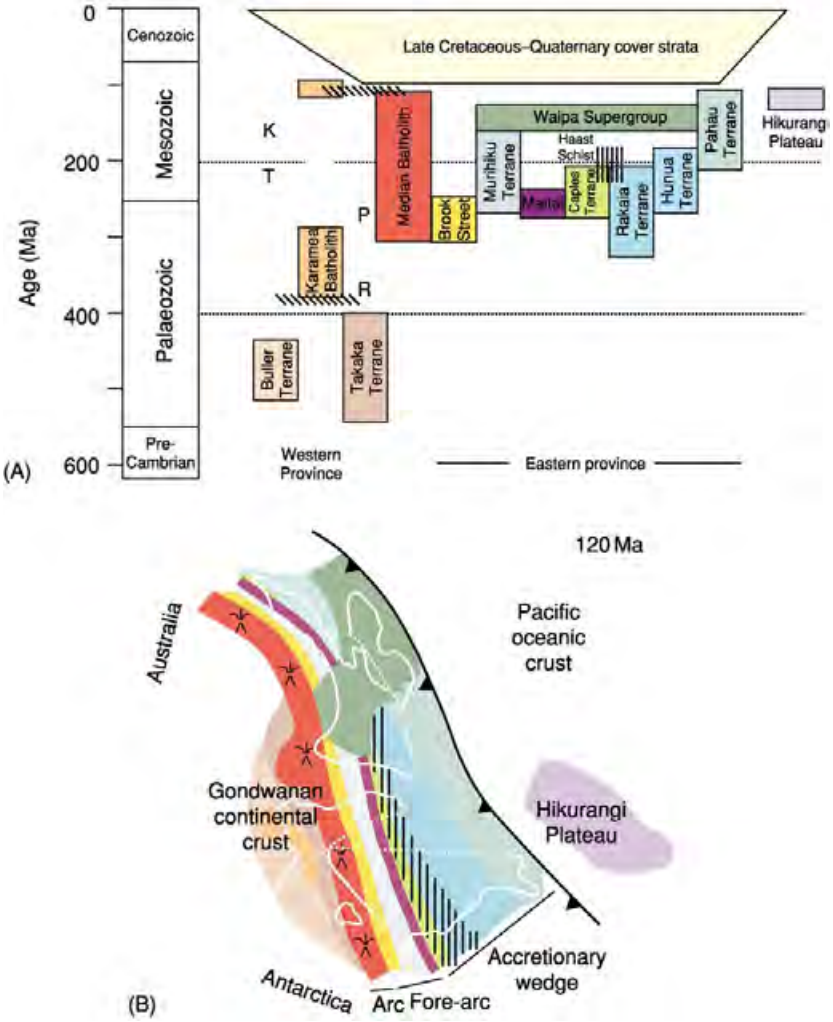


Figure 5 (A) Summary of the age ranges of New Zealand's basement terranes (green, blue, brown, yellow), batholiths (red, orange), and metamorphic rocks (overprint stripes). Gondwana Sequence rocks are shown by letters: K, Kirwans Dolerite; T, Topfer Formation; P, Parapara Group; R, Reefton Group. The terranes can be grouped into Eastern and Western provinces. (B) One possible palaeogeographical reconstruction of south-eastern Gondwana at about 120 Ma (the end of the convergent-margin phase). The present-day New Zealand coastlines are shown as white lines; the Alpine Fault and other faults are shown as white dotted lines. (Reproduced with permission from Mortimer N (2004) New Zealand's geological foundations. *Gondwana Research* 7: 261–272. © International Association for Gondwana Research.)

were laid down and deformed towards the end of Cretaceous subduction.

These nine terranes make up the bulk of the New Zealand volcanosedimentary basement. Smaller tectonostratigraphic units can be regarded as components of the larger terranes. The Median Tectonic Zone is still used by some New Zealand geologists to describe a zone of terrane shards and igneous complexes of uncertain status and correlation that lies between the Brook Street and Takaka terranes (see Plutonic Rocks, below).

Overlap Sequences

Overlap sequences of varying ages can be recognized by their lesser deformation and metamorphism and distinctive petrofacies, as compared with older immediately underlying rocks. Lateral correlatives are used to constrain models of terrane amalgamation and accretion. The only New Zealand rocks that have been correlated with autochthonous Gondwanan sequences occur in two small outliers in northern South Island (marked G in Figure 2). Four units – the Reefton Group (marine Devonian), Parapara Group (marine Permian–Triassic), Topfer Formation (nonmarine Triassic), and Kirwans Dolerite (a Middle Jurassic low-titanium tholeiite sill intrusion) – indicate that the Buller and Takaka terranes had been accreted to Gondwana by the end of the Palaeozoic. In North Island, a postulated Late Jurassic overlap sequence – the Waipa Supergroup, possibly sourced

from the Median Batholith – may indicate that most of the basement terranes had accreted by this time. The post-105 Ma cover sediments provide a firm minimum age for the mutual juxtaposition and accretion of the basement terranes.

Plutonic Rocks

There are three composite regional batholith-sized belts (more than 100 km^2) of plutons in New Zealand, and numerous smaller isolated plutons. The Median Batholith is a composite Cordilleran batholith with intrusive contacts against the Brook Street and Takaka terranes. It comprises dozens of 1–10 km Carboniferous to Early Cretaceous gabbroic to granitic subalkaline I-type plutons. The eastern half of the Median Batholith has also been called the Median Tectonic Zone. The ages and average compositions of the rocks change across the batholith axis: Permian gabbroids dominate the eastern edge, Triassic to early Early Cretaceous dioritoids dominate the central part, and late Early Cretaceous adakitic granitoids are found on the western margin. Roof pendants of petrologically related volcanosedimentary rocks occupy about 5% of the batholith area. How much of the Median Batholith is allochthonous is debatable.

The Karamea–Paparoa Batholith lies entirely within the Buller Terrane. Its constituent plutons are dominated by Devonian–Carboniferous I-type and S-type granites. The Hohonu Batholith, also within the Buller Terrane, represents 105–82 Ma plutonism associated with the change from convergence to rifting.

Metamorphic Overprints

Regionally extensive metamorphic overprints include Devonian and Cretaceous polymetamorphic amphibolite–granulite facies gneisses (formed from Buller, Takaka, Median, and Karamea–Paparoa protoliths). Some of the gneisses are confined to the lower plates of two Cretaceous metamorphic core complexes in the South Island (stars in Figure 2). The Haast Schist of Jurassic–Cretaceous pumpellyite–actinolite to amphibolite facies overprints the Caples, Bay of Islands, and Rakaia terranes. Metamorphism probably took place in the deep parts of a Jurassic–Early Cretaceous accretionary wedge. Exhumation of the Haast Schist belt was episodic, with most of the schist being at the surface by 105 Ma and deeper levels being exhumed along the Alpine Fault from 20 Ma to the present day. The third major kind of regional tectonic overprint is the Early Cretaceous mélange that is present as belts in and between the Rakaia and Pahau terranes. In part this mélange is probably coeval with the Haast Schist mineral growth, but it formed in the shallower parts of the accretionary wedge. The steep mélanges may represent zones of strike-slip deformation.

Episodes of tectonic activity were previously described as orogenies. In the early 1980s it was realized that the Mesozoic Rangitata Orogeny was probably a composite of an older subduction-related event and a younger extension-related event. With the recognition that much of the intrusion, metamorphism, and deformation in the Buller and Takaka terranes ('Tuhua Orogen') occurred during the Cretaceous (of 'Rangitata' age), the orogen terminology became obsolete.

Cover Strata

Late Early Cretaceous to Holocene rocks rest unconformably on all the older basement units. The last 100 Ma of New Zealand's geological history can conveniently be divided into four periods, which are discussed in the following sections.

Late Early Cretaceous Intracontinental Rifting

Estimates of the timing of the end of Palaeozoic–Mesozoic subduction range from 125 Ma to 85 Ma. The youngest clearly subduction-related plutonic suites are 125 Ma old, and the oldest rift-related alkaline volcanics are about 100 Ma old. Detrital zircon dates of about 100 Ma have been obtained from sandstones imbricated in the accretionary wedge of the Pahau Terrane, but ignimbrites that fill extensional half-grabens are 105–100 Ma old. The two metamorphic core complexes, Paparoa and Fiordland, also attest to extreme local continental extension at around 105 Ma. Hypotheses about the reason for the cessation of subduction in the Cretaceous include migration of a spreading ridge along the trench, stalling of spreading outboard of the trench, and collision of the Hikurangi Plateau large igneous province (Figure 1) with the Gondwanan margin. Late Early Cretaceous sedimentary rocks typically comprise synrift non-marine deposits succeeded by passive-margin marine transgressive strata.

Late Cretaceous–Palaeogene Passive Margin

The oldest oceanic crust adjacent to the Challenger and Campbell Plateaus is about 85 Ma old. Spreading in the Tasman Sea ceased at about 55 Ma. New Zealand moved north (Figure 3) due to continuing spreading on the Pacific–Antarctic Ridge. Marine basins developed across Zealandia as a result of post-rift thermal subsidence. The maximum marine inundation of Zealandia, with widespread limestone deposition, occurred in the Oligocene, but local fluvial and coal deposits dating from throughout the

Late Cretaceous–Cenozoic (albeit in different parts of North and South islands) indicate that Zealandia was never entirely submerged.

Miocene–Pliocene Active Margin Development

By earliest Miocene times, a new plate boundary (in broadly the same place as the present-day boundary; Figure 1) had propagated through Zealandia and cut across the trend of the basement terranes. The development of the Pacific–Australian plate boundary had profound geological consequences. The Northland and East Coast allochthons, consisting of Late Cretaceous–Palaeogene ophiolitic and sedimentary rocks, were thrust onto North Island from the north-east at the end of the Oligocene (*ca*. 25 Ma). Subduction-related arc volcanism started at about 25 Ma and became widespread in Northland and the Coromandel between 15 Ma and 5 Ma. On South Island, intraplate stratovolcanoes developed along the east coast. The Neogene succession is generally thick, clastic-dominated, and regressive.

Quaternary

Although they are part of the continuing Neogene volcanotectonic phase, Quaternary rocks are shown separately in Figure 2 because of their wide areal extent and strong association with present-day landforms. The subduction-related Quaternary volcanoes of North Island have erupted extensive ignimbrite sheets, which blanket the older rocks. In South Island large fluvioglacial outwash plains issue from glacially eroded valleys.

Conclusions

The Phanerozoic geological history of New Zealand can be interpreted in terms of the progressive Pacificwards growth of Gondwana by terrane accretion and batholith intrusion at an obliquely convergent margin. Continental growth was terminated by widespread extension in southern Gondwana from about 105 Ma and was followed by seafloor spreading from about 85 Ma in the Tasman Sea and Southern Ocean. The entire basement and its cover of passive-margin Late Cretaceous to Palaeogene sediments were subjected to renewed deformation in the Neogene with the inception of the modern Australia–Pacific plate margin.

See Also

Antarctic. **Australia:** Tasman Orogenic Belt. **Gondwanaland and Gondwana**. **Large Igneous Provinces**. **Oceania (Including Fiji, PNG and Solomons)**. **Plate Tectonics**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges.

Further Reading

Adams CJ, Campbell HJ, Graham IJ, and Mortimer N (1998) Torlesse, Waipapa and Caples suspect terranes of New Zealand: Integrated studies of their geological history in relation to neighbouring terranes. *Episodes* 21: 235–240.

Ballance PF (ed.) (1993) *South Pacific Sedimentary Basins*. Sedimentary Basins of the World 2. Amsterdam: Elsevier.

Bradshaw JD (1989) Cretaceous geotectonic patterns in the New Zealand region. *Tectonics* 8: 803–820.

Cole JW (1986) Distribution and tectonic setting of Late Cenozoic volcanism in New Zealand. *Bulletin of the Royal Society of New Zealand* 23: 7–20.

King PR (2000) Tectonic reconstructions of New Zealand: 40 Ma to the present. *New Zealand Journal of Geology and Geophysics* 43: 611–638.

Korsch RJ and Wellman HW (1988) The geological evolution of New Zealand and the New Zealand region. In: Nairn AEM, Stehli FG, and Uyeda S (eds.) *The Ocean Basins and Their Margins*, vol. 7B, pp. 411–482. New York: Plenum Press.

Lee DE, Lee WG, and Mortimer N (2001) Where and why have all the flowers gone? Depletion and turnover in the New Zealand Cenozoic angiosperm flora in relation to palaeogeography and climate. *Australian Journal of Botany* 49: 341–356.

Mortimer N (2004) New Zealand's geological foundations. *Gondwana Research* 7: 261–272.

Sutherland R (1999) Basement geology and tectonic development of the greater New Zealand region: an interpretation from regional magnetic data. *Tectonophysics* 308: 341–362.

NORTH AMERICA

Contents

Precambrian Continental Nucleus
Continental Interior
Northern Cordillera
Southern Cordillera
Ouachitas
Southern and Central Appalachians
Northern Appalachians
Atlantic Margin

Precambrian Continental Nucleus

W Bleeker, Geological Survey of Canada, Ottawa, ON, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

North America is a large continent and much of it is ancient (Figure 1). Fringed along several of its coastlines by younger mountain ranges, the broad interior of the continent is underlain by crust that ranges in age from >4.0 Ga to <1 Ga. This ancient crust (Figure 2) formed and progressively amalgamated during several broad 'orogenic' or mountain-building cycles. Although recording net crustal growth, this protracted history of orogenic cycles was far from linear – it involved numerous crust-forming (e.g., growth of volcanic islands) and crust-destroying (e.g., partial subduction) events, punctuated in time. It also involved a complex choreography of tectonic collisions between small lithospheric plates, subsequent rifting and dispersal events, re-aggregation, associated deformation, partial melting and metamorphism, and large-scale rearrangements along faults. Ultimately, at about 1 Ga, this protracted crustal evolution culminated in the growth of a large supercontinent known as Rodinia. In essence, the ancient core of continental North America, which is generally referred to as Laurentia, is a large fragment of this vast late Proterozoic supercontinent (*see* **Precambrian:** Overview).

Between 800 and 600 Ma, supercontinent Rodinia started to break up, resulting in the plate tectonic dispersal of about a dozen large continental fragments. This event liberated the large North American fragment (Laurentia) out of the parental landmass of Rodinia. Since this breakup, the west coast of Laurentia (present coordinates) has been a long-lived active margin facing oceanic plates and colliding island arcs, whereas its eastern margin was modified by yet another major collision and rifting cycle, first forming the Appalachian-Caledonian Mountain (*see* **North America:** Southern and Central Appalachians) Belt, and finally rifting, starting at *ca.* 200 Ma, to form the present-day Atlantic Ocean basin. (*see* **North America:** Atlantic Margin) These events have modified the margins of Laurentia and added marginal terranes of mostly younger crust, such as the Coast Mountains of British Columbia, large parts of Alaska (*see* **North America:** Northern Cordillera), and Gondwana-derived terranes along the eastern seaboard. Nevertheless, North America remains dominated by Precambrian crust, with a mean isotopic age >2 Ga.

The ancient crustal core of the continent, parts of which are exposed in the Canadian Shield (Figure 2), is underlain by subcontinental mantle lithosphere of above average thickness, which locally reaches down to a depth of 200–300 km into the hotter, convective mantle. This 'mantle keel' developed during or shortly after the amalgamation events in the crust, either by thermal growth (i.e., cooling from the top downwards) or by lateral accretion of buoyant slabs of depleted peridotite. The lithospheric keel is currently a reservoir of diamonds, a high-pressure polymorph of carbon. Transported to the surface by exotic volcanic rocks known as kimberlites, these valuable gemstones are now being produced from several mines across northern Canada (Figure 3). The mantle keel is mechanically coupled with the overlying crust, forming a thick and somewhat cooler lithosphere that enhances the strength and

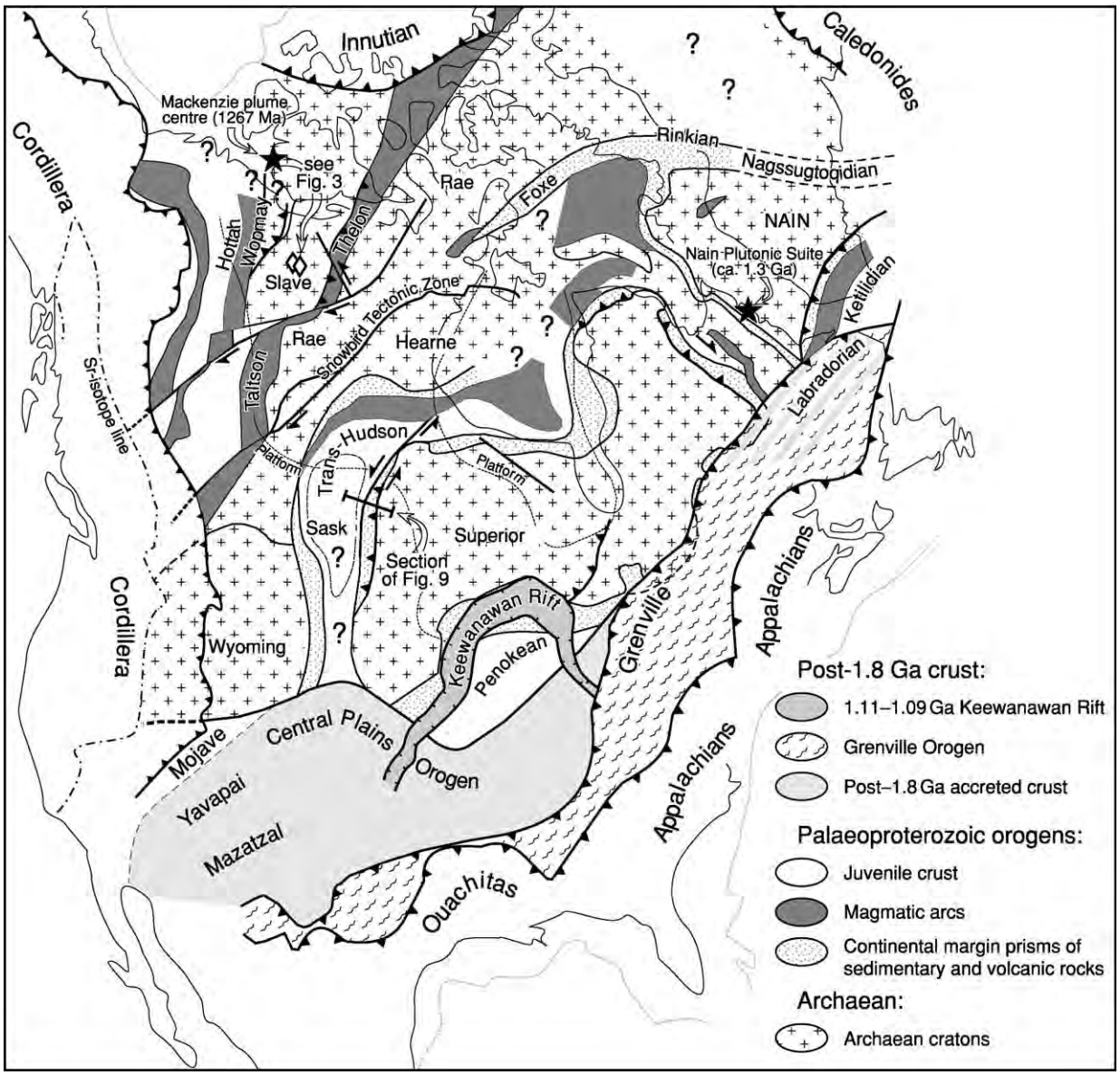


Figure 1 Simplified tectonic map of the Precambrian basement of North America, with Greenland in predrift position (modified after Hoffman, 1988; and Ross *et al.*, 1991).

long-term stability of the ancient continental core of Laurentia.

Modern Plate Tectonic Context and Extent of Laurentia

The continent of North America is an integral part of the North American plate, one of about 12 large lithospheric plates that presently form the segmented outer mechanical and thermal boundary layer of the Earth. These plates float on a solid but convecting mantle. Due to forces acting on their bounding surfaces (e.g., 'slab pull', but also 'ridge push', 'basal drag', and frictional forces), plates drift and interact laterally. Typical absolute plate velocities on the modern Earth are on the order of 1–10 cm year, with the North American plate moving in a west-south-westerly direction at *ca.* 2 cm year (centre of the continent), around a rotation pole located in north-western Equador.

Along its leading edge, the North American plate overrides subducting oceanic plates (Juan de Fuca, Cocos), which fuel arc volcanism along the Cascade and Central American margins. Elsewhere, relative plate motion, with respect to the main Pacific plate, is oblique or sideways, resulting in a transform fault boundary with right-lateral motion (the San Andreas and related faults). Collectively, these processes cause

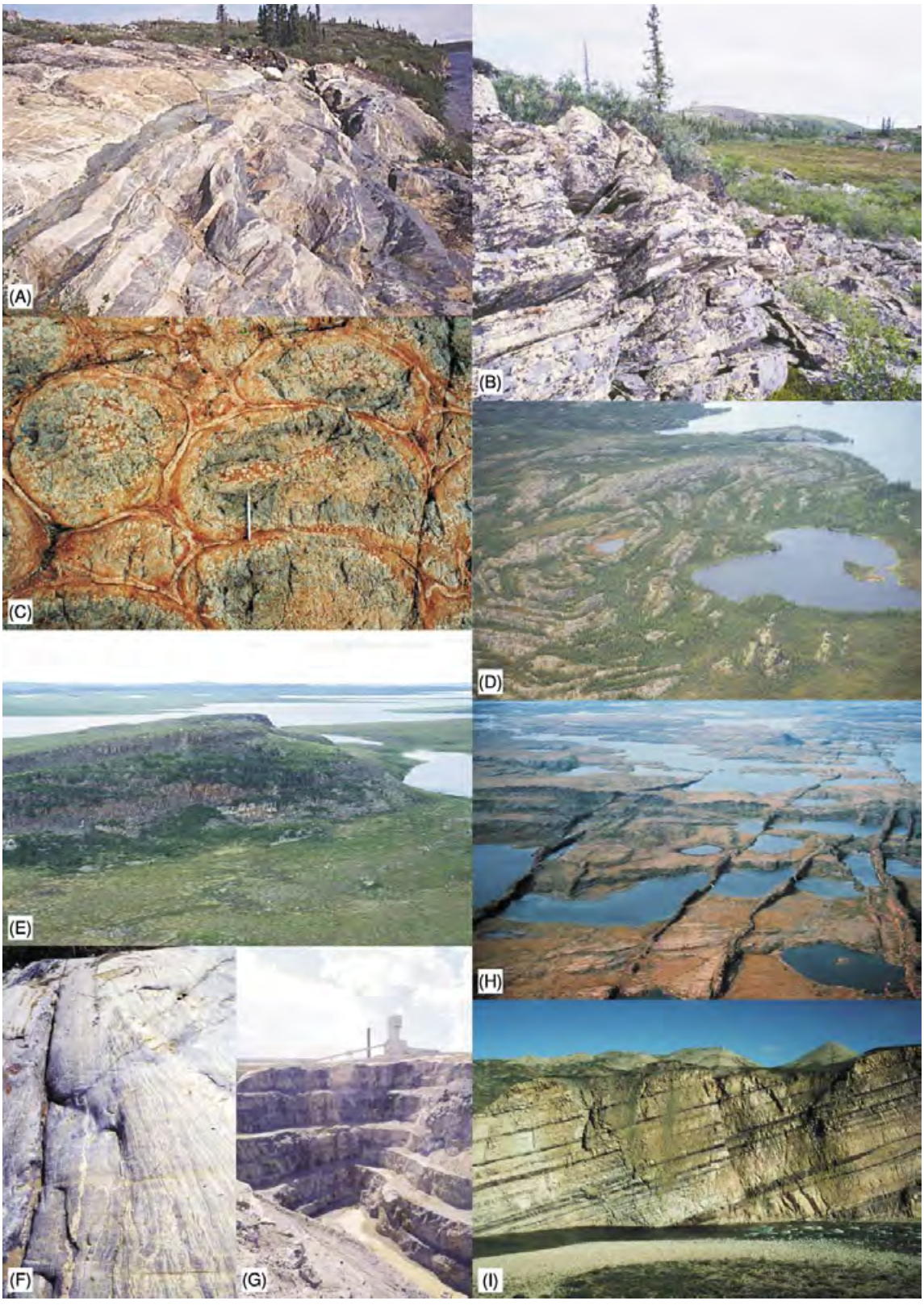


Figure 2 Field photos of representative elements of the Precambrian geology of North America. (A) Earth's oldest intact rocks, the Acasta gneisses, along the western edge of the Slave Craton. *Ca.* 4.03 Ga gabbros and tonalitic gneisses cut by 3.6 Ga isoclinally folded granite sheets and younger mafic dykes. (B) *Ca.* 2.8 Ga fuchsitic quartzites overlying ancient basement gneisses at the base of the Neoarchaean Yellowknife Supergroup, Slave craton. (C) *Ca.* 2710 Ma variolitic pillow basalts of Fe-tholeiitic composition, a typical component of Archaean greenstone belts; Yellowknife greenstone belt, Slave Craton. (D) Multiply deformed, *ca.* 2660 Ma metaturbidites

Figure 3 North America's first diamond mine, the Ekati Mine of the central Slave Craton. Main picture shows the flat barren lands of the central Slave Craton, with several pipe-like kimberlite bodies being excavated. Inset (upper left) shows a close-up of one of the partially excavated pipes. Several mm-size gem quality diamonds are shown on upper right (photos provided by David Snyder and Grant Lockhart).

deformation and modification of the Precambrian crust and lithosphere, which locally extends up to 1000 km inboard of the margin.

The eastern margin of the North American plate is the slow-spreading mid-Atlantic ridge, which runs north through Iceland and to the east of Greenland into the Arctic basin. Hence, Greenland, almost entirely underlain by Precambrian crust, is geologically part of North America and, more specifically, part of its Precambrian core, Laurentia. With progressive northward opening of the North Atlantic rift, Greenland started to rift apart from the main part of Laurentia at about 80 Ma, opening up the Labrador Sea. This rifting was subsequently aborted in favour of propagation of the mid-Atlantic ridge to east of Greenland, along the present trace of the ridge. Here, ridge propagation split off a fragment of Precambrian crust from Laurentia, which remained attached to the Eurasian plate and now underlies parts of Scotland, Ireland, and the adjacent continental shelf. This Precambrian sliver, known as the Lewisian Gneiss Complex, is equally part of Laurentia. Other Precambrian crustal elements may have rifted and drifted off Laurentia with the opening of a proto-Arctic basin (i.e., prior to the Innutian Orogeny, see **Figure 1**), but here relationships are presently less well understood.

The stable Precambrian lithosphere underlying much of North America responds to far-field boundary forces by gentle flexing and arching. As a consequence, Precambrian crust in the north-central parts of the continent is gently arched up, giving rise to the physiographic province of the Canadian Shield. The latter is characterized by scattered rocky outcrops of generally low relief (**Figure 2**), although, locally, rift

of the Burwash Formation, Slave Craton. Turbidites such as these are typical for the uppermost units of Archaean greenstone belts. Here they have been folded into regional mushroom interference patterns by two upright fold generations. (E) Shallow-dipping strata of the *ca.* 2.0 Ga Coronation Supergroup onlapping onto peneplaned Archaean basement (flat foreground), western margin of Slave Craton. These strata represent some of the first true passive margin sequences in the geological record. (F) Archaean–Palaeoproterozoic unconformity, metamorphosed to upper amphibolite facies, along the highly deformed western margin of the Superior Craton. Archaean gneissic layering (right) truncated by primary layering and foliation in basal quartzitic schists of the *ca.* 2.0 Ga Ospwagan Group (left). (G) Thompson Mine, exploiting one of the largest komatiite-associated nickel deposits in the world, hosted by the highly deformed Palaeoproterozoic Ospwagan Group, along the western margin of the Superior Craton. (H) Proterozoic mafic dykes of the 1267 Ma giant radiating Mackenzie swarm, northern Canadian Shield. (I) Tilted strata of Proterozoic fluvial sandstones overlying the *ca.* 1.8 Ga crustal collage of Laurentia; 100 m high cliffs along the Brock River, North-west Territories. Deposits such as these form widespread overlap assemblages as early as 1.75 Ga.

flank uplift has created significant new relief, exposing up to 1000 m-high sections through Precambrian rocks, particularly along the coasts of Labrador, Baffin Island, and Greenland. Minor rejuvenation of relief is currently taking place along arctic coastlines in response to post-glacial rebound. Elsewhere, peneplaned Precambrian basement dips gently beneath a thin Phanerozoic sedimentary cover, giving rise to interior platforms (*see* **North America:** Continental Interior). Presence or absence, and thickness, of this sedimentary cover was, of course, further modulated by changes in global sea-levels, either expanding or shrinking the extent of interior seaways.

In a number of localities, subsidence of the Precambrian lithosphere has been more pronounced, resulting in downwarped intra-continental basins filled with thicker sedimentary successions. One example is the Hudson Bay Basin in the middle of the Canadian Shield, which has been subsiding intermittently during the Phanerozoic. The processes responsible for this intermittent subsidence are poorly understood. Much better understood is the Western Canada Basin, a broad foreland basin forming in response to downflexing of Precambrian lithosphere in advance of the marginal load of the North American Cordillera.

In addition to the broad exposure of the Canadian Shield, crust of Precambrian age is also exposed in more restricted tectonic or plateau uplifts within the Cordilleran and Appalachian-Alleghanian mountain chains that fringe Laurentia. Important examples are the Laramide uplifts in the western United States (e.g., the Wind River Ranges in Wyoming) (*see* **North America:** Southern Cordillera), thrust sheets in the Rocky Mountains, and domal 'core complexes' throughout the internal Cordillera. Deep incision of the Colorado River in the uplifted Colorado Plateau also exposes an important window of Precambrian crust. The Blue Ridge Mountains of the south-eastern United States and the Long Range inlier of Newfoundland are examples of major thrust-related uplifts of Precambrian basement of Grenvillian age in the Appalachians (*see* **North America:** Northern Appalachians). In general, the subsurface edge of Precambrian basement can be traced with isotopic ratios of younger granitoids. An example is the $^{87}Sr/^{86}Sr$-isotopic line of the Cordillera (Figure 1).

Precambrian Nucleus of North America: General Structure

The basic architecture of the Precambrian nucleus of North America is that of a 'collage' comprising crustal elements of different ages (Figure 1). Many of the crustal elements involved in this collage were exotic with respect to each other, having originated in different places around the globe before being amalgamated. Other elements represent crustal growth along the margins of previously assembled continental crust. In places, the crustal collage is truncated by younger rifted margins, providing evidence that, originally, the collage must have extended well beyond its currently preserved margins.

Although the internal structure of the collage is complex, the complexity is far from random. In some ways, the overall architecture can be compared with the 'nested' structure of 'Russian dolls' (Figure 4):

- North America, in its present form, is a fragment of the latest supercontinent in Earth history, Pangaea (Figure 5) (*see* **Pangaea**), which existed from about 300 to 200 Ma and is currently still in the process of progressive breakup and dispersal, although a number of plates have already started their collisional history (e.g., India with Eurasia).
- The North American fragment of Pangaea hosts within itself a large fragment of the earlier supercontinent Rodinia (Figure 4), which was amalgamated from 1200 to 980 Ma. As discussed earlier, this fragment of Precambrian crust, several thousand kilometres in diameter, is commonly referred to as Laurentia.
- On a smaller scale, Laurentia hosts within itself 6–7 large fragments of Archaean (>2.5 Ga) crust, the Archaean 'cratons' (*see* **Precambrian:** Overview). Among these, the Superior, Wyoming, and Slave cratons are some of the better-known examples. These crustal fragments are called cratons because they show long-term stability, having been affected by younger deformation only around their edges. The typical length scale of these Archaean cratons is *ca.* 1000 km. The Superior Craton is the largest preserved Archaean craton on Earth.
- Furthermore, the larger Archaean cratons are typically composite, consisting of a number of domains with disparate crust formation ages. Included among these domains are ancient crustal fragments that are dominated by gneissic granitoids with ages of crystallization or inheritance of >2.9–3.0 Ga. Such ancient gneiss domains (e.g., the Central Slave Basement Complex of the Slave Craton; or the North Caribou Terrane of the Superior Craton; see Figure 4) have typical length scales of 100–300 km.
- Finally, these ancient gneiss domains are themselves heterogeneous. Embedded within them are found Earth's oldest rocks, including, for instance, the 4.03 Ga Acasta gneisses of the Slave Craton (Figures 2A and 4). Individual examples of these

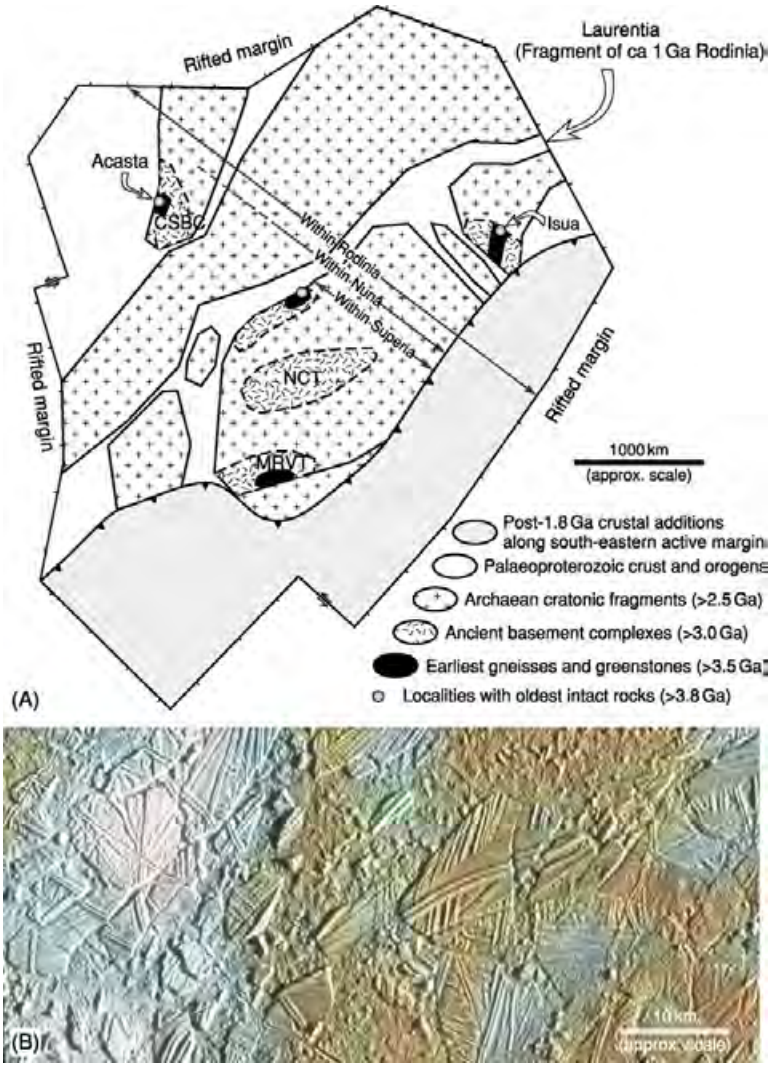


Figure 4 (A) Cartoon illustrating the systematic fractal architecture of the Precambrian crust. The Precambrian record consists of a collage of 'nested' fragments across a variety of scales. At the largest scale, Laurentia itself represents a rifted fragment of Rodinia, a *ca* 1 Ga supercontinent. Its margins formed between 780–600 Ma, when Rodinia started to break up. At the smallest scale, we find individual gneiss complexes consisting of some the oldest intact rocks on Earth (e.g., Acasta, Isua), nested within larger Archaean cratons. The overall fractal structure is one of preservation and reflects repeated cycles of fragmentation and re-aggregation, with smaller older fragments preserved in larger younger fragments. Abbreviations: CSBC, Central Slave Basement Complex, Slave Craton; NCT, North Caribou Terrane, Superior Craton; MRVT, Minnesota River Valley Terrane. (B) An identical fractal pattern of nested fragments, but at smaller scales, is seen in this picture of the icy crust of Europa, one of Jupiter's large moons (source: NASA, Galileo Orbiter PIA01127). Linear features are ridges formed by upwelling and subsequent freezing of water along cracks.

Early Archaean gneiss complexes are typically preserved at the 1–10 km scale, in some cases ranging up to *ca*. 100 km.

Hence, the structure of the Precambrian nucleus of North America, and that of the geological record in general, is self-similar (fractal) in nature, repeating a basic motif across a variety of scales, ranging from the size of modern plates to that of individual ancient gneiss complexes or the greenstones belts embedded within them. As plate tectonics is the dominant process in shaping this pattern on the modern Earth, it is tempting to conclude that plate tectonics must have been equally dominant since at least *ca*. 4.0 Ga, the time of the Earth's oldest preserved crust (Figure 6). One might also conclude that plates have grown in size over time. These conclusions seem reasonable but there are a number of important caveats.

The similarity in basic pattern strongly suggests repeated fragmentation and dispersal events of pre-existing crust, with fragments becoming incorporated in new crustal collages through time. However, similarity in pattern and implied kinematic processes

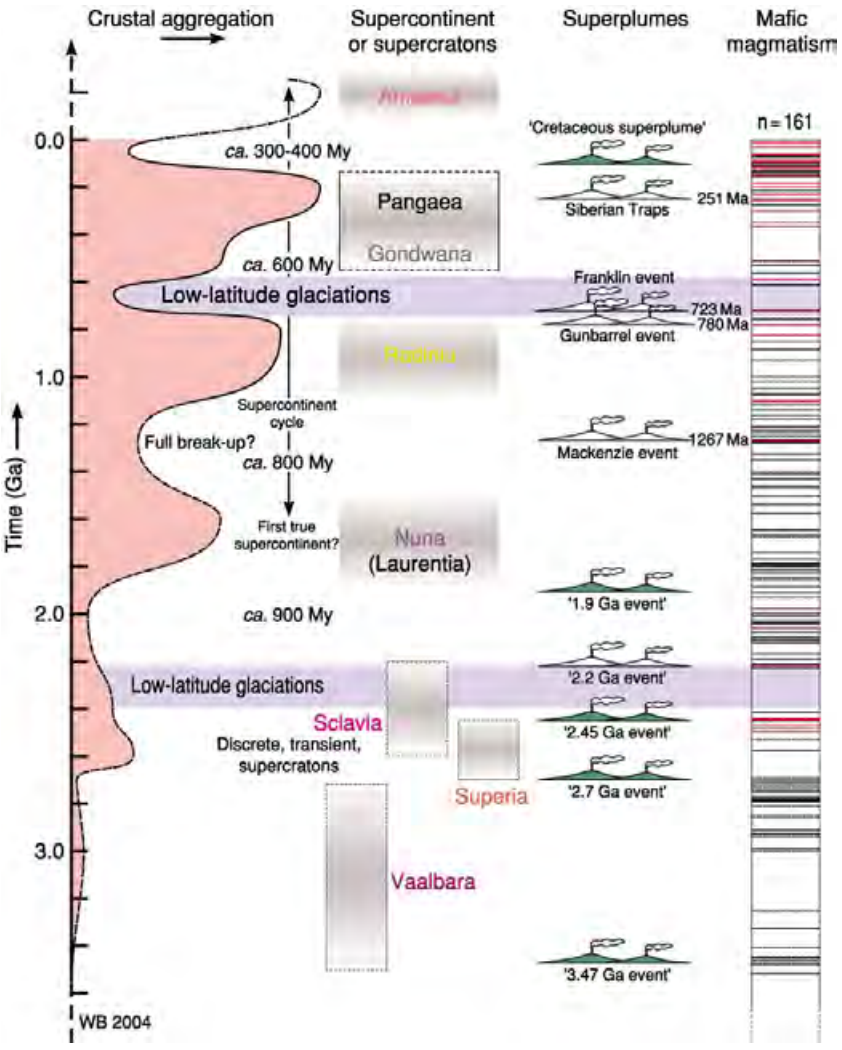


Figure 5 The history of crustal aggregation states (supercratons, supercontinents) through time (vertical axis; modified after Bleeker, 2003). Mid-Proterozoic Nuna, including the large 1.8 Ga core of Laurentia, was probably the first true supercontinent in Earth history. The Late Archaean may have been characterized by several discrete, transient, aggregations referred to as supercratons: Vaalbara, Superia, Sclavia and possibly others. The diachronous break-up of these supercratons, in the Palaeoproterozoic, spawned the present ensemble of *ca.* 35 Archaean cratons, which now are variably incorporated into younger crustal collages. Since the assembly of Nuna, the time gaps between successive crustal aggregation maxima appear to have become shorter. Note the correlation of intervals of global glaciation with two periods of continental breakup and dispersal, and with possible minima in the frequency of mafic magmatic events in the continental record (legend for the latter: red line, well-established mantle plume event; black line, other mafic magmatic event; dashed line, poorly dated event; modified after Ernst and Buchan (2001).

does not necessitate similarity at a dynamic process level. Modern lithospheric plates tend to be large and relatively rigid, and negative buoyancy of subducting, old and cold oceanic plates ('slab pull') is a major component in the overall force balance driving plate motions. At present, it is controversial whether in a hotter Archaean Earth, with 2–4 times the present heat production and a more substantial fraction of the primordial heat budget still preserved inside the Earth, thinner and smaller lithospheric 'plates' interacted the same way as their modern counterparts. A key question is whether the return flow of oceanic mafic-ultramafic material was by rigid slab subduction (i.e., 'plate tectonics') or whether residues, after tonalitic melt extraction, descended back into the convecting mantle by more disorganized gravitational sinking ('drip tectonics')?

Attempts to answer this question from the geological record, as preserved in Archaean cratons, leaves little doubt that the cratons preserve a record of interaction between different crustal fragments. Nevertheless, Archaean cratons lack a fully diagnostic set of criteria confirming modern plate behaviour (e.g., passive margins, flexural basins, accretionary prisms, uncontested ophiolites, high pressure-low temperature metamorphic rocks and

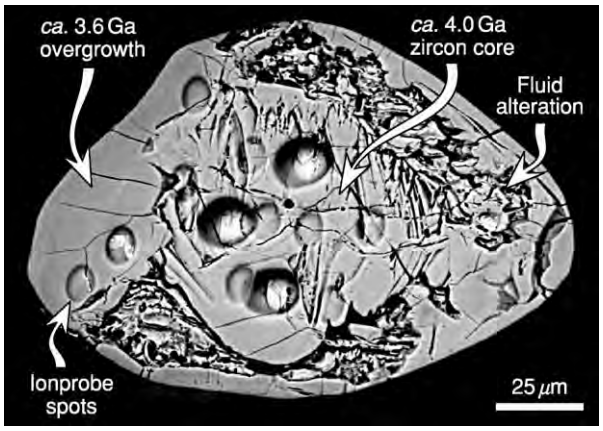


Figure 6 Backscatter electron image of a *ca.* 4.0 Ga zircon crystal from the Acasta gneisses, Earth's oldest intact rocks (see also **Figure 2A**). Metamorphic rims formed at *ca.* 3.6 Ga. Oval spots are scars from the ion probe beam (from Stern and Bleeker, 1998).

paired metamorphic belts, etc.). Hence, it is useful to consider alternative explanations. In doing so, it is also instructive to consider other planetary bodies and their unique pathways of lithospheric evolution. For a comparison with the fractal pattern of repeated fragmentation and amalgamation in the Earth's crust, the icy crust of Europa, one of Jupiter's large moons (**Figure 4B**), is of particular interest (*see* **Solar System: Jupiter, Saturn and Their Moons**). Europa's crust of water-ice, overlying an interior ocean, shows a fractal pattern that is essentially similar to that of Earth's crust, requiring repeated fragmentation events, upwelling and freezing of water along cracks, and lateral mobility of fragments. Yet it is not the product of plate tectonics driven by rigid slab subduction of dense silicate rocks.

Both on Europa and Earth, the fractal pattern of smaller, older fragments embedded in younger crustal collages, is an artefact of the process by which fragments are created and preserved. With each repeated breakup event, crustal fragments can only become smaller, while their number will increase. On the other hand, erosion, tectonic slivering, and partial subduction will, over time, reduce the number of fragments or hide them in the lower crust. Furthermore, a fragment of a certain size and age can only be preserved in a younger fragment that is either the same size or larger. Hence, the negative correlation between fragment size and age, as illustrated in **Figure 4**, is primarily a function of how continental crust is preserved, and may not provide information on whether continental plates have grown in average size over geological time.

Nevertheless, it is intriguing that few of the Archaean cratons come close in size to even the smallest of the major continental plates today. The Superior Craton of North America, the largest extant Archaean Craton, with a surface area of *ca.* $1.57 \times 10^6\,km^2$, is about half the size of the Arabian Plate. Indeed, several independent lines of evidence suggest that Archaean continental aggregations may have been smaller and more transient than their modern counterparts:

- A hotter Archaean Earth, with a substantially higher radiogenic element budget, favours smaller and faster plates to dissipate the increased heat production and avoid catastrophic heating of the planetary interior.
- The record of $^{87}Sr/^{86}Sr$ isotopic compositions of Precambrian seawater, as measured from marine carbonates, shows that prior to 2.6–2.5 Ga, Sr ratios were buffered by dominant interaction with oceanic crust and young mantle-derived mafic rocks. Only after *ca.* 2.5 Ga do we see a rapid rise in the $^{87}Sr/^{86}Sr$ ratios, reflecting more significant input from weathering of aged continental crust.
- Much of the modern detrital sediment load to sedimentary basins and continental shelf-slope systems is provided by large river systems draining large continental hinterlands. The fact that there are few examples of large sedimentary basins of a shelf-slope affinity preserved in the Archaean record, suggests that river systems and their continental hinterlands were smaller, and that large, emerged continental plates were rare or absent.
- As attested by Laurentia, once large cratonic landmasses have aggregated and become underlain by a lithospheric keel, they tend to resist breakup. So, as noted above, the observation that none of the Archaean fragments, with their lithospheric keels, approaches modern continents in size may be significant after all.
- Characteristics of the supercontinent cycle (**Figure 5**) suggest that the successive time lags between breakup and dispersal of one supercontinent (e.g., Rodinia) and the re-aggregation of a subsequent supercontinent (e.g., Pangaea) have become shorter over the last 2.5 billion years. The easiest explanation for this observation is that, following supercontinent breakup and dispersal, modern continental plates, due to their large average size, quickly run out of room and start colliding. More substantial time lags in the past suggest smaller continental plates.
- And finally, Earth's present rate of heat loss is about twice the rate of internal heat production, while at the same time heat production (mainly from K, Th, U) is undergoing slow exponential decay. Thermal arguments predict therefore that,

over time, the lithosphere will slowly thicken and stiffen, evolving ultimately to a relatively immobile 'one-plate' state. This state has already been reached by our closest sister planet, Venus, which is only marginally smaller than Earth (Venus' radius is 95% that of Earth).

Precambrian Nucleus of North America: A Systematic Overview

From 2.0 Ga to 1.8 Ga, over a time-span of about 200 million years, the core of Laurentia formed by progressive amalgamation of 6–7 large fragments of Archaean crust and intervening island arcs into a broad orogenic collage (Figure 1). Crustal shortening and thickening accompanying the various accretion and collision events led to uplift and exhumation and, ultimately, stabilization of the newly formed crust, as indicated by the initiation of widespread intracontinental sedimentary basins that overstep the boundaries between different crustal blocks as early as *ca* 1.75 Ga. Interestingly, the time-frame of relatively rapid crustal aggregation and growth implied for Laurentia is of a similar order to that for the progressive development of the most recent supercontinent, Pangaea, which started with assembly of the large southern continent Gondwana in the latest Proterozoic and terminated with final collision of Gondwana and Laurasia in the Carboniferous to Permian (i.e., an overall time-frame of *ca*. 250 Ma).

The widespread 2.0–1.8 Ga orogenic events in Laurentia, known in North America as the Hudsonian Orogeny, have counterparts in most other continents (e.g., the Amazonian Orogeny in Brazil; the Svecofennian Orogeny of the Baltic Shield; the Capricorn Orogeny of Western Australia). Through the collective amalgamation of a large number of disparate Archaean cratons or microcontinents into a global collage, this first-order orogenic event may have led to the first true supercontinent in Earth history at about 1.75 Ga named Nuna (Figure 5). Through further crustal growth along its external margins, and final continent–continent collision to form the Grenville Orogen along its southeastern margin (Figure 1), Nuna evolved into *ca*. 1 Ga Rodinia.

Clearly, the Archaean cratons are the dominant building blocks in the Palaeoproterozoic assembly of Laurentia, and the core of Nuna. Hence, they form a logical starting point for a brief systematic overview of the Precambrian crust.

The Archaean Cratons

Individual Archaean cratons that constitute the core of Laurentia are: the Superior, Slave, and Wyoming cratons; the highly reworked Rae and Hearne cratons, collectively known as the Churchill structural province; the enigmatic Sask Craton, which underlies much of the Palaeoproterozoic Trans-Hudson Orogen in central Canada; and the Nain or North Atlantic Craton underlying southern Greenland (Figure 1). These cratons represent a subset out of a total ensemble of about 35 large Archaean crustal fragments preserved around the world.

The independent nature of some of the component cratons of Laurentia is still a matter of debate. Clearly, the well-studied Superior and Slave cratons are two pieces of crust that are exotic relative to each other and most likely originated from unrelated ancestral landmasses – the Late Archaean supercratons Superia and Sclavia, respectively (Figure 5). The ancestry of some of the other cratons is less clear.

A putative suture between the Hearne and Rae cratons, the Snowbird Tectonic Zone (Figure 1), remains controversial. Hence, it is possible that the Hearne and Rae cratons represent a contiguous fragment of Archaean crust. The Hearne Craton is locally overlain by an Early Proterozoic cover sequence, the *ca*. 2.4–2.1 Ga Hurwitz Group, which shows similarities to the Huronian Supergroup overlying the southern Superior Craton. Both sequences contain evidence for *ca*. 2.3 Ga low-latitude glaciations. Futhermore, both cratons are cross-cut by *ca*. 2450 Ma mafic dykes. Palaeomagnetic data from these dykes suggest that the Hearne Craton may have originated, at 2450 Ma, from just south of the 'Huronian margin' of the Superior Craton (Figure 7). Thus, both cratons likely originated from within supercraton Superia, possibly along with others such as the Karelia, Nain and Yilgarn cratons.

Whether the Wyoming Craton, underlying much of the north-western United States, is truly distinct or just a southern continuation of the Hearne Craton also remains unclear. The proposed suture between the Hearne and Wyoming cratons is buried underneath thick platform strata. The Sask Craton may be an exotic piece of crust or, alternatively, a partially detached piece of either the neighbouring Hearne or Superior Craton, incorporated in the Trans-Hudson Orogen. Interestingly, many of these fundamental questions can be resolved, in principle, with detailed palaeomagnetic studies and high-precision age dating of the Palaeoproterozoic mafic dyke swarms that cut the cratons, in conjunction with stratigraphic comparisons of the Archaean cratons and the Palaeoproterozoic cover sequences that overlie them.

The Slave Craton

As an example of the Archaean components of Laurentia, the Slave Craton is described in more detail.

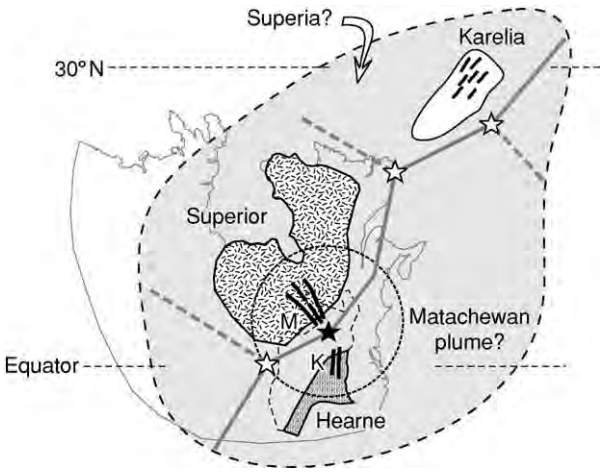


Figure 7 Possible reconstruction of the relative position of the Superior and Hearne crustal fragments at *ca.* 2450 Ma, prior to the inferred break-up of their ancestral supercraton, Superia. Both cratons are cut by *ca.* 2450 Ma mafic dykes that are petrographically similar, the Kaminak dykes (K) of the northern Hearne and the Matachewan dykes (M) of the southern Superior, respectively. Palaeomagnetic data from both dyke swarms allow the reconstruction shown here. Both dyke swarms and the inferred break-up of Superia may have been consequences of the *ca* 2450 Ma Matachewan plume (black star). Karelia, another Archaean craton now embedded in the Baltic Shield, also hosts *ca* 2450 Ma mafic dykes and, thus, may have been part of Superia. Also shown is a possible topology of intracontinental rift zones along which supercraton Superia broke up, liberating individual cratons. Plumes or hotspots along this rift zone (open stars) may have had ages spreading out over a 100–200 million year interval, similar to those along the mid-Atlantic rift during progressive break up of Pangaea.

This well-exposed fragment of Archaean crust measures about 500 (E–W) by 700 (N–S) km and is surrounded by orogens of Palaeoproterozoic or younger age (Figures 1 and 8).

The west-central part of the craton is underlain by an ancient gneiss complex that hosts some of the oldest intact rocks in the world, the *ca.* 4.03 Ga Acasta Gneisses (Figure 2A). This basement complex is overlain by a thin, *ca.* 2.8 Ga quartzite and banded iron formation sequence (Figures 2B and 8), which in turn is overlain by the main cycle of *ca.* 2.73–2.62 Ga volcanic and sedimentary rocks (Figures 2C,D and 8). Very similar stratigraphic relationships are seen in other cratons of Laurentia and around the world.

The Neoarchaean supracrustal rocks of the Slave Craton, known as the Yellowknife Supergroup, were intruded by voluminous granitoid rocks that range in age and affinity from *ca.* 2.71–2.67 Ga, synvolcanic, tonalite-granodiorite plutons to *ca.* 2.58 Ga, post-tectonic, granite plutons. The supracrustal and pre-tectonic granitoid rocks are deformed by several discrete folding events, resulting in complex fold interference structures (Figure 2D) and high-amplitude basement-cored domes. The overall result is an archetypical Archaean 'granite-greenstone terrane'.

The younger granites are part of a craton-wide, *ca.* 2.59–2.58 Ga 'granite bloom', which transported heat, partial melt fractions, and aqueous fluids, as well as most of the incompatible heat producing elements (K, U, Th), from the hot lower crust to the upper crust. This critical and irreversible step allowed the lower crust to cool and to mechanically couple with the mantle lithosphere. Shortly following this terminal granite bloom, Slave crust stabilized and became 'cratonic'.

Following cratonization, from 2.58 Ga to 2.2–2.0 Ga, Slave crust was gradually exhumed by *ca.* 10–15 km (on average) through slow uplift and erosion. This is indicated by the onlap of unconformably overlying Palaeoproterozoic cover sequences onto exhumed Archaean basement (Figure 2E). Starting at about 2230 Ma, the Slave Craton, still within the larger context of its host supercraton Sclavia, was intruded and weakened by numerous mafic dyke swarms and anorogenic alkaline complexes. These events eventually led to rifting and breakup of Sclavia and independent drift of the Slave Craton as a small microcontinent. In general, these Early Proterozoic rifting and dispersal events resulted in some of the first true sediment-rich passive margin sequences in the geological record (e.g., the Coronation Supergroup overlying the western margin of the Slave Craton, Figure 2E).

Proterozoic Orogens: the Glue of the Laurentian Collage

The oldest Palaeoproterozoic orogen within Laurentia is the Taltson-Thelon Orogen along the western margin of the Rae Craton. It is marked by a *ca.* 2.02–1.95 Ga magmatic arc, with a conspicuous aeromagnetic signature that allows it to be traced for several thousand kilometres (Figure 1). This magmatic arc suggests significant consumption of oceanic lithosphere below the western edge of the Rae Craton, leading to collision and eventual suturing of the Slave Craton to the Rae. Partial subduction of Slave crust below the Rae led to flexural subsidence of the Slave and its overlying platform sequence, resulting in one of the best examples of a Palaeoproterozoic foreland trough—the Kilihigok Basin. The short flexural wavelength of this basin suggests a low flexural rigidity of Slave lithosphere at 1.9 Ga, and hence argues against a thick lithospheric keel at that time.

During the Slave–Rae collision, subduction is also active on the western side of the Slave Craton, underneath an outboard Hottah Arc. This arc accreted to

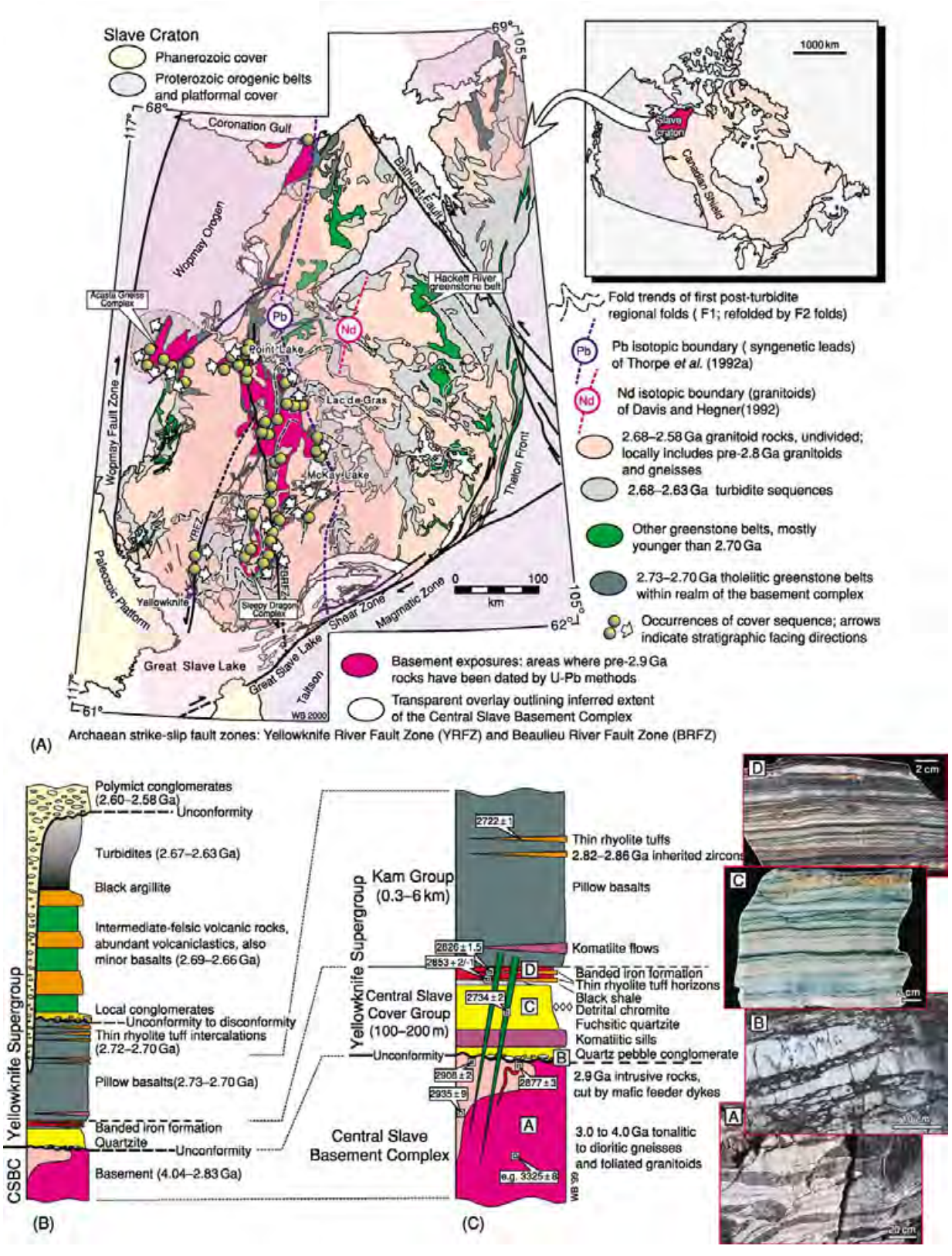


Figure 8 Geology and generalized stratigraphy of the Slave Craton. (A) Geological map; transparent overlay (light purple) outlines minimum extent of the Mesoarchaean to Hadean basement complex of the Slave Craton. (B) Generalized stratigraphy for large parts of the craton. (C) Typical stratigraphy immediately overlying the Mesoarchaean basement complex; photos A–D illustrate different units in the section.

the Slave Craton, followed by a subduction polarity flip and building of a new arc on the Hottah–Slave Suture at *ca*. 1.88 to 1.86 Ga.

Elsewhere, the eastern margin of the Rae-Hearne block (already assembled or always contiguous?) started to interact with outboard island arcs, representing the early stages of the Trans-Hudson Orogen (Figure 1). Similar processes were occurring around the margins of the other cratons, some still at remote locations (e.g., the onset of arc accretion along the southern margin of the Superior Craton, initiating the Penokean Orogen). Then, in a spectacular climax of orogenic activity, between about 1850 Ma and 1800 Ma, the remaining cratons gradually aggregated, collided, and amalgamated. This final stage involved the putative closure of the wide Manikewan Ocean between the Rae-Hearne and Superior cratons to form the >3000 km-long Trans-Hudson Orogen through the centre of the Canadian Shield. Terminal collision in Trans-Hudson Orogen (Figure 9) involved northward indentation of the large Superior Craton into a northern hinterland of the Rae-Nain-Hearne-Slave Collage. This northward indentation caused oroclinal bending of the hinterland and tectonic escape of crustal blocks along large strike-slip faults. The northward movement also caused late-stage transcurrent motion on the lateral margins of the Superior indentor (Figure 1).

Post-1.8 Ga Growth and Modification of Laurentia

Following the climactic growth of Laurentia at *ca*. 1.8 Ga, orogenic activity transferred to the external margins of Laurentia, in the west and the south-east. A complete assessment of growth along the western margin needs to involve a western landmass (most likely a combined Australia-Antarctica continent and its amalgamation history), which has since rifted off.

In the south-east, 1.8 Ga orogenic activity must have left major tracts of oceanic lithosphere along the 'trailing edges' of the plates that carried the Archaean cratons into the Laurentian collage. This oceanic lithosphere was subsequently consumed. A number of successive arcs formed, either as outboard

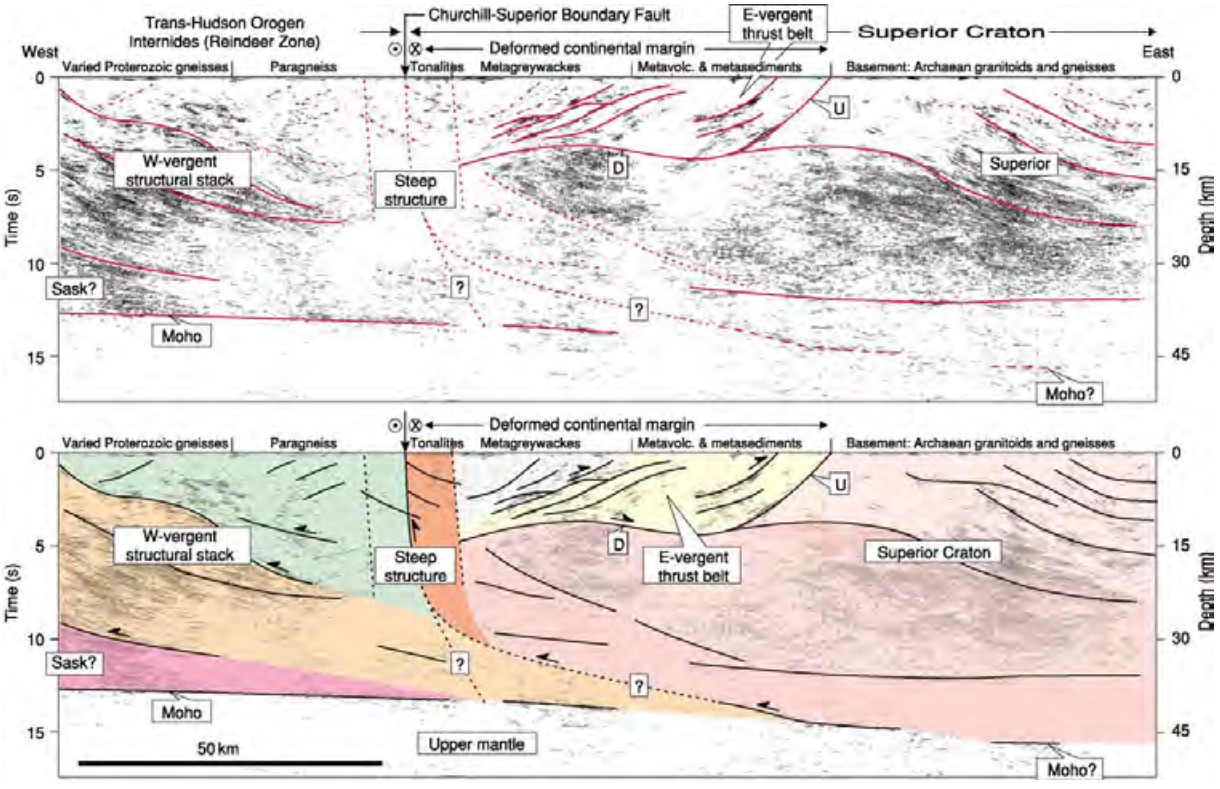


Figure 9 Representative seismic reflection profile (upper diagram) and interpreted cross-section (lower diagram) through one of the major *ca*. 1.8 Ga sutures of Laurentia, the 'Churchill-Superior Boundary Zone', along the western margin of the Superior Craton (see Figure 1 for location of profile). Early deformation imbricated the continental margin of the Superior Craton, including the *ca*. 2.0 Ga Ospwagan Group (Figure 2F and G) into an E-vergent fold-thrust belt. During terminal collision, vergence changed and the buoyant Superior margin started to override juvenile Palaeoproterozoic rocks of the internal zone of the Trans-Hudson Orogen. Oblique convergence degenerated into late-stage sinistral strike-slip deformation along the steep Churchill-Superior Boundary Fault (modified after White *et al*., 2002; see also Bleeker, 1990).

island arcs or, at other times, as magmatic arcs on the growing margin of Laurentia. These long-lived active margin processes led to progressive outward growth of and accretion to Laurentia, culminating with the final collisional history of the *ca*. 1.19–0.98 Ga, high-grade, Grenville Orogen (Figure 1).

Significant pulses of outward growth that can be recognized along various parts of the margin are the Yavapai (1.79–1.71 Ga) and Mazatzal (*ca*. 1.70–1.62 Ga) belts of the south-western United States (Figure 1). The latter is broadly coeval with the Labradorian event of north-eastern Laurentia (Figure 1), which involved accetion of a 1.70–1.66 Ga outboard arc to the margin of Laurentia at about 1655 Ma. The Labradorian Arc and a number of younger arcs (e.g., the Pinwarian Arc) were severely reworked during terminal Grenville collisions. The latter involved final suturing of Laurentia to a southeastern landmass (Amazonia?) to form Rodinia.

During its long-lived Proterozoic active margin evolution, lasting from about 1.8 Ga to 1.0 Ga, Laurentia likely faced a 'Pacific-type' ocean to the south-east (present coordinates). Inboard of this margin, several major magmatic events suggest the upwelling of hot mantle underneath the growing continent. Similar processes ultimately led to the rifting and breakup of Laurentia (as part of Rodinia), starting at about 780 Ma. Important examples of mid-Proterozoic igneous events are several pulses of granite-rhyolite magmatism concentrated in south-eastern Laurentia; the intrusion of the 1.3 Ga Nain plutonic suite in Labrador; the 1267 Ma giant Mackenzie dyke swarm; several more localised mafic dyke swarms; and the basalt-rhyolite volcanism of the Keewanawan event in the mid-continent rift (Figure 1).

Glossary

Nuna Earth's first true supercontinent that formed by progressive amalgamation of numerous microplates and intervening island arc terranes towards the end of the Palaeoproterozoic era, from about 1.9 Ga to 1.7 Ga.

Sclavia Late Archaean ancestral landmass of the Slave Craton and allied cratons. This landmass or supercraton amalgamated during Late Archaean orogenic activity, cratonized shortly after 2.6 Ga, and experienced rifting and break-up between 2.2 Ga and 2.0 Ga, spawning several microplates, including its type craton, the Slave Craton of the Canadian Shield. Existence of this late Archaean supercraton is indicated by the Slave Craton being a cratonic fragment surrounded by rifted margins.

Superia Late Archaean ancestral landmass of the Superior Craton and allied cratons. This landmass or supercraton amalgamated during Late Archaean orogenic activity, cratonized shortly between 2.68 Ga and 2.62 Ga, and experienced rifting and break-up starting at 2.48 Ga, spawning numerous microplates, including its type craton, the Superior Craton of the Canadian Shield. Existence of this Late Archaean supercraton is indicated by the Superior craton being a cratonic fragment surrounded by rifted margins.

Supercraton One of several Late Archaean cratonic landmasses, not necessarily a single supercontinent, that on Palaeoproterozoic break-up spawned the present ensemble of *ca*. 35 Archaean cratons with their rifted margins.

See Also

Earth Structure and Origins. **Gondwanaland and Gondwana**. **North America:** Atlantic Margin; Continental Interior; Northern Appalachians; Northern Cordillera; Southern and Central Appalachians; Southern Cordillera. **Pangaea**. **Plate Tectonics**. **Precambrian:** Overview. **Solar System:** Jupiter, Saturn and Their Moons.

Further Reading

Anderson DL (2002) How many plates? *Geology* 30: 411–414.

Bleeker W (1990) New structural-metamorphic constraints on Early Proterozoic oblique collision along the Thompson nickel belt, northern Manitoba. In: Lewry JF and Stauffer MR (eds.) The Early Proterozoic Trans-Hudson orogen of North America. Geological Association of Canada Special Paper 37, pp. 57–74.

Bleeker W (2002) Archaean tectonics: a review, with illustrations from the Slave craton. In: Fowler CMR, Ebinger CJ, and Hawkesworth CJ (eds.) *The early Earth: physical, chemical and biological development*, Geological Society of London Special Publication 199, pp. 151–181. UK: Bath.

Bleeker W (2003) The late Archean record: a puzzle in ca. 35 pieces. *Lithos* 71(2–4): 99–134.

Buchan KL and Ernst RE (2004) *Diabase dyke swarms and related units in Canada and adjacent regions*. Geological Survey of Canada Map No. 2022A, scale 1:5 000 000, Ottawa, Canada.

Davies GF (1999) *Dynamic Earth: plates, plumes and mantle convection*. Cambridge: Cambridge University Press.

de Wit MJ and Ashwal LD (1997) *Greenstone Belts*. Oxford Monographs on Geology and Geophysics 35, Oxford: Oxford University Press.

Ernst RE and Buchan KL (2001) Large mafic magmatic events through time and links to mantle-plume heads. In: Ernst RE and Buchan KL (eds.) *Mantle plumes: their identification through time*, Geological Society of America, Special Paper 352, pp. 483–575. Colorado, Boulder.

Gower CF and Krogh TE (2002) A U-Pb geochronological review of the Proterozoic history of the eastern Grenville Province. *Canadian Journal of Earth Sciences* 39: 795–829.

Grotzinger JP and Royden L (1990) Elastic strength of the Slave Craton at 1.9 Gyr and implications for the thermal evolution of the continents. *Nature* 347(6288): 64–66.

Heaman LM (1997) Global mafic magmatism at 2.45 Ga: remnants of an ancient large igneous province? *Geology* 25(4): 299–302.

Hoffman PF (1988) United plates of America, the birth of a craton: Early Proterozoic assembly and growth of Laurentia. *Annual Reviews in Earth and Planetary Science* 16: 543–603.

Hoffman PF (1989) Precambrian geology and tectonic history of North America. In: Bally AW and Palmer AR (eds.) *The geology of North America – An overview,* The Geology of North America, vol. A, pp. 447–512. Doubler, Colorado: Geological Society of America.

Hoffman PF (1991) Did the breakout of Laurentia turn Gondwanaland inside-out? *Science* 252(5011): 1409–1412.

Hoffman PF (1992) Supercontinents. In: Encyclopedia of Earth System Science 4, pp. 323–328. London: Academic Press Inc.

Ross GM, Parrish RR, Villeneuve ME, and Bowring SA (1991) Geophysics and geochronology of the crystalline basement of the Alberta basin, western Canada. *Canadian Journal of Earth Sciences* 28: 512–522.

Stern RA and Bleeker W (1998) Age of the world's oldest rocks refined using Canada's SHRIMP: the Acasta gneiss complex. *Geoscience Canada* 25: 27–31.

Wheeler JO, Hoffman PF, Card KD, *et al.* (1996) *Geological map of Canada.* Geological Survey of Canada 'A' Series Map No. 1860A, scale 1:5 000 000. Ottawa, Canada.

White D, Lucas SB, Bleeker W, Hajnal Z, Lewry JF, and Zwanzig HV (2002) Suture-zone geometry along an irregular Paleoproterozoic margin: the Superior boundary zone, Manitoba, Canada. *Geology* 30(8): 735–738.

Continental Interior

D F Merriam, University of Kansas, Lawrence, KS, USA

© 2005, Elsevier Ltd. All Rights Reserved.

> PB King wrote "The [Continental] Interior Lowlands are differentiated from the Laurentian [Canadian] Shield mainly by their retention of a sedimentary cover, so that the tectonic boundary between the two divisions is arbitrary and the differences are not fundamental."
>
> (*The Tectonics of Middle North America*, 1951)

Introduction

The vast country of essentially flat-lying sediments rimmed on three sides by the old Appalachian, Ouachita, and Rocky Mountain mountain chains and to the north the poorly exposed, heavy glaciated, low-lying Canadian Shield (nucleus) composed of Precambrian crystalline igneous and metamorphic rocks, comprises the North American Continental Interior (Figure 1). This immense area, also known as the Central Stable Region, is drained by the mighty Mississippi River and tributaries – and contains the large, relatively shallow, freshwater Great Lakes. This physiography accurately reflects the tectonic provinces (Figure 2). Occasionally, bits of the basement poke through the veneer of this essentially gently-dipping sedimentary cover in the Black Hills of South Dakota, Wichita Mountains of Oklahoma, Ozarks of Missouri, and Baraboo Range of Wisconsin. The shape and form of the sedimentary rocks over the area conform well to the shape and form of the Precambrian basement foundation on which they lie.

Although the topography is subdued and the spectacular scenery is formed for the most part, by erosional remnants or incised valleys, it reflects geological conditions of the region and thus the landforms are a key to understanding recent developments. This part of North America forms the physiographic provinces of the Interior Lowlands and the Great Plains – the bread basket of the continent.

The nucleus of the North American continent is the Precambrian Canadian Shield – the craton, a stable feature around and on which subsequent events took place (*see* **North America:** Precambrian Continental Nucleus). The craton itself had a long and complex history but because of poor exposure and structural complexities it has been difficult to decipher. With a few notable exceptions, the craton is made up of ancient igneous and metamorphic crystalline rocks. The notable exception is the Midcontinent Rift System (MRS), extending from Minnesota to Oklahoma, which is an elongated, younger, faulted, and down-dropped trough containing extrusive igneous rocks interspersed with thick, locally derived, immature sediments (Figure 3). Near the end of the Precambrian, a prolonged period of erosion beveled and smoothed the old fractured/faulted surface prior to it being covered by the younger sedimentary rocks.

Figure 1 Index map of North American physiographic provinces showing location of Interior Lowlands and Great Plains of Continental Interior wedged in between the Appalachian Mountains on the east and the Rocky Mountains on the west. (From Kay and Colbert (1965), John Wiley & Sons.)

This part of the Earth was relatively stable during the next several hundred million years, and as the seas advanced and retreated over the low relief surface, they deposited fairly uniform marine sediments, layer upon layer, over extensive areas. The type and distribution of the sediments responded to the global change in sea-level as well as to changes in local conditions. All of these events were recorded in the sedimentary section, which, as the history unfolded, was sprinkled liberally with long intervals of no record. Because of the sensitivity to conditions in the accumulation of sediments in this type of environment, much of the history is not chronicled; in fact, it has been estimated that as much as 85% of time in any one place is not represented by the rock record. This situation is aptly described as layer-cake geology.

The Foundation

The foundation is the craton, which was formed by adding on (accretion of) material to the original

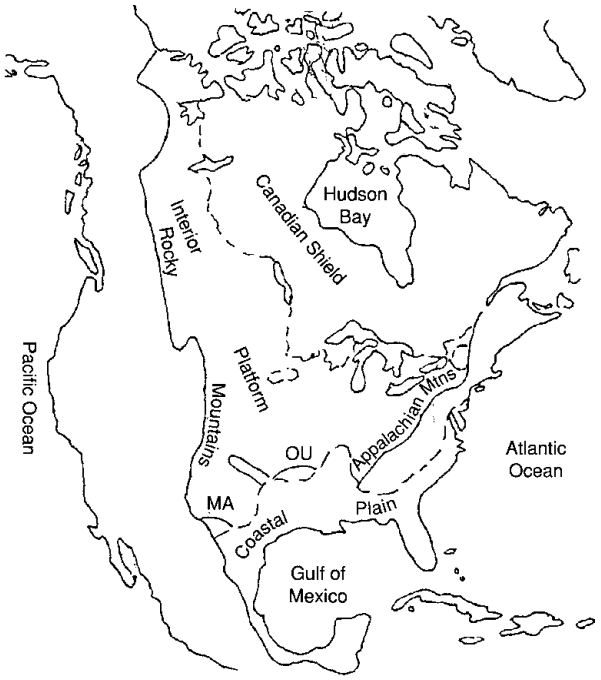


Figure 2 Simplified tectonic map of the North American Continent (MA = Marathon Uplift; OU = Ouachita Mountains). (Modified from Bally, Scotese and Ross (1989) GSA, The Geology of North America, v. A.) Note physiographic provinces mimic tectonic ones.

primaeval nucleus (Archaean age; 3600 to 2600 Ma) so that the younger Precambrian crystallines (Proterozoic age; 2500 to 900 Ma) rim the nucleus to the south (Figure 3). The Midcontinent Rift System (MRS; 1200 to 1000 Ma) cuts across the older Precambrian terrain. This foundation forms part of the continental crust which ranges in thickness from 40 to 50 km (Figure 4).

In the Continental Interior south of the Canadian Shield, the Precambrian basement complex is exposed in only a few places. Thus, the knowledge of this time period in the Earth's history is known from the thousands of wells drilled in the search for petroleum, water, and mineral resources, plus, of course, geophysical data.

The Sedimentary Veneer

Overlying the Precambrian foundation is the sedimentary veneer, the composition and configuration of these younger sedimentary deposits giving clues to the conditions under which they were deposited and accumulated. Thus, information is contained in the rocks about the location of sea margins, climate, and life at that time.

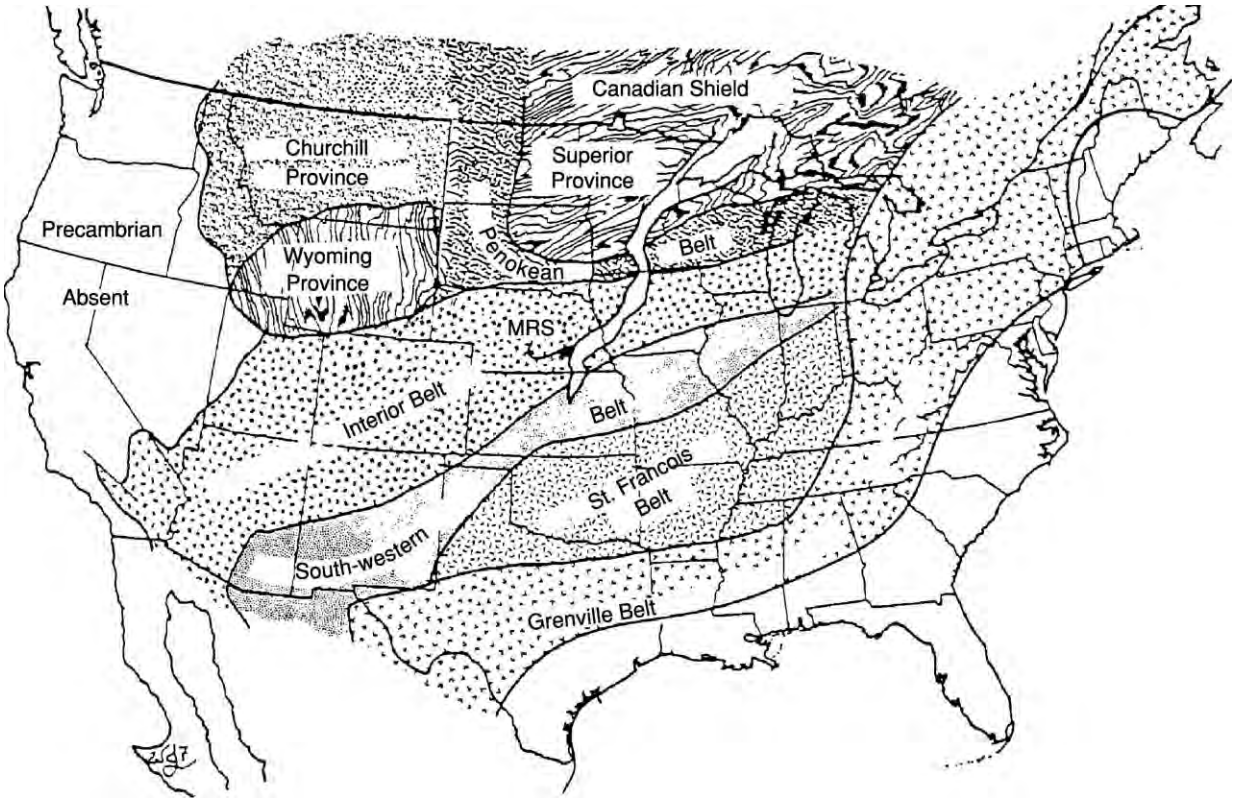


Figure 3 Proterozoic crustal provinces in North America. Archaean: Canadian Shield (>2500 Ma); Proterozoic: Penokean Belt (1900–1820 Ma), Interior Belt (1780–1690 Ma), South-western Belt (1680–1610 Ma), St Francois Belt (1480–1380 Ma), Greenville Belt (1000–550 Ma). Midcontinent Rift System (MRS). (Compiled by Van Schmus and Bickford (eds.) (1993) GSA, The Geology of North America, v. C-2.)

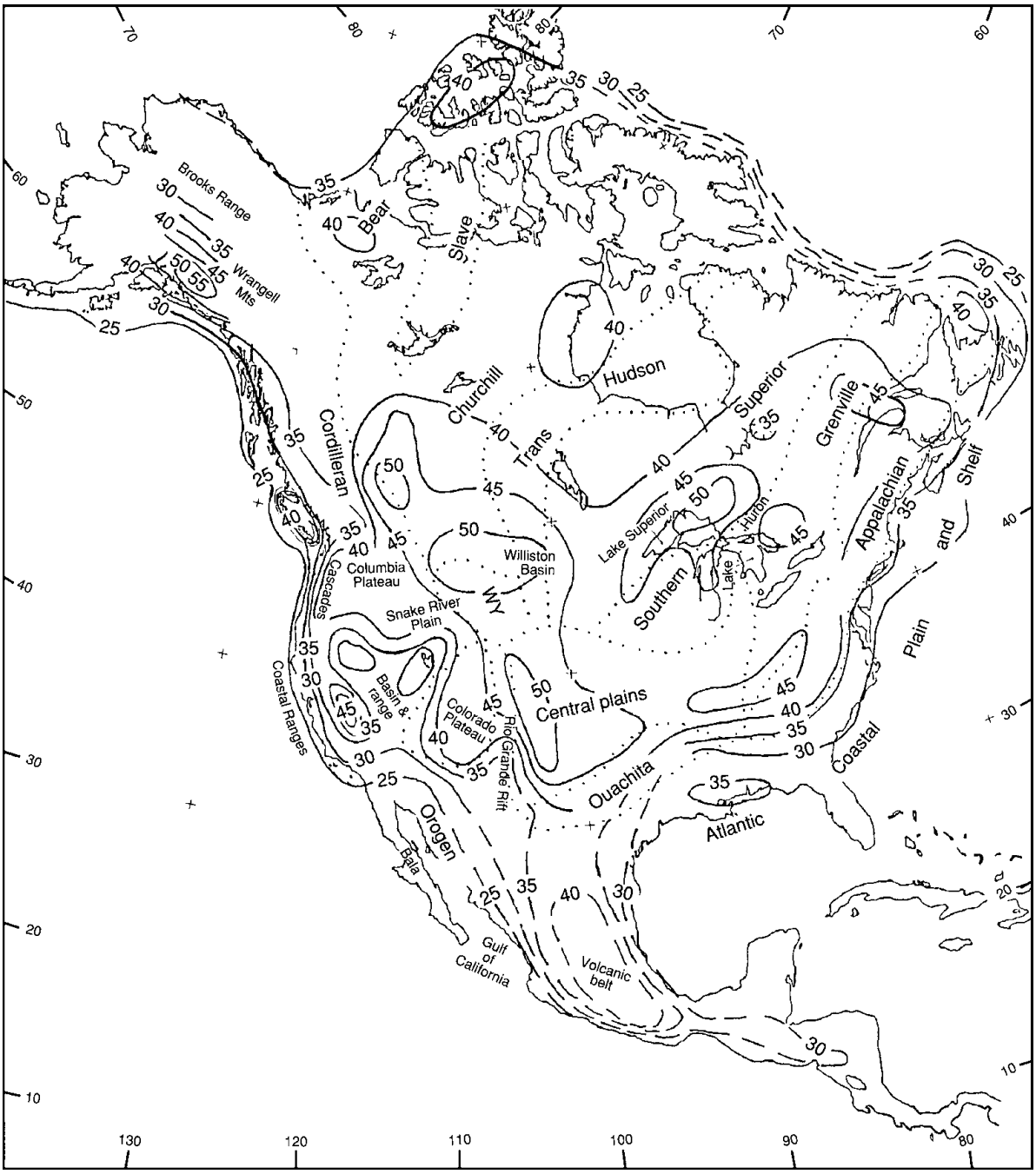


Figure 4 Crustal thickness of North America superposed on simplified geologic province map. Thickness in Continental Interior ranges from 40 to 50 km; Contour Interval = 5 km. (From Houston (1993) GSA, The Geology of North America, v. C-2.)

The original major geological time divisions of the rock sequence were established in Europe and equivalents correlated to North America in the nineteenth Century (Figure 5). It was soon learned, however, that the classic classification of rock units did not always coincide with the major divisions on the craton; these divisions, separated by major gaps in the record, became known as sequences. Sequences (*see* **Sequence Stratigraphy**) are major cycles in the Earth's history recording events on and adjacent to the craton. In a simplistic view, they are records of a significant advance and retreat of the sea lapping on the edge of the craton, ideally starting with a sequence of nonmarine clastics, followed by marine sediments,

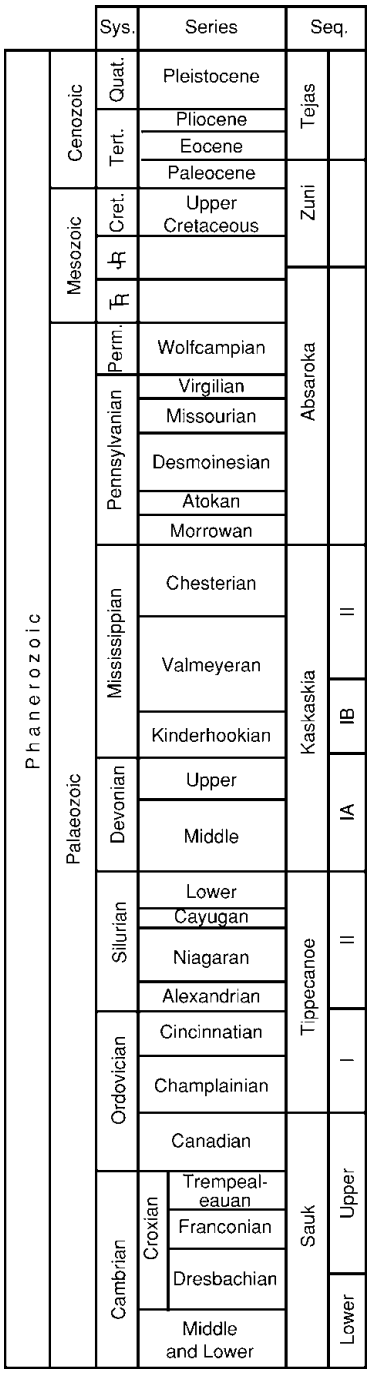


Figure 5 Correlation chart showing relation of sequences and subsequences to standard chronostratigraphic scheme for Phanerozoic. (From Sloss (1988) GSA, The Geology of North America, v. D-2.)

culminating with evaporites and clastics, and ending with a hiatus (gap), then followed by another cycle (Figure 6). All or parts of the sequence may or may not be present at any one location, depending on local conditions.

The craton was never very high above sea-level, a few tens or a couple of hundred metres at most, so that slight changes in sea-level brought inundations or withdrawals of the sea over wide expanses. The seas that spread over the craton, thus, were never very deep, and as the craton gradually foundered, the sediments accumulated. In concert with the rise and fall of sea-level, tectonic forces were busy forming and shaping mountain belts around the edges of the craton. On the craton, the tectonic forces were transmitted through the rigid basement rocks, causing a differential vertical movement of the fractured basement blocks as they readjusted to the stress. A series of uplifts and basins formed as a result of these structural movements and the sediments responded accordingly, filling in the low areas and either not being deposited on the high areas, or deposited and then stripped off. Most of the basins are shallow but some on the margins, such as the Anadarko Basin of Oklahoma and Denver Basin of Colorado, are deeper structural troughs with corresponding thicker sections of sedimentary rocks.

The Phanerozoic Geological Record

As the seas advanced over the low-lying craton, sediments accumulated and thus preserved a record of the history of that episode. The first Phanerozoic sediments on the craton were detrital material derived from erosion of the craton itself. In fact, the exposed part of the craton contributed clastics to the sequence during each episode. The Sauk sequence (latest Proterozoic to Early Ordovician) saw a major inundation as the seas advanced over the craton (Figure 7). The Tippecanoe sequence (Middle Ordovician to Early Devonian) was less widespread and confined mostly to the eastern part of the craton and distribution of the Kaskaskia sequence was similar. However, the early part of the Absaroka sequence (latest Mississippian to Early Permian) was more extensive, and represents a calumination of events, as all but the most northern part of the craton was affected.

Near the end of the Palaeozoic, major changes in the sedimentation and structural history (*see* **North America:** Southern Cordillera) of the central part of North America took place as the result of the Ancestral Rocky Mountain, Ouachita, and Alleghany/Appalachian orogenies (*see* **North America:** Southern and Central Appalachians). The next major inundation of the continent took place in the Zuni sequence, when Cretaceous seas covered much of the craton. Again, major changes in the structural history near the end of the Mesozoic, with the rise of the Rocky Mountains to the west, mark the end to

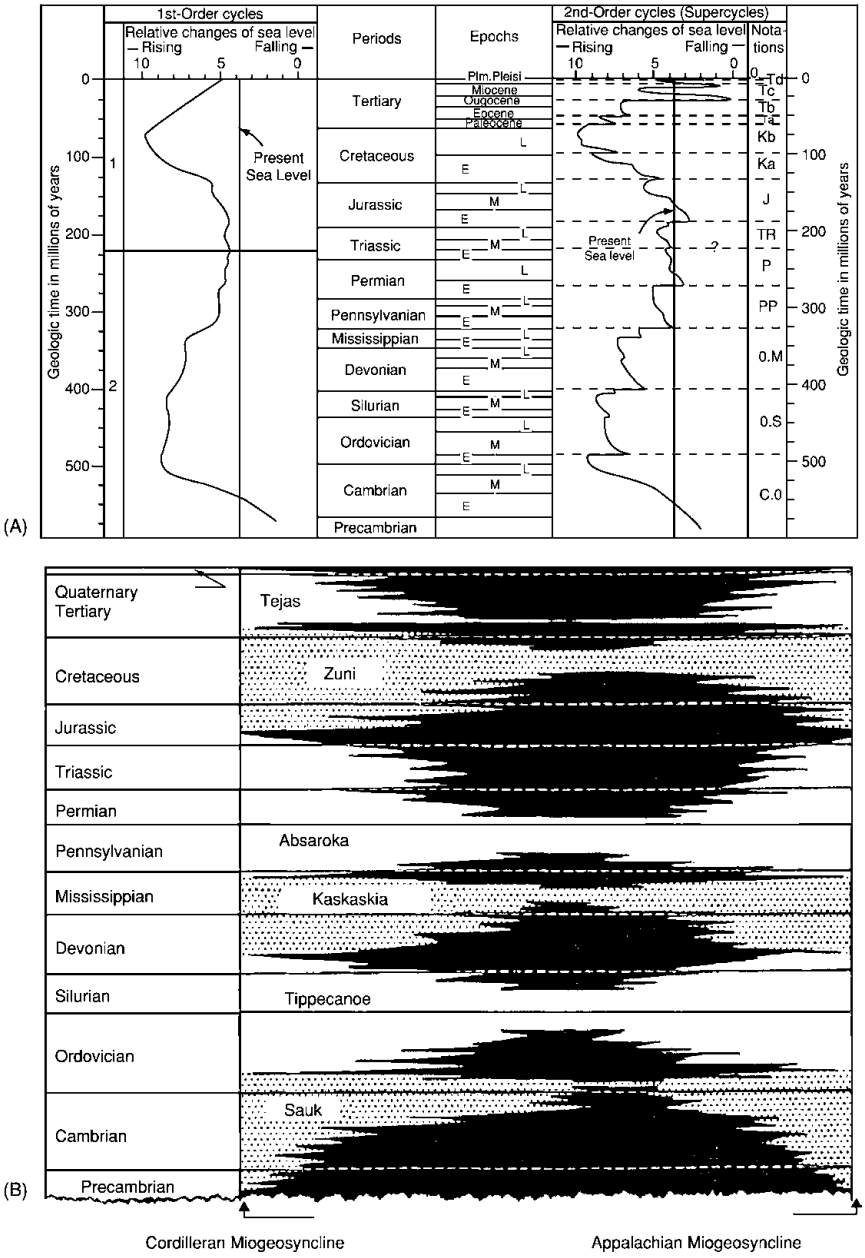


Figure 6 (A) First- and second-order global cycles of relative change of sea-level during Phanerozoic time. (Adapted from (1989) GSA, The Geology of North America, v. A.) (B) Sloss' (1963) diagram of relation of sequences to craton. Black areas are gaps (hiatuses) and white and stippled areas represent deposition of sediments during successive depositional episodes. Note the stratigraphic section is more complete off the craton.

the sea incursions on the continent, so that the Tejas sequence is represented only by terrigenous sediments – material derived from the erosion of the newly-formed Rocky Mountains. This is now the situation remains today – the interior of the continent is high and dry. The end of each episode was marked by withdrawal of the seas and prolific erosion eradicating parts or all of the recorded history of the immediate and preceding chapters.

Several episodes of igneous intrusions into the sediments in the southern part of the Continental Interior took place during the later Cretaceous and Early Tertiary. These small igneous bodies were intruded from the mantle at great depths rapidly along zones of weakness in the continental crust into the sedimentary section. They are of interest, not only for their composition, time of emplacement, and implication for the tectonics, but because some of them, on

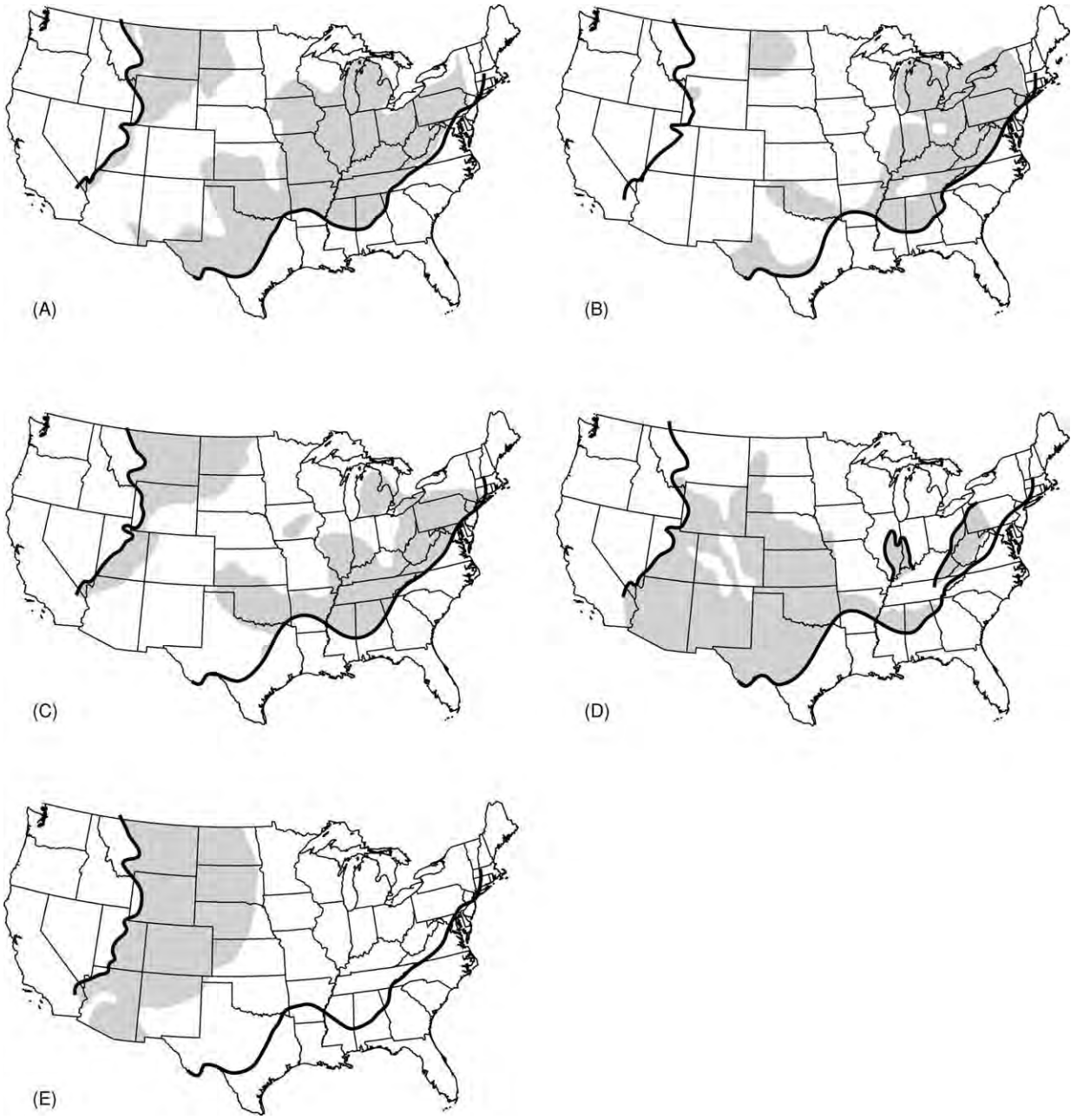


Figure 7 Location of sedimentary sequences preserved on North American craton by sequence. (Modified from Sloss (1988) GSA, The Geology of North America, v. D-2.) (A) Sauk sequence; (B) Tippecanoe sequence; (C) Kaskaskia sequence; (D) Absaroka sequence; and (E) Zuni sequence.

their way up from the depths, incorporated bits and pieces of mantle, continental crust, and Phanerozoic sediments.

The beginning of the last chapter of the story is concerned with the large continental glaciers that covered much of the north-eastern and east-central part of North America (*see* **Tertiary To Present:** Pleistocene and The Ice Age). Numerous advances and retreats of the glaciers are recorded in the remains they left behind – glacial debris (outwash and till; Figure 8A and B) and large erratics (out-of-place foreign boulders; Figure 8C) with thick deposits of loess (wind-blown glacial dust; Figure 8D). In places this forms a cloak over the area, the icing on the layer cake, so to speak.

The glaciers had disrupted the inplace drainage from the end of the Tertiary so that with their retreat, the drainage reoriented to its present

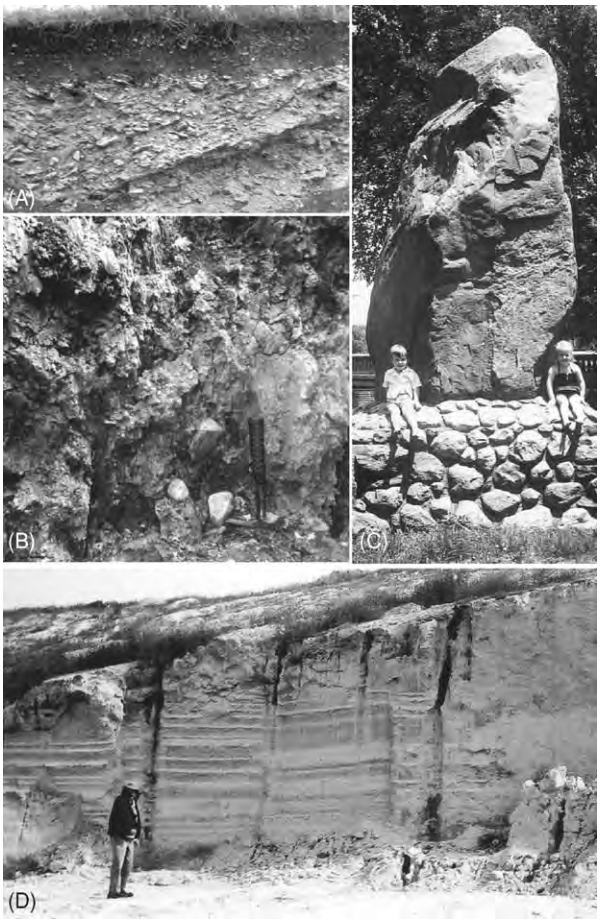


Figure 8 Glacial debris. (A) Glacial outwash deposit, which covers extensive areas of the Continental Interior; (B) glacial gumbo till deposited directly from the glacier; (C) large glacial erratic in Lawrence, Kansas; (D) thick loess deposit (glacial wind-blown dust) in central Nebraska.

position. Drainage south of the southern extent of the glaciers was profoundly affected with overloads of water and material, and this latest episode of history is evidenced by the underfit streams, stream piracy, and extensive stream and river terrace development. The final adjustments are taking place today as the Continental Interior returns to 'normal'.

The Sediments

The succession of stratigraphic units on the craton could be and have been described as monotonous, but they are anything but boring. Special features of these unique sedimentary units are their; (i) lateral persistence; (ii) extreme thinness, the tremendous ratio of width to thickness; (iii) sharpness of boundaries; and (iv) cyclic nature.

The contained micro- and macro-fauna and flora (or lack thereof) in these sediments not only gives information on conditions under which the sediments formed, but provides information on relative age for correlation from one area to another, where the sediments are not continuous. Understanding life as it occurred in ancient times is based mainly on understanding the occurrence and distribution of similar forms today.

In addition to the usual sedimentary types, such as sandstone, siltstone, shale, limestone, dolomite, anhydrite, and gypsum, that occur in the Phanerozic sequence, there are four rock types that are formed under special circumstances: chalk, salt, black shale, and coal. Chalk is made up of tiny micro-organisms that accumulated in the open sea in warm water (**Figure 9A and B**). Major chalk deposits throughout the Earth are prominent in the Cretaceous (Zuni) and the Kansas chalk is famous for its marine vertebrate fauna. Salt is formed when and where marine waters are restricted and evaporate (**Figure 9C**). Salt is used here as an all-inclusive term for minerals precipated from the evaporation of sea-water. The Silurian (Tippecanoe) and Permian (Absaroka) are two of those times in Earth history when vast areas of salt were formed on the craton under these conditions.

The black shale accumulates under confined conditions at the bottom of the sea, where because of lack of water circulation and reducing conditions, all forms of carbonate and other soluble material are dissolved (**Figure 9D**). Radioactive material is concentrated in the black shales, making them distinctive subsurface marker beds on gamma-ray wireline logs and easy to correlate. On the North American craton, black shales are prevalent in the Late Devonian to Early Mississippian (Kaskaskia) and numerous thin but persistent ones in the Middle and Upper Pennsylvanian and Lower Permian (Absaroka). Lignite and coal form near sea-level and are thus a good palaeoenvironment indicator (**Figure 10A and B**). In the Interior Province, several basins contain bituminous coal, mostly Pennsylvanian in age, whereas in the Northern Great Plains it is mostly sub-bituminous coal or lignite, and Cretaceous or Tertiary in age. The plant material grew in the shallow water along the shore, dying in place, accumulated, and was buried, eventially forming peat. As the peat was buried, the material compacted to coal and in doing so created a handy indicator to determine the depth of burial. However, it never has been buried deep enough in the Continental Interior, or subjected to tectonic forces, to create the highest grade of coal, anthracite.

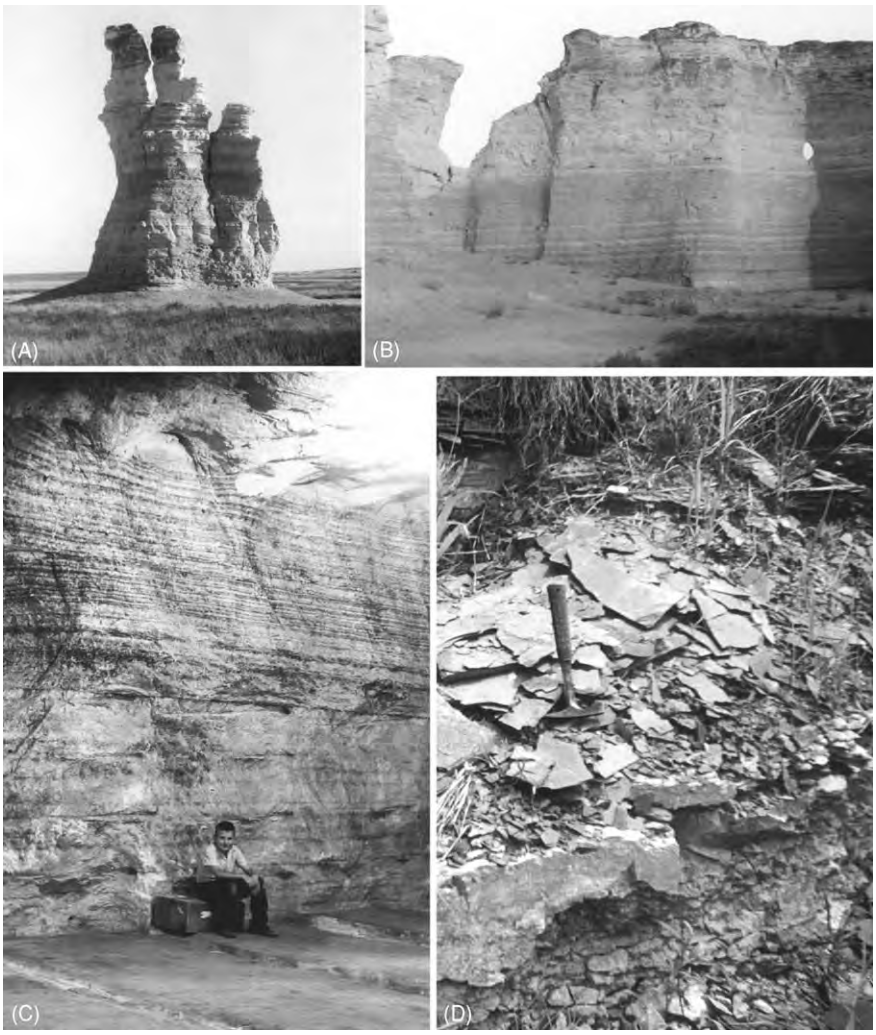


Figure 9 Examples of special rock types in the Continental Interior. (A) Cretaceous chalk in western Kansas; (B same exhibiting rhythmic layering of chalk and bentonite; (C) Permian salt in mine in central Kansas (note: rhythmic layering); and (D) Pennsylvanian thin-bedded, fissile black shale.

Several other less widely distributed rock types are distinctive and informative: volcanic ash (Figure 10C) and its altered form bentonite, fossil soils (palaeosols), and extensive thin-layered chert beds (Figure 10D). Volcanic ash is recognised easily in the Tertiary and its altered form of bentonite in the Cretaceous.

The bentonites occur in distinctive thin bands over large areas and form instant time-lines. Multiple fossil soils are widespread in the Upper Permian and indicators of subareal exposure and climatic conditions at the time. In the Lower Permian, thin but persistent nodular chert beds occur and individual beds can be traced for many miles, indicating uniform conditions in the sea over a wide area. And lastly, the mantle of Pleistocene glacial material (either deposited from the glaciers or derived from their action) is an indicator of climatic conditions. These masses of thick, heavy ice deformed the crust and modified the landscape extensively.

Accumulation time of sediments can be estimated by applying known sediment accumulation rates of today to their thickness or using the relation of total sediment thickness to the known time period. Because of the limiting factors in both methods, the results are only an indication.

Erosion, usually by water and wind, and the rates at which it takes place, is of interest; generally the rates are higher in young terrains and less in the older ones, such as the Continental Interior. The obvious exception to the erosion rate, of course, is in those areas that are farmed and thus subject to the plough and other forms of tillage that break the surface and hasten the destruction of the land. Erosion rates generally are steady, but accelerated in times of crisis such

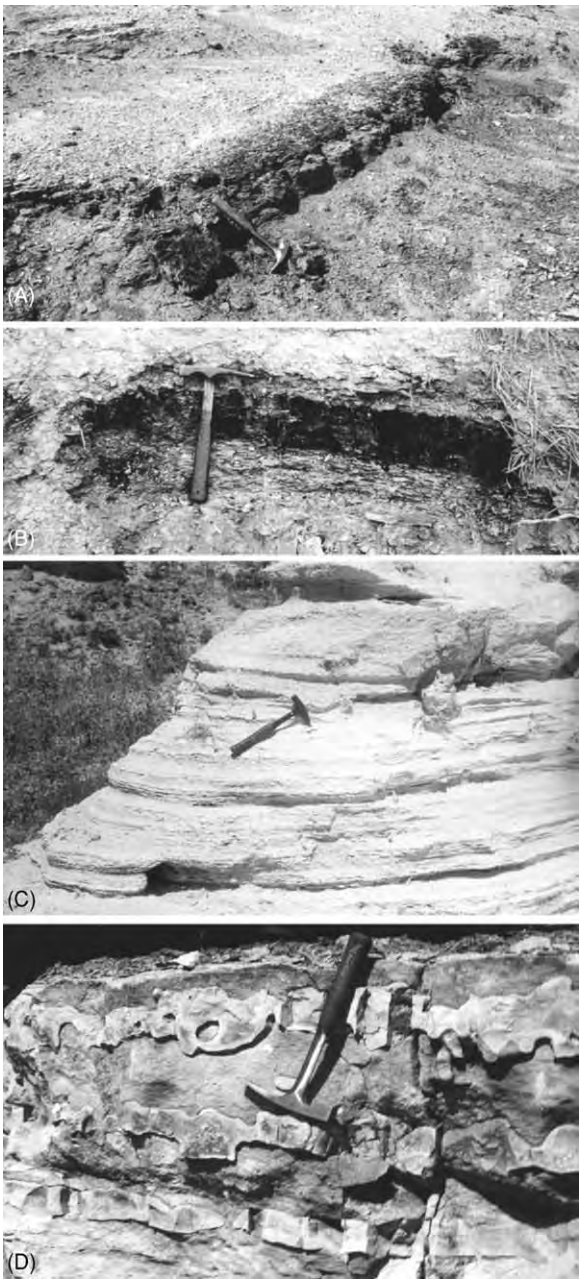


Figure 10 Other distinctive rock types that occur in the Continental Interior. (A) Lignite from Cretaceous of central Kansas; (B) thin bituminous coal (Upper Pennsylvanian); (C) Pliocene volcanic ash, a distinctive geochemical signature makes it an excellent marker bed for correlation; (D) layered chert in limestone in the Flint Hills of central Kansas.

as floods or devastating wind storms. More 'normal' rates generally prevail in between catastrophic events.

Rates of denudation can be determined by measuring the amount of material carried by streams and rivers in a unit of time. This usually is reported as so many centimetres reduction in the land surface per millennium. Because of the relatively high elevation of the North American Continent today, these rates are high and probably are not representative of the past. Another approach to denudation rates is to determine how much material was present and now evoded away and to divide that amount by the time interval. Neither approach is totally satisfactory, but both give a fair estimate.

The Sedimentary Sequence

With special conditions, unique rock sequences occur. The North American Craton is one place where these special sequences – rhythmites and cyclothems – are preserved. Rhythmites are alternations of two sediment types, such as the chalk and bentonite that developed in the Cretaceous (Zuni), or siltstone and the shales that are present in the Pennsylvanian (Absaroka). Cyclothems are repetitious sequences that occur repeatedly in the Permo-Pennsylvanian (Absaroka). The classic cyclothem records a single advance and retreat of the sea in an area, giving rise to a symmetrical rock cycle of non-marine to marine to non-marine sediments; however, the ideal seldom occurs. Because the transgressive part of the cycle usually is better represented (or better preserved) than the regressive part, most cycles are asymmetrical and many are incomplete or have components repeated.

Bundles of cyclothems have been termed megacyclothems and an example is given in Figure 11. Origin of cyclothems has been proposed as being the result of: (i) change of sea-level; (ii) tectonic movements; or (iii) climatic factors; the truth is that all three are probably responsible to some degree. In Kansas alone, in just the Middle and Upper Pennsylvanian (Absaroka) section, there are parts or all of about 40 cyclothems. Their formation probably was the result of changes in sea-level as a result of the waxing and waning of southern-hemisphere glaciers. Sediments were accommodated spatially by the craton gradually sinking in relation to sea-level, perhaps partially as a result of the weight of the added sediments and the glaciation was caused by changes in the climate.

Structural Development

The development of the structural elements on the craton to their present disposition has not been spectacular, but has been incremental and persistent. The movement is recorded, not only by the sediments, but by the gaps in the record – the unconformities. Much of the regional structural development on the craton has been by epeirogenesis; that is, up-and-down movement and tilting of the area in response to orogenic (tectonic) activity elsewhere. Four major unconformities indicate four major changes in the structural regimen – the end

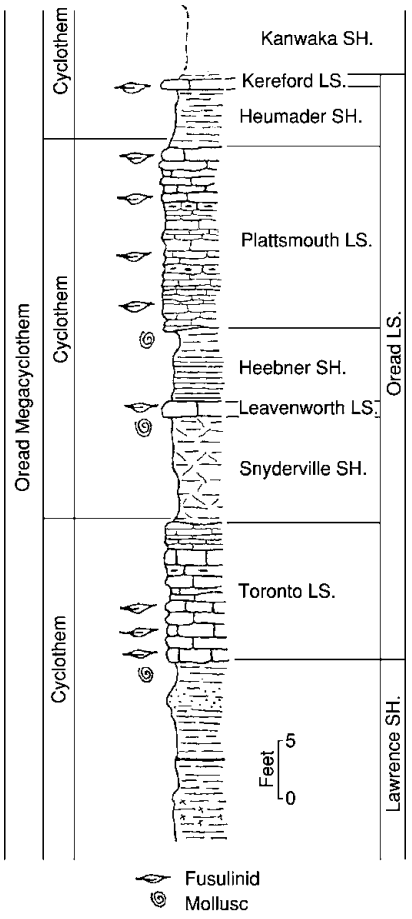


Figure 11 Example of Pennsylvanian megacyclothem from Kansas, showing individual cyclothems in the Oread Limestone Formation (Virgilian, Pennsylvanian). (Adapted from Merriam (1963) Kansas Geological Survey Bulletin 162.)

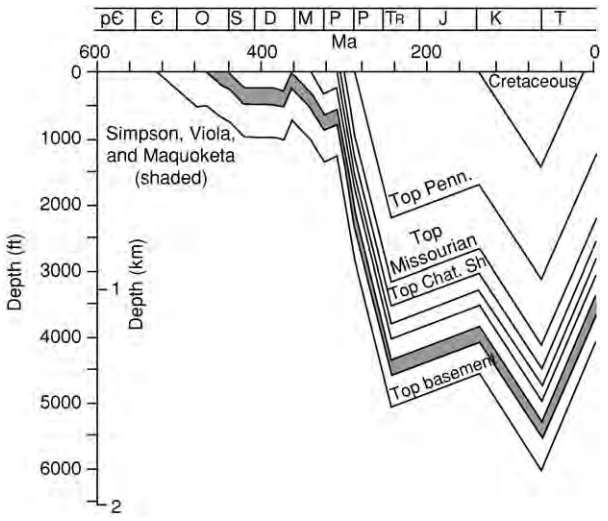


Figure 12 Burial-history diagram of the Salina Basin (central Kansas; for location see Figure 13. Diagram shows amount of sediment deposited and amount removed giving depth of burial of sediments. (From Newell and Hatch (1997) Kansas Geological Survey Bulletin 240.) The diagram illustrates that the basin subsided and received sediments from Late Cambrian to end of the Devonian when it was uplifted. Major subsidence commenced in the Mississippian, with minor uplift near end of the Mississippian and continued from then, with minor ups and downs until the present.

Table 1 Different procedures used to determine overburden removed. (Adapted from Merriam and Förster (2001) Kansas Academy of Science Program and Abstracts.)

- compaction-depth relations
- vitrinite reflectance
- oil field overpressure
- maturation models
- thermal modeling
- mineralogy and rock textures of intrusives
- fluid inclusions
- geologic restoration

of the Precambrian, the Ouachita Orogeny (Mississippian), the Allegheny or Appalachian Orogeny (Late Palaeozoic), and the Laramide Orogeny (Cretaceous).

An important part of the story on the structural development of the Continental Interior is knowing how deep and when the sediments were buried. The ups and downs of the craton at any location is profiled by constructing a two-dimensional burial-history diagram (Figure 12). Another question is how much material was removed at an unconformity. There are numerous ways in which this can be done, but a combination of several approaches for any area is the most satisfactory (Table 1). This information then can be used together with the sediment distribution to understand the history.

Cratonic Structures

The sediments are preserved more completely in the basins than on the adjoining uplifts, as would be expected. The major basins, uplifts, and structural discontinuities in the Continental Interior are shown in Figure 13. A series of cross-sections of the Continental Interior are shown in Figure 14. These cross-sections give the best impression of the subsurface layer-cake nature and the relation of stratigraphic units to one another. Most impressive is the thinness of the sedimentary cover (when viewed in perspective (Figure 15); indeed, the sediments are but a facade over the Precambrian foundation. The structures shown in Figure 13 are those that formed near the end of the Mississippian and in the Early Pennsylvanian during the Alleghany/Appalachian orogenies and since have been slightly modified.

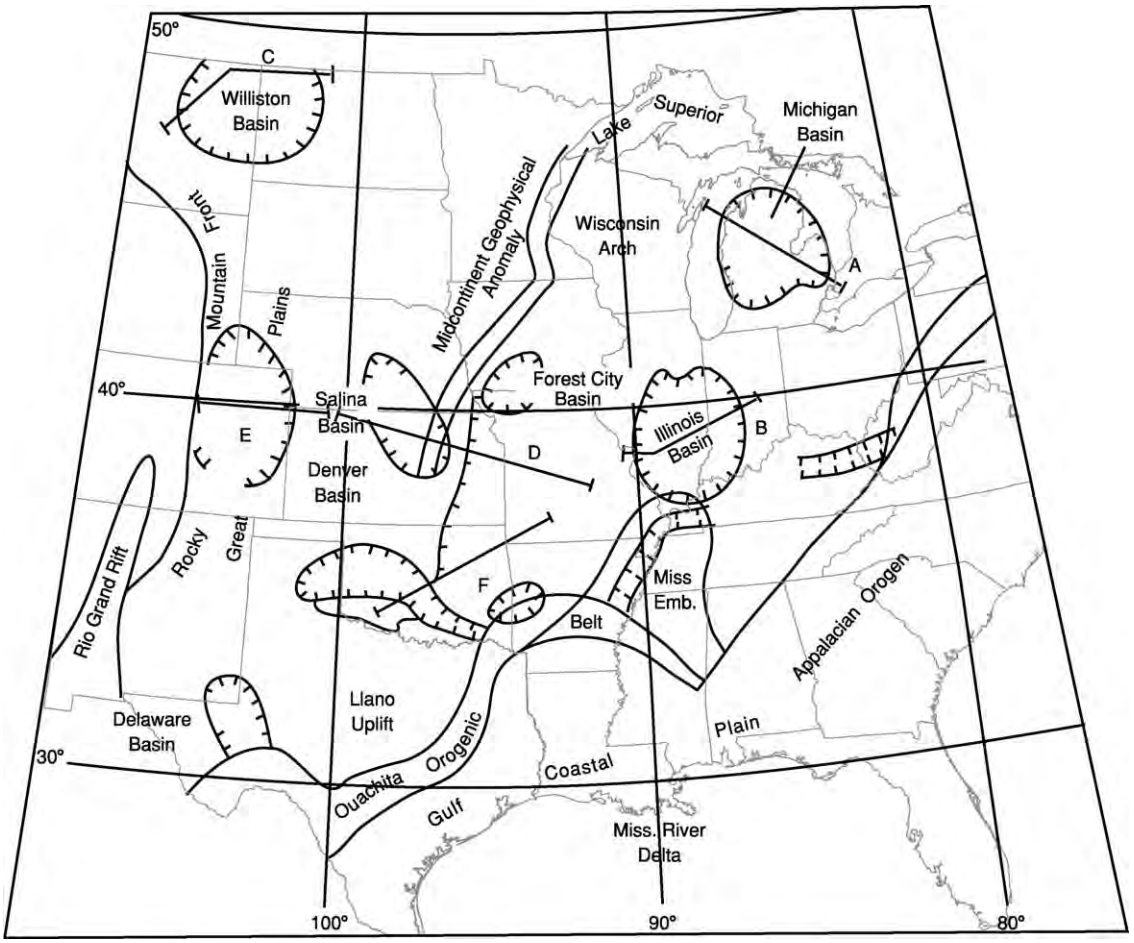


Figure 13 Location of major structural features in the Continental Interior. A, New York-Alabama Lineament; B, Ouachita Foldbelt; C, Wichita Uplift; D, Anadarko Basin; E, Ardmore Basin; F, Aricoma Basin; G, Nemaha Uplift. (From Hinze and Braile (1988).)

There is no apparent spatial pattern of the basins; they seemingly formed wherever there was a weakness in the foundation. Through time, several of these regional structures were inverted, (that is, basins were uplifted to become positive areas, or vice versa), and some positive features foundered to form depressions.

The Precambrian basement also is the key to the location of the minor structures of the region. These secondary features show a preferential orientation: one set north-east–south-west and the other north-west–south-east, with a ancillary set north–south and east–west. These structures were formed by the vertical displacement of basement fault blocks in response to outside tectonic forces in which the overlying sediments were draped over the offset blocks by differential compaction of the unconsolidated, soft sediments. These anticlines (plains-type folds) are present in the sedimentary cover on cratons worldwide (Table 2; Figure 16). After initial formation of the structural regimen in Late Mississippian to Early Pennsylvanian time (Absaroka), continued movement of the basement blocks accentuated the folds. This incremental movement has continued from the Late Palaeozoic until today.

The incremental movement of the fault blocks is recorded in the Permo-Pennsylvanian (Absaroka) sedimentary section by thinning of beds over the structure and by convolute features known as seismites (Figure 17). The seismites, formed by dewatering of soft sediments, were triggered by palaeoearthquakes. By noting the geographic and stratigraphic occurrence of the seismites, it is possible to determine the relative intensity and size of the area affected by the palaeoquake. Earthquakes are the result of the adjustment of the basement to stress and are recorded on the craton today. In historic times, several large earthquakes occurred – the New Madrid quake (1811–1812) in Missouri and Manhattan quake (1867) in Kansas are two examples; fortunately there was little to damage at those early dates.

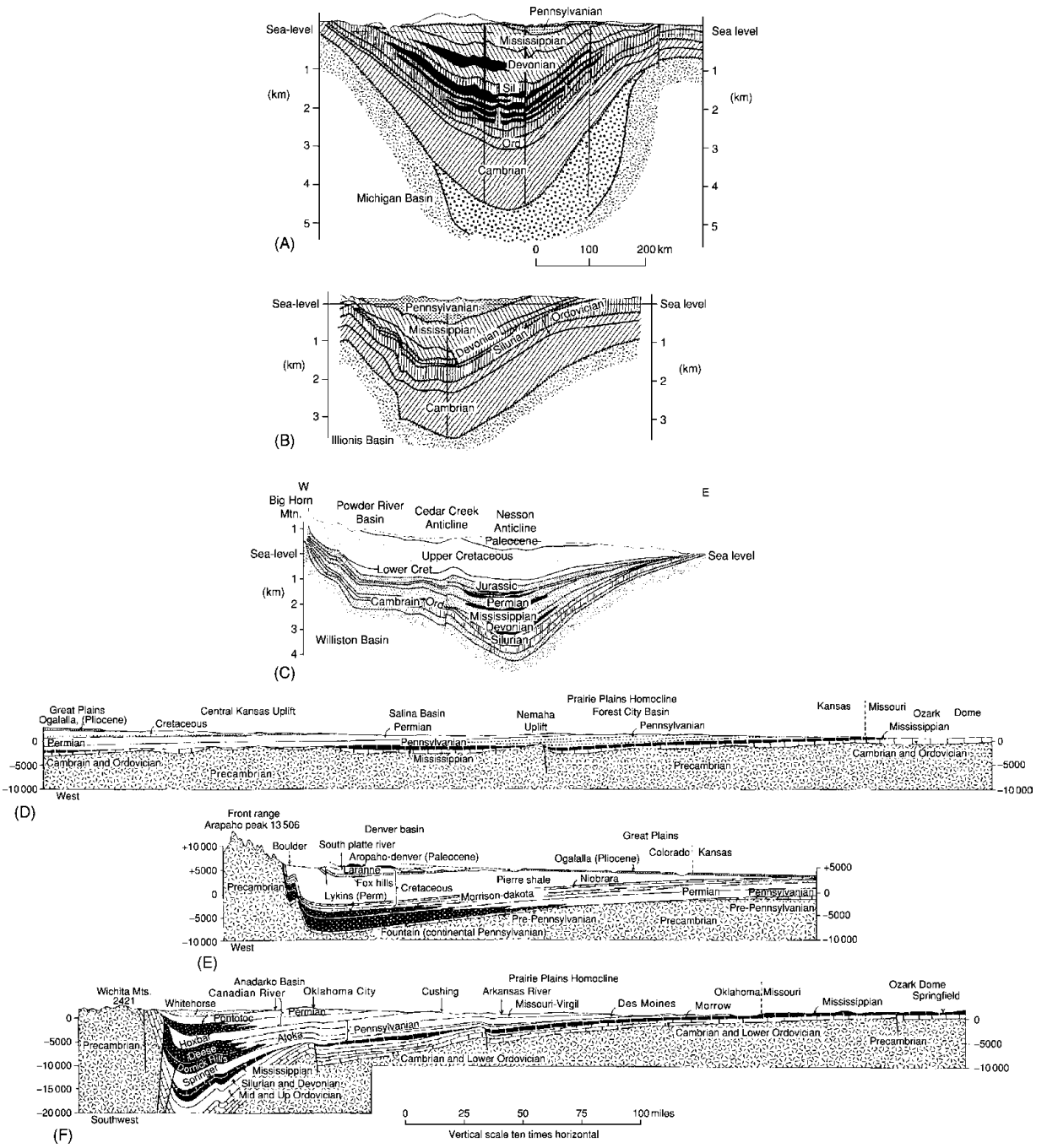


Figure 14 Cross-sections of selected areas in the Continental Interior. A Michigan Basin; B Illinois Basin; C Williston Basin; D Ozark Dome west to Great Plains; E Great Plains through Denver Basin to Rocky Mountains; F Ozark Dome south-west through the Anadarko Basin to Wichita Mountains. Denver and Anadarko basins are two of the deepest troughs in the Continental Interior. (A, B, and C from Bally (1989) Phanerozoic basins of North America, *in* The Geology of North America, v. A; E, F, and G from King (1951) The tectonics of middle North America: Princeton University Press.)

Mineral Resources

The Continental Interior has long been known as a rich source of metallic minerals, mostly produced from Precambrian rocks. Gold has been produced from the Precambrian at Lead, South Dakota and the Porcupine and Kirkland Lake districts in Ontario, Canada. The extensive Lake Superior (Michigan) copper deposits occur in Precambrian rocks and have been known for

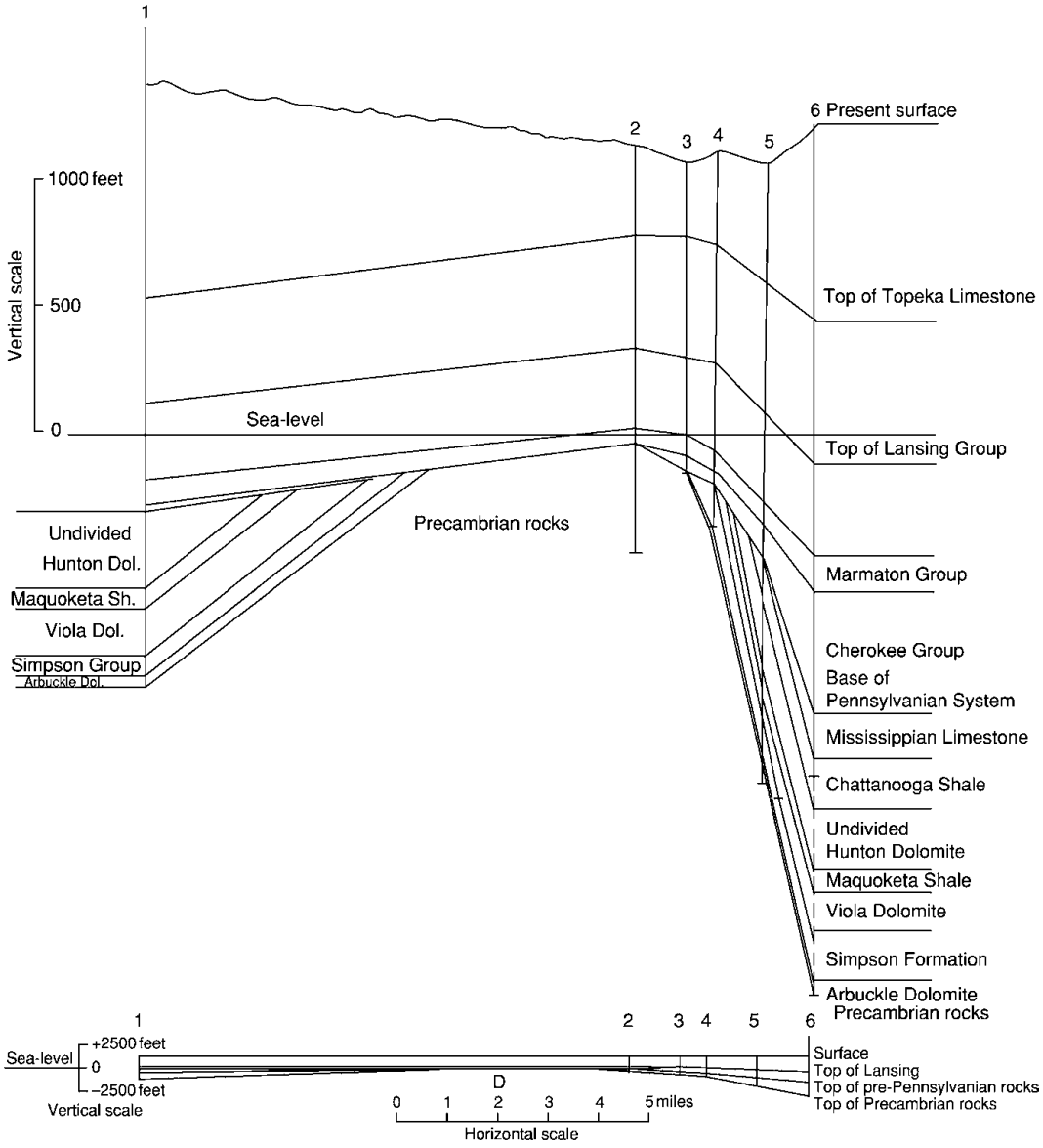


Figure 15 Cross-section of the Nemaha Anticline, a major structural feature showing the result of exaggeration (upper) to at scale (lower). Exaggeration is used by geologists to show relationships better in flat-lying stratigraphic units (from Lee, 1956, Kansas Geological Survey Bulletin 121).

Table 2 Feature of plains-type folds. (Modified from Powers (1925) Geological Society of America Bulletin, v. 36.)

- Structural noses are dominant structure at shallow depth (ie in surface units)
- Most folds are small in areal extent and amount of structural closure
- Closure increases downward and structure becomes sharply defined
- Amount of closure is proportional to size in many structures

centuries and were utilized by the Indians. Copper was also produced at Ducktown, Tennessee from Cambrian metamophosed sediments. The occurrences of lead and zinc deposits in south-eastern Missouri were known from the eighteenth century. The Tri-State District centred on Joplin, Missouri, described as the greatest zinc district on Earth, occurs in Mississippian carbonates. The rich Precambrian iron-ore deposits of the Mesabi Range in Minnesota and in the adjacent states of Wisconsin, and Michigan have been utilised for a century. Sudbury, Ontario, Canada is a major world source of nickel that occurs in a Precambrian complex.

Extensive sub-bituminous and bituminous coal deposits occur in the Midcontinent area and is mined both underground and in strip pits where shallow. Petroleum seeps in Miami County, Kansas were used by the Indians for medicinal purposes, and pioneers

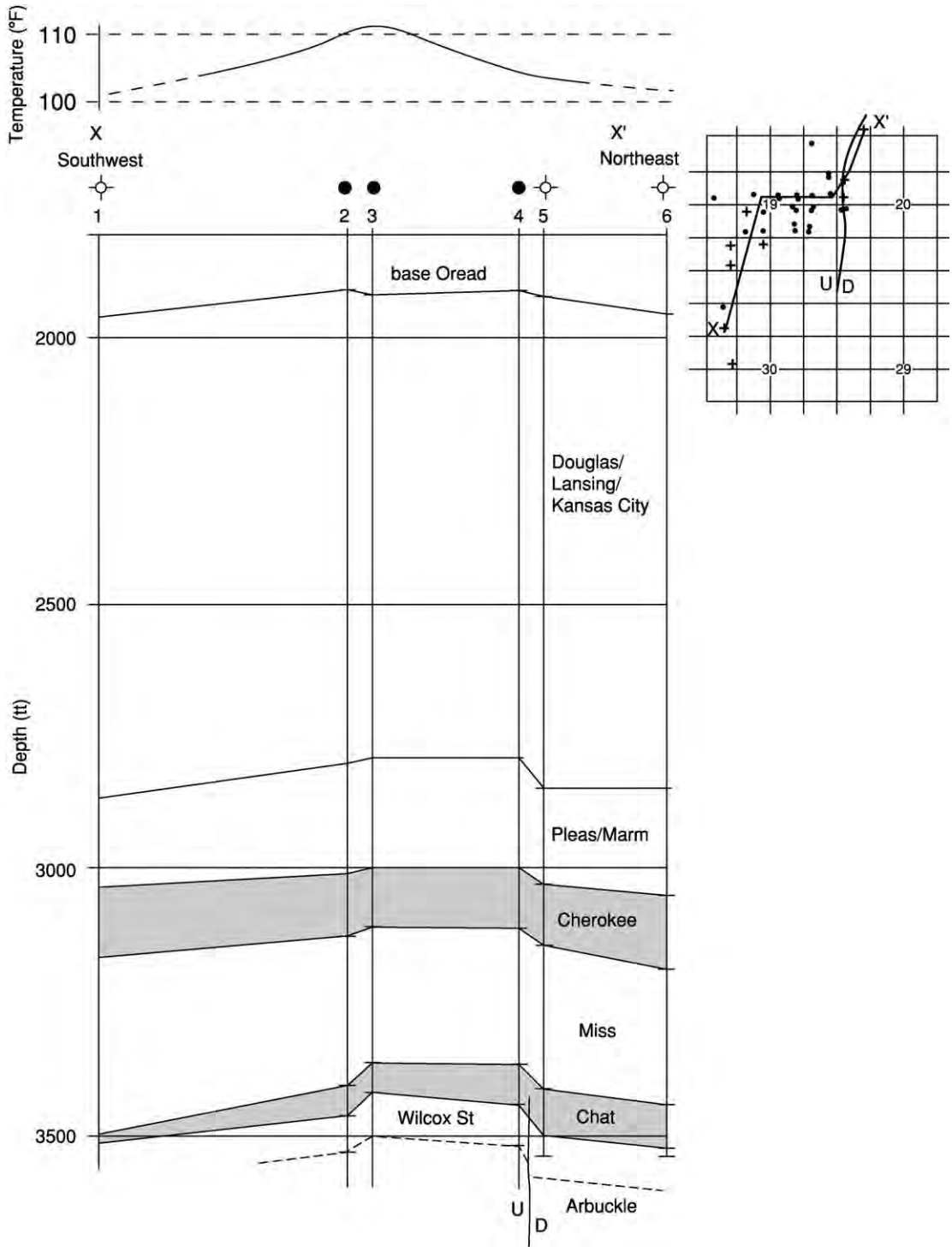


Figure 16 Cross-section of typical 'plains-type fold' (Slick-Carson oil field, T32S, R3E, Cowley County, Kansas). Note the anticline closure increases with depth, faulted on east flank, and temperature high over crest of anticline; all characteristics of oil-producing plains-type folds.

on the Santa Fe and Oregon trails on their way west used it for lubricating their wagon wheels. Petroleum was discovered in a well drilled in 1860 near Paola, Kansas, opening the large Midcontinent Oil and Gas Province shortly after the Drake well was drilled in 1859 in Pennsylvania. Natural gas was exploited for town lighting, starting in about 1870.

Gold discoveries in the Black Hills of South Dakota and in the Rocky Mountains in Colorado spurred the gold rushes, with thousands of get-rich

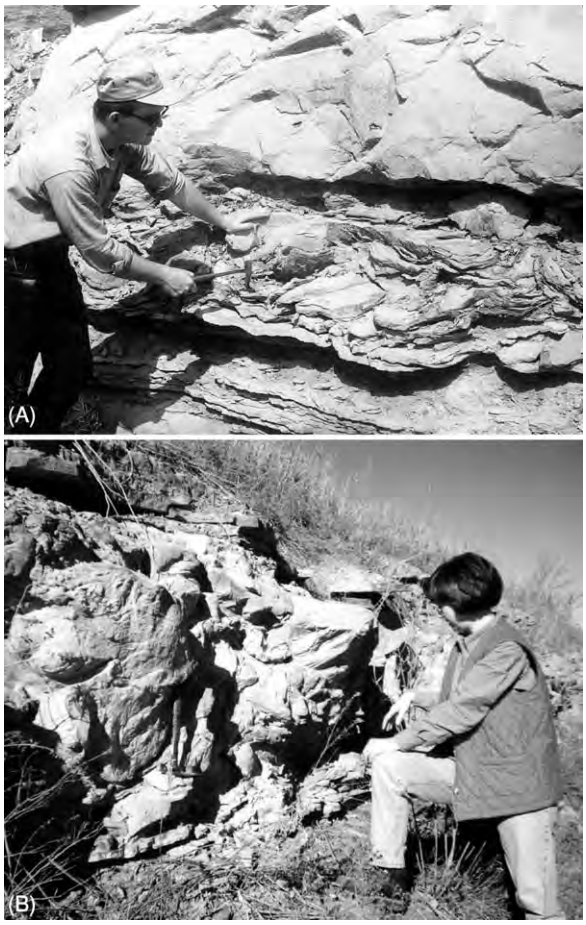


Figure 17 Sedimentary soft-sediment deformation indicating incremental movement of Precambrian fault blocks. (A) Soft-sediment deformation (seismite) in sandstone in Douglas Group (Upper Pennsylvanian) in south-eastern Kansas; (B) seismite in Upper Pennsylvanian Douglas Group clastic section in Montgomery County, south-eastern Kansas.

gold seekers making their way across the plains to the 'promised land'. As the area was populated, building material and water were needed and many other resources besides metals were exploited, including gypsum, salt, fluorite, clay, lime, glass sand, sand and gravel, building and crushed stone, and fertilizer minerals.

See Also

North America: Precambrian Continental Nucleus; Southern and Central Appalachians; Southern Cordillera. **Sequence Stratigraphy**. **Tertiary To Present:** Pleistocene and The Ice Age.

Further Reading

Frazier WJ and Schwimmer DR (1987) *Regional stratigraphy of North America*. New York: Plenum Press.

Houston RS (ed.) (1993) The Wyoming Province. In: Reed JC Jr (ed.) *Precambrian conterminous US: The Geology of North America*. Geology Society of America, vol. C-2, pp. 121–170.

Merriam DF (1963) The geologic history of Kansas. *Kansas Geological Survey Bulletin* 162: 317.

Rascoe B Jr and Hyne NJ (eds.) (1989) Petroleum geology of the Midcontinent. *Tulsa Geology Society Special Publication* 3: 162.

Sloss LL (ed.) (1988) *Sedimentary cover – North American craton: US: The Geology of North America*. Geology Society of America, v. D-2, p. 506.

Van Schmus WR and Bickford ME (eds.) (1993) Transcontinental Proterozoic provinces, In: Reed JC Jr (ed.) *Precambrian conterminous US: The Geology of North America*. Geology Society of America, v. C-2, pp. 171–334.

Northern Cordillera

J W H Monger, Geological Survey of Canada, Vancouver, BC, Canada and Simon Fraser University Burnaby, BC, Canada
R A Price, Queens University, Kingston, ON, Canada
W J Nokleberg, United States Geological Survey, Menlo Park, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The northern Cordillera extends some 3500 km north-westward from latitudes 47–48° N in Washington–Montana, USA, through Canada, and westwards across Alaska to terminate physiographically, but not geologically, at the Bering and Chukchi Seas between Alaska and the Russian Far East. Its southern boundary is delineated by the southern limits of the Olympic Mountains and North Cascade Ranges in north-western Washington, the northern Columbia Plateau in central Washington, and a system of west–north-west-trending faults in northern Idaho and western Montana called the 'Lewis and Clark line'.

The northern Cordillera follows an ocean–continent boundary created by about 540 Ma, emerged as a mountain belt between 185 and 60 Ma, and is still growing. A noteworthy feature is that about three-quarters of its surface, and all of Alaska, is underlain

by an orogenic collage composed of allochthonous rocks accreted to the ancient continental margin of western North America.

Physiography

The northern Cordillera contains three major physiographical divisions, the Rocky Mountains System, the Intermontane Plateau System, and the Pacific Mountains System, each comprising subsidiary ranges or other physiographical features (Figure 1; Appendix A). It is flanked on the east by the Interior Plains System, and on the north by a continental shelf and slope less than 200 km wide that terminates in water depths of about 2000 m in the Arctic Ocean. The Rocky Mountains System contains many summits that exceed 2000 m in elevation, with Mount Robson near latitude 53° N the highest peak at 3955 m. Between latitudes 48° and 60° N, the great valley called the Rocky Mountains Trench divides the system longitudinally, but its continuation north of this, called the Tintina Trench, extends into the Intermontane Plateau System in east-central Alaska. The Intermontane Plateau System has lower relief than flanking systems, and is mostly between 1000 and 2000 m in height, but in Alaska includes extensive lowlands under 300 m. Summit elevations in the Pacific Mountains System are mostly between 2000 and 3000 m, except for the Alaska Range in southern Alaska and Saint Elias Mountains in south-western Yukon where, respectively, Mount McKinley (or Denali) at 6193 m and Mount Logan

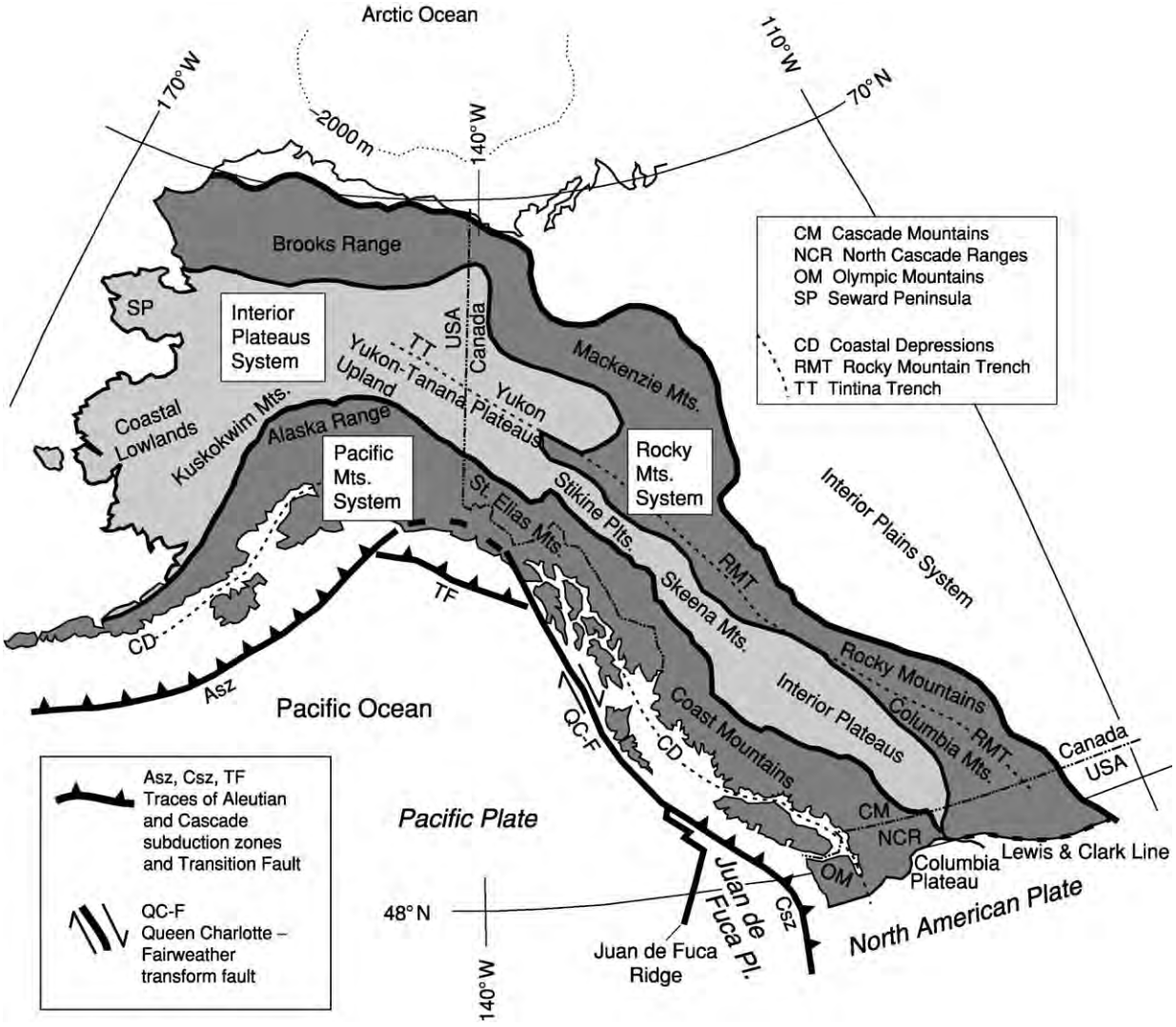


Figure 1 Major physiographical systems of the northern Cordillera, some subsidiary physiographical features, and plate boundaries.

at 6050 m are the highest peaks on the continent. The Pacific Mountains System contains a longitudinal but discontinuous coastal depression, oceanward of which are many islands and the Olympic Mountains in the south and Saint Elias Mountains in the north. The continental shelf and slope along the Pacific Ocean are about 200 km wide; the base of the continental slope is about 2000 m deep south of latitude 60° N, but westwards across the northern Gulf of Alaska its depth increases to 7000 m in the Aleutian Trench off the western Aleutian Islands. The Bering and Chukchi Seas are mostly less than 200 m deep, except for a 3000 m deep basin north of the western Aleutians.

The Cordilleran ice-sheet covered most of the region several times between 2 Ma and 12 Ka, was up to 2.5 km thick in places, and shaped local landforms. Parts of central Alaska, the eastward extension of this region into Yukon, and the Arctic coastal plain remained largely ice-free. Today, ice cover is restricted to mostly retreating glacier complexes mainly at higher elevations, but which descend to sea-level in parts of southern and south-eastern Alaska.

Neotectonics

The northern Cordillera occupies a 'soft' plate margin up to 1200 km wide landward of an offshore convergent–transform–convergent plate boundary (Figure 1). Earthquakes are concentrated near most plate boundaries, but are scattered across the entire region. South of latitude 51° N, the small, young (≤9 Ma) Juan de Fuca and Explorer oceanic plates converge with the North American Plate at rates of 2–4.6 cm year^{-1} along the currently aseismic, locked Cascadia subduction zone, 200–300 km landward of which is the active Cascade magmatic arc. Between latitudes 51° and 60° N, the plate boundary is the dextral transform Queen Charlotte–Fairweather Fault, west of which the enormous Pacific Plate moves northwards to converge with the North American Plate at rates ranging from 5.7 cm year^{-1} near the transform to 8 cm year^{-1} south of the western Aleutian Islands. The Pacific Plate descends beneath Alaska along the Aleutian subduction zone, and above it the Aleutian–Wrangell magmatic arc extends eastwards some 3800 km from the Komandorsky Islands near Kamchatka, along the Aleutian Island chain, and into the mountains of southern Alaska and south-westernmost Yukon. The North American Plate continues for some 2000 km west of the Bering Strait to terminate against the Eurasian Plate near longitude 140° E.

There is widespread evidence for neotectonic activity in the northern Cordillera in addition to earthquakes and volcanoes. Seismically imaged folds and thrust faults in Late Tertiary and Quaternary strata are in the accretionary wedges above the Cascadia and Aleutian subduction zones, and faults and folds occur in the Arctic shelf-slope region off northern Alaska and Canada. On-land surface evidence for latest Tertiary to Holocene deformation is sparse, possibly because erosion by ice eliminated the surface expression of many small-scale tectonic features. Segments of the Denali Fault in southern Alaska and south-western Yukon are active dextral strike-slip faults, as are other fault segments in central Alaska. Glaciated valleys crossing the Fairweather Fault in south-eastern Alaska are dextrally offset, and neotectonic faults occur on Vancouver Island in south-western British Columbia. In western Washington, east–west-trending Holocene folds and reverse faults in the coastal depression are related to north–south compression, as are structures involving Late Miocene lava flows in the Columbia Plateau, east of the Cascade arc.

On a larger scale, regional differential uplift of up to several kilometres in the last 5–10 million years in the Alaska Range, Saint Elias Mountains, Coast Mountains, and Olympic Mountains is documented by elevated Late Tertiary erosion surfaces, changes from wet to dry floras in regions now in rain-shadows inland of the coastal mountains, palaeodrainage patterns, and fission track studies. Differential regional uplift is not well documented elsewhere, but major reversals of drainage in upper reaches of the Fraser and Columbia Rivers in central and south-eastern British Columbia, respectively, possibly reflect Neogene differential uplift of parts of the Rocky Mountains System, in addition to documented blockage by temporary ice dams and landslides during waning stages of Cordilleran glaciation. The late differential uplift means that physiographical system boundaries only approximately coincide with major bedrock boundaries, many of which formed between the Early Jurassic and the Neogene.

Crustal Thickness

Seismic refraction and reflection profiles across the northern Cordillera show the crust to be nearly 50 km thick under the south-eastern Canadian Cordillera and parts of northern Alaska (Figure 2). Below the Intermontane Plateau System in Alaska and northern Canada, it is about 35 km thick, and under the southern Canadian Cordillera only 30 km thick. Above subducting lithosphere in southern Alaska and south-western Canada, the crust is apparently underplated by material detached from the lower plate and is over 50 km thick, whereas landward of the Queen Charlotte–Fairweather transform fault it is only about 25 km thick.

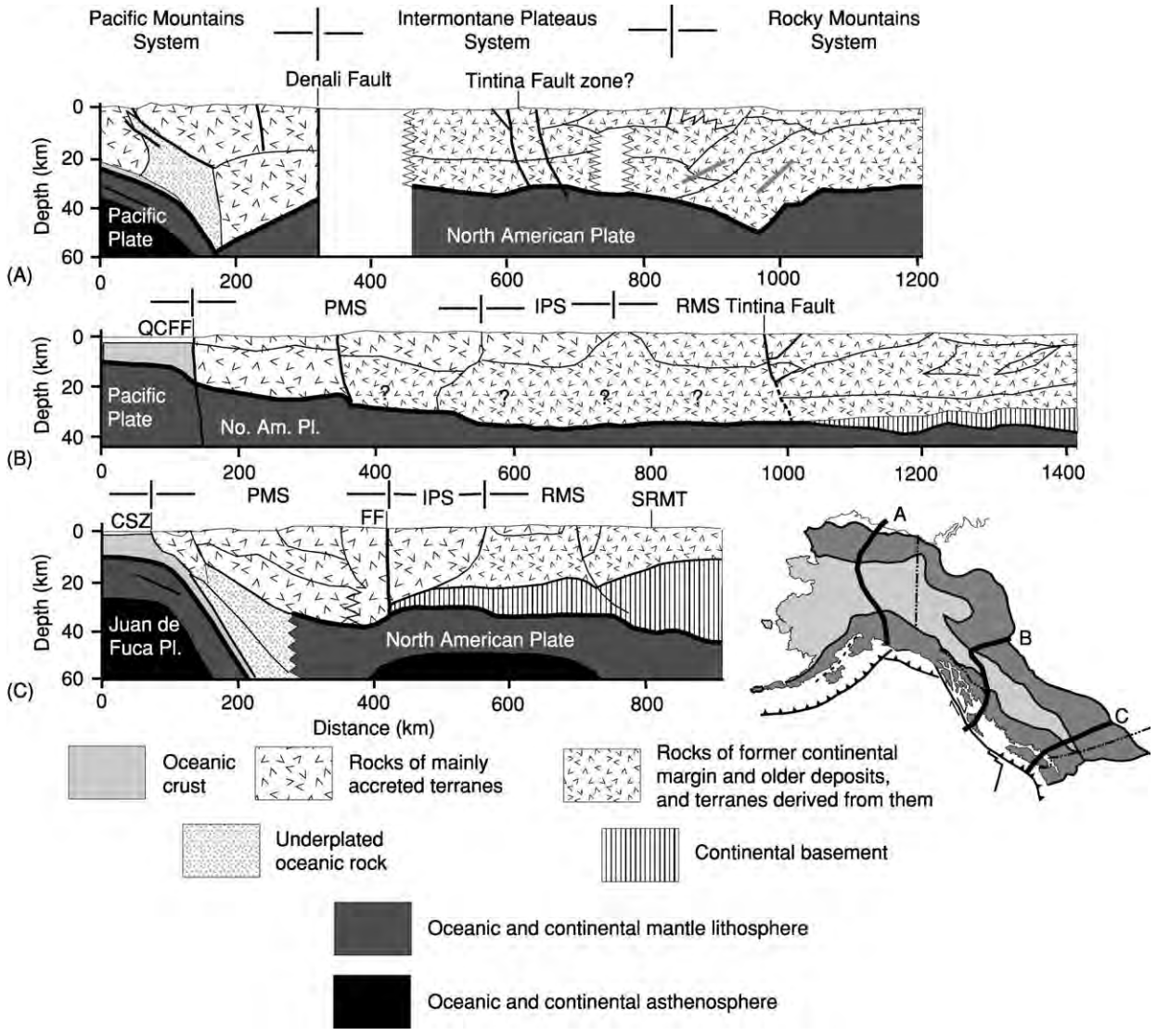


Figure 2 Crustal thickness, crustal origins, and lithospheric structures interpreted from seismic reflection and refraction profiles and surface geology across the northern Cordillera. Locations on map: (A) Trans-Alaska Crustal Transect; (B) Lithoprobe Northern Canadian Cordillera Transect; (C) Lithoprobe Southern Canadian Cordillera Transect. CSZ, Cascadia subduction zone; QCFF, Queen Charlotte–Fairweather Fault; SRMT southern Rocky Mountains Trench. Modified from Clowes RM and Hammer PTC (2000) (*see* Further Reading).

Bedrock Features

Two fundamentally different divisions of older rocks in the northern Cordillera are separated by a boundary that snakes back and forth across the physiographical boundary between the Rocky Mountains and Intermontane Plateau Systems. In most of the Rocky Mountains System, autochthonous or parautochthonous rocks were deposited on or near continental basement. Much of the remainder is an orogenic collage of allochthonous terranes, many of which show little geological or isotopic evidence of continental influence. Superimposed on the older rocks are features formed during the evolving Cordilleran mountain belt in Mesozoic and Tertiary time (Figures 3 and 4).

Autochthonous and Parautochthonous Rocks

Rocks of the ancient continental margin, and deposited on or near it, form five divisions based on age and the tectonic setting in which they formed.

1. The continental basement that extends under the eastern Cordillera from Montana–Idaho–eastern Washington to north-western Yukon is a Palaeoproterozoic orogenic collage composed of fragments of Archaean (4.0–2.6 Ga) continental crust embedded in metamorphosed remnants of

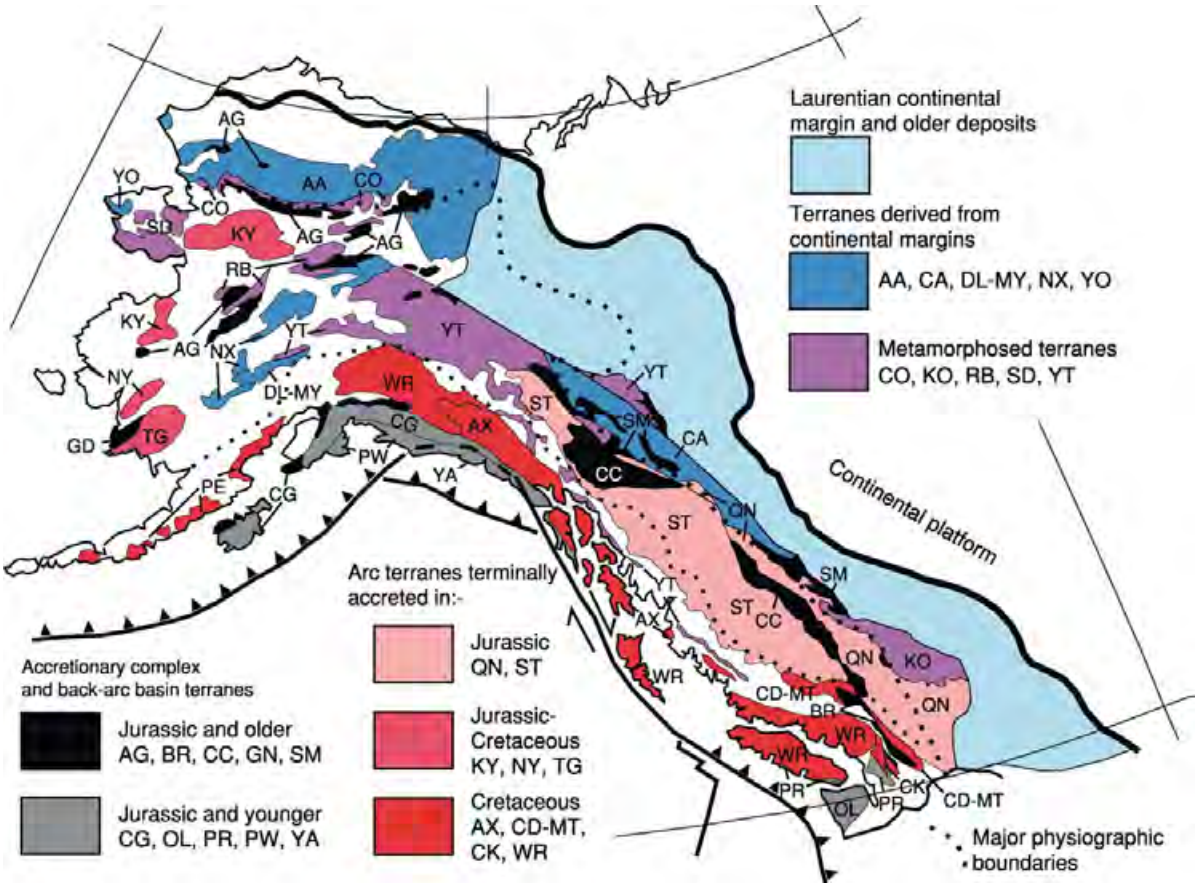


Figure 3 Distribution, origins, and times of accretion of terranes. Terrane names, abbreviated herein as AG, BR, etc., and brief terrane descriptions are in Appendix B.

Palaeoproterozoic (2.3–1.8 Ga) magmatic arcs and accretionary wedges. Although mostly buried, it is exposed locally in south-eastern British Columbia (Monashee Complex) and northern Idaho (Priest River Complex). The continental basement slopes beneath the eastern Rocky Mountains System as a layer about 35 km thick, but thins abruptly and appears to be no more than a few kilometres thick beneath the southern Intermontane Plateau System, and is absent below the Pacific Mountains System.

2. Palaeoproterozoic (1.8–1.7 Ga) basinal to platformal deposits (the Wernecke Supergroup) in northern Canada and east-central Alaska (Tindir Group) are rift-related and as much as 13 km thick. However, it is uncertain whether they formed in an intercontinental basin or during continental break-up. An angular unconformity between folded and slightly metamorphosed rocks of the Wernecke Supergroup and younger (1.4–1.0 Ga) Mesoproterozoic basinal to platformal deposits brackets the age of the Racklan Orogeny. Mesoproterozoic (1.5–1.4 Ga) strata (Belt-Purcell Supergroup) in Montana, Idaho, and adjoining parts of British Columbia and Alberta form a north-west-trending prism up to 20 km thick. The lower part consists of deep-water turbiditic quartzite, siltstone and argillite, and voluminous mafic sills; the upper part comprises shallow-water siliciclastics and stromatolitic carbonates deposited in intertidal and floodplain settings. The rocks apparently accumulated within and near an intracontinental rift basin.
3. Neoproterozoic (780–550 Ma) strata are the oldest rocks to occur more or less continuously along the entire northern Cordillera. Glaciogenic deposits (Sturtian?) at the base of the Windermere Supergroup contain mafic volcanic rocks with ages of 780–760 Ma. These date the initiation of rifting that led eventually to dispersal of components of the Neoproterozoic Rodinian supercontinent and creation of what became the margin of Laurentia. The overlying succession, up to 10 km thick, consists of shale with intercalated quartz-and-feldspar 'grit' turbidites, and grades upwards into slope deposits, carbonate, and evaporite shelf deposits.

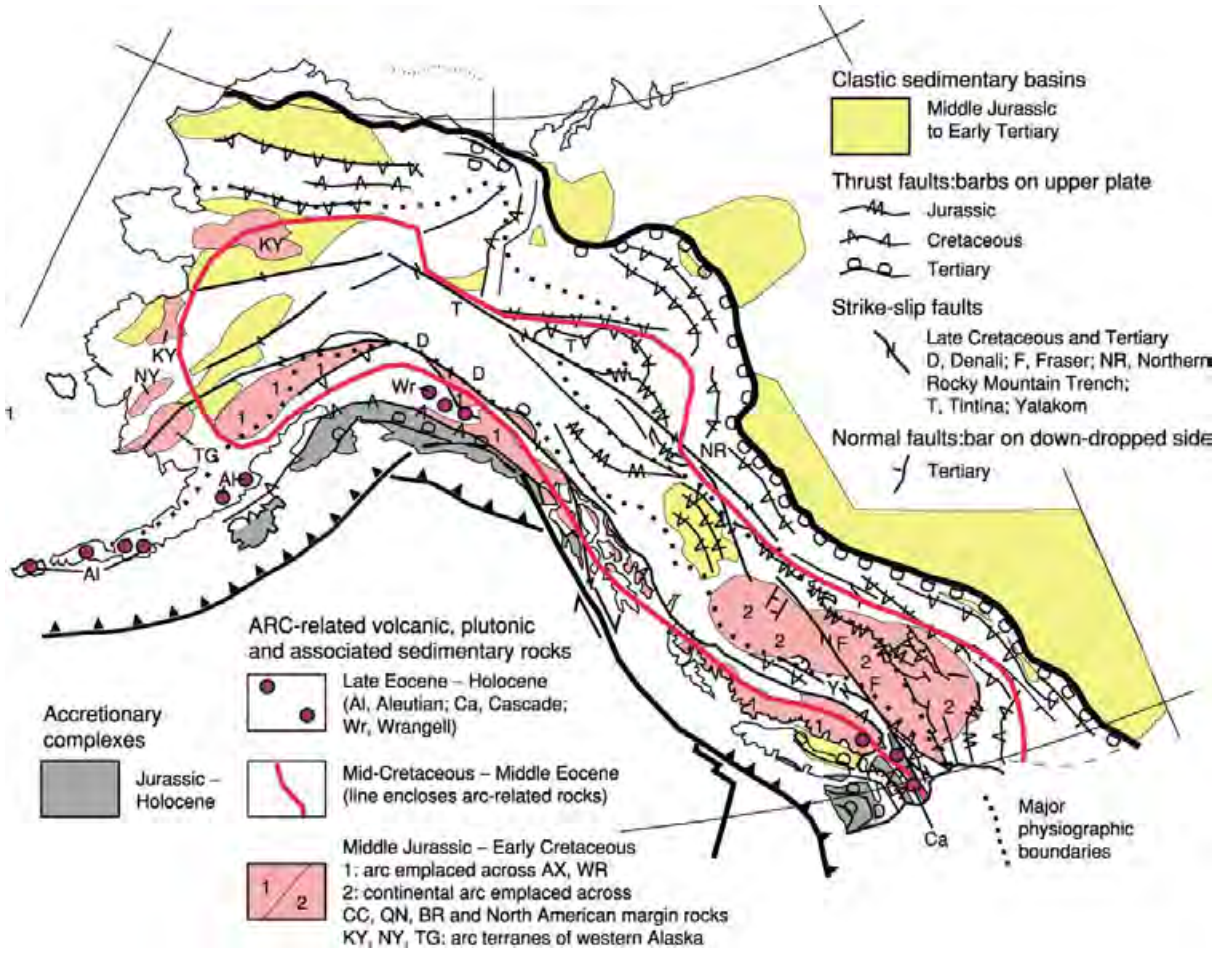


Figure 4 Features accompanying Cordilleran orogenesis: magmatic arcs and associated accretionary complexes; basins resulting from contraction, crustal thickening, uplift, and erosion; and syn- and post-accretionary structures. Terrane names, abbreviated herein as AX, WR etc; are in Appendix B.

Younger glaciogenic sediments (Varanger equivalent?) occur locally in the upper Windermere. A second major rifting episode at the top of the succession is recorded by an angular unconformity beneath Cambrian rocks and by coarse siliciclastic deposits and associated magmatic rocks dated at 550–600 Ma. Rifting culminated in complete supercontinental separation by about 540 Ma at latitudes 48–55° N, but took place earlier further north.

4. Early Cambrian to Late Jurassic (540–150 Ma) miogeoclinal strata were deposited along the newly formed continental margin of Laurentia. In the south, the continental shelf appears to have been relatively narrow and contains abundant shallow-water carbonates, but in the north it is broader and deeper water sediments are more abundant, characteristics that have been ascribed to asymmetric rifting. Cambrian to Late Devonian carbonates with locally interbedded shales exhibit lateral facies changes from incomplete platform successions about 2 km thick below the plains and in the eastern mountains, to shallow-water accumulations up to 10 km thick that were deposited on the former continental shelf. Beyond the shelf, the strata are replaced by calcareous shale and shale, and further west by deep-water carbonaceous shale with local intercalated mafic volcanic rocks and coarse-grained siliciclastic deposits. Devonian–Mississippian and Permian–Triassic siliciclastic rocks, widespread along the Laurentian continental margin from northern Alaska to north-eastern Washington, have distinctive detrital zircon and geochemical signatures indicating their derivation from northern source areas. The predominance of latest Palaeozoic and Early Mesozoic siltstone and shale over carbonate may reflect relocation of this margin to higher latitudes as a result of the northward drift and rotation of Laurentia.
5. Latest Jurassic to Palaeogene (150–40 Ma) clastic deposits were eroded from older rocks that

were uplifted and exposed above sea-level during folding and thrusting towards, and over, the edge of the continental platform, and accumulated in foreland basins along the margins of the advancing fold and thrust belt. The sediments, up to 4 km thick, consist of deltaic and alluvial sandstone, shale, conglomerate, and coal, as well as interfonguing marine sediments deposited during three transgressive–regressive cycles of latest Jurassic and Early Cretaceous, mid-Cretaceous, and Late Cretaceous and Palaeogene ages.

Rocks of the Orogenic Collage

Major components of the orogenic collage have been called terranes. These are fault-bounded regions of upper crustal rocks, detached from their lower crust and mantle lithospheres, and with different geological histories from one another. Many terranes contain palaeomagnetic, palaeobiogeographical, and/or detrital zircon signatures very different from those of neighbouring terranes and from the north-western (in present co-ordinates) Laurentian margin. Most terranes in the Pacific Mountains and Intermontane Plateau Systems south of latitude 60° N are remnants of arcs and accretionary complexes whose generally low Sr_i and positive ε_{Nd} values indicate the mantle origins of the magmatic rocks within them. Others, mainly in the western Rocky Mountains System and Intermontane Plateau System north of latitude 60° N, are displaced former continental margin deposits (**Figure 3**; Appendix B). Evidence of terrane provenance and amounts of displacement can be gathered by comparing these signatures with those in coeval rocks of the ancient continental margin, which has been oriented approximately longitudinally since the Late Palaeozoic, and thereby provides latitudinal control.

Continental margin terranes mostly formed along the Laurentian margin. The Arctic Alaska Terrane in the northernmost Cordillera consists of Neoproterozoic to Jurassic continental margin strata. The terrane has affinities with parts of the Innuitian Orogen of Arctic Canada in that it contains rocks deformed, metamorphosed, and intruded during the Devonian Ellesmerian Orogeny. However, there is disagreement about whether or not it underwent large amounts of counterclockwise rotation or translation relative to the rest of North America during the Early Cretaceous rifting and seafloor spreading that created adjacent parts of the Arctic Ocean basin. The Coldfoot, Kootenay, Seward, and Yukon–Tanana terranes contain metamorphosed Neoproterozoic and Palaeozoic strata formed in distal continental margin settings, and some contain Middle Devonian to Early Carboniferous arc magmatic rocks and underwent latest Palaeozoic metamorphism. Possible exceptions to a Laurentian provenance are the Dillinger, Mystic, and Nixon Fork terranes in Intermontane Alaska, whose Early and Middle Palaeozoic stratigraphy and faunas are similar to those of the Siberian Craton, and possibly parts of the Kootenay Terrane, for which a peri-Gondwanan provenance has been suggested.

A key question, whose answer constrains the potential amount of displacement of magmatic arc terranes relative to the former continental margin, is whether subduction was beneath the continent or away from it. Mid-Mesozoic and younger magmatic arcs evidently faced away from the continent, because contemporary accretionary complexes are located oceanward of them, with the exception of the Koyukuk–Nyak–Togiak Arc terranes in western Alaska, which are outboard of the oceanic Angayucham terrane. The facing directions of Early Mesozoic arc rocks in the Quesnel Terrane and of the Middle Jurassic arc in the Stikine Terrane were also away from the continent, but those of Early Mesozoic and older arcs of the Stikine, Alexander, and Wrangellia terranes are unknown. It has been suggested that the facing direction of the Late Palaeozoic arc in the Quesnel Terrane in southern British Columbia was away from the continent, but, conversely, that of a Late Palaeozoic arc in the Yukon–Tanana Terrane, where Carboniferous and Permian subduction evidently involved the Slide Mountain Terrane, is thought to have been towards it. Middle and Late Devonian arc rocks extend into continental margin deposits in the south-eastern Canadian Cordillera, suggesting that at that time subduction was beneath the continent.

Terranes mainly in the Intermontane Plateau System south of latitude 60° N have counterparts as far south as central California and north-western Nevada. Permian to earliest Jurassic faunas in the Quesnel, Slide Mountain, and Stikine terranes differ from those in rocks deposited on the ancient continental margin in Canada, but are comparable to coeval faunas in the western conterminous USA and Mexico. Middle Permian to Middle Triassic faunas in the accretionary complex called the Cache Creek Terrane, located between the Stikine and Quesnel terranes, are similar to faunas found in eastern Asia and in rocks of Permian Tethys.

The Wrangellia and Alexander terranes in the Pacific Mountains System of Canada and Alaska contain Palaeozoic faunas most similar to those in the Canadian Arctic and Siberia. Their Early Mesozoic faunas are more comparable to those further south in the western Cordillera, although Wrangellia also contains an Early Jurassic ammonite of a characteristic Russian species unknown elsewhere from North America. The Alexander Terrane features a

distinctive record of Neoproterozoic and Early Palaeozoic arc magmatism and orogeny; suggestions have been made for its origin in regions as far apart today as eastern Australia and the Barents Sea, with the latter currently favoured.

Northern Cordilleran Mountain Building

The final accretion of terranes to the former continental margin started in the Early Jurassic and was associated with the subduction of oceanic lithosphere beneath the western North American Plate, accompanying arc magmatism, and the collapse of back-arc basins. By the Late Cretaceous, almost the entire region was above sea-level.

For much of its length, the northern Cordillera features two tracts of Mesozoic and Palaeogene high-grade metamorphic rocks that are loci of the greatest amounts of burial, crustal thickening, differential uplift, and erosion and/or tectonic exhumation. The metamorphic rocks are located in internal parts of the Rocky Mountains and Pacific Mountains Systems and are flanked on both sides by far less metamorphosed rocks in much of the remainder of the Cordillera (Appendix A). Some metamorphic rocks in Intermontane Alaska and Yukon (in the Ruby, Seward, and Yukon–Tanana terranes) are older. Starting in the latest Early Jurassic ($\leq$185 Ma), terranes (Cache Creek, Quesnel, Stikine, and Slide Mountain) in the Intermontane Plateau System in Canada were thrust towards the continent and deeply buried former continental margin rocks that are exposed today in internal parts of the Rocky Mountains System. Metamorphism along the southern Arctic Alaska Terrane and in parts of Intermontane Alaska similarly appears to be related to Middle Jurassic–Early Cretaceous thrusting of oceanic rocks and arc terranes (Angayucham, Koyukuk, Nyac, and Togiak) on to Arctic Alaska. Beginning in the mid-Cretaceous ($\sim$95 Ma), the terranes (Alexander, Wrangellia) near the present coast tectonically overlapped the previously accreted terranes, mainly preserved in the Intermontane Plateau System, along a boundary located in the internal part of the Pacific Mountains System.

Compressional structures in the remainder of the northern Cordillera can be directly related to those in the two tracts of high-grade metamorphic rocks (Figure 4). External parts of the Rocky Mountains System contain Cretaceous and Paleocene thrust faults that carried miogeoclinal and older deposits for distances of up to 200 km over the continental platform. In the Pacific Mountains System, Late Cretaceous to Holocene folds and thrust faults mostly verge oceanward. In the Intermontane Plateau System in Canada, oceanward-verging folds and thrust faults are Jurassic, whereas those directed towards the continental interior are mainly Cretaceous. Convergent deformation is contemporaneous with uplift, erosion, and deposition in marine to non-marine basins that are both internal and external to the Cordillera (Figure 4).

In western Alaska, Cretaceous and Palaeogene convergence between the North American and Eurasian Plates may have driven the 40–60° counterclockwise rotation of south-western Alaska. Rotation, combined with dextral displacement on major faults in Intermontane Alaska, created the curvilinear grain of the Alaskan Cordillera and its expanded western termination.

Late Cretaceous, Tertiary, and Holocene dextral strike-slip faults lace the northern Cordillera and apparently reflect oblique subduction and partial coupling of the western North American Plate margin to the northward-moving Pacific Plate and to its precursor, the Kula Plate, which had been entirely subducted by $\sim$37 Ma (Figure 4). The total amount of displacement on these faults, as determined from geological features offset across them, approaches 1000 km. In the north, there are displacement amounts of 425 km on the Tintina Fault and up to 400 km on the Denali Fault, and, in the south, about 300 km on combined Fraser and Yalakom Faults. A major unresolved discrepancy exists between the above numbers and those arising from palaeomagnetic studies on Late Cretaceous rocks ($\leq$85 Ma), which suggest dextral displacements of as much as 3500 km for parts of the western Cordillera.

In addition to the above, a sparse record of Early Cretaceous sinistral, Early to Middle Jurassic dextral, and Triassic–Jurassic sinistral strike-slip faulting was possibly responsible for the fragmentation, displacement, and duplication of arc terranes in the collage prior to their final accretion and further disruption by dextral faulting. In the Pacific Mountains System, a Middle Jurassic–Early Cretaceous island arc founded on Wrangellia and Alexander terranes was apparently displaced southward in Early Cretaceous time. There is about 800 km of overlap between it and the northern end of the contemporary continental arc in the southern Intermontane Plateau and internal Rocky Mountains Systems (Figure 4). In addition, Early Mesozoic oroclinal bending, or left-lateral strike-slip faulting, has been invoked to account for apparent duplication of a Quesnel–Stikine Arc and enclosure of the associated Cache Creek Accretionary Complex (Figure 3).

Palaeogene (58–40 Ma) normal faults are widespread south of latitude 55° N in the Intermontane System and in parts of the Pacific and Rocky Mountains Systems. Many faults are oriented north–northeasterly or north-easterly, at angles to the prevailing

general grain of this part of the Cordillera, and some clearly are associated with movement on dextral strike-slip faults. Normal faults bound graben and half-graben structures containing coeval clastic and volcanic rocks, and many, especially in south-eastern British Columbia and adjoining parts of Idaho and Montana, contributed to the tectonic exhumation of formerly deeply buried rocks in metamorphic core complexes, in some of which Palaeoproterozoic continental basement is exposed.

Economic Deposits

Hydrocarbon deposits occur mainly in the Rocky Mountains System and include oil and gas originating from Palaeozoic and Early Mesozoic strata, and coal in foreland basins. Coal was formerly exploited in Late Mesozoic and Early Tertiary basins in the western Cordillera, and oil and gas are known in rocks of the Pacific continental shelf.

Metallic mineral deposits occur widely. Interior parts of the Rocky Mountains System contain widespread lead, zinc, and silver deposits, two of which, the former Sullivan Mine in south-eastern British Columbia and the active Red Dog mine in north-western Alaska, are some of the largest known. The Intermontane Plateau and Pacific Mountains Systems contain several large copper and molybdenum deposits, with associated golds in porphyry systems in arc-related plutons. Placer gold deposits were of historical importance, but currently are less so.

Evolution of the Northern Cordillera

The origins of the northern Cordillera extend back some 770 Ma to rifting events that marked initial stages in the break-up and dispersal of components of the Rodinian supercontinent. By Early Cambrian time (~540 Ma), rifting had culminated in the seafloor spreading that opened the distant ancestor of the Pacific Ocean basin and created the Laurentian continent–ocean boundary that now is embedded in the Rocky Mountains System.

The intraplate continent–ocean boundary persisted until the Middle Devonian (~390 Ma), when it was succeeded by a convergent plate boundary that initially generated arc magmatic rocks within distal continental margin strata in a belt that extends from California to the Canadian Arctic. Arc magmatism has continued until the present, but has varied greatly in character and volume. From the Early Carboniferous until the Early Mesozoic (355–185 Ma), the period spanning amalgamation and eventual break-up of the supercontinent called Pangaea, the convergent plate boundary lay well offshore and featured island arc chains (Quesnel, Stikine) separated from the continental margin by back-arc basins (Slide Mountain). The region offshore western Pangaea probably was similar to the present western Pacific Ocean basin.

The accretion of arc, back-arc basin, and accretionary prism rocks and former continental margin fragments to the old continental margin started in the Jurassic at about 185 Ma. The accretion records the initiation of Cordilleran mountain building and coincided with the break-up of Pangaea and the advance of the North American continent towards the trench on its western margin. It resulted in the relocation of the continental margin some 500 km oceanward of its original position by the Late Cretaceous (~90 Ma), when the Cordillera probably resembled parts of the modern Andes.

During and following accretion, the collage and parts of the former continental margin were sliced and shuffled by strike-slip faults associated with oblique plate convergence, stretched and dismembered by extension and, in Alaska, apparently rotated counterclockwise. Most of the major physiographical features of at least the Pacific Mountains System, and probably much of the remainder of the northern Cordillera, probably emerged in the later Neogene ($\leq$12 Ma).

Appendix A: Northern Cordilleran Physiographical Systems and Underlying Bedrock Geology

The names Rocky Mountains, Intermontane Plateau, and Pacific Mountains Systems, used for the major physiographical divisions in Alaska, are approximately, but not completely, equivalent to the Eastern, Interior, and Western Systems used in British Columbia.

Features east and north of the Northern Cordillera: Interior Plains System, Arctic Coastal Plain, and Arctic Ocean

Rocky Mountains System
External part

Some major physiographical units: Montana and Canadian Rocky Mountains, Mackenzie Mountains, northern Brooks Range.

Bedrock: (1) Archaean and Palaeoproterozoic continental basement; (2) Mesoproterozoic (locally Palaeoproterozoic) clastic, carbonate, and minor magmatic rock associated with rift-related basins; (3) Neoproterozoic clastic, minor carbonate, and mafic magmatic rocks formed during rifting of the Rodinian supercontinent; (4) Cambrian to Jurassic carbonate and shale deposited on platform, shelf, and slope of Laurentian continental margin; and

(5) latest Jurassic to Palaeogene marine to non-marine clastics deposited in foreland basins concomitantly with folding and thrusting towards and over the continental platform.

Rocky Mountains and (in part) Tintina Trenches
Internal part

Some major physiographical units: Columbia, Cassiar, and southern Brooks Range.

Bedrock: (1) local Palaeoproterozoic continental basement; (2) Neoproterozoic rift-related clastic and volcanic rocks; (3) Palaeozoic, mainly deep-water, clastic rock, rift- and arc-related volcanic and plutonic rocks; (4) local Late Palaeozoic to Early Jurassic magmatic and sedimentary rocks formed in island arcs and marginal basins; (5) Middle Jurassic to Paleocene plutonic rocks; and (6) local Palaeogene magmatic and sedimentary rocks; rocks typically metamorphosed up to high grades and complexly deformed (mainly) during Middle Jurassic to earliest Tertiary compression, and also (in the south) by Palaeogene extension and (in the north) dextral strike-slip faulting.

Intermontane Plateau System

Some major physiographical units: northernmost Columbia Plateau, Interior Plateaus, Skeena Mountains, Yukon Plateaus, Yukon–Tanana Upland, Kuskowim Mountains, Coastal Lowlands, Seward Peninsula.

Bedrock: (1) variably metamorphosed Neoproterozoic and Palaeozoic continental margin deposits in Alaska and Yukon; elsewhere (2) Devonian to Early Jurassic sedimentary and magmatic rocks formed in island arcs, accretionary complexes, and back-arc basins; (3a) Middle Jurassic to Palaeogene continental magmatic arc rocks, and (3b) marine and non-marine clastic deposits eroded from tectonically uplifted regions; (4) local Neogene flood basalt; rocks deformed by compression mainly in the Mesozoic and extension–transtension in the Early Tertiary.

Pacific Mountains System
Internal part

Some major physiographical units: North Cascade Ranges (in Washington) and Cascade Mountains (in British Columbia), Coast Mountains, northern Alaska Range.

Bedrock: (1) dominant (~80%) Middle Jurassic to Palaeogene plutonic rocks (called the Coast Plutonic Complex); (2) Palaeozoic to Holocene volcanic and sedimentary rocks formed mainly in magmatic arcs, but locally in accretionary complexes; most rocks metamorphosed up to high grades and complexly deformed mainly in mid-Cretaceous to Palaeogene time.

Coastal depressions
External part

Some major physiographical units: Olympic Mountains, Insular Mountains, Saint Elias Mountains, Alaska Ranges, Aleutian Range, Aleutian Islands, continental shelf and slope.

Bedrock: (1) latest Proterozoic to mid-Cretaceous volcanic and sedimentary rock formed mainly in island arc and (in Triassic) oceanic plateau settings; (2) mid-Cretaceous and younger clastics eroded mainly from Coast and Cascade Mountains region; (3) Early Palaeozoic to Palaeogene plutons; (4) Early Jurassic to Holocene clastic-rich accretionary complexes in part underlying continental shelf and slope; most rocks folded and thrust oceanwards in Cretaceous to Holocene time.

Features west and south of the Northern Cordillera: Pacific Ocean here underlain by Juan de Fuca and Pacific oceanic plates

Appendix B: Summary Descriptions of Northern Cordilleran Terranes

The abbreviations in parentheses are used in Figure 3; most palaeogeographical references are made with respect to the North American continental interior.

Continental Margin Terranes

Displaced Neoproterozoic, Palaeozoic, and, locally, Early Mesozoic sedimentary rocks.

Arctic Alaska (AA): former North American continental margin deposits, locally with Devonian plutons, possibly rifted away by counterclockwise rotation of ~70° from its former position near the Canadian Arctic Islands in Early Cretaceous time (≤130 Ma).

Cassiar (CA): former North American continental margin deposits, sliced off and translated northwards by 425–2000 km on right-lateral strike-slip faults.

Coldfoot (CO), Kootenay (KO), Ruby (RB), Seward (SD), and Yukon-Tanana (YT): derived mainly from Neoproterozoic and Early Palaeozoic rift and lower slope deposits, and Middle Devonian to Early Carboniferous continental margin arcs; metamorphosed

and multiply deformed in Palaeozoic (?), Mesozoic, and Palaeogene times.

Dillinger (DL), Mystic (MY), and Nixon Fork (NX): small terranes in Intermontane Alaska whose stratigraphy and Early and Middle Palaeozoic faunas resemble those of the Siberian (or North Asian) craton margin.

Accreted Terranes

Arc terranes terminally accreted in the Jurassic

Quesnel (QN) and Stikine (ST): Devonian to Permian clastic rock and carbonate, arc-related volcanic, and plutonic rocks; local Palaeozoic ultramafic rocks, basalt, chert, pelite, and carbonate formed in marginal basins or ocean floor (=SM?); overlying Late Triassic and Early Jurassic arc-related magmatic and clastic rock and minor carbonate. Permian, Triassic, and Early Jurassic faunas most similar to those further south in the western conterminous USA and north-western Mexico; contrarily, palaeomagnetic data suggest little or no northward displacements. ST apparently duplicated QN in the Early Mesozoic by oroclinal bending or left-lateral strike-slip faulting and enclosed the associated accretionary complex (CC).

Arc terranes terminally accreted in the Cretaceous

Alexander Terrane (AX): latest Proterozoic and Early Palaeozoic magmatic arc and associated clastic rock and carbonate, and Late Palaeozoic and Early Mesozoic rift-related clastic and volcanic rock and carbonate. AX and WR apparently linked by Carboniferous plutons and overlapped by Middle Jurassic–Early Cretaceous arc rocks. Palaeomagnetic data suggest an Early Palaeozoic palaeogeographical location either near eastern Australia or the Barents Sea; Silurian and Devonian faunas favour the latter location; Triassic palaeolatitudes of AX comparable with those of Alaskan WR.

Cadwallader (CD) and Methow (MT): Upper Triassic magmatic arc rocks and associated sedimentary rock overlying Permian ocean floor rocks; unconformably overlain by Early Jurassic to Early Cretaceous marine arc-derived clastic and local volcanic rocks; overlying Late Cretaceous continental arc volcanics carry palaeomagnetic signatures indicating $\sim 35^\circ$ of northward displacement.

Chilliwack (CK): Devonian to Early Jurassic arc-related volcanic and clastic rocks and carbonate; in general, stratigraphy and faunas resemble ST and QN; possibly a fragment of ST, sliced off and translated southwards by sinistral strike-slip faults in the Early Cretaceous.

Koyukuk (KY), Nyac (NY), and Togiak (TG): Late Jurassic–Early Cretaceous arc rocks in western Alaska associated with the subduction of AG and GN oceanic terranes.

Wrangellia (WR): Devonian to Middle Jurassic magmatic arc rocks, with associated clastic rock and carbonate, and conspicuous Upper Triassic plume-related (?) tholeiitic basalt overlain by carbonate; Permian faunas comparable with some in Canadian Arctic and Urals; Early Jurassic faunas comparable with those elsewhere in the western Cordillera; palaeomagnetic data from Triassic strata on Vancouver Island indicate little or no latitudinal displacement, whereas those from southern Alaska indicate about 20° northward translation; interpreted as due to Tertiary separation of the two regions; WR linked to AX by Middle Jurassic and possibly Carboniferous.

Accretionary complex and back-arc basin terranes; ordered from continental interior to ocean Unless noted otherwise, all are characteristically highly disrupted and most contain melanges; pillow basalts, local alpine-type ultramafic rocks, and rare blueschists, in places associated with eclogite. Pre-Late Jurassic accretionary complexes in the Cordillera typically contain abundant radiolarian chert, whereas younger ones contain abundant clastic rocks including sandstone.

Angayucham (AG), Goodnews (GN), and Slide Mountain (SM): remnants of oceanic and marginal basins. SM is Carbonifereous to Triassic; AG and GN range in age from Devonian to Late Jurassic and appear to be accretionary complexes related to KY, NY, and TG arc terranes in western Alaska; SM in Yukon is associated with a Late Palaeozoic arc in YT and contains blueschist; further south, SM appears to be a back-arc basin, lacking melanges, eclogites, and blueschists, and separating QN from the former continental margin. The late Palaeozoic palaeomagnetic record of SM suggests northward displacement of $\sim 20^\circ$ with respect to the craton, in accord with displacements indicated by its Middle Permian fauna, which is similar to that in the south-western USA and north-western Mexico.

Cache Creek (CC) and Bridge River (BR): CC is Early Carboniferous to Early Jurassic and contains distinctive large masses of Late Carboniferous to Late Triassic shallow-water carbonate that probably formed as atolls capping seamounts, as do several similar complexes of this age in the circum-Pacific region. Middle Permian to Middle Triassic faunas

are similar to those in eastern Asia and the region occupied by Permian Tethys. CC contains Triassic blueschist and is interpreted as the accretionary complex accompanying QN and ST magmatic arcs. BR is similar to CC, but is of Early Carboniferous to Late Jurassic age, lacks distinctive large carbonate bodies, and was finally accreted in the Cretaceous.

Chugach (CG) and Pacific Rim (PR): CG is a long-lived Mesozoic accretionary complex. The inboard part of CG is a chert-rich assemblage that, in one locality, contains a limestone block with Permian Tethyan fauna and Early Jurassic blueschist; it is the accretionary complex associated with an Early Mesozoic arc on WR. The outboard part of CG is a Late Jurassic and Cretaceous clastic-rich assemblage with local Cretaceous blueschist; PR is similar but local; it is the accretionary complex accompanying the Middle Jurassic–Early Cretaceous arc founded on WR and AX.

Olympic (OL) and Prince of Wales (PW): Eocene to Oligocene; OL contains the youngest Cordilleran accretionary complex exposed on land; PW lies outboard of CG.

Yakutat (YA): Late Mesozoic accretionary complex overlain by Tertiary marine and continental clastic rocks; probably displaced northwards along the continental margin on the Queen Charlotte–Fairweather transform fault.

See Also

North America: Southern Cordillera. **Palaeomagnetism**. **Plate Tectonics**. **Tectonics:** Mountain Building and Orogeny. **Terranes, Overview**.

Further Reading

Armstrong RL (1988) Mesozoic and early Cenozoic magmatic evolution of the Canadian Cordillera. In: Clark SP Jr., Burchfield BC, and Suppe JE (eds.) *Processes in Continental Lithospheric Deformation, Geological Society of America, Special Paper 218*, pp. 55–91. Denver, Colorado: Geological Society of America.

Clowes RM and Hammer PTC (2000) Comparison of lithospheric structures across the Alaskan and Canadian Cordillera. In: *GeoCanada 2000, GAC-MAC-CGU-CSEG-CSPG Conference, Calgary, Alberta, Canada, Conference CD* Abstract 529, Calgary: Alberta.

Coney PJ, Jones DL, and Monger JWH (1980) Cordilleran suspect terranes. *Nature* 228: 329–333.

Engebretson DC, Kelley KP, Cashman HP, and Richards MA (1992) 180 million years of subduction. *GSA Today* 2: 93–100.

Enkin RJ, Mahoney JB, Baker J, Reisterer J, and Haskin ML (2003) Deciphering shallow magnetic inclinations: 2. Implications from Late Cretaceous strata overlapping the Insular/Intermontane superterrane boundary. *Journal of Geophysical Research* 108(B4): 2186.

Gabrielse H and Yorath CJ (eds.) (1991) Geology of the Cordilleran orogen in Canada. In: *Geological Survey of Canada, Geology of Canada 4*, and *Geological Society of America, The Geology of North America G-2*, Ottawa, Ontario: Geological Survey of Canada and Denver, Colorado: Geological Society of America.

Gehrels GE and Ross GM (1998) Detrital zircon geochronology of Neoproterozoic to Permian miogeoclinal strata in Alberta and British Columbia. *Canadian Journal of Earth Sciences* 35: 1380–1402.

Ghosh DK (1995) Nd-Sr isotopic constraints on the interactions of the Intermontane Superterrane with the western edge of North America in the southern Canadian Cordillera. *Canadian Journal of Earth Sciences* 32: 1740–1758.

Nolkelberg WJ, Parfenov LM, Monger JWH, *et al.* (2000) *Phanerozoic Tectonic Evolution of the Circum-North Pacific. United States Geological Survey Professional Paper 1626.* Menlo Park, CA: United States Geological Survey.

Oldow JS, Bally AW, Ave Lallement HG, and Leeman WP (1989) Phanerozoic evolution of the North American Cordillera: United States and Canada. In: Bally AW and Palmer AR (eds.) *The Geology of North America – An Overview, Geological Society of America, The Geology of North America A,* pp. 139–232. Denver, Colorado: Geological Society of America.

Plafker G and Berg HC (eds.) (1994) *The Geology of Alaska. Geological Society of America, The Geology of North America G-1.* Denver, Colorado: Geological Society of America.

Samson SD and Patchett PJ (1991) The Canadian Cordillera as a modern analogue of Proterozoic crustal growth. *Australian Journal of Earth Sciences* 38: 595–611.

Scotese CR, Nokleberg WJ, Monger JWH, *et al.* (2001) *Dynamic Computer Model for the Metallogenesis and Tectonics of the Circum North Pacific. United States Geological Survey Open File* 01-26, version 1.0 (CD). Menlo Park, CA: United States Geological Survey.

Smith PL, Tipper HW, and Ham DM (2001) Lower Jurassic Amaltheidae (Ammonitina) in North America: paleobiogeography and tectonic implications. *Canadian Journal of Earth Sciences* 38: 1439–1449.

Southern Cordillera

A W Snoke, University of Wyoming, Laramie, WY, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Spanish term 'cordillera' refers to a series of parallel ranges or chains of mountains. The word was first applied in the western hemisphere to the mountain ranges of western South America, i.e., Las Cordilleras de los Andes. In the western United States, the Rocky Mountains and Coast Ranges and mountains between are collectively called the Cordillera. In Mexico, the cordillera includes the Sierra Madre Occidental, Del Sur, and Oriental systems, as well as the Sonoran Basin and Range Province and the Baja California peninsula (Figure 1). The focus in this article is on the southern half of the western North American Cordillera, including the mountain belt from ~47° N to ~16° N, an area that extends >4500 km. Thus this region is referred to as the 'Southern Cordillera'.

Diverse physiographic provinces are included in the Southern Cordillera. Where crossed by the Fortieth Parallel, it includes, from east to west, the Rocky Mountains, the Basin and Range Province, the Sierra Nevada, the Great Valley, and the Coast Ranges (Figure 1). Farther south, the Colorado Plateau is a prominent physiographic province between the Rocky Mountains and the Basin and Range Province, whereas in the north-western United States the Cenozoic volcanic provinces (Snake River Plain, Columbia Plateau, and Cascade Range) are prominent. An important topographic feature in the Southern Cordillera is the tract of high elevation that constitutes the Rocky Mountains and adjacent Colorado Plateau and Great Plains. Most of the 14 000-foot or greater mountain peaks in the conterminous United States are within the Southern Rocky Mountains, with the exception of Mt Whitney (14 494 ft; ~4421 m) in the Sierra Nevada and Mt Shasta (14 162 ft; ~4319 m) and Mt Rainier (14 410 ft; ~4395 m) in the Cascade Range.

Crustal thickness in the Southern Cordillera is strikingly variable. The Colorado Rocky Mountains and Sierra Nevada both exceed 50 km in thickness. The thinnest continental crust is in the north-western and north-eastern parts of the Basin and Range Province (~25–20 km) and around the northern part of the Gulf of California (Salton Trough), where the crust of south-eastern California and south-western Arizona has been rifted.

The fundamental nature of the Southern Cordillera is a Mesozoic and Cenozoic history as a convergent margin. The subduction of oceanic lithosphere along the Pacific margin and partial subduction of continental lithosphere in the interior of the orogen led to the development of accretionary complexes, magmatic belts, a mobile igneous and metamorphic core zone, and a foreland belt of thin- and thick-skinned contractional deformation. In contrast, the Late Cenozoic history of this circum-Pacific orogen is characterized by partitioning of deformation and magmatism into discrete magmatotectonic zones related to ongoing subduction of oceanic lithosphere, evolution of the San Andreas transform-fault system, north-westward migration of the Mendocino triple junction, and intraplate continental lithosphere extension.

Precambrian Framework

The nucleus of North America is the Precambrian craton, commonly referred to as 'Laurentia'. This Precambrian craton consists of seven Archaean microcontinents welded together along Palaeoproterozoic collisional orogenic belts. The Archaean Wyoming province is exposed in basement-involved, Laramide (Late Cretaceous and Early Palaeogene) uplifts of the Rocky Mountain foreland. The Wyoming province is bordered by Proterozoic orogenic belts on its northern, eastern, and southern margins, whereas its western margin, which extends into north-eastern Nevada and under the Snake River Plain, is a rifted margin formed during Neoproterozoic and Early Cambrian breakup of the Late Mesoproterozoic supercontinent 'Rodinia'. Along the southern margin of the Wyoming province, a Palaeoproterozoic geosuture (Cheyenne belt) is well exposed in the Medicine Bow Mountains of south-eastern Wyoming, and is interpreted as a Palaeoproterozoic collisional zone between a rifted continental margin (Archaean Wyoming province and overlying Palaeoproterozoic rocks) and a Palaeoproterozoic oceanic supra-subduction complex. This collisional orogenic event, the Medicine Bow Orogeny, developed during the interval 1.78–1.74 billion years ago (Ga). It was apparently the harbinger of a series of Palaeoproterozoic accretion events, which continued until ~1.65 Ga. This long history of lateral accretion of

Figure 1 Major physiographic provinces of the United States and Mexican Cordillera. Base map © 1992, Raven Maps & Images.

chiefly juvenile Palaeoproterozoic crust yielded a zone of accretion at least 1200 km wide, now extensively exposed in parts of Colorado, New Mexico, and Arizona. In the south-western United States, these accretion events have been traditionally referred to as the Yavapai and Mazatzal orogenies.

By ~1.6 Ga, a supercontinent (i.e., 'Nuna', after P F Hoffman) had been assembled that included Laurentia and other ancient elements that are now part of the Greenland and Baltic shields. This supercontinent experienced multiple magmatic and tectonic events in the next 500 million years. In the first event, a broad belt of Mesoproterozoic magmatism and associated tectonism developed across much of the supercontinent. In the Rocky Mountains, this magmatism is well represented by ~1.4-billion-year-old plutonic complexes that are widespread south of the Cheyenne Belt. Plutons of similar age are common in the Basin and Range Province of Arizona, in the Mojave Desert of south-eastern California, and in the subsurface of West Texas and eastern New Mexico. Near the end of the Mesoproterozoic, the North American Craton experienced both rifting (manifested by the Mid-Continent Rift System) and severe contraction, a product of the collisional Grenville Orogeny (~1.1–0.9 Ga). That orogeny culminated in assembly of the Late Mesoproterozoic supercontinent Rodinia, the reconstruction of which remains controversial.

Breakup of Rodinia, and Neoproterozoic to Mid-Palaeozoic Sedimentation

The breakup of Rodinia was followed by development of a thick wedge of Neoproterozoic through mid-Palaeozoic continental terrace deposits along the Cordilleran and Appalachian margins. In what is now the Canadian Cordillera, rifting began ~750 Ma and evolved into a Neoproterozoic passive margin where the Windermere Supergroup was deposited. In the Southern Cordillera, initial rifting was incomplete, and only during a second rifting event was the rift-to-drift transition achieved. Tectonic subsidence studies indicate that the drift phase in the US Cordillera did not occur until the latest Neoproterozoic or earliest Cambrian (~600–555 Ma). Nevertheless, Neoproterozoic clastic deposits, including glacial diamictite, can be traced from the Mackenzie Mountains in the Canadian Cordillera to the Death Valley region in southern California. The Neoproterozoic rocks in the Southern Cordillera are considered rift deposits that accumulated in isolated basins, but by the Early Cambrian, a continuous, Atlantic-type, passive margin existed virtually the length of the western North American Cordillera.

The Neoproterozoic deposits were the initial stratigraphic units of an enormous wedge of chiefly clastic and carbonate rocks that accumulated along the rifted, western margin of the Southern Cordillera through Late Devonian time. This wedge of continental margin deposits is often referred to as the 'Cordilleran Miogeocline', and it reached ~10 000 m in total thickness. The miogeocline is separated from a partially equivalent, but considerably thinner, cratonic sequence by the 'Wasatch Line', interpreted as a hinge line in the depositional framework of Southern Cordillera. This fundamental boundary is also interpreted as the eastern limit of Neoproterozoic, syndepositional faulting related to initial rifting of Rodinia in the Southern Cordillera. The Phanerozoic cratonic sedimentary sequence is characterized by disconformities and in some cases complete periods are unrepresented (e.g., the Silurian over most of Wyoming). Subsequently, the Wasatch Line represents an important boundary during Pennsylvanian–Permian basin development (e.g., Oquirrh Basin), coincides with a regional ramp during foreland fold-and-thrust belt development, and is the approximate eastern margin of the Cenozoic Basin and Range extensional province. Furthermore, a segment of the Intermountain seismic belt follows the Wasatch Line.

Another fundamental boundary in the Southern Cordillera is the $Sr_i = 0.706$ line, a boundary based on the initial $^{87}Sr/^{86}Sr$ ratio in Mesozoic and Cenozoic igneous rocks. This isotopic boundary has commonly been interpreted as the western extent of Precambrian basement rocks (Figure 2) and therefore also roughly correlates with the western extent of miogeoclinal sedimentary rocks.

Palaeozoic Orogenies

Passive margin sedimentation ended in Late Devonian time when Cambrian through Devonian oceanic rocks of the Roberts Mountains allochthon were thrust onto the continental shelf during the Antler Orogeny, best documented in north–central Nevada. Slope-and-rise sedimentary rocks and seafloor mafic volcanic rocks were thrust upon coeval, shallow-water shelf strata along a regional thrust fault (Roberts Mountains Thrust). Detailed palaeontological and stratigraphic studies have shown that the Roberts Mountains allochthon consists of fault-bounded packets of oceanic-facies rocks imbricated into a tectonic wedge. A great clastic wedge (Antler Flysch) was shed eastward from the resulting highlands into a broad foredeep that included much of eastern Nevada and extended into Utah. Upper Pennsylvanian and Permian limestone and clastic rocks unconformably overlie the Roberts Mountains

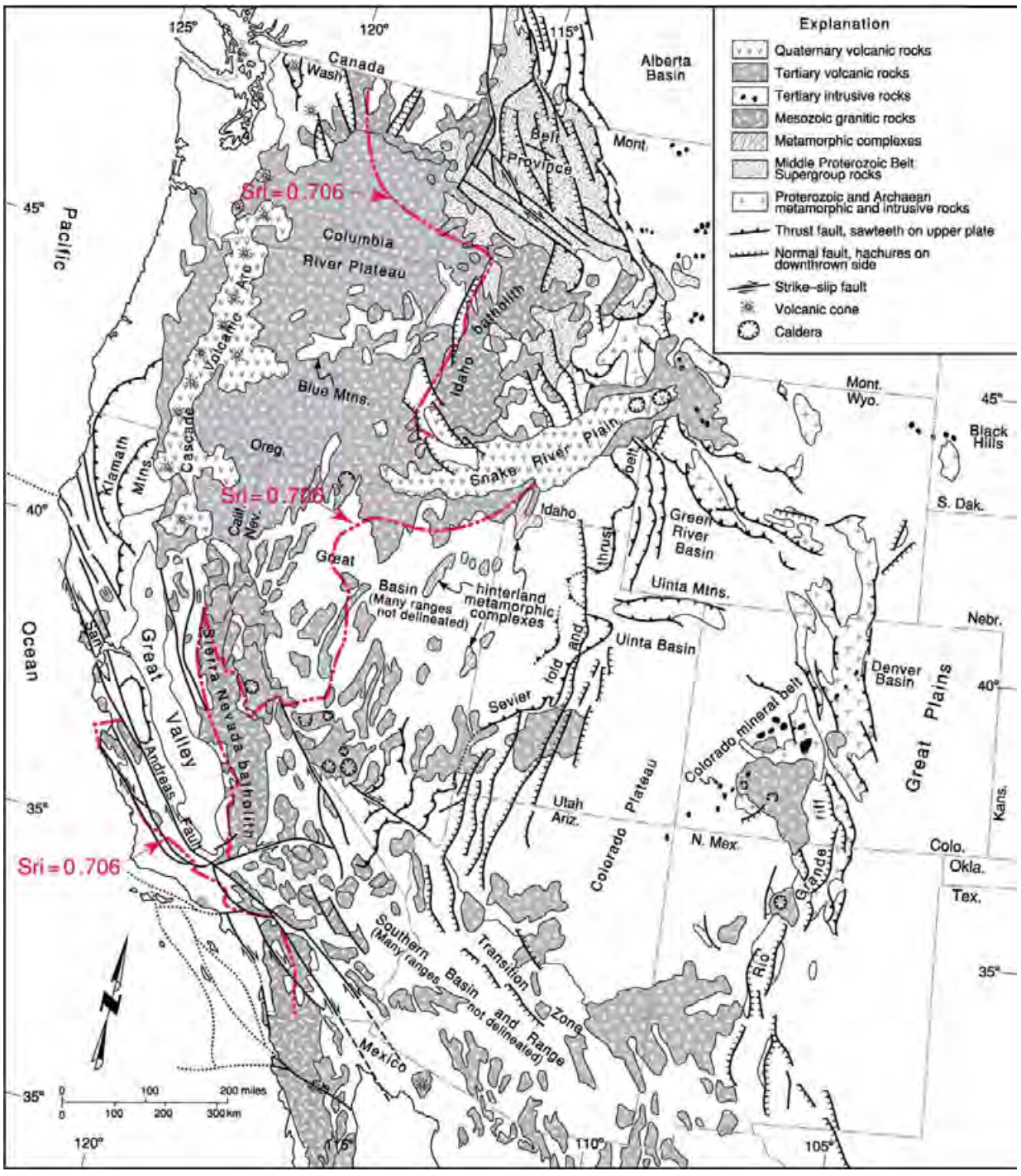


Figure 2 A tectonic map of the United States Cordillera, showing selected geological/tectonic features.

allochthon. Similar relationships occur to the north-east in the Pioneer Mountains of central Idaho and to the south-west in roof pendants in the Sierra Nevada batholith. Various plate tectonic models have been suggested for the Antler Orogeny and considerable debate has focused on the Antler Magmatic Arc that plays a role in virtually all of them. Some have argued that the Antler Arc subsided after collision with the western North American continental margin and was subsequently buried by younger rocks of later orogenic cycles. Others have stated that Devonian oceanic-arc rocks in the eastern Klamath Mountains and northern Sierra Nevada are the magmatic arc elements involved in the Antler Orogeny; however,

there is debate about the facing direction of the arc. One hypothesis is that the arc faced and migrated south-eastward during progressive rollback of the subducted slab. Still another model suggests that the arc faced westward, and the emplacement of the Roberts Mountain allochthon was related to the collapse of a back-arc basin.

Near the Permian–Triassic transition, a tectonic event similar to the Late Devonian Antler Orogeny occurred in the Southern Cordillera. The field relationships for this event are known best in north-western Nevada. The deep-water chert–argillite-limestone–greenstone of the Havallah sequence was thrust eastward onto autochthonous, shallow-water Upper Palaeozoic strata along the Golconda Thrust. This Golconda allochthon is composed of numerous fault-bounded slices of rock, such as the Roberts Mountains allochthon. This contractional orogeny is referred to as the 'Sonoma Orogeny', and again there is debate over the facing direction of the oceanic arc involved in this orogenic event. One school of thought interprets the tectonic setting for the Sonoma Orogeny as an oceanic arc–continental margin collision involving incipient subduction of continental crust beneath an east-facing oceanic arc (i.e., subduction directed westward). Others view the Sonoma Orogeny as another example of the collapse of a back-arc basin developed behind a west-facing oceanic arc (i.e., subduction directed eastward), remnants of which are preserved in the northern Sierra Nevada and eastern Klamath Mountains. An anomaly of the Sonoma Orogeny, unexplained by any model, is the fact that it did not create an extensive foreland basin.

Far to the east of the Antler and Sonoma orogenies, and approximately between them in time of origin, are structural and sedimentological effects related to the amagmatic, Pennsylvanian–Permian, intraplate Ancestral Rocky Mountains Orogeny. Manifestations of the Late Palaeozoic Ancestral Rockies are best demonstrated in the present-day Colorado Rocky Mountains and environs, but structural and sedimentological effects related to the Ancestral Rockies Orogeny can be traced from southern Oklahoma to northern Nevada. The uplifts supplied salmon pink to red arkosic sandstones, which grade into marine strata; these deposits are important components of the Late Palaeozoic stratigraphic section of the central Rocky Mountain region. The Ancestral Rockies Orogeny is commonly interpreted as an intraplate orogeny related to the Late Palaeozoic collision of the South American–African plates of Gondwana with the southern margin of Laurentia during the development of the supercontinent Pangaea. Orogenic effects of this continent–continent collision are manifested in the Marathon–Ouachita Orogeny in the south–central United States and are part of an extensive Late Palaeozoic orogenic system that can be traced from West Texas to central Europe.

Truncation of the Cordilleran Miogeocline and Pre-Cenozoic Strike–Slip Faulting along the South-western Margin of the Cordillera

In a 1969 synthesis of the plate tectonic evolution of California and environs, Warren Hamilton noted the apparent truncation of the south-west-striking Cordilleran miogeocline in southern California and north-western Mexico. Although Hamilton favoured Late Palaeozoic to Triassic truncation, subsequent studies have demonstrated that in Early to Middle Pennsylvanian time, the depositional framework in the Death Valley region experienced a fundamental change in orientation from north-east/south-west to north-west/south-east. This important change in orientation of Late Palaeozoic depositional facies in south-eastern California is interpreted as the result of a sinistral transform fault zone that was initiated during the Early or Middle Pennsylvanian and continued to be active into the Early Mesozoic. In Late Triassic time (Norian), this north-west/south-east strike was maintained during the initiation of the Cordilleran continental-margin magmatic arc that can be traced from southern Arizona into the eastern Sierra Nevada. Late Palaeozoic magmatism is recognized in eastern Mexico where it invades Gondwanan crust; whereas scarce Late Permian plutonic rocks occur in the western Mojave Desert and environs. How these igneous rocks relate to the development of the continental-margin magmatic arc is uncertain, but the occurrences in the Mojave Desert area suggest magmatism shortly after the establishment of a newly formed continental margin bounded by a transform-fault system.

Another widely cited tectonic hypothesis concerning major strike–slip displacement along the south-western margin of the Southern Cordillera postulates a major Late Jurassic, sinistral, transform-fault boundary designated the 'Mojave–Sonora Megashear'. This regional fault zone has been considered to be significant in the translation of part of northern Mexico (e.g., 'Caborca block') into its present position in the Southern Cordillera after the initial rifting and breakup of Pangaea and during the development of the Gulf of Mexico. However, the recognition that Palaeozoic depositional and structural trends were initially truncated in the Southern Cordillera in Late Palaeozoic time has dramatically reduced the potential significance of the Mojave–Sonora Megashear in reshaping the south-western

margin of the Southern Cordillera. Furthermore, detailed field studies in southern California have failed to locate the trace of the Mojave–Sonora Megashear in that region, although Late Jurassic (?) faults possibly related to it have been identified in Sonora, Mexico.

Possibly significant Early Cretaceous, dextral strike–slip faulting has been suggested in the wall rocks of the Sierra Nevada batholith; and younger, right-slip, crystal-plastic shear zones are known from the Sierra Nevada batholith. In one speculative model, a large-scale, Early Cretaceous dextral strike–slip fault system is hypothesized to have extended from the Mojave Desert region to the western margin of the Idaho batholith, suggesting significant northward translation of a large tract of the accreted terranes of the Western Cordillera. The region inferred to have been displaced northward following its accretion to western North America includes the Blue Mountains (north-eastern Oregon), Klamath Mountains (south-western Oregon and north-western California), and northern Sierra Nevada (Figure 1).

Early Mesozoic Continental to Oceanic Magmatic Arc

The Middle Pennsylvanian–Middle Early Triassic oblique truncation of the continental margin of the Southern Cordillera was subsequently overprinted by the development of a north-west/south-east-trending magmatic arc. It can be traced from southern Arizona, where it is built on continental (sialic) crust, to the eastern Klamath Mountains, where it is built on oceanic (ultramafic to mafic) crust. The transition from continental to oceanic arc is inferred to occur at about 39° N latitude in the Sierra Nevada. This Early Mesozoic magmatic arc is thought to have been west-facing with an eastward-dipping subduction zone. Late Triassic (~220 Ma) blueschist-facies metamorphic rocks in the northern Sierra Nevada, Klamath Mountains, and near Mitchell, Oregon (inlier of the Blue Mountains Province) are interpreted as the innermost subducted rocks of an accretionary complex that developed seaward of the magmatic arc. Its development initiated a long-lived convergent plate boundary zone along the Southern Cordillera. This west-facing magmatic arc served as the 'backstop' for the accretion of numerous tectonostratigraphic terranes that were added to the western North American continental margin from mid-Jurassic through Early Tertiary time.

Accreted Terranes

The 'terrane concept' was conceived and initially applied by W Porter Irwin (United States Geological Survey) to explain complex geologic relationships in the south-eastern Klamath Mountains of California. Subsequently, the concept has been utilized throughout the western North American Cordillera as well as in other orogens (e.g., the Appalachians and the Caribbean region). A tectonostratigraphic (or lithotectonic) terrane is an allochthonous, fault-bounded assemblage of rocks with a different geological history than that of adjacent rock units. Tectonostratigraphic terranes can consist of continental-margin features such as a displaced continental margin or part of an original fringing island arc and/or its subduction complex. Oceanic features such as plateaus, seamounts, or even back-arc basins may occur as discrete terranes in accretionary orogens. The tectonostratigraphic terranes in the western North American Cordillera are tectonic slices of such crustal elements and are not lithospheric or even crustal sections. The juxtaposition and amalgamation of such slices produced the Cordilleran 'collage' of terranes that characterizes the western part of the orogen.

In the Southern Cordillera, Phanerozoic tectonostratigraphic terranes are the basic 'building-blocks' of continental accretion west of the $Sr_i = 0.706$ line (i.e., the inferred western margin of Laurentia in the Southern Cordillera). Beginning in the late 1970s and extending to the present, many geological studies have focused on determining the affinity of various tectonostratigraphic terranes that are part of the Cordilleran Orogen. Clearly some of these terranes are 'pericratonic', having originated near the western margin of Laurentia, whereas other terranes are truly 'exotic' to the Cordillera. These far-traveled terranes represent major additions to western North America.

Ophiolites, sometimes dismembered, are an important component of some of the accreted, tectonostratigraphic terranes of the Southern Cordillera. In the Klamath Mountains of north-western California and south-western Oregon, well-preserved ophiolitic sequences range in age from Early Palaeozoic through Late Jurassic. The oldest are in the eastern parts of the Klamath Mountains and younger sequences are to the west. Some examples are the Trinity Subterrane, North Fork, Rattlesnake Creek, and Western Klamath terranes. The Josephine Ophiolite of the Smith River Subterrane (part of the Western Klamath Terrane) is the best-studied ophiolite sequence in western North America. To the north-east in the Blue Mountains of north-east Oregon, the Baker Terrane includes important ophiolitic sequences. In east–central California, ophiolitic sequences commonly occur as tectonic slices along the Foothills fault system of the western Sierra Nevada metamorphic belt. Many,

if not all, of these Cordilleran ophiolites developed in supra-subduction-zone settings, which indicate origins by rifting and spreading within oceanic arcs. This intra-arc extension may have been a result of oblique subduction and broad-scale transtension within the arc, similar to the present rifting and spreading in the Andaman Sea north of Sumatra. In contrast, the rifting and spreading may have been more orthogonal, such as the ongoing propagating rift and spreading centre related to the opening of the Lau Basin behind the Tofua (Tonga) Arc in the south-western Pacific. This supra-subduction interpretation of Cordilleran ophiolites implies proximity to, and temporal overlap with, oceanic-arc deposits, and has been demonstrated throughout the Cordillera by detailed geological mapping coupled with geochemical and geochronological studies. A particularly well-documented example of this relationship is the deposition of the Upper Jurassic Rogue and Galice formations and development of the Late Jurassic Josephine Ophiolite—all within the western Klamath Terrane. The Galice Formation, consisting of slaty metashales and metagreywackes with subordinate metaconglomerate and metavolcanic rocks, lies depositionally on both the Rogue Formation and Josephine Ophiolite. This situation suggests a close proximity in space and time between an oceanic arc (Rogue Formation), its adjacent sedimentary apron (Galice Formation), and the development of a complete ophiolite sequence (i.e., Josephine Ophiolite).

Another type of tectonostratigraphic terrane common in the western North American Cordillera is the accretionary complex terrane. Some of these terranes include ophiolitic components (e.g., the Baker Terrane in north-eastern Oregon), but most are either chert rich or clastic rich and are characterized by tectonic melange, broken formation, and/or olistostromal deposits. Fossiliferous rocks in some accretionary complex terranes provide some of the most reliable palaeogeographic data obtainable from tectonostratigraphic terranes. In the Cordillera, limestone blocks (of tectonic or olistostromal origin) yield fossils long recognized as 'exotic' to the western North American Cordillera. For example, Permian Tethyan fossils have been discovered in the Klamath and Blue Mountain provinces (North Fork, Rattlesnake Creek, and Baker terranes) and are especially characteristic of the Cache Creek Terrane of British Columbia, Canada. The significance of the Tethyan fauna in the Cordillera is still debated by palaeontologists. However, the association of these exotic limestones with blueschist-facies blocks strongly suggests that these terranes represent accretionary complexes that incorporated various rock types during the subduction of Pacific Ocean crust along the margin of western North America. The polarity of the original magmatic arcs related to these accretionary complexes is commonly controversial. Some accretionary complexes in the Southern Cordillera may be composites produced by collision of oppositely facing arcs. A possible example is the Baker Terrane (north-east Oregon) that apparently developed during Late Palaeozoic through Early Mesozoic time between the partly coeval Olds Ferry and Wallowa arcs. Such a situation is analogous to the arc–arc collision presently ongoing in the Molucca Sea between the oppositely facing Sangihe and Halmahera island-arc systems.

The largest accreted tectonostratigraphic terrane in the Southern Cordillera is the Guerrero Superterrane of western Mexico; its full geographic extent is still poorly known. The Guerrero Superterrane includes various volcanogenic terranes ranging in age from Late Jurassic to Early Cretaceous. They yield geochemical and isotopic data that suggest an intraoceanic arc setting. However, presence of a possible basement unit (the Arteage Complex with negative initial ε_{Nd} values) unconformably(?) overlain by the volcanogenic units suggests a composite history for the superterrane. In contrast, some tectonicists view the Arteage Complex as an underthrusted subduction complex overridden by the Guerrero Superterrane Arc. In many recent tectonic models, the Guerrero Superterrane is interpreted as an exotic, east-facing, intraoceanic arc accreted to mainland Mexico in mid-Cretaceous time. After collision with the Mexican continental margin, a polarity reversal occurred, with the development of a west-facing, continental-margin magmatic arc. In effect, it is the southern continuation of the Cretaceous arc of the Peninsular Ranges of southern California and northern Baja California. Some workers have argued that the Guerrero Superterrane arc is only part of an enormous, east-facing intraoceanic arc that included the Great Arc of the Caribbean. The latter is considered to be the source of the numerous allochthonous, arc-related terranes that rim the circum-Caribbean region from Cuba to Tobago and along the northern margin of South America.

Jurassic Magmatic and Tectonic Events

The Jurassic history of the western North American Cordillera is particularly complex and includes various events along the continental margin as well as orogenic effects within the interior of the continent. How these Jurassic orogenic processes interrelate is a major unresolved problem in Cordilleran tectonics. A classic Late Jurassic orogeny of the continental

margin is the Nevadan Orogeny originally defined by Eliot Blackwelder in 1914. Recent studies have suggested that the age of this orogeny can be tightly bracketed in the Klamath Mountains, at ~150 Ma. However, in light of the plate tectonic paradigm, coupled with recognition that terrane accretion is commonly progressive along a continental margin, the regional significance of a tightly defined orogenic event has been questioned. The classic definition of the Nevadan Orogeny is based on rock relationships in the western Sierra Nevada of California and the Klamath Mountains of Oregon, in comparison to the California Coast Ranges. These include folds and associated cleavage in Kimmeridgian–Oxfordian (Upper Jurassic) metasedimentary rocks of the Mariposa and Galice formations and the nearby association of relatively undeformed Tithonian (uppermost Jurassic stage) sedimentary rocks of the Knoxville Formation in the Coast Ranges. This classic definition was reinforced with the recognition that Early Cretaceous granitic plutons cut these structural features in the Sierra Nevada.

The plate tectonic setting of the Late Jurassic Nevadan Orogeny is especially controversial with respect to the inferred geological evolution of the western Sierra Nevada as compared to that of the Klamath Mountains. Many tectonicists have favoured a collision between a west-facing Early Mesozoic arc (i.e., an eastern arc) and an exotic, east-facing oceanic arc (i.e., a western arc) to explain the geological relations manifested in the western Sierra Nevada metamorphic belt. The accreted western arc would be separated from the eastern arc by a composite accretionary complex, in part manifested by the Palaeozoic and Triassic Calaveras Complex east of the Sonora fault. In the Klamath Mountains, structural relationships involving west-directed regional thrusting preclude the accretion of an exotic east-facing oceanic arc. In this light, the most widely accepted tectonic model for the Nevadan Orogeny in the Klamath Mountains is the collapse of a back-Arc basin behind a west-facing oceanic arc (Rogue–Chetco Arc). In this model, collapse could be related to increased coupling across the subduction system (i.e., the arc became contractional after an earlier extensional history). Resolution of these two contrasting tectonic scenarios has not been completely achieved, although some favour a polarity reversal of the western arc from east facing in the western Sierra Nevada to west facing in the Klamath Mountains. Such polarity reversals do occur along the strike of modern intra-oceanic arcs. However, this compromise requires drastically different tectonic settings for the deposition of the Galice and Mariposa formations, which have been recognized as probable stratigraphic and temporal equivalent units for over 100 years.

Cretaceous Palaeogeographic Belts and Transition to an Andean-Type Continental Margin

Late Cretaceous palaeogeography of western North America was dominated by three tectonic belts that extended virtually the length of the Cordillera: (1) an Andean-type volcanic–plutonic magmatic arc on the west (i.e., Cordilleran Magmatic Arc), (2) a thin-skinned fold-and-thrust belt on the east (initiated in mid-Cretaceous time), and (3) a hinterland (zone of metamorphism, magmatism, and deformation) (Figure 3). Another extensive feature of western North America during the Cretaceous was the Western Interior Basin. At its maximum development, the Western Interior Basin extended from the Arctic Ocean to the Gulf of Mexico.

Magmatism related to the development of the Cordilleran Magmatic Arc began in the Middle Triassic, but was only developed in the continental crust of Mexico, Arizona, and south-eastern California. From east–central California north, Late Triassic through Middle Jurassic magmatism is recorded in

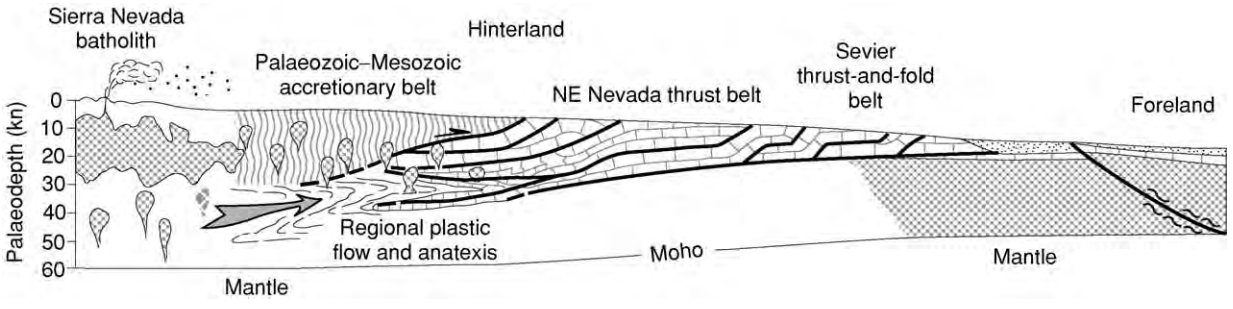


Figure 3 Schematic Late Cretaceous cross-section of the western United States Cordillera, showing the relationship between the Sevier fold-and-thrust belt and fractured Rocky Mountain foreland on the east, metamorphic and igneous rocks of the hinterland, and Sierra Nevada batholith on the west. Reproduced (but modified) with permission from Miller EL and Gans PB (1989) Cretaceous crustal structure and metamorphism in the hinterland of the Sevier thrust belt, western U.S. Cordillera. *Geology* 17: 59–62.

tectonostratigraphic terranes not yet completely attached to the continental margin. During the Middle and Late Jurassic, magmatism spread eastward to include areas well within Laurentian crust (e.g., the Ruby Mountains of north-eastern Nevada). Also, during this time the Sundance Sea occupied an incipient foredeep in the Western Interior Basin and important terrane accretion took place in the western Sierra Nevada, Klamath Mountains, and Blue Mountains.

The geological history of the Southern Cordillera during the early part of the Cretaceous Period (i.e., Neocomian) is poorly understood. Exposures of Neocomian magmatic and sedimentary rocks are rare in the western North American Cordillera. The time interval ~140–125 Ma is commonly referred to as 'the Early Cretaceous magmatic lull'. Global sea-level was low during this time, and in the Rocky Mountain region and environs a regional unconformity exists between the chiefly fluvial deposits of the Upper Jurassic Morrison Formation and overlying mid-Cretaceous strata of the Western Interior Basin. In contrast, widespread Late Cretaceous magmatism in the Sierra Nevada and shortening in the foreland fold-and-thrust belt (i.e., Sevier Orogenic Belt of south-eastern Idaho, western Wyoming, and northern Utah) were synchronous with the culmination of seaway advancement. The fold-and-thrust belt is characterized by sled-runner thrust faults that are directed eastward and merge into a décollement near the top of the crystalline Precambrian basement. Major thrust systems are younger to the east, and the shortening history extends from ~119 to ~52 Ma. The Late Cretaceous culmination of magmatism in the Cordilleran Magmatic Arc (~100–85 Ma) coincides with intense shortening in the fold-and-thrust belt. One tectonic model argues for a close tie between Cretaceous magmatism in the California arc and continental lithospheric underthrusting beneath the arc (so-called A-type subduction). Although Sevier belt thrusting was initiated in mid-Cretaceous time, earlier (Jurassic) back-arc shortening is well documented in central Nevada (e.g., the Luning-Fencemaker fold-and-thrust belt).

While Late Cretaceous magmatism occurred in the Cordilleran Arc and shortening took place in the Sevier Orogenic Belt, regional metamorphism, large-scale folding, and the generation of peraluminous granitic rocks (commonly containing two micas, garnet, ±sillimanite) characterized the deep crust of the hinterland (Figure 3). The regional thrust faults of the Sevier Belt root into this igneous–metamorphic infrastructure, and remobilized Precambrian basement rocks were involved in both the plastic deformation and deep-crustal anatexis. Juvenile magmatic additions from the mantle did not play a significant role in the petrogenesis of the granitic rocks. Tectonic thickening coupled with dehydration melting during late decompression were the key processes responsible for metamorphism and anatexis in the Sevier hinterland. The deep-crustal rocks of the Sevier hinterland are exposed in windows developed during Tertiary crustal extension.

Laramide Orogeny

Distribution of igneous activity in the Cordillera of the western United States changed dramatically near the end of the Cretaceous Period. At ~75 Ma, magmatism in the Sierra Nevada batholith ended, and magmatism migrated eastward into the Rocky Mountains and beyond. This change in the geographic distribution of igneous activity in the Southern Cordillera is roughly synchronous with the Laramide orogeny of the Rocky Mountains. In the Rocky Mountain foreland, the Laramide Orogeny is characterized by the development of basement-involved uplifts and adjacent deep basins during Late Cretaceous through Early Eocene time (Figures 3 and 4). This 'thick-skinned' tectonic style is dramatically different from the contemporaneous 'thin-skinned' style of the Sevier fold-and-thrust belt. Although both deformations may be broadly related to plate tectonic processes near the continental margin, the local decoupling levels for these foreland deformation belts are different. The decoupling zone for the Sevier fold-and-thrust belt lies above the Precambrian basement except in the more interior (western) parts of the belt, whereas seismic-reflection studies show the decoupling zone for the basement-involved Laramide uplifts to be rooted into the deep crust.

The typical plate tectonic explanation for both the Late Cretaceous change in geographic distribution of Cordilleran magmatism and development of a thick-skinned tectonic style in the eastern Rocky Mountains involves a change to a shallow-dipping subduction mode. In this interpretation, subduction occurs at a shallow depth beneath the overriding plate in a wide segment of the orogen that extended from south-western Montana to central New Mexico (Figure 4). Why such segmentation of the subduction system would occur along the strike of the orogen is a major unanswered question in Cordilleran tectonics. The Andean orogenic system of South America is commonly proposed as a present-day analogue for the Late Cretaceous–Early Tertiary orogen of western North America. In the present Andean system, segmentation of the subduction system may be related to the character of the subducting plate; i.e., the presence of an oceanic plateau or aseismic ridge. The age of the subducted oceanic lithosphere may also be important

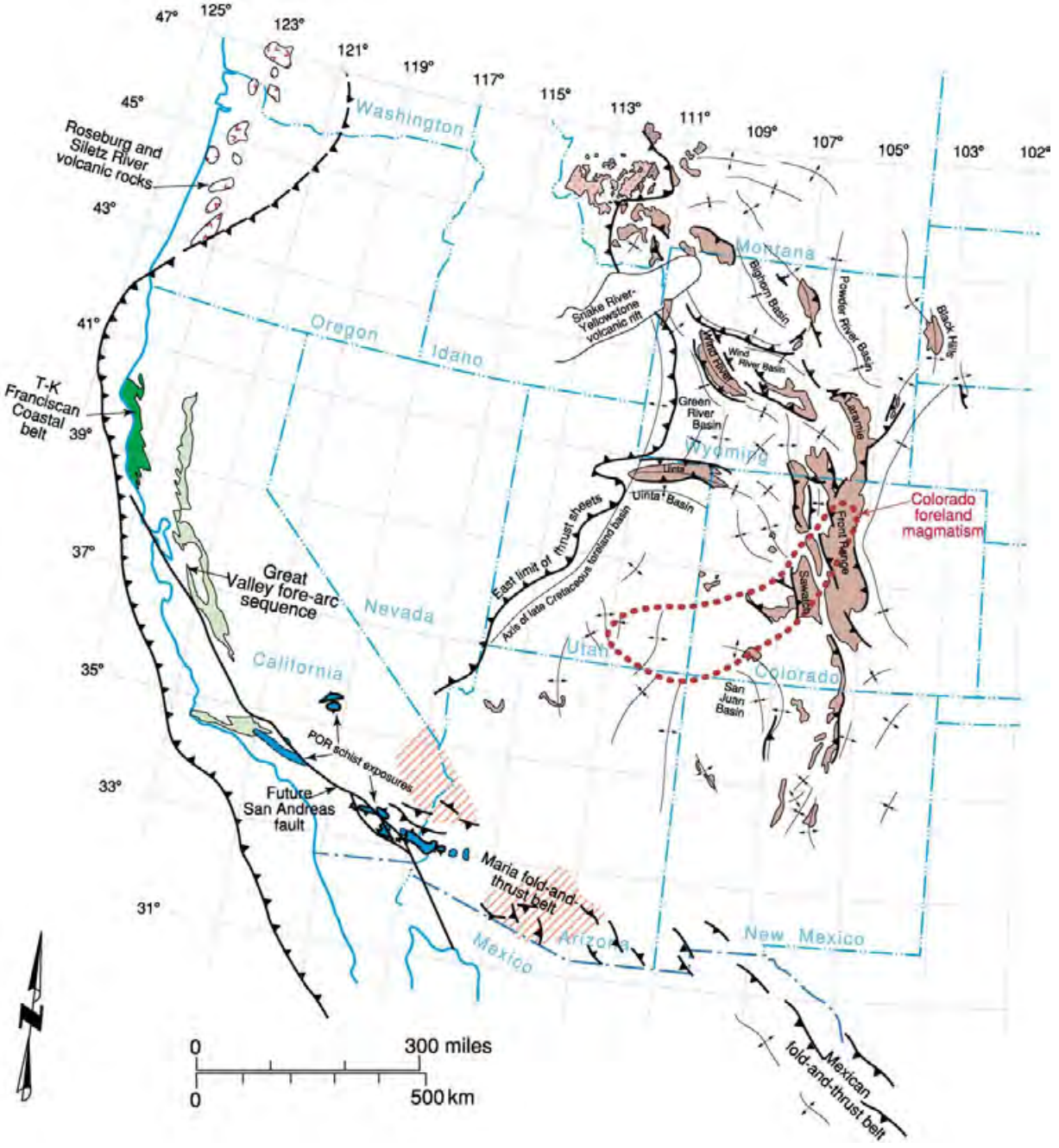


Figure 4 Distribution of the effects of the Late Cretaceous–Early Palaeogene Laramide Orogeny in the south-western part of the western North American Cordillera. T–K, Tertiary–Cretaceous; POR, Pelona–Orocopia–Rand; oblique red lines denote areas of late Cretaceous, high-grade, regional metamorphism.

in the segmentation process, because young lithosphere is more buoyant than old, dense lithosphere. In the southern Andes, the absence of Quaternary volcanic centres in the western part of the orogen and development of basement-involved uplifts (Sierras Pampeanas) in the eastern part of the orogen coincide with the subduction of the Juan Fernandez Ridge and an eastward shift in subduction-related seismicity.

Additional evidence for a shallow-dipping subduction during Late Cretaceous–Early Palaeogene time in the Southern Cordillera is provided by a series of windows in southern California that expose oceanic rocks (Pelona, Orocopia, and Rand schists) beneath an upper plate of continental-affinity rocks. These windows are the product of Cenozoic deformation (both extensional and contractional) and provide

exceptional exposures of underthrusted, Late Cretaceous–Early Tertiary oceanic rocks that apparently underlie much of southern California and adjacent south-western Arizona.

Post-Laramide, Early Cenozoic Magmatic and Tectonic History

The post-Laramide Cenozoic history of the Southern Cordillera is characterized by new patterns of magmatism and tectonic strain. The throughgoing tectonic belts, which characterized Late Mesozoic time, were replaced by domains of extension, contraction, and strike–slip deformation. During mid-Eocene time, a broad belt of magmatism extended from southern British Columbia into central Idaho and north-western Wyoming, and a roughly contemporaneous zone of magmatism existed in southernmost Arizona and New Mexico and extended farther south into Mexico (Sierra Madre Magmatic Zone). These zones of Eocene magmatism were separated by a broad amagmatic corridor in the west–central United States that became the site of a large, Middle Eocene lake system. Accompanying the mid-Eocene magmatism in the Pacific north-west, metamorphic core complexes developed in areas of large-magnitude crustal extension. Typical complexes are characterized by a hanging wall of upper crustal rocks, sometimes including syntectonic volcanic and sedimentary deposits, separated from a footwall of mid-crustal igneous and metamorphic rocks by a plastic-to-brittle, normal-sense shear zone. Younger rocks are commonly structurally emplaced on older rocks, and brittle deformation features, including low-angle detachment faults, are superposed on the crystal-plastic deformation of the normal-sense, mylonitic shear zone. This northern belt of magmatism and accompanying localized, large-magnitude crustal extension migrated southward in Late Eocene and Early Oligocene time. Initiation of core-complex development in the eastern Great Basin was later than in the Pacific north-west, and large-magnitude crustal extension continued into the Early Miocene in the Ruby–East Humboldt and Snake Range core complexes (eastern Nevada). During the Late Eocene through Early Miocene, enormous amounts of volcanic ejecta erupted as ash-flow tuff sheets in the Great Basin (Nevada and western Utah). This 'ignimbrite flare-up' has significant implications for the crustal composition of the Great Basin, including substantial mafic magmatic intra- or underplating of the extended crust of the region. Volcanic ash from these enormous eruptions spread eastward in the upper atmosphere and formed a conspicuous air-fall component in post-Laramide, Late Eocene to Miocene strata of the Rocky Mountains (especially in Wyoming and environs). By Early Miocene time, the northern and southern magmatic zones had merged, and a continuous Neogene magmatic arc could be traced from the early Western Cascades arc into the Mojave–Sonoran Volcanic Zone. Numerous examples of Miocene core-complex development are well documented from south-eastern California across southern Arizona and into Sonora, Mexico. Still younger examples of core-complex development (Late Miocene to Pleistocene) are present in other areas of large-magnitude extension in Southern Cordillera, such as the ongoing rifting of continental crust in the northern Gulf of California, Mexico, and in the Salton Trough, California.

Elevations on the Colorado Plateau range from approximately 1.5 to 3.5 km, with the highest elevations typically associated with igneous centres such as the San Francisco volcanic field in northern Arizona. In deep canyons (e.g., the Grand Canyon), the elevation is considerably less than 1.5 km. The average elevation of the plateau is ~2 km, and the crustal thickness is ~45 km. Stratified rocks exposed on the plateau indicate that the area was near sea-level for much of the Phanerozoic and that uplift occurred after the deposition of Upper Cretaceous marine sedimentary rocks. The western and southern margins of the plateau are delineated by normal-fault systems related to the Basin and Range Province, whereas its northern and eastern margins merge into the eastern Rocky Mountains. The south-eastern margin of the Colorado Plateau in central New Mexico is delineated by normal faults related to the Rio Grande rift. The processes that facilitated uplift of the Colorado Plateau remain controversial, as well as the age or ages of uplift. One tectonic model relates the uplift of the plateau to eastward, intracrustal flow toward the Colorado Plateau from the overthickened, Sevier hinterland (now part of the Basin and Range Province). Another model argues for uplift related to lithospheric attenuation as a by-product of shallow-dipping subduction associated with the Laramide Orogeny. Still other tectonic models favour a polyphase uplift history: initially during the Laramide Orogeny and subsequently in the Late Cenozoic as part of a regional uplift, including the Southern Rocky Mountains and Great Plains. Clearly, the cause of the uplift of the Colorado Plateau remains a major unresolved problem in Southern Cordilleran tectonics.

As the San Andreas Fault (transform) system developed off the west coast of Mexico, and the triple junction between the North American, Pacific, and Juan de Fuca plates (Mendocino triple junction) migrated north-westward, the Neogene magmatic arc was shut off at its south end. In a broad area east of the late Western Cascade Arc, the Columbia River

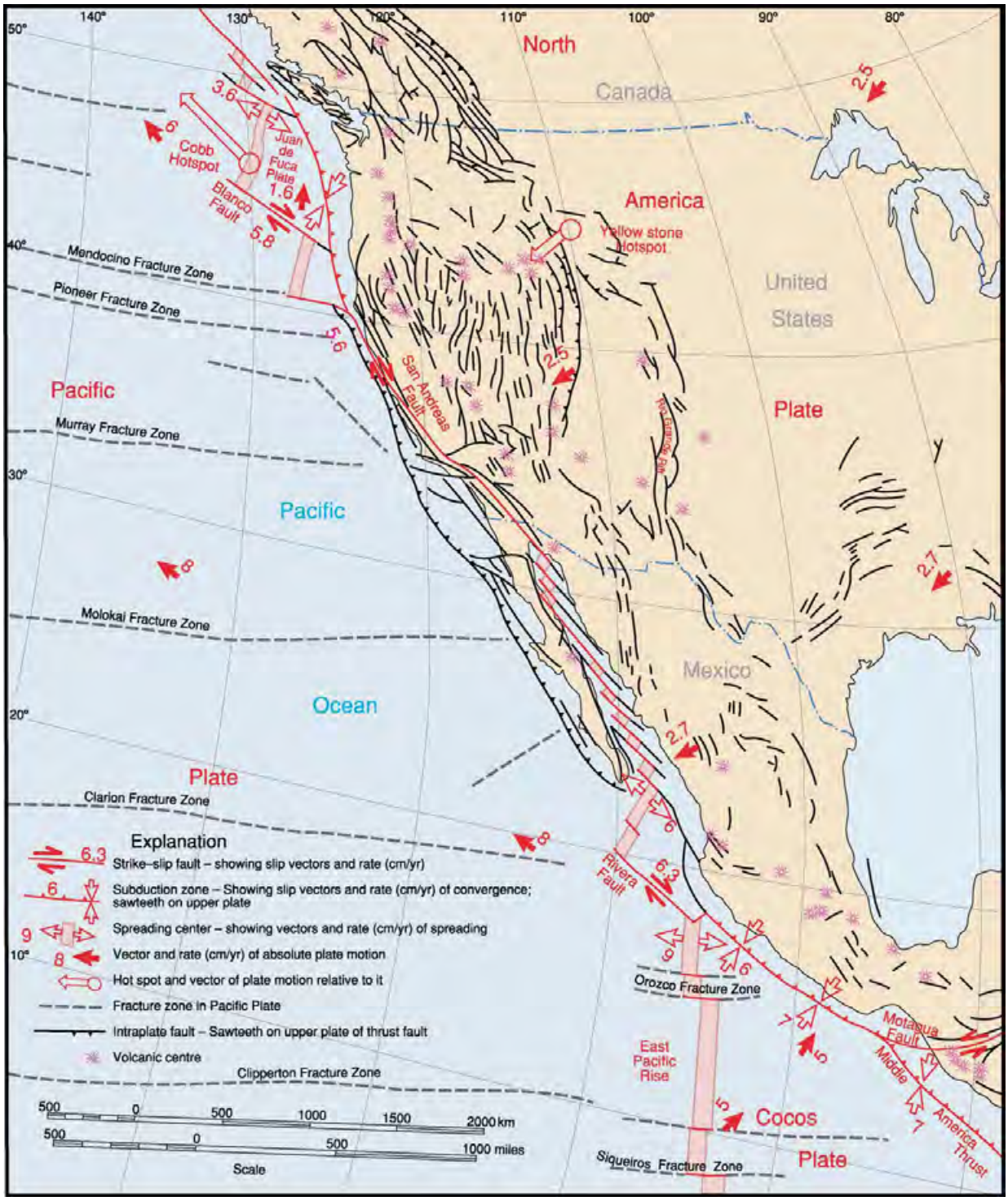


Figure 5 Present-day plate tectonic setting of the southern segment of the Cordilleran Orogen of western North America and adjacent north-eastern Pacific Basin. Reproduced (but modified) with permission from Drummond KJ, *et al.* (1982) *Pacific Basin sheet of plate-tectonic map of the circum-Pacific region (scale 1:20 000 000)*. Tulsa, OK: American Association of Petroleum Geologists.

flood basalt province developed between ~17 and 14 Ma. Basaltic dike swarms of this age are present in north–central Nevada and indicate the initiation of rifting in that area that eventually produced the Late Cenozoic Basin and Range Province.

Late Cenozoic Tectonic/Volcanic Systems and Seismicity

The Late Cenozoic tectonic evolution of the Southern Cordillera is dominated by four large-scale, tectonic/volcanic systems: (1) the San Andreas transform Fault and north-westward migration of the Mendocino triple junction, (2) the crustal extension in the Basin and Range Province and Rio Grande rift, (3) the north-east propagation of the Yellowstone hotspot track and concurrent north-west-propagation of the Newberry volcanic trend, and (4) the Cascadia subduction zone and volcanic arc. The development of these features was in part concurrent, and elements of all four systems are presently active. A relationship between the San Andreas transform-fault system and crustal extension in the Basin and Range Province has been suggested in several analyses of the Late Cenozoic tectonic history of the Southern Cordillera. However, effects associated with plate boundary slip are significant only in the south-western Basin and Range Province (e.g., eastern California shear zone). Late Cenozoic, regional crustal extension characteristic of the Basin and Range Province reflects the removal of the Farallon slab and subsequent asthenospheric upwelling coupled with north-west retreat of the Pacific Plate with respect to the interior of the North American Plate. At ~5 Ma, the Baja California peninsula was rifted from mainland Mexico when the San Andreas fault system shifted inland from a former position on the Pacific side of the peninsula. The north-western separation of Baja California from the mainland is an example of the interplate transfer of continental lithosphere.

The present-day seismicity of the Southern Cordillera is concentrated in belts at least partly related to fundamental plate tectonic boundaries (Figure 5); these boundaries are either strike–slip or convergent. Off the western coast of Oregon, seismicity is related to the Cascadia subduction zone and the Blanco Fracture Zone. Farther south, in western California and offshore, the Mendocino triple junction and San Andreas Fault system are the loci of present-day seismicity. Still farther south along the western coast of Mexico, seismicity is related to convergence associated with the Middle America Thrust.

Other prominent belts of seismicity are within the North American Plate and include the eastern California shear zone and its extensions into western Nevada (e.g., Walker Lane). The seismicity associated with this intraplate tectonic zone is interpreted as distributed strain associated with the plate boundary zone. Still another prominent zone of intraplate seismicity is the Intermountain seismic belt. It can be traced from southern Utah into western Montana and is particularly prominent in the Yellowstone National Park area. This belt of seismicity reflects a combination active normal faulting along the eastern margin of the Basin and Range Province (e.g., Wasatch front) as well as magmatism related to the Yellowstone hotspot.

See Also

North America: Northern Cordillera. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges.

Further Reading

Armstrong RL and Ward PL (1993) Late Triassic to earliest Eocene magmatism in the North American Cordillera: implications for the western interior. In: Caldwell WGE and Kauffman EG (eds.) *Evolution of the Western Interior Basin, Geological Association of Canada Special Paper 39*, pp. 49–72. St John's, NL: Geological Association of Canada.

Burchfiel BC, Lipman PW, and Zoback ML (eds.) (1992) *The Cordilleran Orogen: Conterminous U.S. The Geology of North America*, vol. G-3. Boulder, CO: Geological Society of America.

Coney PJ, Jones DL, and Monger JWH (1980) Cordilleran suspect terranes. *Nature* 288: 329–333.

DeCelles PG (2004) Late Jurassic to Eocene evolution of the Cordilleran thrust belt and foreland basin system, western U.S.A. *American Journal of Science* 304: 105–168.

Dickinson WR (2000) Geodynamic interpretation of Paleozoic tectonic trends oriented oblique to the Mesozoic Klamath-Sierran continental margin in California. In: Soreghan MJ and Gehrels GE (eds.) *Paleozoic and Triassic Paleogeography and Tectonics of Western Nevada and Northern California, Geological Society of America Special Paper 347*, pp. 209–245. Boulder, CO: Geological Society of America.

Dickinson WR (2002) The Basin and Range province as a composite extensional domain. *International Geology Review* 44: 1–38.

Dickinson WR and Lawton TF (2001) Carboniferous to Cretaceous assembly and fragmentation of Mexico. *Geological Society of America Bulletin* 113: 1142–1160.

Drummond KJ, *et al.* (1982) *Pacific Basin sheet of plate-tectonic map of the circum-Pacific region (scale 1:20 000 000)*. Tulsa, OK: American Association of Petroleum Geologists.

Hamilton W (1969) Mesozoic California and the underflow of Pacific mantle: *Geological Society of America Bulletin* 80: 2409–2430.

Hamilton W (1978) Mesozoic tectonics of the western United States. In: Howell DG and McDougall KA (eds.) *Mesozoic*

Paleogeography of the Western United States, Pacific Coast Paleogeography Symposium 2, pp. 33–70. Los Angeles, CA: The Pacific Section, Society of Economic Paleontologists and Mineralogists.

Miller EL and Gans PB (1989) Cretaceous crustal structure and metamorphism in the hinterland of the Sevier thrust belt, western U.S. Cordillera. *Geology* 17: 59–62.

Moores EM (1998) Ophiolites, the Sierra Nevada, "Cordilleria," and orogeny along the Pacific and Caribbean margins of North and South America. *International Geology Review* 40: 40–54.

Oldow JS, Bally AW, Avé Lallemant HG, and Leeman WP (1989) Phanerozoic evolution of North American Cordillera: United States and Canada. In: Bally AW and Palmer AR (eds.) *The Geology of North America—An Overview, The Geology of North America*, vol. A, pp. 139–232. Boulder, CO: Geological Society of America.

Sedlock RL, Ortega-Gutiérrez F, and Speed RC (1993) *Tectonostratigraphic Terranes and Tectonic Evolution of Mexico, Geological Society of America Special Paper 278*. Boulder, CO: Geological Society of America.

Severinghaus J and Atwater T (1990) Cenozoic geometry and thermal state of the subducting slabs beneath western North America. In: Wernicke BP (ed.) *Basin and Range Extensional Tectonics near the Latitude of Las Vegas, Nevada, Geological Society of America Memoir 176*, pp. 1–22. Boulder, CO: Geological Society of America.

Wernicke BP, Christiansen RL, England PC, and Sonder LJ (1987) Tectonomagmatic evolution of Cenozoic extension in the North American Cordillera. In: Coward MP, Dewey JF, and Hancock PL (eds.) *Continental Extensional Tectonics, Geological Society of London Special Publication 28*, pp. 203–221. London: Geological Society of London.

Ouachitas

K C Nielsen, The University of Texas at Dallas, Richardson, TX, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Ouachita orogenic belt extends approximately 2100 km along the southern margin of North America (Figure 1). These Palaeozoic sedimentary rocks are associated with the rifted cratonic margin developed during Late Proterozoic to Early Cambrian time. Two promontories, Alabama and central Texas, and two re-entrants, Arkansas/Oklahoma and West Texas, characterized this margin. The Ouachita Mountains of Arkansas/Oklahoma and the Marathon uplift of West Texas developed in these re-entrants. While stratigraphic evidence suggests some regional adjustments coeval with the earlier stages of Appalachian tectonics, the Early Palaeozoic succession of the Ouachita orogen primarily records a long period of stable carbonate deposition along the continental shelf and slow deep-water clastic sedimentation off the shelf. This pre-orogenic phase ended abruptly in the Lower Carboniferous. At that time, the Ouachita basin began subsiding rapidly in response to closure of the Iapetus Ocean and the assembly of Pangaea (*see* **Pangaea**). A thick succession of mud-rich turbidites filled the basin, beginning at various times during the Early Carboniferous (Mississippian). The oldest deformation of this sequence is inferred from the Mississippian clastic wedge of the Black Warrior basin. Additionally, some authors have suggested that early folding documented in the metasediments of the uplifted Ouachita facies may correlate to early subduction in the Ouachita basin. The Late Carboniferous (Pennsylvanian) synorogenic sediments continued filling the basin and were translated inboard over the coeval shelf sequence, forming a generally north-verging fold and thrust belt. Subsidence within the associated foreland basins suggests a diachronous character to this advancing thrust system, beginning in the Late Mississippian in the east and in the Middle Pennsylvanian in the foreland of Oklahoma. Stratigraphic constraints indicate that thrusting ended in the Late Pennsylvanian in Oklahoma while thrusting continued in West Texas until the Lower Permian. The classic Permian carbonate succession unconformably overlies the deformed Marathon sequence. During the later stages of the deformation, basement cored uplifts developed in the Ouachita Mountains of Arkansas and Oklahoma, creating large north-east-trending uplifts. Timing of these uplifts is debateable, but suggests post-thrust deformation and a significant shift in regional shortening, possibly coincident with the later stages of the Alleghanian orogen of the southern Appalachian Mountains.

Unconformable Mesozoic units overlie approximately 80% of the Ouachita Orogenic Belt. Two large exposures, the Ouachita Mountains of Oklahoma and Arkansas and the Marathon uplift of West Texas, provide the bulk of our knowledge about the belt (Figure 1). Geophysical and well data provide the remainder of the database. Two important references serve as the foundation for discussion about the Ouachita orogenic belt and the

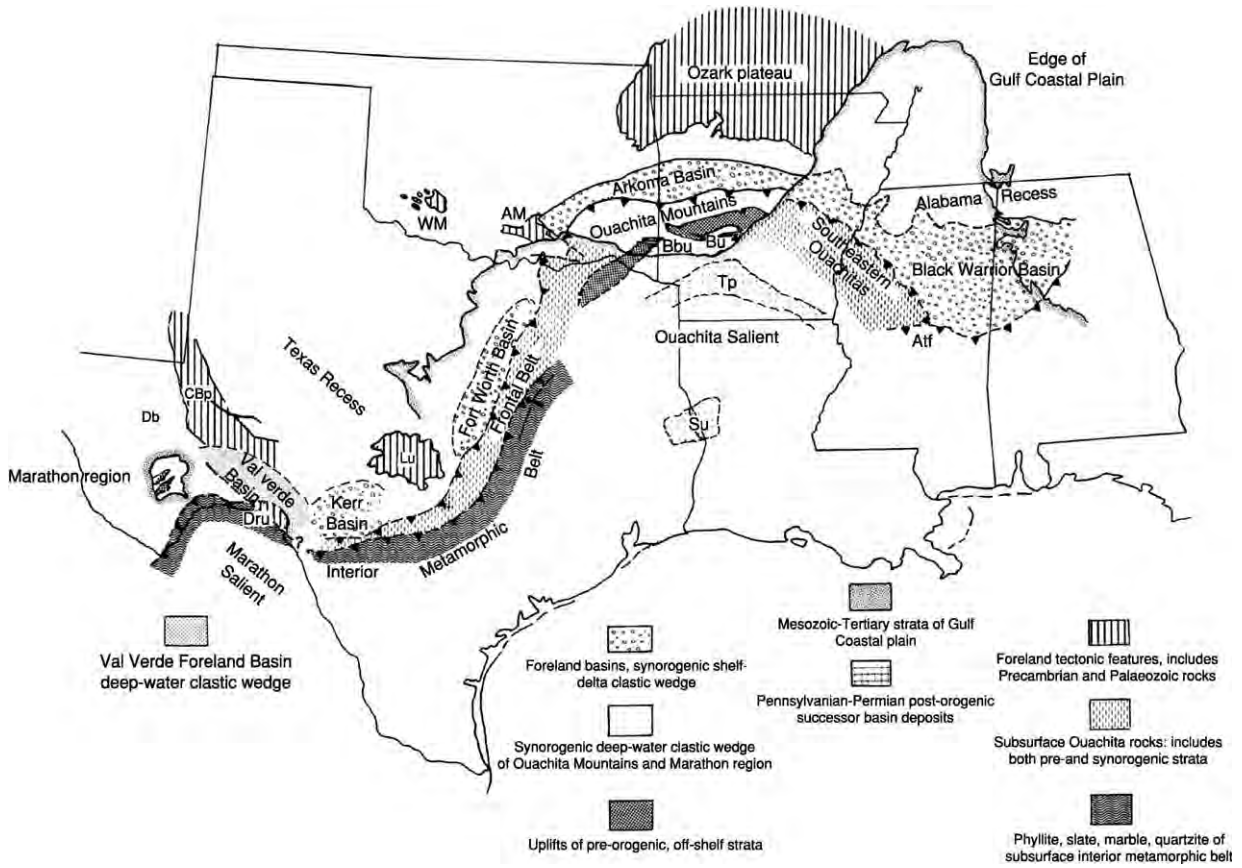


Figure 1 Tectonic elements of the Ouachita Orogen. Abbreviations (AM = Arbuckle Mountains, Atf = Appalachian Tectonic Front, Bbu = Broken Bow Uplift, Bu = Benton Uplift, CBp = Central Basin Platform, Db = Delaware Basin, Dru = Devils River Uplift, Tp = Texarkana Platform, Su = Sabine Uplift, WM = Wichita Mountains). (After Viele and Thomas, 1989.)

interested reader should consult these. The seminal work of Flawn and his colleagues in 1961 provided a complete review of the mountain belt and has been a cornerstone for all subsequent work. 28 years later, the Geological Society of America published a thorough update of the Ouachita Orogen. This discussion is synthesized from these sources.

Cambrian Continental Margin

The trace of the Ouachita Orogen is irregular (Figure 1) and this spatial distribution is directly related to the Late Proterozoic–Early Cambrian rifting of the North American Craton. From central Alabama and the junction with the Appalachian Mountains, the fold and thrust belt extends WNW across the Mississippi embayment into Arkansas and Oklahoma. There the trend shifts to the SSW, extending from Oklahoma into central Texas. At that point the trend shifts westerly continuing into West Texas, only to turn south again near the Marathon Uplift. The mountain belt extends a few hundred kilometers into Mexico before it is obscured by a Mesozoic tectonic overprint. This orthogonal pattern of recesses and salients reflects the rifting geometry along the southern margin of North America during the Late Precambrian and Early Cambrian. The spreading ridge for this rifting is inferred to be subparallel to the east coast of North America and to the south-trending margin in Texas (Figure 1). The connection between Alabama and Oklahoma and between central and West Texas are considered to have been transform boundaries. The Southern Oklahoma aulacogen is coeval with this rifting event and continues the WNW trend of the transform boundary into the craton. Additional Late Precambrian intracratonic rifting has been documented in the Mississippi Valley; the grabens trend NNE from the Ouachita Front through eastern Arkansas and western Tennessee. Igneous rocks exposed in the southern Oklahoma aulacogen reveal a 570–525 Ma bimodal volcanic package consistent with the rift model. Approximately 1 km of Cambrian clastic units in the Mississippi Valley graben indicate additional regional extension. However, outside these rift zones, including the Ouachita Orogen, there is little evidence of this rift sequence.

Stratigraphy

A fairly consistent lithostratigraphic succession characterizes the Palaeozoic sequence of the Ouachita Orogen. These sediments were deposited in two distinct phases. A pre-orogenic package consists of carbonate shelf sediments north and west of the continental margin and a deep-water shale, sandstone, and chert sequence on the continental slope or rise south and east. The second phase is dominated by syn-orogenic turbidites filling the basin and fluvial-deltaic sediments covering the shelf.

Pre-Orogenic Sequence

A classic carbonate shelf sequence accumulated on the craton over the Precambrian basement during the Early to Middle Palaeozoic (Figure 2). A thin basal sandstone unit is in contact with the Precambrian basement. This sandstone is overlain by a transgressive carbonate facies. In the Southern Oklahoma aulacogen, this carbonate succession is approximately 3.5 km thick and coeval sequences on the craton is approximately 1 km thick (Figure 2). Within the Arkoma Basin, this sequence is 1.2 km thick and about 2 km thick in the Black Warrior Basin. In many places, such as the Arkoma Basin, this pre-orogenic package records nearly continuous deposition and consistent thickness. The section in the Black Warrior Basin records influences of the Taconic Orogeny (Upper Ordovician) along the south-east margin of the basin and reveals more rapid deposition of the Devonian chert sequences along the south-west margin.

Within the Ouachita Basin, south and east of the rift margin, approximately 2–4 km of 'Ouachita facies' accumulated in generally deep marine settings (Figure 2). These shale, sandstone, and chert units currently form most of the exposures in the central uplifts of the Ouachita Mountains. Basinal pre-orogenic units have been subdivided into a lower shale-dominated section with two sandstone units followed by an upper black siliceous shale and chert succession with one interbedded sandstone (Figure 2). The change in lithology is interpreted as a function of progressively slower depositional rate and a similar subsidence history can be seen off the east coast of North America on the Blake Plateau. Current directions for the turbidites in the pre-orogenic sequence indicate that much of the deposition was coming from the north, but that the Silurian Blaylock Sandstone was sourced from the south-east: therefore the basin may have been two-sided during the pre-orogenic phase. The transition between shelf and deep water deposits is missing in the Oklahoma salient, due to subsequent shortening of the sequences. In the Marathon Region, there is a thinner pre-orogenic sequence that can also be described in two phases (Figure 2). The early units, in contrast to the Ouachita Mountains, include a lower turbidite, an overlying deep-water carbonate sequence (Marathon Formation) with overlying deep-water shale, calcarenite, conglomerates, and boulder beds. The upper phase is a chert and limestone sequence, similar in character to the chert sequence in the Ouachita Mountains. There has been significant debate about whether these cherts were deposited in shallow or deep-water settings. There is little argument about the fact that they record a long period of very slow sedimentation.

Syn-Orogenic Sequence

The tectonic picture changed dramatically in the Lower Carboniferous and this is reflected in the stratigraphic sequence. The subsidence rate and sediment input increased. The result was a rapid change up-section to deep-water shales and sandstones with interbedded debris flows. Within the Ouachita Basin, this contact appears to have been gradational, although some chert breccias have been found near the base of the section. These turbidite sequences filled the basin from east to west throughout the Carboniferous. Thickness estimates vary, but generally range from 12–14 km of section in the Ouachita Mountains where four major units are identified: Stanley and Jackfork Groups, Johns Valley, and Atoka Formations. The Stanley Group is the oldest of these flysch units, extending from the Middle to Late Mississippian. The characteristic lithologies include shale, sandstone, tuff, impure chert, and siliceous shale. Olistostromal deposits are frequent and several thin volcaniclastic units are found near the base of the Stanley Group around the Broken Bow Uplift. Gradationally overlying the Stanley Group, the Jackfork Group is characterized by prominent ridge-forming sandstone units. The Johns Valley Formation is distinguished from the underlying, gradational Jackfork Group by the presence of abundant exotic boulders and associated contorted shale. These boulders are correlated to the shelf sequence along the North American margin as well as the pre-orogenic Ouachita facies. These boulder beds are concentrated along the north-west margin of the thrust belt. The overlying unit is the widely spread Atoka Formation. Within the Ouachita Basin, the Atoka Formation is a typical turbidite, with nearly equal proportions of sandstone and shale. The maximum thickness of 8.5 km is found in the frontal Ouachita Mountains, but 2.5–4.9 km of coeval deltaic sandstones are found across the Arkoma Basin. The Atoka Formation thickens across regional, down-to-the-south normal faults, generated by tectonic loading of the continental

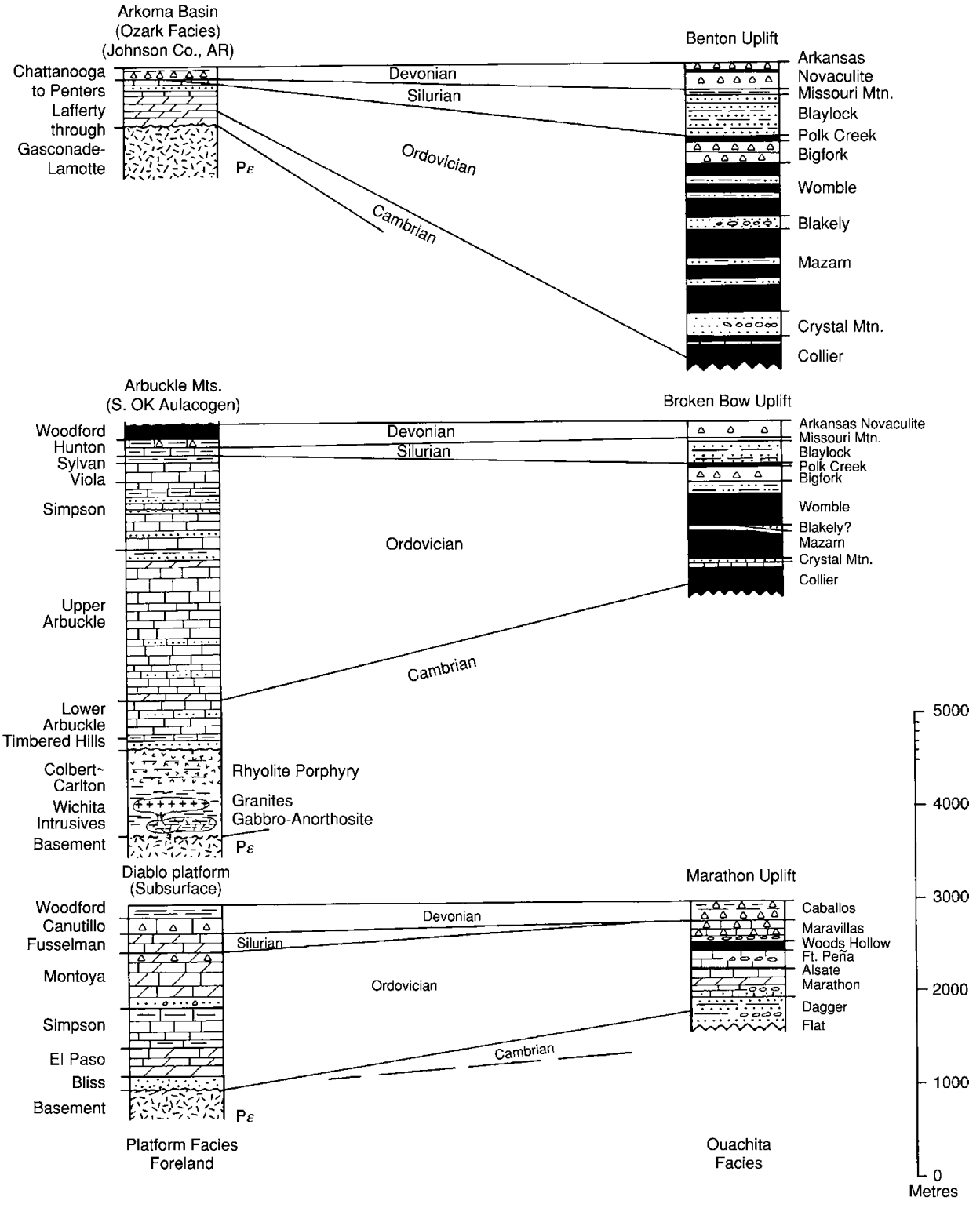


Figure 2 Representative lithological columns of the pre-orogenic sequences at various settings along the Ouachita Orogen. (After Arbenz, 1989.)

margin. A northward migrating depositional centre characterizes the syn-orogenic sequence. The Stanley Group is 3.3 km thick just north of the Broken Bow Uplift. The Jackfork Group reaches a maximum thickness of 2.1 km along the axis of the Ouachita Mountains, and the Atoka Formation is thickest in the frontal Ouachita Mountains. While the entire section is allochthonous, these displaced depocentres are believed to be in relative order. Within the thrust belt, the Atoka Formation is the youngest deformed unit, but in the Arkoma Basin the Atoka Formation is overlain by additional deltaic deposits of Late Pennsylvanian Carboniferous age. The syn-orogenic sequence of the Marathon Uplift is also dominated by turbidite deposition but the sequence is thinner (Figure 3). The lower turbidite is 2.0 km thick in the eastern portion of the basin and thins to approximately 100 metres of shale in the west. There are abundant soft sediment deformational features in the section and these indicate extension parallel to the north-west trending palaeocurrents. The Tesnus Formation extends through the Mississippian into the Early Pennsylvanian and is overlain by the Dimple Limestone. This thin carbonate turbidite contains a clast assemblage of cratonic shelf rocks. The palaeocurrent data confirms a north-west source. The Haymond Formation is a repetitive siliciclasitc turbidite package that maintains a consistent 1.2 km thickness across the Marathon Basin. In this case, the palaeocurrents are trending WSW. The Haymond Formation includes important boulder beds that yield Early Palaeozoic rocks representing the pre-orogenic sequence and a full range of exotic rock types including various igneous clasts, some of which are Siluro-Devonian. These exotic clasts are commonly interpreted to have come from a southern source. The youngest unit in the Marathon Basin is Latest Pennsylvanian molasse type of sediments of the Gaptank Formation.

Regional Subdivisions

The Ouachita Orogen can be subdivided into four parts. The eastern section extends from central Mississippi north-westwards into Arkansas. This section is entirely in the subsurface. North of the tectonic front, a passive margin carbonate shelf of Cambrian to Early Mississippian age is overlain by Late Mississippian to Middle Pennsylvanian shallow marine to deltaic sediments. This sequence dips beneath the tectonic front which translates deep-water undifferentiated mudstones, cherts, and sandstones of the Ouachita facies.

The Ouachita Mountains represent the largest outcrop area along this orogenic belt (Figure 1). Similar to the pattern documented to the east, the carbonate shelf and overlying Carboniferous clastic sequences are structural below the allocthonous Ouachita facies. A minimal estimate for this translation is in the order of 100 km and maybe much greater. The Ouachita Mountains are commonly subdivided into the frontal imbrication zone between the Choctaw/Y-City and Windingstair faults, the central zone extending southward to the Broken Bow-Benton Uplifts, and then the Athens Plateau south of the uplifts. The imbrication zone is characterized by closely spaced thrust sheets made up of the younger Carboniferous units. The central zone, on the other hand, contains older Carboniferous units folded into a few long wavelength (8–16 km), north vergent synclines each truncated by a large thrust fault. The uplift areas trend ENE obliquely across the general east–west structural fabric of the Ouachita Mountains. These uplifts expose intensely deformed Ouachita facies and the Lower Stanley Group. The deformational history is more complicated than seen in the overlying central zone with at least three folding events and associated cleavage development recognized. Two distinctive aspects of these older rocks include a significant development of southerly verging folds and a disharmonic relationship to the large-scale folds of the central zone. The uplifts are also the location of low-grade metamorphism as reflected in recrystallized cherts (novaculite), chlorite growth, vitrinite reflectance, and ubiquitous quartz veins. The maximum temperature for these metasediments approach 300°C in the southern Broken Bow and eastern Benton Uplifts (Figure 4). The Athens Plateau has a structural style similar to the central zone to the north-west, northerly verging faults-truncating the Carboniferous sequence.

Within Texas, the Ouachita Orogen has been traced almost 1000 km in the subsurface. The structural subdivisions of the Ouachita Mountains have been extended (Figure 1). In this case, the 'frontal zone' includes all of the unmetamorphosed northerly verging structures inboard of the uplifts. The 'interior' zone is correlated to the uplifts of the Ouachita Mountains and includes deformed metasediments presumed to be equivalent to the pre-orogenic sequence. The Athens Plateau sequence is defined as the 'southern Carboniferous province' in the subsurface, but is only documented immediately south of the exposed Ouachita Mountains. The frontal zone varies in width and can be followed south to central Texas. The most distinctive structural element in the frontal zone is the leading thrust as this boundary is clearly proven by drilling. The interior zone extends from the Broken Bow Uplift, but is not continuous across north-east Texas (Figure 1). In general, the

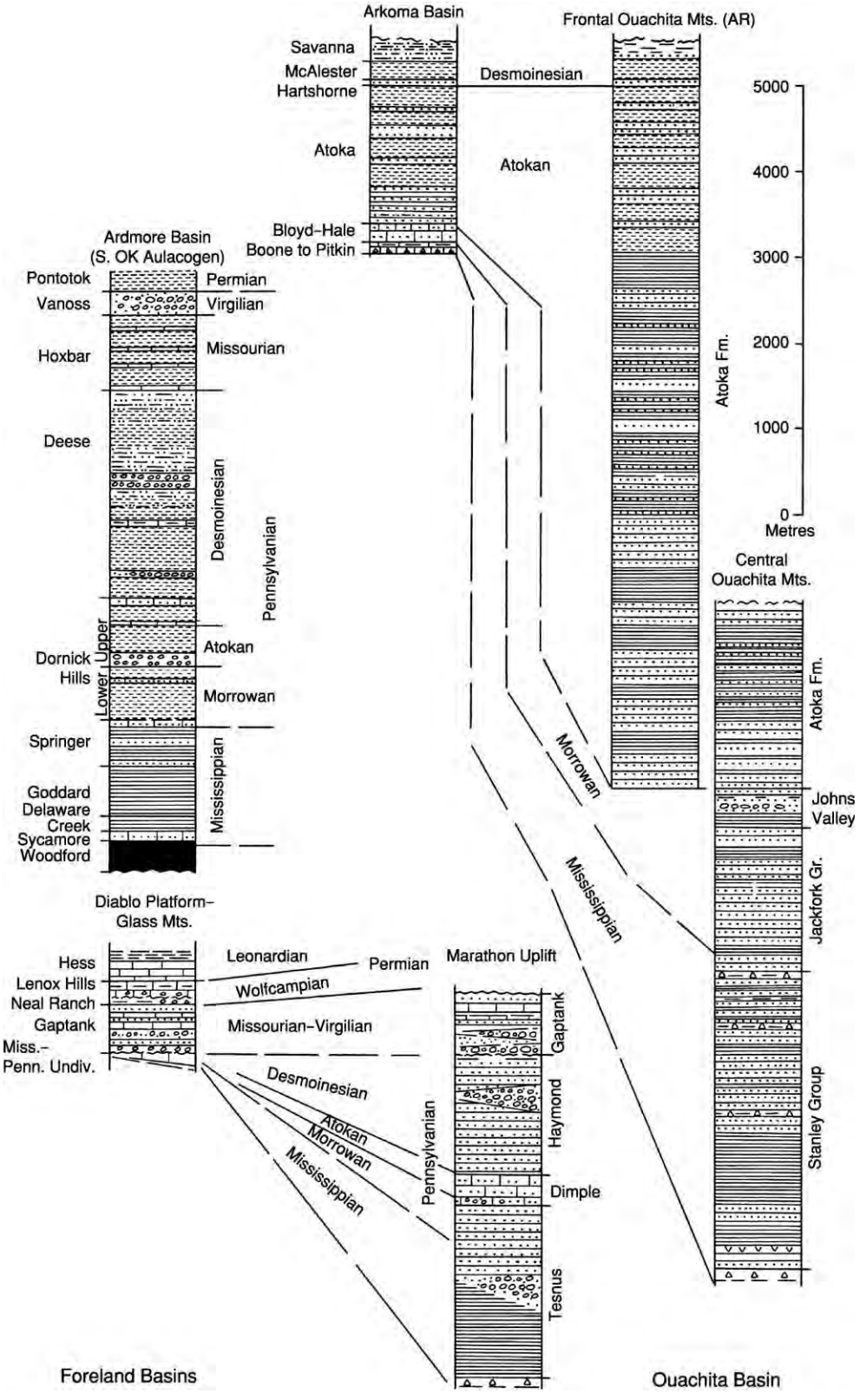


Figure 3 Representative lithological columns of syn-orogenic sequences showing the transition from the craton to the Ouachita Basin. (After Arbenz, 1989.)

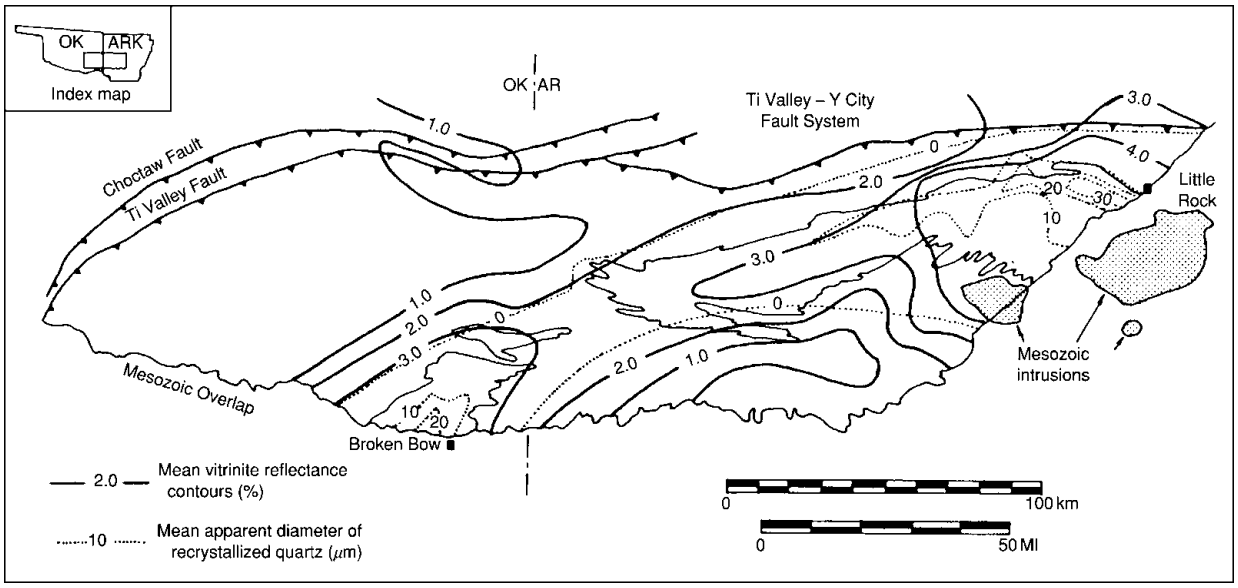


Figure 4 Thermal maturity of exposed strata of the Ouachita Mountains. (After Viele and Thomas, 1989.)

interior zone follows the frontal zone and together they form an 80 km wide belt. The frontal zone appears to be cut out in south-central Texas only to reappear east of the Marathon Uplift. Two notable uplifts are documented along this subsurface trend: the Waco Uplift along the south-trending section and the Devils River Uplift along the westerly-trending section. The first is interpreted to be a segment of the North American Craton thrust inboard carrying an overlying allocthonous section of Ouachita facies. The second uplift is an approximately 120 km by 50 km northwest-trending Precambrian block juxtaposed to the Carboniferous foreland basin.

The last segment of the Ouachita Orogen is located in West Texas and includes the exposures of the Marathon Uplift and two smaller outcrops to the south-west (Figure 1). Similar to the Ouachita Mountains, the entire outcrop belt of Marathon Uplift is allochthonous. This area is correlated to subsurface frontal zone. The stratigraphic sequence is similar to that of the Ouachita Mountains. The folding style is directly related to the nature of the lithologic package, long wavelength folds in the thick flysch sequence and complex, tight folding and faulting in the pre-orogenic sequence. The measured shortening has been utilised to restore this belt to a palaeogeographic position as much as 200 km to the south-east.

Five foreland basins are an integral part of the Ouachita Orogen (Figure 1). From east to west, these include the Black Warrior, Arkoma, Fort Worth, Kerr, and Val Verde basins. The Black Warrior Basin is critical in that it is bordered on two sides by convergent zones; the Ouachita thrust faults cut the southwest side of the basin and the Appalachian thrust belt on the south-east intersects these faults. The Black Warrior Basin is a homocline dipping gently towards the tectonic front and extending beneath the front. This homocline rises up to a large domal structure, the Nashville Dome, to the north. The pre-orogenic sequence rests on Precambrian basement and is dominated by carbonate shelf deposition. The Upper Mississippian-Pennsylvanian syn-orogenic clastic wedge progrades north-eastward. This section thickens towards the tectonic front and is consistent with tectonic loading at a convergent margin. The other foreland basins have similar geometries and pre-orogenic successions. They differ primarily in the nature of the syn-orogenic sequence, both in timing and depositional environments.

Regional Geophysics

The Bouguer anomaly map represents the most extensive geophysical data set for the region (Figure 5). A positive anomaly associated with the interior zone can be followed from central Arkansas through Texas into the Marathon region. The extension of this belt from Arkansas into Mississippi is more obscure. While uplifted basement blocks, mafic intrusions, and metamorphism contribute to the anomaly, the anomaly requires a major transition in crustal structure

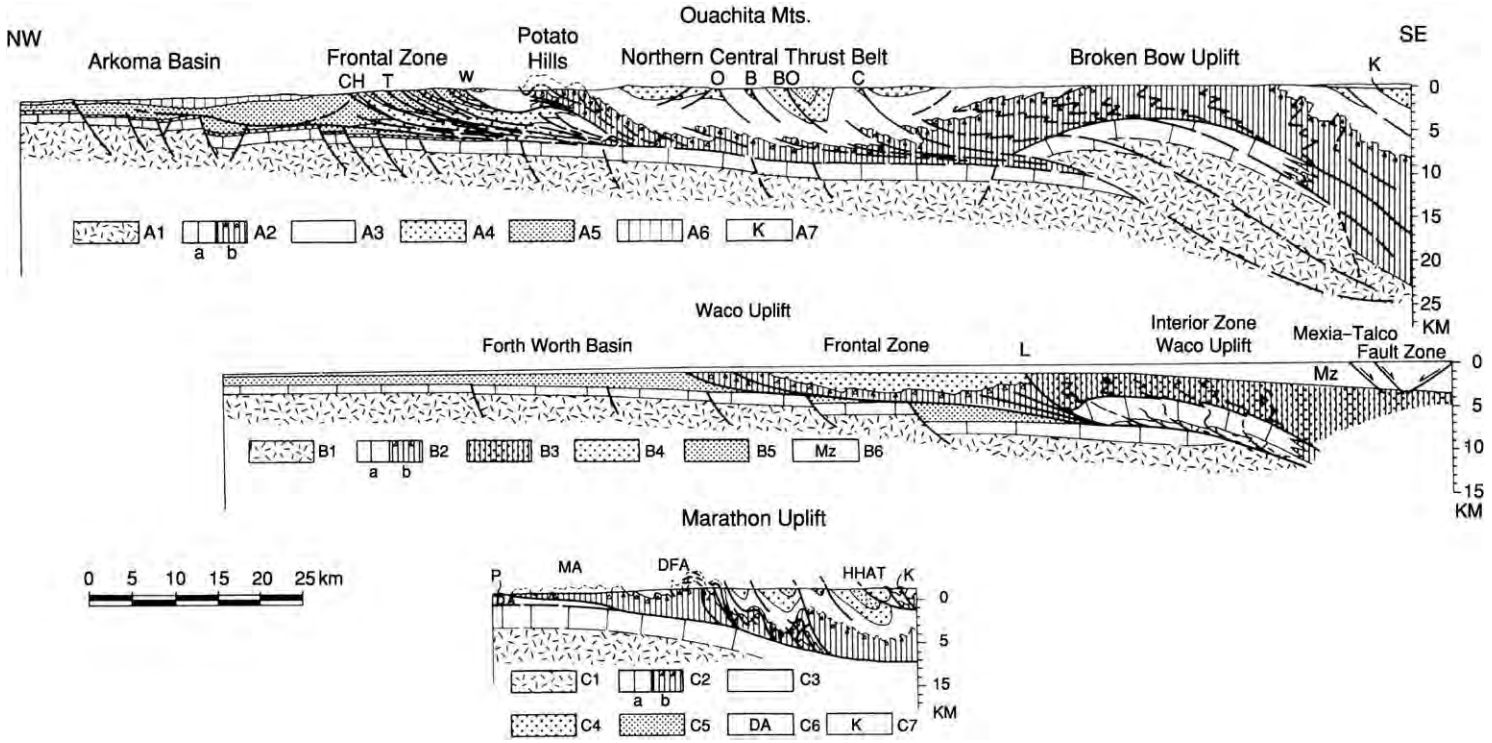


Figure 5 Generalized cross-sections across the Ouachita Orogen in Oklahoma and Texas. (From Hatcher *et al.*, 1989) (Modified from Arbenz, 1989.)

across this trend. In other words, it is proposed that this anomaly represents the Early Palaeozoic continental margin. Of similar amplitude is the northwest-trending positive anomaly associated with the Southern Oklahoma aulacogen (Figure 6). Allowing for the uplifted Precrambrian blocks and mafic intrusions, this implies significant crustal modification during rifting. The aulacogen trend is clearly developed from the Panhandle area of Texas southeastward to the Ouachita Orogen where the two trends form a distinctive X-pattern. Closed gravity minima are observed for three of the foreland basins: Arkoma, Fort Worth, and Kerr basins. Gravity lows are noted in the Black Warrior and Val Verde basins, but the pattern is more subdued. Of these Basins, the Arkoma Basin is the most dramatic (~80 mgal). Interestingly, the lowest values of this minimum are located just north of the Broken Bow Uplift where a steep gradient is also apparent. This pattern argues for a thick sedimentary section in the central Ouachita Mountains, some of it possibly below the allocthonous Ouachita facies. The Benton, Broken Bow, and Waco uplifts are actually located on the gradient inboard of the interior zone anomaly. Thus, these uplifts have a similar density to the Ouachita facies; that they are not deeply rooted in the basement and have been translated significant distances. In addition, the uplifts probably reflect late stage compressional deformation along the transitional crust of the continental margin. In contrast, the Devils River Uplift does reveal a distinct positive anomaly related to the large Precambrian block.

Regional seismic data is widely scattered. Early refraction and surface wave studies document a normal continental crust north and west of the margin and an ambiguous noncratonal crust model south of the margin. The COCORP line in southwestern Arkansas is the longest line across the fold and thrust belt. Crustal models based on the COCORP and additional data to the south show an increase in velocities, which have been interpreted to be interleaved granitic and basaltic crust. In this area, the Early Palaeozoic transitional crust is thought to be preserved; and the deformed Palaeozoic section is detached and transported northward over the continental margin. Seismic reflection data support the allocthonous, northerly verging geometry of this fold and thrust belt. Additional features identified in

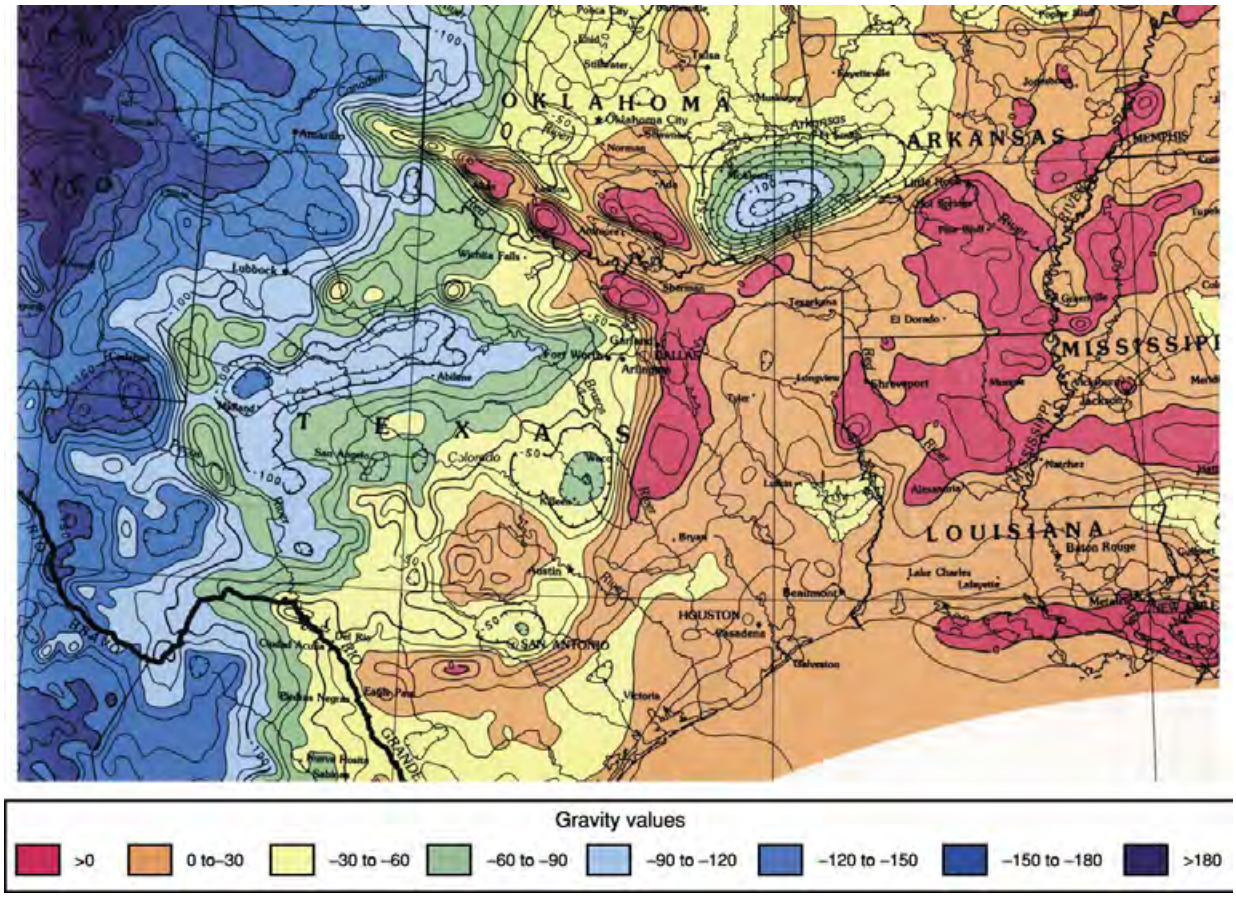


Figure 6 Bouguer Anomaly Map of south-central North America. Contour Interval is 10 Milligals. (After Keller *et al.*, 1989.)

the COCORP line are the basement cored uplift, Early Palaeozoic sections overlying a presumed Precambrian basement, allocthonous Ouachita facies structurally overlying the Early Palaeozoic section, and a tectonically thickened section of Carboniferous units south of the uplift. A final feature identified in the subsurface is a successor basin of Late Pennsylvanian sediments in unconformable contact with the deformed Carboniferous units south of the Ouachita Mountains (Figure 1).

Tectonic Synthesis

Representative cross-sections provide a clear picture of the structural style of the Ouachita Orogenic Belt (Figure 6). Common to these sections is the dominant thin skin style of deformation. The amount of translation is minimally 50 to 100 km; the greatest of these displacements appear to be in the salients. The foreland basin reveals the tilt of the basement and the pre-orogenic sequence towards the tectonic zone. Also, the strongly asymmetric nature of the foreland basin is outlined. In the Ouachita Mountains, the extremely thick sedimentary succession north of the Broken Bow Uplift is illustrated along with the large folds of the central zone. The basal detachment is above the shelf sequence, possibly with a Carboniferous section also in the footwall. Commonly, the detachment is interpreted to extend over the basement uplift. The pattern is similar in the Waco Uplift, although the basement block is much thinner. The Marathon cross-section is limited to the exposed uplift and reflects a style similar to the frontal zone of the Ouachita Mountains, but well data has extended the thrust further to the south where some metamorphic rocks have been associated with the 'interior zone'.

The Ouachita Orogen represents the imprint of the collision of North America, South America, and possibly several small continental blocks and/or island arcs. The specific nature of these blocks is debatable; however, there is agreement about the general timing. Initial subsidence is recorded by the shift from the Devonian chert and siliceous shale deposition to the Carboniferous turbidite sequence. In the Ouachita Mountains, this transition is in the Lower Mississippian. Similar timing is envisioned for the Caballos Novaculite of the Marathon region. Coincident with the change in the basin, the shelf areas of the Black Warrior Basin experienced a change from the Lower Mississippian shallow marine chert and limestone to an Upper Mississippian clastic wedge. This Mississippian clastic wedge prograded northeastward, reflecting a source to the south-west. In the Arkoma Basin, however, the shelf was stable with Mississippian shallow marine limestone grading southward into mudstones. A second clastic wedge in the Black Warrior Basin of Pennsylvanian age is more extensive, and merges with a another wedge prograding from the Appalachian tectonic front. Within the Arkoma Basin, deltaic sedimentation on a stable shelf continued until Middle Pennsylvanian (Lower Atokan). This was followed by the development of down-to-south normal faults and syn-orogenic deposition. Thrust faults truncate the Pennsylvanian sediments in the foreland basins. The Arkoma Basin is divided into a southern compressional zone and a northern extensional zone related to plate loading and normal fault development. The youngest units of the Arkoma Basin involved with the Ouachita Orogenesis are mid-Desmoinesian (Late Pennsylvanian). Coincident with this compression is the development of a successor basin south of the exposed Ouachita Mountains. Desmoinesian shallow marine deposits rest unconformably on deformed Ouachita Carboniferous units. In the Marathon area the compressional deformation is subdivided into two phases. The first phase corresponded with the folding of the Dimple Limestone, formation of the Val Verde foreland Basin, and the transition syn- to post-orogenic sediments (Gaptank Formation). The timing of this phase is Middle to Upper Pennsylvanian. The younger phase deformed these folded and faulted Pennsylvanian units as part of the development of the Dugout Creek Thrust. This younger phase extended into the Early Permian (Wolfcampian).

While initial subsidence appears to be nearly coincident along the length of the Ouachita Basin, the compressional events began earlier in the east (Middle Mississippian). It has been suggested that this early phase was related to a collision with an island arc along the south-east margin of North America. The 'two tier' fold style developed in the Ouachita Mountains is interpreted as reflecting deformation in two distinctly different settings. The earliest deformation involved Cambrian through to Early Mississippian basinal sediments. Tight folds and faults developed in a growing accretionary wedge above the oceanic and transitional continental crust. The remainder of the Carboniferous flysch sequence was then deposited over these deformed units and the entire accretionary prism was thrust onto the southern margin of North America in the Late Pennsylvanian. As mentioned above, the final stages of thrust faulting were Desmoinesian in the Ouachita Mountains. The principal thrusting began about this time in West Texas and continued into the Early Permian. Therefore a diachronous model is commonly applied to the Ouachita Mountains, closing from east to west. Thrust faults of the Appalachian Alleghenian

Orogeny truncate the Ouachita trends in the Black Warrior Basin. The final stages of deformation in the Ouachita Mountains are also documented in the rotated folds and cleavages of the Broken Bow and Benton uplifts. These basement cored structures trend ENE and reflect a rotation of the principal shorten direction. It has been proposed that these uplifts represent the far-field effects of the Alleghenian collision to the south-east. Various microplates have been proposed to account for the translation of the Ouachita facies inboard over the coeval shelf sequences. Collectively these have been identified as the Sabine Plate and represent elements trapped between South and North America during the assembly of Pangaea. Subsequent Mesozoic rifting associated with the development of the Gulf of Mexico have unfortunately obscured these relationships.

See Also

North America: Northern Cordillera; Southern Cordillera; Southern and Central Appalachians. **Pangaea**.

Further Reading

Arbenz JK (1989) The Ouachita system. In: Bally AW and Palmer AR (eds.) *The Geology of North America: An Overview,* v. A, pp. 371–396. Boulder, Colorado: Geological Society of America, The Geology of North America.

Ethington RL, Finney SC, and Repetski JE (1989) Biostratigraphy of the Paleozoic rocks of the Ouachita orogen, Arkansas, Oklahoma, west Texas. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States,* v. F-2, pp. 563–574. Boulder, Colorado: Geological Society of America, The Geology of North America.

Flawn PT, Goldstein A Jr., King PB, and Weaver CE (1961) *The Ouachita system*. Austin, University of Texas, Bureau of Economic Geology Publication number 6120, p. 401.

Hatcher RD Jr., Thomas WA, and Viele GW (1989) *The Appalachian-Ouachita Orogen in the United States,* v. F-2, p. 767. Boulder, Colorado: Geological Society of America, The Geology of North America.

Keller GR, Kruger JM, Smith KJ, and Voight WM (1989) The Ouachita system: A geophysical overview. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States,* v. F-2, pp. 689–694. Boulder, Colorado: Geological Society of America, The Geology of North America.

King PB (1937) *Geology of the Marathon region, Texas,* U.S. Geological Survey Professional Paper 187, p. 148.

Lillie RJ, Nelson KD, de Voogd B, *et al.* (1983) Crustal structure of Ouachita Mountains, Arkansas: A model based on integration of COCORP reflection profiles and regional geophysical data. *American Association of Petroleum Geologists Bulletin* 67: 907–931.

Lowe DR (1989) Stratigraphy, sedimentology, and depositional setting of pre-orogenic rocks of the Ouachita Mountains, Arkansas and Oklahoma. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States;* v. F-2, pp. 575–590. Boulder, Colorado: Geological Society of America, The Geology of North America.

McBride EF (1989) Stratigraphy and sedimentary history of Pre-Premian Paleozoic rocks of the Marathon uplift. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States:* v. F-2, pp. 603–620. Boulder, Colorado: Geological Society of America, The Geology of North America.

Morris RC (1989) Stratigraphy and sedimentary history of post-Arkansas Novaculite Carboniferous rocks of the Ouachita Mountains. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States:* v. F-2, pp. 591–602. Boulder, Colorado: Geological Society of America, The Geology of North America.

Muehlberger WR and Tauvers PR (1989) Marathon fold-thrust belt, west Texas. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States,* v. F-2, pp. 673–680. Boulder, Colorado: Geological Society of America, The Geology of North America.

Nicholas RL and Waddell DE (1989) The Ouachita system in the subsurface of Texas, Arkansas, and Louisiana. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States;* v. F-2, pp. 661–672. Boulder, Colorado: Geological Society of America, The Geology of North America.

Nielsen KC, Viele GW, and Zimmerman J (1989) Structural setting of the Benton-Broken Bow uplifts. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States:* v. F-2, pp. 635–660. Boulder, Colorado: Geological Society of America, The Geology of North America.

Thomas WA (1989) The Appalachia-Ouachita orogen beneath the Gulf Coastal Plain between the outcrops in the Appalachian and Ouachita Mountains. In: Hatcher RD (ed.) *The Appalachian-Ouachita Orogen in the United States:* v. F-2, pp. 537–553. Boulder, Colorado: Geological Society of America, The Geology of North America.

Viele GW and Thomas WA (1989) Tectonic synthesis of the Ouachita orogenic belt. In: Hatcher RD (ed.) *The Appalachian-Orogen in the United States:* v. F-2, pp. 695–728. Boulder, Colorado: Geological Society of America, The Geology of North America.

Southern and Central Appalachians

R D Hatcher, Jr, University of Tennessee, Knoxville, TN, USA

© 2005, Elsevier Ltd. All Rights Reserved.

> The Appalachian Chain is the most elegant on Earth, so regularly arranged that its belts of formations and structures persist virtually from one end to the other.... But the apparent simplicity is deceiving; actually it is full of guile, and its geology has aroused controversies as acrimonious as any of those in our science.
>
> (Philip B King, 1970)

Introduction

The exposed Appalachians extend from Alabama to Newfoundland, with buried components extending eastward and southward beneath the Gulf and Atlantic coastal plains, and the Atlantic continental shelf (Figure 1). The southern and central Appalachians reach their greatest width in Tennessee, the Carolinas, and Georgia. To the north, the chain reaches its narrowest point at the latitude of New York City, and from there it widens northward again into New England. Physiographical provinces from north-west to south-east include the Cumberland–Allegheny Plateau, Valley, and Ridge from Pennsylvania southward, the Blue Ridge from South Mountain in southern Pennsylvania to northern Georgia, and the Piedmont from New Jersey to Alabama. Basins filled with Late Triassic–Early Jurassic sedimentary (and some volcanic) rocks that formed prior to the opening of the present Atlantic Ocean are superposed on the Piedmont from North Carolina northward. Physiographical provinces to the west correspond roughly to geological provinces, but there is little correspondence with geological provinces in the interior of the chain. The southern Appalachians contain the highest mountains in the entire chain. Though there is one mountain (Mt Washington, New Hampshire) in New England with an elevation greater than 6000 ft (~2000 m), and none farther north, there are at least 39 named peaks in the southern Appalachian Blue Ridge of western North Carolina and eastern Tennessee with elevations exceeding 6000 ft. Topographic relief here can exceed 5000 ft (1600 m).

The Appalachians were created following the rifting of the Rodinia supercontinent during three or more Palaeozoic orogenic events: the Ordovician–Silurian Taconian, the Devonian Acadian and Devonian–Mississippian Neoacadian, and the Pennsylvanian–Permian Alleghanian orogenies (Figure 2). All orogenies were diachronous. The character of the orogen changes roughly at the latitude of New York City, from one constructed largely from west to east to the north, to an orogen built from the inside out to the south, with the exception of the Alleghanian Thrust Belt that extends up the Hudson River Valley to just south of Albany, New York (Figure 3).

Philip King doubtlessly had the southern and central Appalachians in mind when he made his observation about the Appalachians. These mountains have been studied for over 150 years, and several fundamental concepts in geological science were developed in the process. The continentward segment of the chain consists of a foreland fold–thrust belt that changes style within; it is fold-dominated from Pennsylvania to south-west Virginia, thrust-dominated from south-west Virginia to Alabama, and fold-dominated again in Alabama (Figure 3). This along-strike change in character is related to changes in mechanical stratigraphy, to relationships of the shape of the original continental margin, and to the shape of the Gondwanan indenter that collided with Laurentia at the end of the Palaeozoic. The internal parts of the chain to the east initially consist of deformed rifted margin sedimentary rocks and recycled Grenville basement. These give way eastward to accreted terranes of distal Laurentian metasedimentary oceanic and arc-affinity material (Tugaloo, Chopawamsic, Potomac, and Baltimore terranes), then a sequence of mid-Palaeozoic metasedimentary rocks (Cat Square Terrane), and a truly exotic peri-Gondwanan (Pan-African) Carolina–Avalon Superterrane of Neoproterozoic to Ordovician arc volcanic and volcaniclastic rocks, including the Smith River allochthon (Figure 3). The subsurface Suwannee Terrane lies south of an east- and west-trending suture zone in southern Georgia and Alabama (Figure 4), and contains Gondwanan-affinity basement and cover sedimentary rocks to the south. The Appalachian collage records a complete Wilson Cycle of the opening and closing of a series of Palaeozoic oceans, starting with rifting of the supercontinent Rodinia and ending with the assembly of the supercontinent Pangaea.

The reference frame used throughout this article is the present geography. However, the equator trended north–south (present geographic frame), bisecting North America during the Early and Middle Palaeozoic. Laurentia rotated counterclockwise during the Middle to Late Palaeozoic, bringing the continent into its present orientation, but still at equatorial

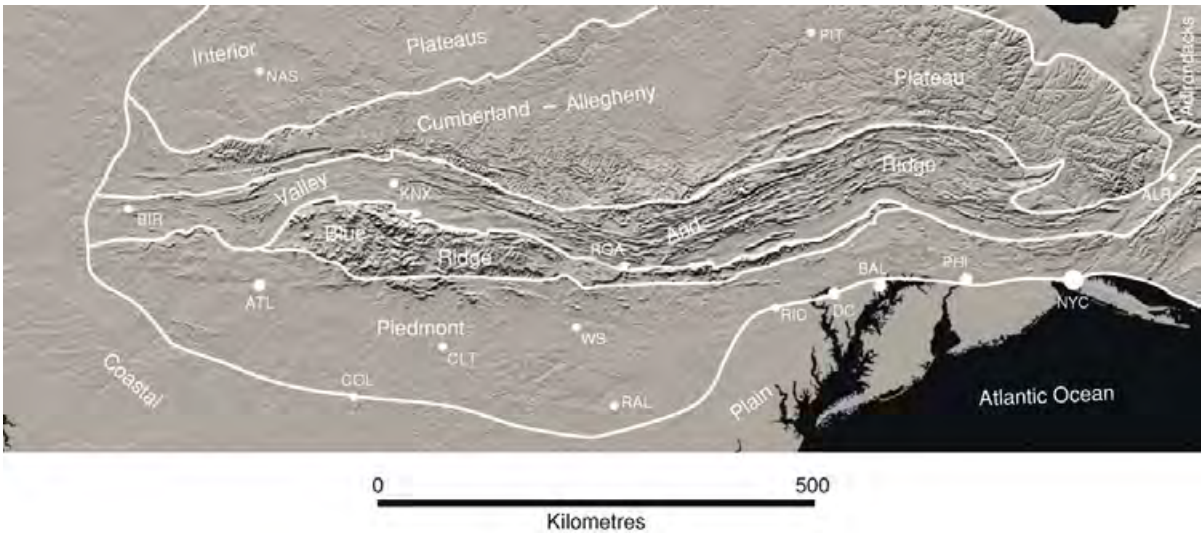


Figure 1 Shaded relief map showing the physiographical provinces of the southern and central Appalachians. ALB, Albany, New York; ATL, Atlanta, Georgia; BAL, Baltimore, Maryland; BIR, Birmingham, Alabama; CLT, Charlotte, North Carolina; COL, Columbia, South Carolina; DC, Washington, DC; KNX, Knoxville, Tennessee; NAS, Nashville, Tennessee; NYC, New York City; PHI, Philadelphia, Pennsylvania; PIT, Pittsburgh, Pennsylvania; RAL, Raleigh, North Carolina; RIC, Richmond, Virginia; ROA, Roanoke, Virginia; WS, Winston-Salem, North Carolina. Image from the United States Geological Survey.

latitude. Most subsequent movement of Laurentia during the Mesozoic and Cenozoic has been northward to its present location north of the equator.

Neoproterozoic to Ordovician Iapetan Margin Development and Destruction

Rifting of Rodinia

Supercontinent Rodinia formed 1.2–1 billion years ago (Ga) as a product of the worldwide Grenvillian Orogeny. The continent remained intact for ~300 million years (My) and then rifted apart at ~750 Ma in western and north-eastern North America, but failed rifting occurred at ~735 Ma in the southern and central Appalachians. Failed rifting, however, produced the alkalic A-type Crossnore plutonic–volcanic suite; the Grandfather Mountain and Mount Rogers Formation marine, non-marine, and glaciogenic sedimentary and bimodal volcanic rocks in North Carolina, Tennessee, and Virginia; and the Robertson River igneous complex in central Virginia. Successful rifting of Rodinia along the entire southern-central Appalachian margin occurred at ~565 Ma, with basins that deposited the Ocoee Supergroup in Tennessee, North Carolina, and Georgia, and the Catoctin volcanic–sedimentary assemblage in Virginia. Neither of these assemblages is physically connected, because initial rifting created an irregular margin in south-eastern Laurentia, leaving basement highs that separated rift basins.

Deposition of the Ocoee Supergroup locally reached thicknesses exceeding 15 km, with most of the thickness contained in the Great Smoky Group. The Ocoee Supergroup contains three groups: (1) the Snowbird Group, which rests unconformably on Grenvillian basement, (2) the Great Smoky Group, which nowhere has been observed in direct contact with basement rocks and is faulted onto the Snowbird Group rocks, and (3) the Walden Creek Group, which conformably overlies both the Great Smoky Group and the Snowbird Group, revealing an extraordinarily complex relationship between the components of the Ocoee Supergroup. The Snowbird Group consists of immature basal sandstone overlain by much cleaner quartz arenite to arkosic sandstone, then a pelitic unit. The Great Smoky Group consists of a rapidly deposited deep-water sequence of immature greywacke and pelite (some of which was deposited in an anoxic environment), minor amounts of cleaner sedimentary rocks, and polymictic vein quartz-dominated conglomerate that also contains quartzite, sandstone, granitoid, black shale rip-ups, and variably textured limestone and dolostone clasts. The Walden Creek Group consists of banded chlorite–sericite slate that contains minor amounts of limestone overlain by shale, clean sandstone, and limestone. The latter unit has yielded soft-bodied wormlike fossils, and possible ostracodes and inarticulate brachiopods, indicating a Cambrian (probably Tommotian) age. Coarse polymictic vein quartz-dominated conglomerate facies also occur within this unit, and again the clasts are

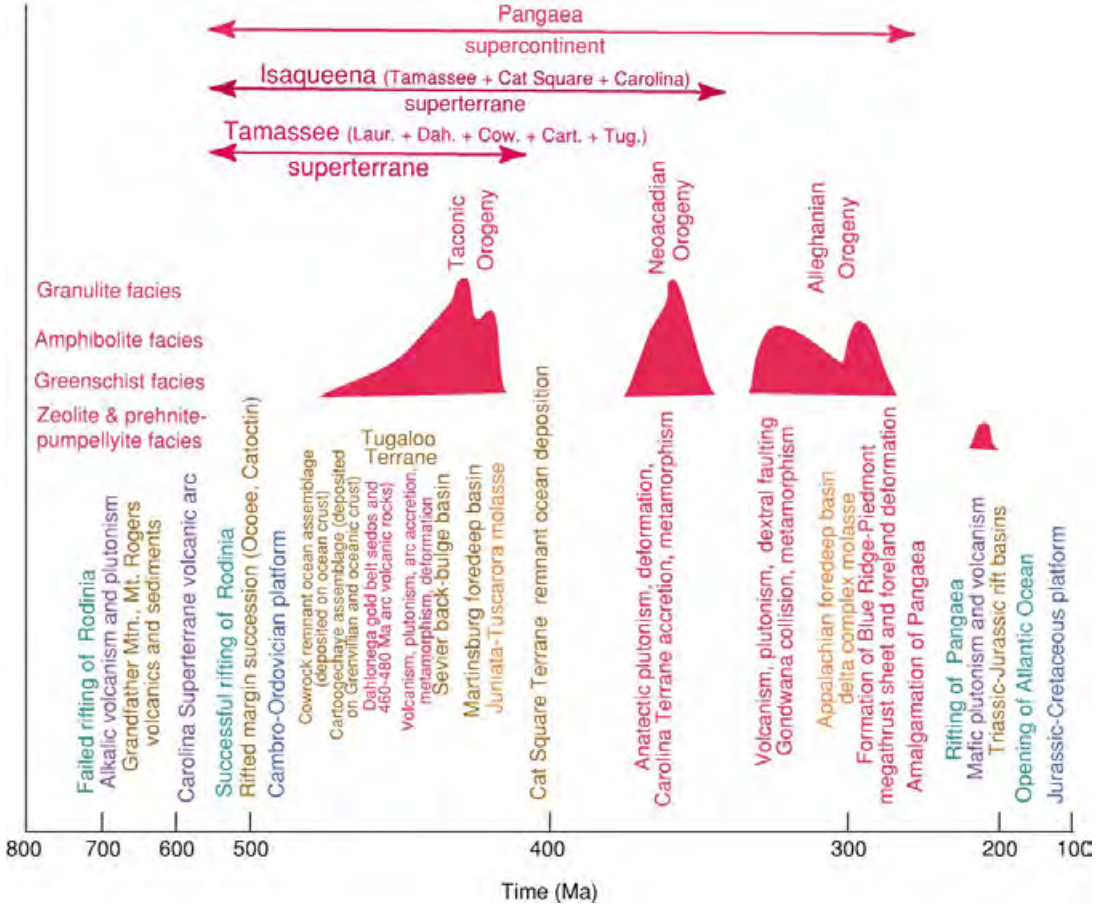


Figure 2 Events distributed through time in the Appalachian Wilson Cycle. High points on the curves denoting orogenies indicate locations of maxima in time and maximum pressure and temperature conditions. Compressional events and related activities are red; arc and rifted margin volcanism is purple; clastic wedges are brown and orange; carbonate platforms are blue; rifting events are green. Note that the horizontal scale is expanded between 500 and 200 Ma.

dominated by vein quartz; several textural and compositional variations of limestone and dolomite; and quartzite, sandstone, and black shale rip-ups. Many of the conglomerates clearly were deposited in channels; armoured mudballs have been observed in at least one palaeochannel. Though the Ocoee Supergroup thins gradually southward to near Cartersville, Georgia, its thickness rapidly decreases northward from its maximum of 15 km in the Great Smoky Mountains National Park, to zero south-east of Johnson City, Tennessee.

In central Virginia, the Ocoee-equivalent ~565-My-old Catoctin Formation consists of basalt and rhyolite, with some interlayered clastic sedimentary rocks that are locally underlain by an older suite of immature rift-related sedimentary rocks (Snowbird Group equivalent?). East of the Blue Ridge anticlinorium in northern North Carolina and Virginia is a sequence of distal, deeper water clastic sedimentary and rift-related volcanic rocks, the Ashe–Alligator Back-Lynchburg sequence. More distal eastward and south-eastward equivalents of these metasedimentary and metavolcanic rocks (Ashe–Tallulah Falls–Sandy Springs) have been recognized in the eastern Blue Ridge and Inner Piedmont from the Carolinas to Alabama. They are overlain by Cambrian–Ordovician(?) metasedimentary rocks (Candler, Chauga River), and by Middle Ordovician(?) mid-oceanic ridge basalt-arc volcanic rocks (Slippery Creek–Poor Mountain–Ropes Creek). Farther east in the Virginia Piedmont is a sequence of felsic–mafic volcanic rocks from ~470 Ma (Chopawamsic) that were intruded by plutons at ~455 Ma, then overlain by fossiliferous Late Ordovician (Ashgill) fine-grained clastic rocks comprising the Chopawamsic Terrane.

The Lower Cambrian Chilhowee Group was deposited on the Laurentian margin during the rift-to-drift transition, and consists of upward-maturing alternating sandstone and shale units. The lowest unit consists of immature sandstone and greywacke with some basalt. The upper part of the Chilhowee Group

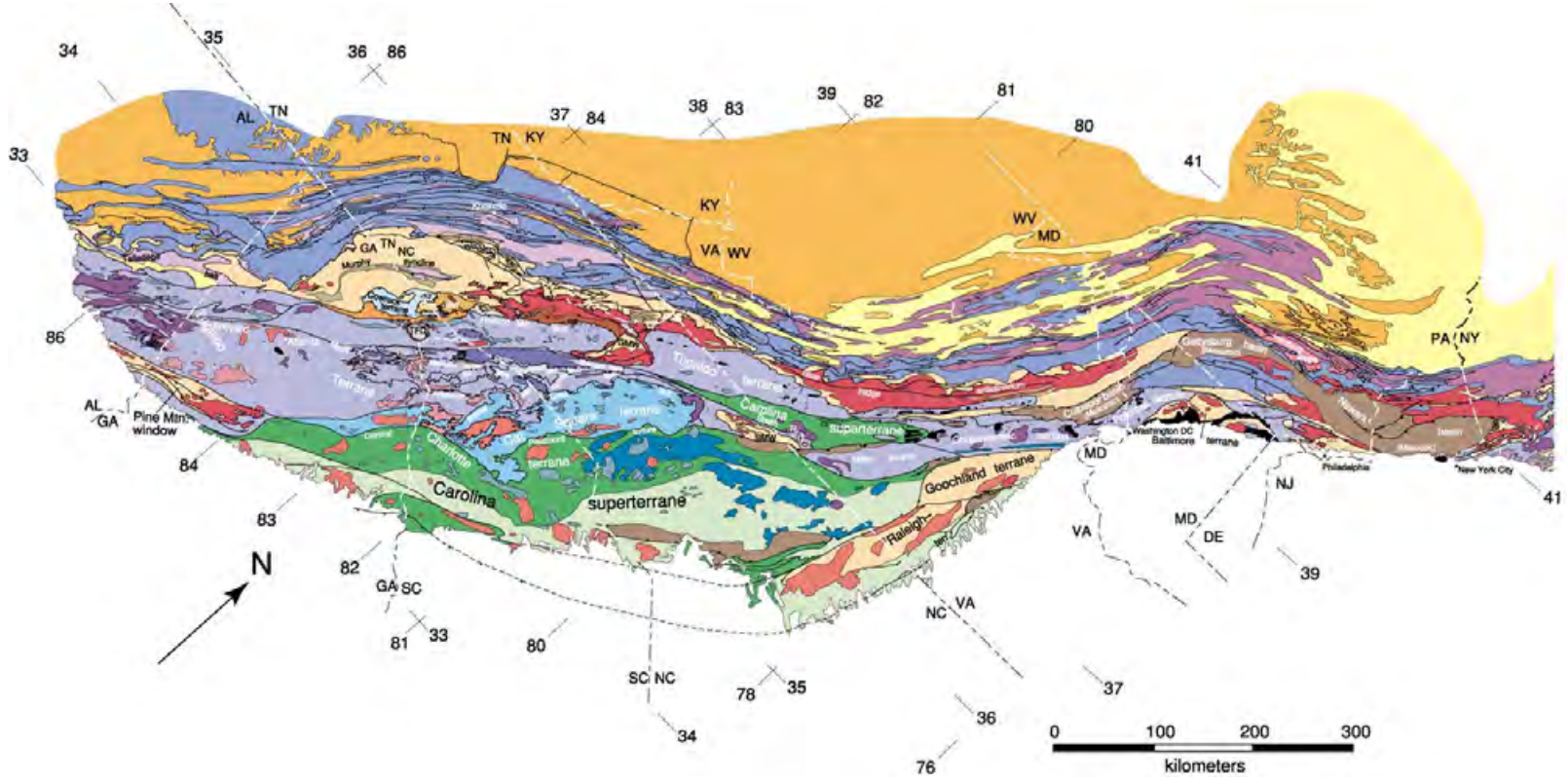


Figure 3 Tectonic map of the southern and central Appalachians. GMW, Grandfather Mountain window; SMW, Sauratown Mountains window; TFD, Tallulah Falls Dome; Cart. Terr., Cartoogechaye Terrane. Colour key: ultramafic rocks are black; Alleghanian clastic wedge is orange (to north-west); Acadian clastic wedge is light yellow (north-eastern and south-western parts of map); Middle Ordovician (Sevier, Blountian) clastic wedge is lavender; Upper Ordovician–Silurian Martinsburg–Tuscarora (type Taconian) clastic wedge is wine; Cambro-Ordovician platform rocks are dark blue; rifted Laurentian margin Neoproterozoic to Cambrian clastic and volcanic rocks are tan; Murphy syncline is olive green with lavender in centre (imbedded in tan in Georgia and North Carolina); ~1.1-Ga (Grenvillian) rocks are red; pre-Grenvillian 1.8-Ga Mars Hill Terrane is brown-orange (next to large red area in the Blue Ridge); Cowrock Terrane is light blue (in Georgia and North Carolina); Cartoogechaye Terrane is light orange (in Georgia and North Carolina); Dahlonega gold belt is burnt orange (Georgia and North Carolina); Hamburg Klippe is pink (north-eastern part of map); high-grade parts of Carolina Superterrane are dark green; low-grade parts of Carolina Superterrane are light green; Neoproterozoic pluton is light grey (north-west of GMW); Neoproterozoic to Early Cambrian plutons (Carolina Superterrane) are dark blue-green; Ordovician plutons are dark purple; likely Ordovician and Silurian plutons are medium purple; Devonian plutons dark gray-green; Carboniferous to Permian plutons are red-orange; Late Triassic–Jurassic sedimentary and volcanic rocks are brown. Scale = 1:2 500 000.

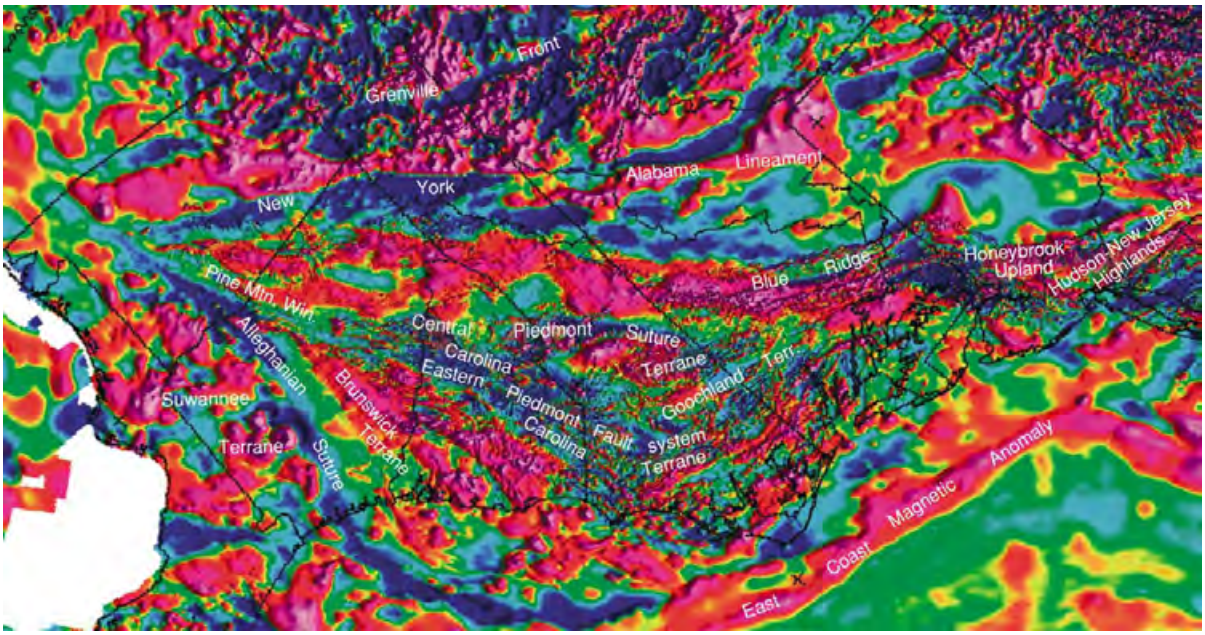


Figure 4 Magnetic map of major crustal features in southern and central Appalachia. Magnetic anomalies are shown in spectral colours, with red indicating highs and blue indicating magnetic lows. Image from the United States Geological Survey.

was probably deposited along the open-ocean margin, because it grades upward into the first successful carbonate bank assemblage. Both the Chilhowee Group and the overlying Shady (–Tomstown–Dunham) Dolomite are continuous throughout the Appalachian margin.

Passive Margin Evolution

Once the first carbonate bank was established, the south-eastern Laurentian margin continued to develop, with pulses of clastic sediment derived from the eroding Grenvillian mountains of progressively lower relief. From Late Cambrian to earliest Middle Ordovician time, carbonate deposition transgressed across the entire margin into the central interior of North America, extending westward as far as present-day western-most Texas, and northward into southern and eastern Canada. This formed the first extensive carbonate platform across much of interior Laurentia. Following this great incursion of shallow-water marine carbonate deposition into interior North America, the entire carbonate platform was uplifted and a karst surface formed over all, except for an area in northern Virginia, Maryland, and Pennsylvania. Carbonate deposition resumed during the Middle Ordovician (Llanvirn). Shortly afterwards, the eastern margin began rapidly to subside as Ordovician volcanic arcs were obducted onto the eastern Laurentian margin (Figure 5). A foredeep formed in front of the arriving thrust sheets, to the west of that a forebulge, and a back-bulge basin formed ~200 km to the west into the platform. This

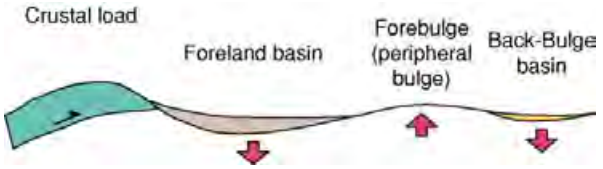


Figure 5 Ordovician loading of the Laurentian margin by obducted volcanic arcs, producing a foredeep basin, forebulge, and a back-bulge basin farther into the continent.

back-bulge basin accumulated >3 km of clastic sediment in the southern Appalachians, forming the Sevier clastic wedge that graded westward into the thin platform limestone sequence. Potassium (K)-bentonite beds (Late Llanvirn) from continuing volcanic activity to the east were deposited throughout the southern and central Appalachians and on the adjacent platform to the west.

The younger (Caradocian) Martinsburg clastic wedge formed a thin successor basin atop the Sevier wedge in the southern Appalachians, but accumulated ~5 km of deep-water turbidic clastics in Pennsylvania, probably in a true foredeep. The Martinsburg assemblage graded upward into the shallower water Ashgill Juniata red-bed succession, and then into the Llandoverian Clinch–Tuscarora quartz arenite and Shawangunk Conglomerate (to the north-east). Though the source of the clastic sedimentary assemblage in the Sevier back-bulge and Martinsburg foredeep basins was clearly to the east, detrital zircon suites contain

no Palaeozoic components. All are dominated by zircons derived from Grenvillian and older Laurentian sources, suggesting provenance from exhumed and cannibalized platform, and rifted margin sedimentary and volcanic rocks. Clasts in Llanvirn polymictic conglomerates in the Sevier Basin sample platform carbonate and older lithologies, but none are derived from sources in the Taconian Orogen. Detrital zircon suites from the rifted margin succession (Ocoee Supergroup, Ashe Formation, and equivalents) across the southern Appalachian Blue Ridge and western Inner Piedmont (Tugaloo Terrane) are all dominated by zircons from ~1.1 Ga, with smaller components of zircons from 1.3–1.5 Ga. Dahlonega gold belt detrital zircons, though also dominated by 1.1- to 1.4-Gy-old components, contain components from 2.7–2.9 Ga, 1.9–2.2 Ga, and 650–700 Ma. These rocks were traditionally considered only a more distal facies of the Ocoee Supergroup, but detrital zircons suggest they may be more exotic. In addition, the Dahlonega gold belt contains an arc complex from 460 and 480 Ma.

Taconic Events

The Taconic Orogeny is clearly diachronous in the southern and central Appalachians, with earlier pulses beginning in the Late Arenig in the southern Appalachians and the younger event—the type Taconic Orogeny—beginning sometime in the Llandeilo or Caradoc in the central Appalachians. Corresponding suites of plutons, penetrative deformation, and metamorphism to conditions as high as granulite facies were a product of subduction, obduction, and arc accretion. The 455- to 460-My-old plutonic–volcanic suite (Hillabee, Ropes Creek, Poor Mountain, Chopawamsic, Baltimore) in the southern and central Appalachians has a mid-oceanic ridge basalt (MORB) to arc affinity. At that time, the Neoproterozoic to Cambrian rifted margin assemblage was A-subducted beneath the arc, penetratively deformed, and metamorphosed to mid-crustal metamorphic assemblages.

During the Taconic events, the Laurentian margin underwent A-subduction and several oceanic and arc assemblages were obducted onto the margin, with accompanying emplacement of the Cowrock, Cartoogechaye, and Tugaloo Terranes in the southern Appalachians, and the Chopawamsic, Potomac, and Baltimore Terranes in the central Appalachians (Figure 3). In New Jersey and eastern Pennsylvania, Grenvillian basement was remobilized in the core of the recumbent Musconnectong Nappe. Basement was also remobilized to the north-east (Hudson–New Jersey Highlands) and south-west, forming external massifs in the core of the Blue Ridge from Virginia to Georgia. Internal massifs in the eastern Virginia Goochland Terrane, the Tugaloo Terrane west of the Brevard fault zone in the Carolinas and north-eastern Georgia, and possibly in the Pine Mountain window in Georgia and Alabama were remobilized, although remobilization of basement in the latter may be younger (see later).

A suite of thrust sheets carrying deep-water sedimentary rocks was emplaced onto the New England and Canadian Maritimes margin at this time, locally carrying oceanic crust and mantle in ophiolite sheets (Quebec and Newfoundland). These extend into the central Appalachians as the Hamburg Klippe in New Jersey and the south-eastern Pennsylvania Piedmont. None has been identified farther south. Other premetamorphic thrusts formed at this time from the Pennsylvania Piedmont southward. Amalgamation of old and new crust generated during the Taconian events formed a superterrane along the south-eastern Laurentian margin.

Cat Square Terrane

Deposition continued until post-430 Ma in the still-open remnant of the Theic Ocean that lay between south-eastern Laurentia and the approaching Carolina Superterrane. Here a sequence of immature deep-water sandstones (subgreywacke to greywacke) and pelites that today consist of aluminous schist and biotite gneiss were probably deposited on oceanic(?) crust. Detrital zircons include the usual prominent 1.1- to 1.4-Ga suite, but with the addition of major components of zircons from 600, 500, and 430 Ma, along with small populations of zircons from 1.9–2.2 and 2.7–2.9 Ga. This diverse population suggests sediment was derived from both Laurentia and Carolina. Presence of younger zircons precludes docking of the Carolina Superterrane before the Late Devonian or Early Carboniferous.

Pine Mountain Terrane

The Pine Mountain Terrane in west-central Georgia and eastern Alabama is a complex window exposing 1.1-Gy-old Grenvillian basement and a thin cover sequence of Early Palaeozoic(?) schist, quartzite–dolomite, and a higher schist. The window is framed by the dextral Towaliga and Goat Rock and Dean Creek faults to the north and south, respectively, but is closed by the Box Ankle Thrust at the east end. All of these faults yield Alleghanian ages, but the Box Ankle Fault was emplaced at sillimanite-grade pressure and temperature (P–T) conditions, whereas the Towaliga and Goat Rock(?) faults moved during garnet-grade conditions, and the Dean Creek Fault moved under chlorite-grade conditions (the same as the Modoc, farther east). Detrital zircons from the

cover-sequence quartzite inside the window yield the usual dominance of ages from 1.1 Ga, but a significant component of zircons from 1.9–2.2 Ga occur there, suggesting a Gondwanan provenance for this sequence and basement. If this is a rifted block of Gondwanan basement, the emplacement kinematics would be very complex, because of the Late Palaeozoic ages of all of the faults framing the window. This basement block and its cover have traditionally been considered Laurentian and the cover sequence is thought to be metamorphosed platform rocks. It remains a suspect terrane.

Peri-Gondwanan Carolina Superterrane

Coeval with the rifting of Rodinia was the development of a number of Peri-Gondwanan terranes in the intervening oceans separating Laurentia from Gondwana. All of these terranes developed proximal to Gondwana, formed composite superterranes, and were accreted to the Laurentian assemblage during the mid-Palaeozoic. Peri-Gondwanan components located in the southern and central Appalachians consist of the Carolina Superterrane and outlier Smith River allochthon (Figure 3). The Neoproterozoic component of the Carolina Terrane consists of a mafic to felsic volcanic arc assemblage and associated volcaniclastic sedimentary rocks. A number of plutons were generated in the range of 550–600 Ma and intruded into the underpinning of the arc complex. This assemblage was metamorphosed during the Cambrian because plutons from 530 Ma cut upper amphibolite to greenschist facies metamorphic rocks. Neoproterozoic Ediacaran metazoan fossils have been identified on bedding planes in low-grade felsic volcanic rocks near Durham, North Carolina.

Overlying the arc complex is a sequence of Cambrian and possible Ordovician clastic sedimentary rocks. A Middle Cambrian Acado-Baltic (*Paradoxides*) fauna occurs in these sedimentary rocks in South Carolina, and Ordovician conodonts have also been reported in another part of the sequence in North Carolina. The high-grade western Carolina Superterrane contains a 350- to 360-My-old metamorphic overprint and numerous younger (Devonian to Carboniferous) granitoids probably related to the mid-Palaeozoic to Late Palaeozoic docking of the Carolina Terrane. The greenschist and lower grade central and eastern Carolina Terrane consists mostly of volcanic and volcaniclastic rocks interrupted by the Alleghanian Kiokee Belt metamorphic core. South-east of this metamorphic core are more low-grade volcanic and volcaniclastic rocks.

Mid-Palaeozoic Sedimentation and Neoacadian Docking of the Carolina Superterrane

Evidence for the Acadian Orogeny from 410–380 Ma in the southern and central Appalachian internides is mostly lacking. A suite of granitoid plutons from 374–382 Ma (Rabun, Pink Beds, Looking Glass, Mt Airy, Stone Mountain) intruded the eastern Blue Ridge of the Carolinas and Georgia, and a suite of anatectic granitoids (Toluca) from 380 Ma exists in the Cat Square Terrane. The Catskill delta spread Middle and Upper Devonian sediment from New England into New York and Pennsylvania, but pinches to the south. Late Silurian–Early Devonian sedimentary rocks comprise the Talladega Group in Alabama. Deposition of the Upper Devonian–earliest Carboniferous Chattanooga–Brallier (and equivalents) black shale unconformably on folded and faulted Silurian and older rocks on the platform may be an indicator of the Neoacadian Orogeny in the internal parts of the southern and central Appalachians.

Detrital zircon ages require that the Cat Square Terrane lying immediately east of the Tugaloo Terrane and west of Carolina (Figure 3) be accreted during the mid-Palaeozoic to Late Palaeozoic. Anatectic granites formed at ~380 and 366 Ma. Metamorphic rims on zircons, not only in the Cat Square Terrane, but also throughout the Tugaloo Terrane, along with deformed 366-My-old plutons in the Cat Square Terrane, indicate peak metamorphism occurred from 360–350 Ma. U/Pb thermal ionization mass spectroscopy (TIMS) ages on monazite yield similar 360 (and 320) Ma ages. Metamorphism to upper amphibolite and granulite facies conditions probably corresponds to time of docking of the Carolina Superterrane, whereby the previously assembled superterrane plus the Cat Square Terrane were A-subducted beneath Carolina, achieving the burial depths necessary to produce wholesale migmatization in the Inner Piedmont and eastern Blue Ridge.

The Inner Piedmont, which now consists of the Cat Square and the eastern Tugaloo Terrane, records a unique pattern of crustal flow that probably is the key to the docking kinematics of the Carolina Terrane. This flow pattern, from south-east to north-west, consists of a north-west-directed suite of thrust sheets that to the west became south-west directed as they were buttressed against the eastern Blue Ridge rocks along the Neoacadian Brevard fault zone. The docking process was probably thus oblique transpressional, beginning in the north and propagating south-westward. As the Carolina Superterrane subducted the components previously joined to Laurentia, the

Inner Piedmont may have decoupled to form an orogenic channel through which the partially melted plastic mass escaped and extruded south-westward.

The western boundary of the Tugaloo Terrane is the Chattahoochee–Holland Mountain fault in the eastern Blue Ridge, which truncates the 374-My-old Rabun Granodiorite. This fault transported migmatitic kyanite to sillimanite-grade Tugaloo Terrane rocks over staurolite–kyanite–sillimanite-grade rocks of the Dahlonega gold belt in southern North Carolina and northern Georgia, and was tightly folded after emplacement. The Chattahoochee–Holland Mountain must therefore be either a Neoacadian or an Alleghanian fault. Though the detrital zircon suites in the Tugaloo Terrane have quite uniform ages and probably reflect a Laurentian source, both Ordovician and Devonian plutons in the western Tugaloo Terrane contain a wide range of inherited components, from dominant 1.1-Gy-old to Archaean zircons. In contrast, almost none of the Ordovician (~466 Ma) and Devonian (~380 Ma) plutons in the eastern Tugaloo and Cat Square Terranes contain any inheritance. The magmas in the western Tugaloo Terrane must have passed through an appreciable thickness of old crust, whereas the magmas crystallizing the plutons in the Inner Piedmont east of the Brevard fault zone, which were generated in more oceanic to arc environments, did not pass through older continental crust on their way to crystallization. Inner Piedmont and eastern Blue Ridge rocks cooled sufficiently to block the metamorphic zircon rims and monazite from 360–350 Ma, and the western Carolina Superterrane recorded $^{40}Ar/^{39}Ar$ cooling from 350 to 360 Ma. Both terranes were reheated at 325 Ma, probably marking the initial collision of Gondwana with Laurentia. The 350- to 360-Ma event actually records the first thermal connection between the Inner Piedmont and the Carolina Superterrane.

Late Mississippian to Permian Alleghanian Zippered Collision with Gondwana and the Amalgamation of Pangaea

Actual collision of Gondwana with crust amalgamated during the Early to Middle Palaeozoic orogenies began the final chapter in the formation of the Appalachian Mountain chain. This collision was oblique, transpressive, rotational, and involved north-to-south collision zipper-like closing of Gondwana with the previously assembled Laurentian–Peri-Gondwanan collage. Initial collision produced a suite of largely S-type granites. Additionally, the Brevard Fault zone was reactivated dextrally under chlorite-grade conditions, and additional dextral faults—the Hylas and other components of the eastern Piedmont Fault system (Roanoke Rapids, Nutbush Creek, Modoc, Augusta, Towaliga, Goat Rock, Bartlett's Ferry)—formed as crustal blocks attempted to escape from the collision zone by moving south-westward from the vicinity of present-day northern Virginia–Maryland–Pennsylvania–New Jersey. In addition to the 325-My-old plutonic suite, a younger suite of plutons from 300–265 Ma is largely confined to the Carolina Terrane. These plutons were likewise generated during various stages of collision. A fault-bounded metamorphic belt, the Kiokee–Raleigh–Goochland Terrane extending from Georgia to the eastern Virginia Piedmont (Figure 3), was also generated during the early and middle stages of collision. Here low-grade Carolina Terrane rocks were metamorphosed to middle to upper amphibolite facies assemblages. In South Carolina and Georgia, the north-west flank of this metamorphic belt consists of the dextral Modoc Fault, whereas the down-to-the-south Augusta fault comprises the south-eastern flank. The Modoc Fault has been traced south-westward to the south-eastern flank of the Pine Mountain window in Georgia and Alabama, where it joins the Dean Creek Fault.

The collision process with its accompanying faulting and other indications of compression caused the interior of the mountain chain to be uplifted beginning in the Late Mississippian (about 325–320 Ma). The uplift process was complex and is recorded by current indicators, variable sediment thickness, and decreased maturity of sediments that demonstrate the orogen was unroofed piecemeal from Alabama to Pennsylvania. At that time, a near-sea level delta formed on the western flank of the chain on the Laurentian margin and the first sediments that contain Palaeozoic detrital zircons were deposited from Alabama to Pennsylvania. The transition from marine to non-marine depositional conditions took place from the Late Mississippian into the Early Pennsylvanian, and abundant land plants formed the coal and the associated sedimentary rocks of the Appalachian Basin.

The final stage of the Alleghanian scenario involved head-on collision of south-eastern Laurentia with Gondwana (Figure 6). The major product of head-on collision is the Blue Ridge–Piedmont megathrust sheet that transported crust amalgamated during all previous Palaeozoic events at least 350 km onto the North American platform. The Brevard Fault remained a suitably oriented weak fault and was again reactivated, this time as an out-of-sequence thrust in the interior of the megathrust sheet The Late Mississippian to Permian clastic wedge was

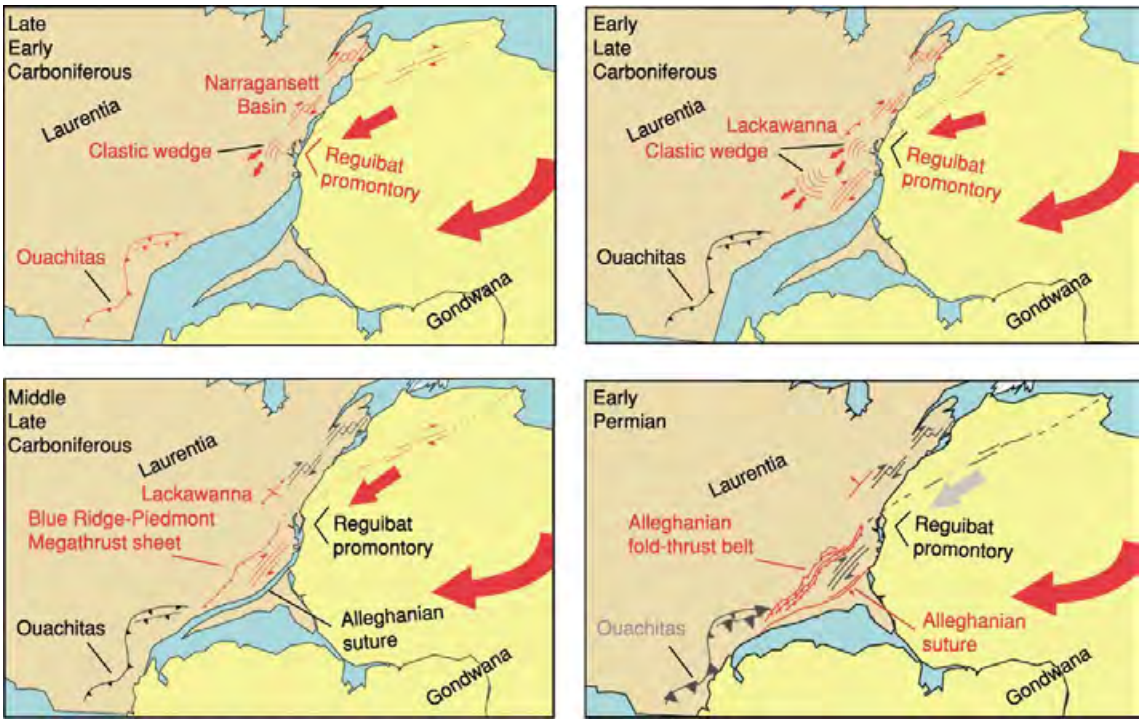


Figure 6 Rotational transpressive collision of Gondwana with Laurentia to form the Supercontinent of Pangaea at the end of the Palaeozoic. Features in red were active at the time indicated. Reproduced with permission from Catalán JRM, Hatcher RD Jr, Arenas R, and García FD (eds.) (2002) *Geological Society of America Special Paper 364*. Boulder, CO: Geological Society of America.

ultimately deformed as the Blue Ridge Piedmont megathrust sheet pushed into the continent, deforming the platform into a foreland fold–thrust belt. In addition, the delta complex that had formed to the west of the rising mountain chain was ultimately deformed. Though final collision was head-on and the megathrust sheet moved westward, uplift of the interior of the chain was non-uniform. The oldest parts of the delta complex are in the southern part of the orogen, possibly related to uplift and erosion of the earlier Ouachita Orogen. This part of the foredeep was superseded to the north by a younger delta complex that fed south-westward along the trough of the foredeep.

The collision was finished about 265 Ma, forming Supercontinent Pangaea and completing the Palaeozoic Wilson Cycle of opening and closing of oceans. The subsurface Suwannee Terrane south of the almost east- and west-trending suture in southern Georgia and Alabama consists of Gondwanan basement and cover Ordovician to Devonian sedimentary rocks of West African affinity. These rocks were left on the Laurentian side in the wake of the breakup of Pangaea. The southern and central Appalachians constitute a microcosm in the assembly of all existing continents along the huge suture zone that extends from south-eastern Laurentia through North-west Africa, western Europe, and the Ural Mountains in Russia.

Breakup of the supercontinent began during the Late Triassic, at ~50 My following the amalgamation of Pangaea, beginning the initial stages of opening of the present Atlantic Ocean. The assembly of Pangaea took place over a span of at least 100 My, but ironically, its breakup occurred over a much shorter time-span. This contrasts with the ≥200-My time-frame for assembling Rodinia and the 300 My that it took for it to break apart.

See Also

Grenvillian Orogeny. **North America:** Precambrian Continental Nucleus; Northern Appalachians. **Pangaea**.

Further Reading

Aleinikoff JN, Zartman RE, Rankin DW, and Burton WC (1995) U–Pb ages of metarhyolites of the Catoctin and Mt. Rogers Formations, central and southern Appalachians: evidence of two pulses of Iapetan rifting. *American Journal of Science* 295: 428–454.

Bream BR, Hatcher RD Jr, Miller CF, and Fullagar PD (2004) Detrital zircon ages and Nd isotopic data from the southern Appalachian crystalline core, GA–SC–NC–TN: new provenance constraints for part of the Laurentian margin. In: Tollo RP, Corriveau L, McLelland J, and

Bartholomew MJ (eds.) *Geological Society of America Memoir 197*, pp. 459–475. Boulder, CO: Geological Society of America.

Faill RT (1998) A geologic history of the north-central Appalachians. Part 1: Orogenesis from the Mesoproterozoic through the Taconic orogeny; Part 2: The Appalachian basin from the Silurian through the Carboniferous. *American Journal of Science* 297: 551–569, 729–761.

Faill RT (1998) A geologic history of the north-central Appalachians. Part 3: The Alleghany orogeny. *American Journal of Science* 298: 131–179.

Hatcher RD Jr (2001) Rheological partitioning during multiple reactivation of the Paleozoic Brevard Fault Zone, Southern Appalachians, USA. In: Holdsworth RE, Strachan RA, Macloughlin JF, and Knipe RJ (eds.) *The Nature and Significance of Fault Zone Weakening: Geological Society of London Special Publication 186*, pp. 255–269. London: Geological Society of London.

Hatcher RD Jr (2002) The Alleghanian (Appalachian) orogeny, a product of zipper tectonics: rotational transpressive continent–continent collision and closing of ancient oceans along irregular margins. In: Catalán JRM, Hatcher RD Jr, Arenas R, and García FD (eds.) *Variscan-Appalachian dynamics: The Building of the Late Paleozoic Basement: Geological Society of America Special Paper 394*, pp. 199–208. Boulder, CO: Geological Society of America.

Hatcher RD Jr, Thomas WA, and Viele GW (eds.) (1989) *The Appalachian–Ouachita orogen in the United States: The Geology of North America*, vol. F–2. Boulder, CO: Geological Society of America.

Hibbard JP, Stoddard EF, Secor DT Jr, and Dennis AJ (2002) The Carolina zone: overview of Neoproterozoic to early Paleozoic peri-Gondwanan terranes along the eastern flank of the southern Appalachians. *Earth Science Reviews* 57: 299–339.

Thomas WA, Becker TP, Samson SD, and Hamilton MA (2004) Detrital zircon evidence of a recycled orogenic foreland provenance for Alleghanian clastic-wedge sandstones. *Journal of Geology* 112: 23–37.

Northern Appalachians

C R van Staal, Geological Survey of Canada, Ottawa, ON, Canada

© 2005, Elsevier Ltd. All rights reserved.

Introduction

The Appalachian Orogen is a remarkably linear, north-east-trending, Palaeozoic mountain belt that generally follows the eastern seaboard of North America, with the segment between Long Island Sound and Newfoundland (Figure 1) being referred to as the Northern Appalachians. Prior to the Mesozoic opening of the Atlantic Ocean, the Appalachians continued into the Caledonides of the British Isles and Scandinavia, forming a long linear mountain chain that was created by the closing of the Iapetus (Cambrian–Early Devonian) and Rheic (Devonian–Carboniferous) Oceans.

Iapetus started to open at the end of the Neoproterozoic (about 570 Ma) as a result of the final breakup of the supercontinent Rodinia, and had achieved a width of about 5000 km by the end of the Cambrian (500–490 Ma). The south-facing margin of ancient North America (Laurentia), which was then situated near the equator in a roughly east–west orientation, defined the northern limit of Iapetus. The northern margin of Gondwana, a large continent centred on the south pole and comprising present-day Africa, Australia, Antarctica, India, and large parts of South America, represented the southern limit of Iapetus (Figure 2). The large Gondwanan landmass was assembled as a result of late Neoproterozoic to Early Cambrian collisions; thus the opening of Iapetus overlapped in time with the final assembly of Gondwana. The Rheic Ocean opened when the Avalonian microcontinent rifted from Gondwana during the Early to Middle Ordovician. The Rheic Ocean was host to several small continental terranes, including Meguma, which were accreted to Laurentia before the final Carboniferous arrival of Gondwana and the formation of the Pangaean supercontinent.

Tectonostratigraphical Divisions

The tectonic architecture and the evolution of the distinctive rock assemblages in the Northern Appalachians are generally described within a framework of tectonostratigraphical zones and subzones. The concept of zonal divisions, based on sharp contrasts in lithology, stratigraphy, fauna, structure, geophysics, plutonism, and metallogeny of Early Palaeozoic and older rocks, was first introduced more than 30 years ago by Harold Williams, following detailed studies of the well-exposed coastal sections in Newfoundland. The identification of these zonal divisions throughout the Northern Appalachians implies a common geological evolution within each zone and highlights the usefulness of each in tectonic analysis.

From west to east, the Northern Appalachians have been divided into the Humber, Dunnage, Gander, Avalon, and Meguma zones (Figure 1). The Humber

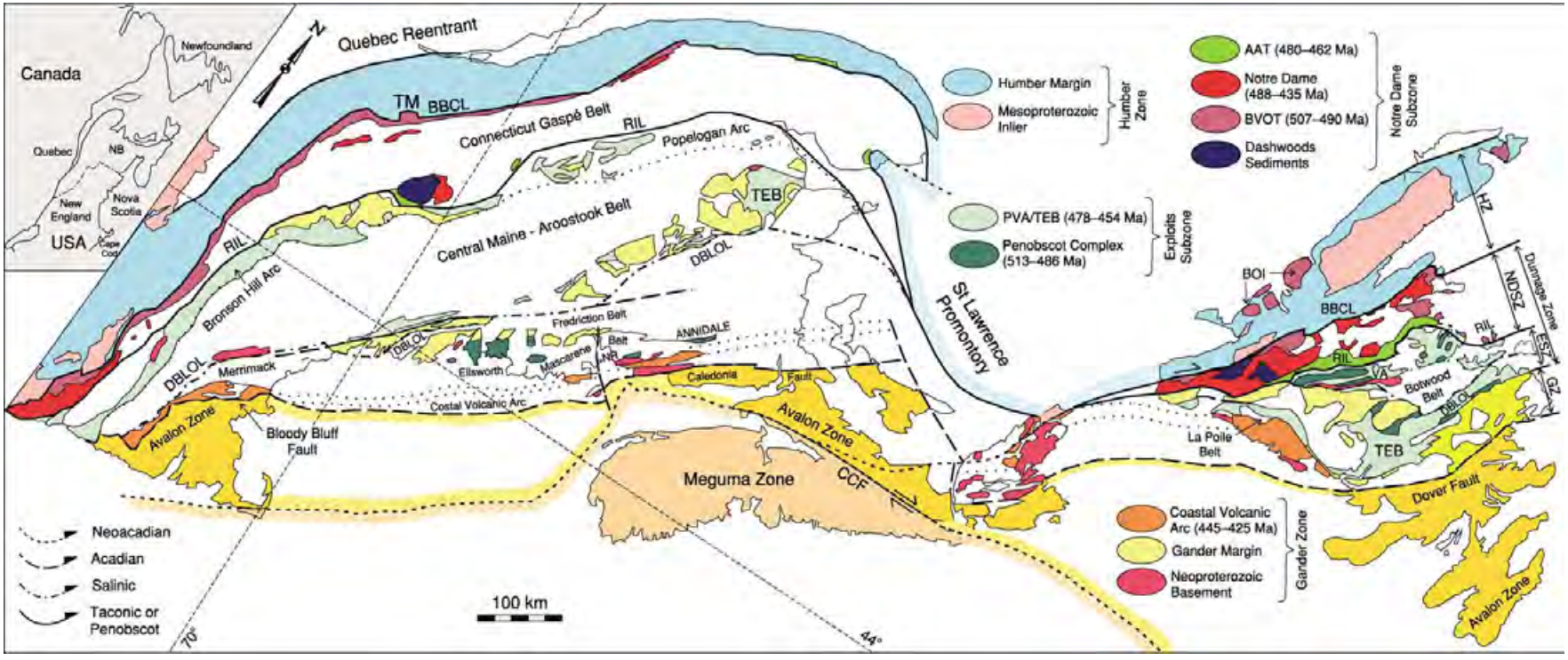


Figure 1 Tectonostratigraphic map of the Northern Appalachians. Zonal divisions are principally based on sharp contrasts in Lower Palaeozoic geology; the main belts of Siluro-Devonian sedimentary and volcanic rocks are outlined in white. The major structures and principal orogenic events responsible for their formation are indicated by line patterns. The extension of structures off the land is based on geophysical interpretation and well data. AAT, Annieopsquotch accretionary tract; BBCL, Baie Verte–Brompton–Cameron Line; BOI, Bay of Islands Ophiolite; BVOT, Baie Verte Oceanic Tract; CCF, Cobequid–Chedabucto fault; DBLOL, Dog Bay–Liberty–Orrington Line; ESZ, Exploits Subzone; GZ, Gander Zone; HZ, Humber Zone; NDSZ, Notre Dame Subzone; NR, New River Belt; RIL, Red Indian Line; PVA, Popelogan–Victoria Arc; TEB, Tetagouche–Exploits back-arc Basin; TM, Thetford Mines Ophiolite; VA, Victoria Arc.

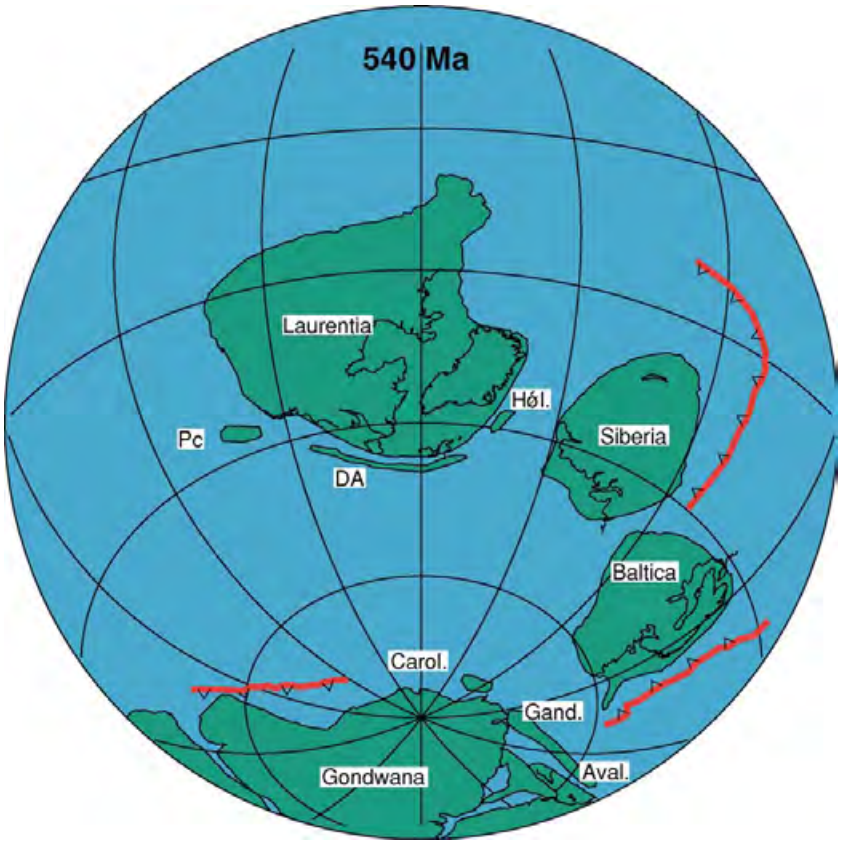


Figure 2 Early Cambrian (around 540 Ma) palaeogeography of the Iapetus Ocean. Seafloor spreading that led to the opening of Iapetus started at around 570 Ma. Aval., Avalonia; Carol., Carolina; DA, Dashwoods; Gand., Ganderia; Hól., Hólanda; Pc, Precordillera Terrane. Meguma was probably situated further east, just outside the field of view, along the north African margin of Gondwana near elements of Armorica. Hólanda and Precordillera are peri-Laurentian microcontinents like Dashwoods but ended up in Baltica and the South American part of Gondwana, respectively, after the closure of Iapetus. Made with the help of Conall MacNiocaill.

Zone represents the peripheral part of the Laurentian Craton, which was involved in Appalachian orogenesis. The Gander, Avalon, and Meguma zones represent peri-Gondwanan microcontinents that accreted to Laurentia during the Early to Middle Palaeozoic (450–380 Ma). The Dunnage Zone contains the remnants of oceanic terranes that formed within the realm of the Iapetus Ocean, and is subdivided into the Notre Dame and Exploits subzones.

Orogenesis (the combined effects of spatially and temporally associated deformation, metamorphism, magmatism and sedimentation) was mainly confined to the central portion of the Northern Appalachians, often referred to as the Central Mobile Belt.

Humber Zone

The Humber Zone is mostly underlain by a Late Neoproterozoic–Ordovician rifted passive margin sequence deposited on Laurentian basement that had experienced multiple orogenic events during the Mesoproterozoic (around 1.5–0.95 Ga), culminating in the Grenville orogeny (around 1.2–0.95 Ga). These basement rocks are exposed in several windows from Newfoundland to southern New York state. Its Late Neoproterozoic–Palaeozoic cover preserves evidence of rifting, development of a warm-water carbonate shelf and slope, and conversion into a convergent margin. Rift-related magmatism took place intermittently during an approximately 70 Ma time-span during the latest Neoproterozoic (620–550 Ma) and was coeval with the deposition of consanguineous clastic sedimentary rocks. Exactly when during this time interval the opening of the Iapetus Ocean began is at present unclear, but palaeomagnetic and geological arguments suggest that it occurred at around 570 Ma. Several lines of evidence suggest that the final stage of rift magmatism (555–550 Ma) and the subsequent Early Cambrian (around 540 Ma) transition from rift to drift sedimentation preserved in Newfoundland, Quebec, and New England (Figure 3), relates to the departure of a small microcontinent, referred to as Dashwoods. This event thus postdates the seafloor spreading associated with the earlier opening of Iapetus. Dashwoods equates

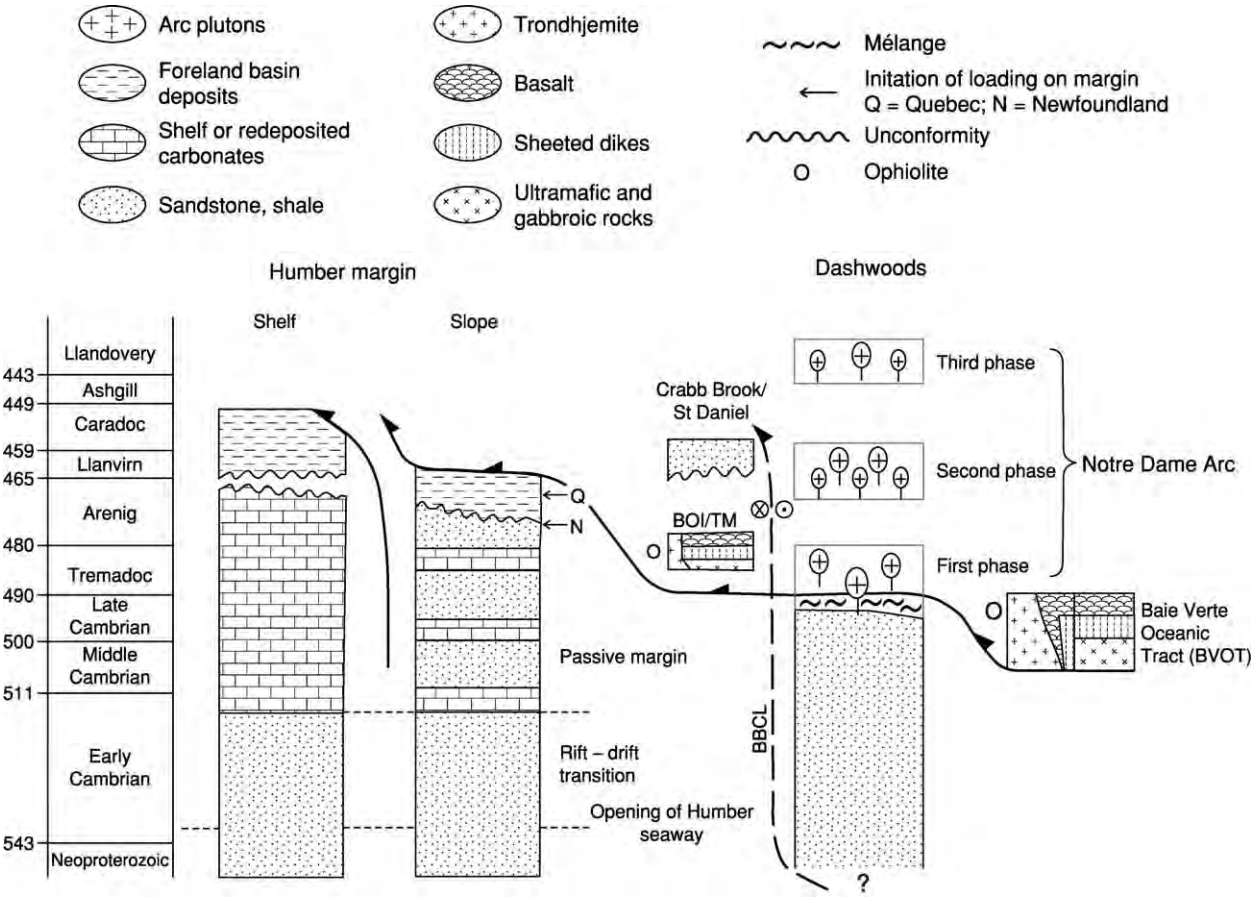


Figure 3 Tectonostratigraphic relationships pertinent to the convergence between the Baie Verte Oceanic Tract and Dashwoods, and between Dashwoods and the Humber margin. East-directed subduction started in the Middle Cambrian immediately outboard of the Dashwoods microcontinent and led to underthrusting of the Dashwoods microcontinent beneath the Baie Verte Oceanic Tract (BVOT). The Bay of Islands (BOI) and Thetford Mines (TM) ophiolites formed during local rifting and concurrent seafloor spreading of the allochthonous BVOT after the Dashwoods microcontinent was subducted beneath the BVOT and the east-directed subduction zone stepped back into the Humber seaway. The Crabb Brook and St Daniel were deposited in piggyback basins formed during obduction of the ophiolites onto the Humber margin. BBCL, Baie Verte-Brompton-Camerons Line.

to the Boundary Mountain Terrane in New England. Spreading between Laurentia and the Dashwoods microcontinent was either very slow or aborted soon after formation, because faunal and palaeomagnetic data suggest that the Dashwoods was separated from Laurentia by only a narrow oceanic seaway (the Humber Seaway). The along-strike extent of the Dashwoods is uncertain. Isotopic evidence for the involvement of Laurentian crust in the generation of Early to Middle Ordovician magmatic rocks of the Notre Dame Arc and Laurentian rift-related clastic rocks positioned in an upper plate setting during the Taconic Orogeny (Figures 1 and 3) suggests that this microcontinent extended southward into New England.

The conversion of the Humber passive margin into a convergent margin is indicated by cratonward migration of a foreland basin formed in response to tectonic loading of the margin by an overriding ophiolite terrane (Figure 3). Foreland-basin development was immediately preceded by uplift and local karst erosion due to the passage of a peripheral bulge across the margin. This process started in the Middle to Late Arenig (around 474 Ma) and lasted at least until the Late Caradoc (around 450 Ma) in Newfoundland. Loading of the margin started slightly later in Quebec and New England (Figure 3) probably owing to the presence of the Quebec Reentrant–St Lawrence Promontory pair (Figure 1).

Dunnage Zone

Palaeomagnetic, fossil, and other geological evidence indicates that the oceanic terranes of the Dunnage zone (Figure 1) can be separated into peri-Laurentian (Notre Dame Subzone) and peri-Gondwanan (Exploits Subzone). Detailed geochemical studies have revealed that virtually all the preserved oceanic rocks have compositions typical of suprasubduction zone

settings (i.e. fore-arc, arc, and back-arc). True oceanic lithosphere formed at mid-oceanic spreading centres, far removed from continental margins and subduction zones, is either very rare or has been lost during subduction.

Notre Dame Subzone The Notre Dame Subzone (Figure 1) lies immediately east of the Humber Zone, from which it is separated by a complex and long-lived fault zone, the Baie Verte–Brompton–Camerons Line. The Notre Dame Subzone comprises several unrelated Middle Cambrian to Middle Ordovician (507–462 Ma) oceanic terranes and an important continental magmatic arc (the Notre Dame Arc), which was intermittently active between 488 and 433 Ma (Figure 3). Rocks of the Notre Dame Subzone can be traced from south-western Connecticut through western New England and Quebec into Newfoundland (Figure 1), where they are best exposed and studied. Rocks belonging to the Notre Dame Arc are commonly referred to as the Shelburne Falls Arc in New England. The eastern boundary of the Notre Dame Subzone with the Exploits Subzone is marked by a major suture (the Red Indian Line) that juxtaposes rocks formed on opposite sides of the Iapetus Ocean (Figure 4).

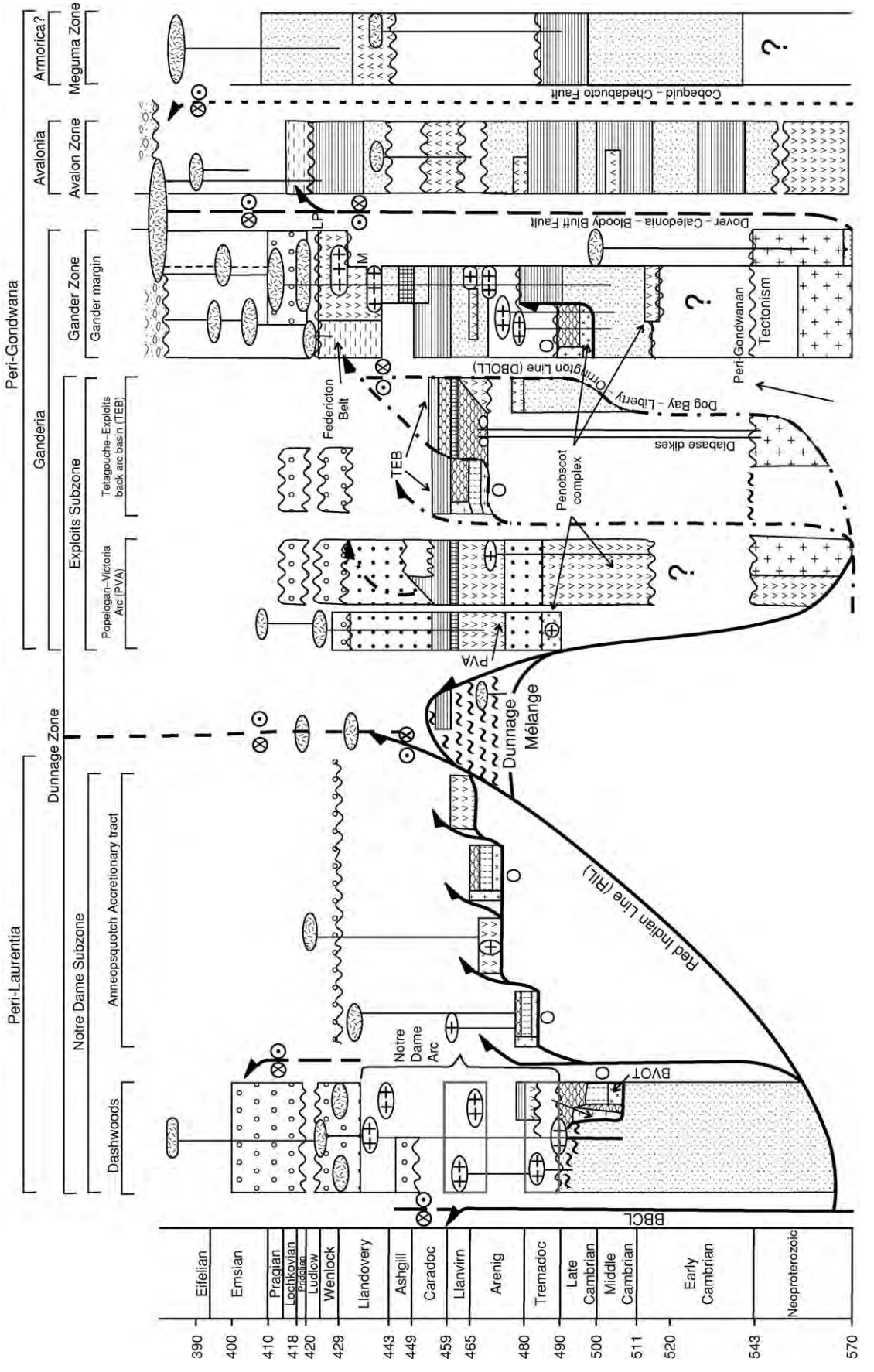
The oldest oceanic rocks in the Notre Dame Subzone occur in the Middle to Upper Cambrian (around 507–490 Ma) Baie Verte Oceanic Tract, which mainly represents an infant arc terrane (wide zone of supra-subduction-zone oceanic lithosphere created during and shortly after the initiation of subduction, before the establishment of a well-defined linear narrow arc axis). The Baie Verte Oceanic Tract was obducted on to the Dashwoods shortly after formation owing to eastward-directed (present coordinates) subduction initiated close to Laurentia. Stitching plutons of the first phase of the Notre Dame Arc (around 488 Ma) indicate that obduction in Newfoundland took place during the latest Cambrian. Rare evidence of Late Cambrian tectonometamorphic events and the age range of Early Ordovician arc plutonism in New England and Quebec suggest that there was little diachroneity in this process along the length of the Northern Appalachians. The first phase of Notre Dame Arc magmatism resulted from the stepping back of subduction into the Humber seaway, transferring the Dashwoods microcontinent from a lower-plate to an upper-plate setting (Figure 3). Closure of the Humber Seaway brought Dashwoods back to Laurentia in the late Arenig (475–470 Ma) and started the Taconic orogeny. The Taconic Orogeny was accompanied by the second phase of Notre Dame Arc magmatism (469–458 Ma) and intense deformation and Barrovian metamorphism of the Dashwoods, including the rocks generated during the first phase of the Notre Dame Arc development.

A younger, unrelated, oceanic terrane, the Annieopsquotch accretionary tract, is situated along the eastern margin of the Notre Dame Subzone, immediately west of the Red Indian Line (Figure 4). The Annieopsquotch accretionary tract comprises a tectonic collage of Arenig–Llanvirn (480–462 Ma) infant arc, arc, and back-arc terranes that formed as a result of west-directed subduction outboard of the Dashwoods. Ultramafic rocks are rarely preserved in the Annieopsquotch accretionary tract, in contrast to the Baie Verte Oceanic Tract, probably because their accretion to the Dashwoods involved underthrusting and structural underplating rather than obduction. Accretion of the Annieopsquotch accretionary tract to Laurentia started at around 470 Ma, suggesting a causal relationship to the Taconic, Laurentia–Dashwoods, collision further west, and ended before 455–450 Ma, when the leading edge of the Exploits subzone (Popelogan–Victoria Arc) started to accrete to Laurentia. The Annieopsquotch accretionary tract is poorly preserved in New England, mainly because the Red Indian Line is here largely buried and/or obscured by Mid-Palaeozoic cover sequences and tectonism. The Red Indian Line approximately follows the eastern boundary of the Connecticut–Gaspé Belt (Figure 1) and is locally marked in Maine by mélanges and ophiolitic rocks.

The Notre Dame Subzone experienced both uplift and subsidence during the Late Ordovician and Silurian, and is locally overlain by a marine to terrestrial Silurian volcanosedimentary sequence. Terrestrial rocks are dominant in Newfoundland, while marine deposits characterize New England and Quebec.

Notre Dame Arc magmatism had ended by the early Silurian and was replaced by a phase of Early to Late Silurian (433–424 Ma) bimodal within-plate volcanism and plutonism; the latter has been ascribed to breakoff of the west-dipping, down-going slab responsible for accretion of the Annieopsquotch accretionary tract, the Exploits Subzone, and the Gander Zone. Early Devonian (Acadian) orogenesis, very penetrative in the adjacent Exploits Subzone and Gander Zone, is weak or absent, showing that the locus of orogenesis had shifted to the south-east by Late Silurian times. Mild orogenesis returned by the Middle to Late Devonian (Neoacadian, 395–350 Ma).

Exploits Subzone The oldest known subduction-related rocks preserved in the Exploits Subzone have Lower Cambrian to Tremadoc (513–486 Ma) ages and compositions typical of an arc or back-arc setting. Some of the rocks have compositions and stratigraphies typical of ophiolite complexes, while


Peri-Gondwana
Ganderia
Avalonia
Armorica?
Meguma Zone
Avalon Zone
Gander Zone
Gander margin
Exploits Subzone
Tetagouche–Exploits back arc basin (TEB)
Popelogan–Victoria Arc (PVA)
Dunnage Zone
Peri-Laurentia
Notre Dame Subzone
Anneopsquotch Accretionary tract
Dashwoods
Cobequid – Chedabucto Fault
Dover – Caledonia – Bloody Bluff Fault
Peri-Gondwanan Tectonism
Federicton Belt
TEB
Ormagton Line (DBOLL)
Dog Bay – Liberty
Diabase dikes
Penobscot complex
PVA
Dunnage Mélange
Red Indian Line (RIL)
Notre Dame Arc
BVOT
BBCL
Eifelian
Emsian
Pragian
Lochkovian
Pridolan
Ludlow
Wenlock
Llandovery
Ashgill
Caradoc
Llanvirn
Arenig
Tremadoc
Late Cambrian
Middle Cambrian
Early Cambrian
Neoproterozoic
390
400
410
418
420
429
443
449
459
465
480
490
500
511
520
543
570

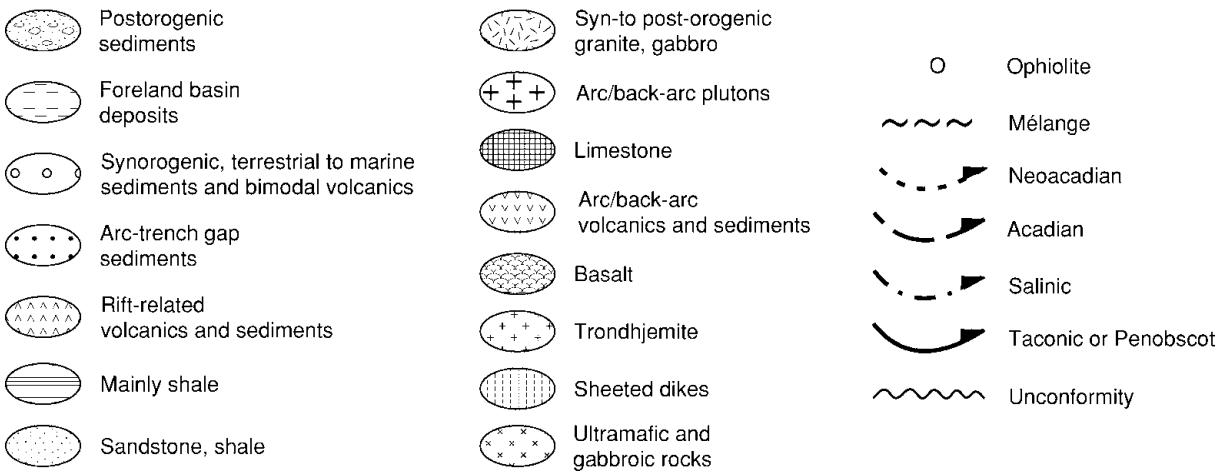


Figure 4 Tectonostratigraphic relationships pertinent to the closure of the main tract of the Iapetus Ocean, the Tetagouche–Exploits back-arc basin, and the oceanic crust that separated Ganderia from Avalonia and Avalonia from Meguma. The Red Indian Line marks the site of the collision between the Notre Dame and Popelogan–Victoria arcs. The line patterns of the major structures correspond to those in **Figure 1**. The arrows on the faults indicate the polarity of subduction and/or thrusting. Obliquity of collision and subsequent strike-slip fault reactivation is indicated by the symbols on each side of the major faults: a point represents moving towards and a cross moving away from the viewer. BBCL, Baie Verte-Brompton-Camerons Line. BVOT, Baie Verte Oceanic Tract. LP, La Poile Belt volcanic rocks; M, Mascarene belt; PVA, Popelogan–Victoria Arc; TEB, Tetagouche–Exploits back-arc basin.

others have close spatial associations with felsic-arc plutonic rocks that are at least as old as 565 Ma (**Figure 4**). The latter probably represent ensialic basement and suggest that at least some of these rocks are not truly oceanic. Coeval and lithologically similar Cambrian arc or back-arc volcanic rocks and associated Neoproterozoic plutonic rocks also occur in the Gander Zone (see below), particularly in southern New Brunswick (e.g. Annidale and New River Belts) and adjacent Maine (Ellsworth Belt) (**Figure 1**). Collectively, these rocks have been interpreted as an arc–back-arc complex, the Penobscot Complex, formed near the leading edge of the Gondwanan margin on which the arenites and shales of the Gander Zone (see below) were deposited. Stitching plutons and isotopic evidence show that the Cambrian oceanic rocks of the Penobscot Complex were obducted onto the continental-margin rocks of the Gander Zone before 474 Ma, during the Penobscot Orogeny (**Figure 4**).

The composite crust of the Penobscot Complex and the Gander Zone sedimentary rocks are disconformably overlain by younger rocks related to the Arenig–Llanvirn ensialic Popelogan–Victoria Arc/Tetagouche–Exploits back-arc system, which was active between 478 and 454 Ma. The Popelogan–Victoria Arc/Tetagouche–Exploits back-arc system has been traced from Newfoundland through central New Brunswick into Maine, mainly on the basis of a remarkably consistent and distinct lithological association of coeval plutonic, volcanic, and sedimentary rocks. Correlations further to the south into Massachusetts become progressively more difficult, owing to a lack of age control and the presence of substantial cover sequences. The Popelogan–Victoria Arc is probably continuous with the Bronson Hill Arc, whereas most of the products of the Tetagouche–Exploits back-arc have been buried beneath the Siluro-Devonian cover sequences of the Central Maine and Merrimack Belts (**Figure 1**). The Popelogan–Victoria Arc had migrated to a mid-Iapetus position by at least the late Llanvirn (around 462 Ma), according to faunal and sparse palaeomagnetic evidence, while the passive-margin side of the Tetagouche–Exploits back-arc, which was deposited on Gander-Zone rocks, was positioned further south owing to subduction-zone retreat and concurrent back-arc spreading. Rifting and seafloor spreading in the Tetagouche–Exploits back-arc is necessary to account for the formation of Llanvirn back-arc oceanic crust (464–459 Ma), which is locally preserved as incomplete ophiolite complexes and large structural slices of highly tectonized back-arc oceanic basalt and gabbro in mélange belts in New Brunswick and Newfoundland. Fossil and palaeomagnetic evidence suggest that the Tetagouche–Exploits back-arc nowhere achieved a width of more than about 1000 km (largely based on the assumption that the pelagic larvae of animals such as brachiopods could traverse oceans up to this width, but no further).

Gander Zone

The Gander Zone (**Figures 1** and **4**) is defined by a distinct monotonous sequence of Lower Cambrian to

Tremadoc (520–480 Ma) arenites and shales, capped by Tremadoc black shale. This sequence can be traced from north-east Newfoundland into New England and is generally accepted to represent the outboard part (outer shelf to slope) of a passive margin built on Gondwanan continental crust. This inference is based on an extensive set of isotope, geochemical, detrital mineral, palaeomagnetic, and fossil data. The sedimentary source area has to be cratonic and contain Archean, Palaeo-, Meso-, and Neoproterozoic rocks. Based on a process of elimination, this craton is generally inferred to be Amazonia, although other Gondwanan sources cannot be ruled out.

With the possible exception of a few outcrops in southern New Brunswick, the basement on which this sedimentary sequence was deposited is not exposed. However, spatially associated Late Neoproterozoic to Early Cambrian arc rocks sporadically exposed in the Gander Zone are, on the basis of distinct isotope, geochemical, and age data, inferred to represent basement. The basement rocks show many similarities in their Neoproterozoic history to coeval rocks of the adjacent Avalon Zone, from which they are separated by a major long-lived ductile–brittle fault zone (Dover–Caledonia–Bloody Bluff Fault), but have experienced a very different Cambrian to Early Devonian evolution. Avalon developed a shale-rich platformal sedimentary succession during the Early Palaeozoic and did not experience orogenesis until the Latest Silurian–Early Devonian Acadian Orogeny. Gander, on the other hand, experienced multiple phases of arc magmatism and orogenesis from the Cambrian to the Silurian (Figure 4). Furthermore, Nd-isotope data indicate that their Precambrian basements are strikingly different. Hence, the Gander and Avalon zones are considered to represent two distinct Gondwanan microcontinents during the Palaeozoic (Ganderia and Avalonia, respectively). Ganderia was accreted to Laurentia in two stages. The leading edge of Ganderia, with the Popelogan–Victoria Arc suprastructure (Figure 4), accreted to Laurentia in the Late Ordovician. Convergence between Ganderia and Laurentia continued, leading to final closure of the Tetagouche–Exploits back-arc by the late Early Silurian (430–425 Ma). During this collision with Laurentia, a Late Ordovician to late Early Silurian (445–425 Ma) extensional arc–back-arc system (coastal volcanic arc/Mascarene–La Poile back-arc basin) was constructed on its southern margin (Figures 1 and 4) owing to convergence between Avalonia and Ganderia. This arc–back-arc system forms a narrow volcanic–plutonic belt that roughly follows the coastline of New England and maritime Canada to southern Newfoundland.

Avalon Zone

The exposed rocks of the Avalon Zone mainly represent the suprastructure of the microcontinent Avalonia. Its distinctive rocks have been traced from the British Isles into Massachusetts, and mainly comprise Neoproterozoic dominantly juvenile arc-related volcanosedimentary successions and associated plutonic rocks, which experienced a complicated and long-lived Neoproterozoic tectonic history that largely predates the opening of Iapetus. Characteristic of the Avalon Zone is the deposition of a Cambrian–Lower Ordovician shale-rich platformal sedimentary sequence that locally includes rift-related volcanic rocks (Figure 4). Fossil and palaeomagnetic data show a strong connection with Gondwana during this period. However, starting in the late Early Ordovician, the data indicate that Avalon started to drift north independently across Iapetus. Early to Middle Ordovician rift-related volcanic rocks (479–460 Ma), preserved in the Avalonian successions of Nova Scotia and New Brunswick, support this assertion. The accretion of Avalonia to Ganderia, and by implication Laurentia, is poorly constrained, because the oldest preserved sedimentary and magmatic (stitching plutons) linkages have Middle to Late Devonian ages. Palaeomagnetic and structural evidence, particularly relating to the Late Silurian (422–418 Ma) start of the inversion of the Mascarene–La Poile back-arc basin on the southern edge of Ganderia and the ages of syntectonic plutons, suggest however that collision started during the latest Silurian and culminated in the mainly Early Devonian (418–395 Ma) Acadian Orogeny.

Meguma Zone

The Meguma Zone represents the most outboard terrane in the northern Appalachians. It is exposed on land only in southern Nova Scotia, but its regional extent is much larger. It has been traced offshore by an impressive set of geophysical and well data from the southernmost part of the Grand Banks south-east of Newfoundland, across the Scotian Shelf and the Gulf of Maine, to southernmost Cape Cod. The basal part of the exposed Meguma zone comprises a thick latest Neoproterozoic to Early Ordovician turbiditic sandstone–shale sequence of the Meguma Supergroup, which was deposited on the continental rise and/or slope to outer shelf of a Gondwanan passive margin. A combination of detrital-zircon, sedimentological, and sparse fossil data suggest a north-west African provenance, but the dataset at present is small and other parts of Gondwana cannot be ruled out as source areas. The Meguma Supergroup is disconformably overlain by Late Ordovician

to Early Devonian shallow-marine dominantly siliciclastic sedimentary rocks, which are locally interlayered with rift-related bimodal Early Silurian (442–438 Ma) volcanic rocks. The Meguma Zone experienced intense orogenesis during the late Early Devonian to Early Carboniferous (395–350 Ma; Neoacadian orogenesis). Fossil evidence (fish and crinoid) suggests that, during the Late Silurian, Meguma was close to Avalonia and/or Baltica, and separated from Gondwana. When combined with the magmatic evidence for Late Ordovician rifting and Neoacadian orogenesis, this suggests that Meguma was a microcontinent or part thereof (perhaps Armorica) during at least the Silurian and Devonian, as Gondwana was not accreted to Meguma until the Carboniferous–Permian Alleghanian Orogeny.

Overview and Summary of the Tectonic Evolution of the Northern Appalachians

The oldest orogenic events associated with the closure of Iapetus are the Late Cambrian–Early Ordovician (495–484 Ma) accretion of the Baie Verte Oceanic Tract to the Dashwoods and the Penobscot complex to Ganderia, with the Baie Verte Oceanic Tract and the Penobscot complex both in upper-plate settings (Figures 3 and 4). Exposed, inferred Ganderian, basement locally contains evidence of older Early Cambrian orogenesis, but this is probably related to peripheral subduction around the Gondwanan continent outside the realm of Iapetus. The accretions of the Baie Verte Oceanic Tract and the Penobscot complex were nearly coeval, but took place on opposite sides of the Iapetus Ocean, when the ocean was at its widest. Because these events took place shortly after the initiation of subduction on either side of the ocean, they signal a major plate reorganization that changed Iapetus from an Atlantic-type ocean into a Pacific-type ocean. Time constraints provided by stitching plutons indicate that both accretionary events were short-lived, forming mélanges and ductile–brittle deformation, and did not lead to any significant regional metamorphism.

The first major orogenic event recorded in the Northern Appalachians was the Early to Middle Ordovician (474–455 Ma) Taconic Orogeny, which principally resulted from closure of the Humber seaway and accretion of the Dashwoods microcontinent to Laurentia. Closure of the Humber Seaway initiated loading of the Humber margin, initially by emplacement of oceanic terranes, some of which are younger than the Baie Verte Oceanic Tract (484–479 Ma). These younger oceanic terranes, which include the Bay of Islands and Thetford Mines ophiolites in Newfoundland and Quebec, respectively, formed either during the initiation of subduction in the Humber Seaway or as a result of localized pericratonic seafloor spreading in pull-aparts or above embayments in the downgoing plate. The Taconic Orogeny culminated in imbrication, folding, and locally intense Barrovian metamorphism, mainly in upper-plate rocks associated with the Dashwoods. Age constraints on Taconic metamorphism (470–455 Ma) are remarkably consistent along the length of the orogen.

Docking of Dashwoods was rapidly followed by accretion of the Annieopsquotch accretionary tract between 470 and 460 Ma and the Popelogan–Victoria Arc at 455–450 Ma to composite Laurentia along its ocean-facing eastern margin. The effects of these accretionary events, which are centred on the Red Indian Line, are difficult to separate from those of the orogenesis accompanying the Dashwoods–Laurentia collision to the west. Since separation of these events on the basis of kinematic arguments is often impossible outside Newfoundland, these events are included in the Taconic orogeny. Structural relationships and the presence of young supra-subduction-zone oceanic rocks (480–464 Ma) in the Annieopsquotch accretionary tract indicate that a new west-dipping subduction zone had formed in the Arenig to the east of Dashwoods. Accretion involved south-east-verging duplex-style thrust complexes, folding, mélange development, and a sinistral oblique reverse shear zones.

The Popelogan–Victoria Arc was formed in the Early Arenig (at about 478 Ma) on the leading edge of Ganderia after the Penobscot orogeny. It was active until the Late Caradoc and formed above an east-dipping subduction zone. Upper-plate extension and rifting as a result of slab rollback dispersed the Popelogan–Victoria Arc into Iapetus and formed a wide (up to 1000 km) Japan Sea-style back-arc basin (Tetagouche–Exploits back-arc) that was partly underlain by Llanvirn oceanic crust.

The opposite subduction polarities associated with the Annieopsquotch accretionary tract (west-dipping) and the Popelogan–Victoria Arc (east-dipping) require that the Red Indian Line marks a collision between two arc terranes (Figure 4). Structural and seismic-reflection data in central Newfoundland indicate that the Popelogan–Victoria Arc was partly subducted beneath the Annieopsquotch accretionary tract. Palaeomagnetic evidence suggests that between 480 and 450 Ma these two subduction zones consumed approximately 3000 km of ocean, implying a minimum closure rate of around 10 cm yr^{-1} in the main Iapetus tract. The remaining tracts of Iapetus (Humber Seaway, Tetagouche–Exploits back-arc, and the ocean

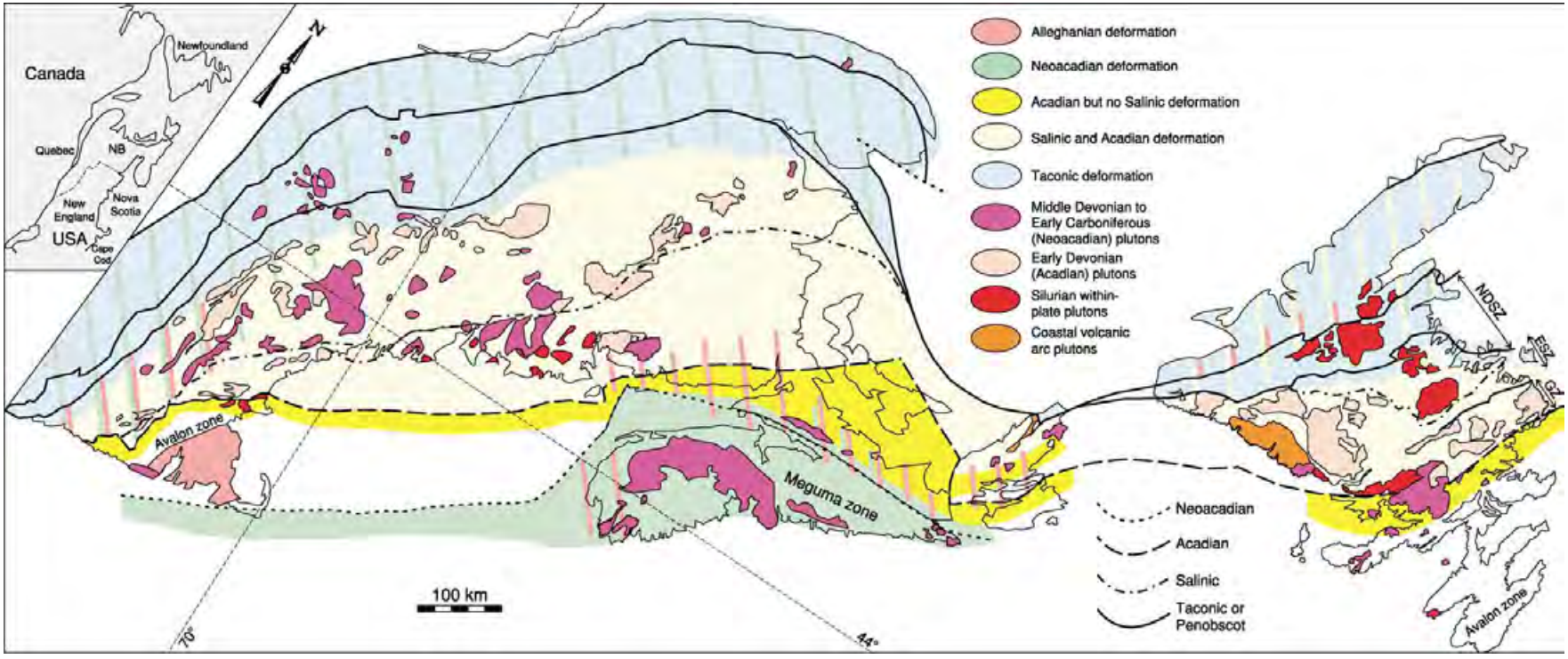


Figure 5 Map of the Northern Appalachians showing the spatial and temporal distributions of deformation, associated metamorphism, and Siluro-Early Carboniferous plutons. For abbreviations see **Figure 1**.

basin separating Avalonia and Ganderia) had a combined width in the order of 2000 km.

After accretion of the Popelogan–Victoria Arc to Laurentia, the west-dipping subduction zone stepped back into the Tetagouche–Exploits back-arc. Late Ordovician–Early Silurian mélanges, blueschists, and foredeep deposits leave little doubt that subduction continued. Closure of the Tetagouche–Exploits back-arc in the Early Silurian (around 430 Ma) accreted the bulk of Ganderia firmly to Laurentia along the Dog Bay–Liberty–Orrington Line. This event, the Salinic Orogeny, was contemporaneous in Newfoundland, New Brunswick, and Maine. Although mainly restricted to rocks east of the Red Indian Line, localized Salinic-age deformation west of this suture is probably related (Figure 5).

Sinistral oblique accretion of Avalonia to Laurentia started at around 421 Ma, shortly after the docking of Ganderia. This caused inversion of the Mascarene–La Poile back-arc basin in southern New Brunswick, Maine, and Newfoundland and led to penetrative orogenesis of Ganderian, including its Siluro-Devonian cover, and, to a much lesser extent, immediately adjacent Avalonian rocks. This collision is the Acadian Orogeny as defined in its type area along the south coast of New Brunswick and neighbouring Maine. It produced polyphase folding including an early phase of large south-east-verging fold nappes and thrusts, a north-west-verging retrowedge also involving fold nappes and ductile thrusts, high-temperature metamorphism, and extensive syntectonic granitic plutonism with bimodal magmatism developing later on. Acadian orogenesis finished before the end of the Early Devonian and did not extend west of the Red Indian Line, other than as localized faulting and development of clastic wedges in foreland basins. Subsequent late Early Devonian to Early Carboniferous Neoacadian orogenesis is present throughout most of the Northern Appalachians (Figure 5) and is at least in part related to the docking of Meguma to the Laurentian margin. The Neoacadian Orogeny was accompanied by a westerly progradation of clastic wedges and granitic magmatism in western New England and maritime Canada and widespread dextral transpression.

The Carboniferous–Permian Alleghanian Orogeny was due to tectonic processes related to the oblique convergence and collision of Gondwana with Laurentia, which is in an upper-plate setting in the Northern Appalachians. Dextral strike-slip faulting and the development of intermontane transtensional basins prevailed during the Early Carboniferous. Transpressional inversion of these basins involved several stages during the late Early Carboniferous to Middle Permian and was accompanied locally by intense deformation and metamorphism near the boundary between the Avalon and Meguma zones. Intense late Alleghanian orogenesis accompanying renewed north-directed underthrusting of Avalonian and Ganderian rocks beneath Laurentia occurred in southern New England, but is absent elsewhere.

The temporal and spatial relationships of the tectonic events discussed above indicate that the Northern Appalachians are an example of a Palaeozoic accretionary orogen, punctuated by the accretion of small suprasubduction-zone terranes and five collisional events related to the arrival of four microcontinents (Dashwoods, Ganderia, Avalonia, and Meguma) and finally Gondwana at the Laurentian margin. Laurentia was thus progressively expanding oceanwards concurrent with an eastward shift of the locus of orogenesis (Figure 5).

Glossary

Boundary Mountain Terrane Equivalent of Dashwoods microcontinent in New England-Southern Quebec. Is inferred to underlie most of the Connecticut-Gaspè Belt

Ganderia Gondwanan-derived microcontinent. Characterized by island arc and back-arc magmatism during the Early Palaeozoic.

See Also

Europe: Caledonides of Britain and Ireland. **North America:** Southern and Central Appalachians. **Plate Tectonics**. **Sedimentary Environments:** Depositional Systems and Facies. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Faults; Folding; Mountain Building and Orogeny; Ocean Trenches.

Further Reading

Barr SM, White CE, and Miller BV (2002) The Kingston Terrane, southern New Brunswick, Canada: evidence for an Early Silurian volcanic arc. *Geological Society of America Bulletin* 114: 964–982.

Dunning G, O'Brien SJ, Colman-Sadd SP, *et al.* (1990) Silurian orogeny in the Newfoundland Appalachians. *Journal of Geology* 98: 895–913.

Fortey RA and Cocks LRM (2003) Palaeontological evidence on global Ordovician–Silurian continental reconstructions. *Earth Science Reviews* 61: 245–307.

Karabinos P, Samson SD, Hepburn JC, and Stoll HM (1998) Taconian orogeny in the New England Appalachians: collision between Laurentia and the Shelburne Falls arc. *Geology* 26: 215–218.

Kerr A (1997) Space–time relationships among Appalachian-cycle plutonic suites in Newfoundland. In: Sinha AK, Whalen JB, and Hogan JP (eds.) *The Nature of Magmatism in the Appalachian Orogen,* pp. 193–220.

Geological Society of America Memoir 191, Boulder, Colorado: Geological Society of America.

Kim J and Jacobi RD (2002) Boninites: characteristics and Appalachian constraints, north-eastern Appalachians. *Physics and Chemistry of the Earth* 27: 109–147.

Robinson P, Tucker RD, Bradley D, Berry HN IV, and Osberg PH (1998) Paleozoic orogens in New England, USA. *Geologiska Föreningens Stockholm Forhandlingar* 120: 119–148.

Swinden HS, Jenner GA, and Szybinski ZA (1997) Magmatic and tectonic evolution of the Cambrian-Ordovician Laurentian margin of Iapetus: Geochemical and isotropic constraints from the Notre Dame Subzone, Newfoundland. In: Sinha K, Whalen JB, and Hogan JP (eds.) *The nature of magmatism in the Appalachian Orogen,* pp. 337–365. Geological Society of America Memoir 191, Boulder, Colorado: Geological Society of America.

Van der Pluijm BA, Johnson RJE, and van der Voo R (1993) Paleogeography, accretionary history, and tectonic scenario: a working hypothesis for the Ordovician and Silurian evolution of the northern Appalachians. In: Roy DC and Skehan JW (eds.) *The Acadian Orogeny,* pp. 27–40. Geological Society of America Special Paper 275, Boulder, Colorado: Geological Society of America.

Van Staal CR, Sullivan RW, and Whalen JB (1996) Provenance and tectonic history of the Gander Margin in the Caledonian/Appalachian Orogen: implications for the origin and assembly of Avalonia. In: Nance RD and Thompson MD (eds.) *Avalonian and Related Peri-Gondwanan Terranes of the Circum-North Atlantic,* pp. 347–367. Geological Society of America Special Paper 304, Boulder, Colorado: Geological Society of America.

Van Staal CR, Dewey JF, Mac Niocaill C, and McKerrow WS (1998) The Cambrian–Silurian tectonic evolution of the northern Appalachians: history of a complex, southwest Pacific-type segment of Iapetus. In: Blundell DJ and Scott AC (eds.) *Lyell: the Past is the Key to the Present,* pp. 199–242. Geological Society Special Publication 143, London: Geological Society.

Waldron JF and van Staal CR (2001) Taconian orogeny and the accretion of the Dashwoods block: a Peri-Laurentian microcontinent in the Iapetus Ocean. *Geology* 29: 811–814.

Whalen JB, Jenner GA, Longstaffe FJ, Gariépy C, and Fryer BJ (1997) Implications of granitoid geochemical and isotopic (Nd, O, Pb) data from the Cambrian–Ordovician Notre Dame arc for the evolution of the Central Mobile belt, Newfoundland Appalachians. In: Sinha K, Whalen JB, and Hogan JP (eds.) *The Nature of Magmatism in the Appalachian Orogen,* pp. 367–395. Geological Society of America Memoir 191, Boulder, Colorado: Geological Society of America.

Williams H (ed.) (1995) *Geology of the Appalachian-Caledonian Orogen in Canada and Greenland.* Geological Survey of Canada, Geology of Canada, no. 6 (also Geological Society of America, The Geology of North America, v. F1).

Williams H, Currie KL, and Piasecki MAJ (1993) The Dog Bay Line: A major Silurian tectonic boundary in northeast Newfoundland. *Canadian Journal of Earth Sciences* 30: 2481–2494.

Atlantic Margin

D R Hutchinson, US Geological Survey, Woods Hole, MA, USA

Published by Elsevier Ltd.

Introduction

The North American Atlantic Continental Margin stretches between Florida and Newfoundland and extends from on land, near the exposed Mesozoic Newark basins, to offshore, near the western limit of the abyssal plain. It is a classic example of a passive or trailing-edge continental margin, in which continental rifting followed by seafloor spreading, sedimentation, and subsidence are the primary controls on margin morphology, tectonics, and sedimentary style. At almost 4500 km long, the margin considered here spans climatic zones from subtropical off the coast of Florida to subarctic off the coast of Newfoundland. During Mesozoic continental separation, the rifting style changed from volcanic, during the North America–Africa separation, to non-volcanic, during the North America–Europe separation. Sediments in the five major offshore basins contain varying volumes and ages of evaporite, carbonate, coarse terrigenous, and fine siliciclastic deposits.

This margin is documented by one of the most thorough high-quality publicly available geological and geophysical datasets of any continental margin in the world. Exploration began in the 1960s, when the new theory of plate tectonics ignited an excitement about Earth science and stimulated a new generation of enthusiastic scientists to rethink how the Earth was formed. The Arab Oil embargo in the 1970s led to the initiation and funding of a sustained period of continental-margin exploration aimed at understanding energy resources in US domestic frontier environments, such as the unexplored continental margins. Finally, the adoption of the United Nations Convention on the Law of the Sea in 1982 provided the legal basis for nations to claim up to 200 nautical miles (or more) for their exclusive economic use. Together, these three events made exploration of the vast

submerged portion of the North American Atlantic Continental Margin possible. This article summarizes the diverse geology of this passive margin discovered during the last 40 years of scientific research.

Morphology of the Margin

The bathymetry of the North American Atlantic Continental Margin (Figure 1) shows a continental shelf that dips gently seawards, an abrupt shelf–slope break at a depth of about 100–200 m, a steep and sometimes dissected continental slope, and, finally, a gently sloping rise that merges with the abyssal plains of the deep ocean. The Blake Plateau is an exception to this morphology. It is a submerged platform at a depth of about 800–1200 m, which ends abruptly in a near-vertical escarpment that descends to more than 5000 m. The morphology at the north and south ends of the margin reflect modification by climatic zonation. For example, between New England and Newfoundland, the margin was shaped by Pleistocene glaciation (*see* **Sedimentary Processes:** Glaciers): banks, channels, ridges, knolls, rocky outcrops, and troughs are observed. In contrast, off the coast of Florida, subtropical reef growth, carbonate deposition, submarine carbonate lithification (*see* **Minerals:** Carbonates), and carbonate solution in deep water have led to oversteepened slopes.

Superimposed on the broad climatic effects are features resulting from oceanographic, geological, and tectonic modifications. The oceanographic processes that modify the seafloor are primarily tides, waves, and currents (for example, creating mobile sand bodies atop Georges Bank and around Sable Island). In deeper water, the Blake Outer Ridge (Figure 1) was formed by a huge sediment drift deposited at the confluence of southwards-flowing and northwards-flowing deep currents.

Perhaps the most striking features of the North American Atlantic Continental Margin are the many submarine canyons that incise the continental slope and rise between Cape Hatteras and Newfoundland.

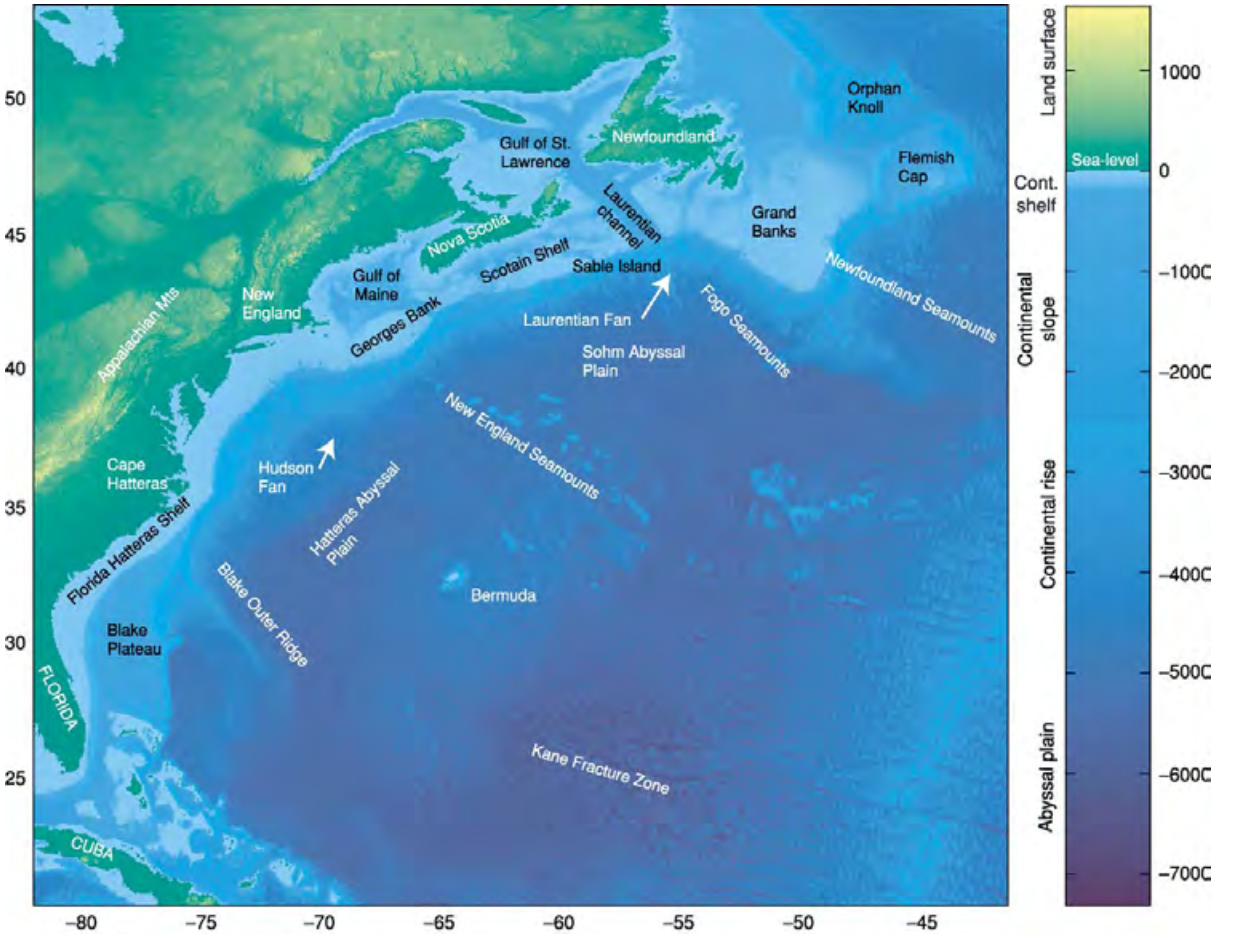


Figure 1 Bathymetric map of the North American Atlantic Continental Margin. Bathymetry was generated using ETOPO2 values with shaded relief. (ETOPO2 provides 2′ × 2′ depth and elevation values, from NOAA/National Geophysical Data Center.)

More than 200 canyons can be identified on the upper slope off the US margin. Many of these canyons also indent the shelf edge. The largest canyon system, the Hudson Shelf Valley and Canyon, can be traced back to the mouth of the Hudson River (Figure 2). Abundant slumps, debris-flow deposits, and other evidence of mass movement are seen on the sides of many canyons. Fewer canyons exist on the lower slope because the canyons coalesce there. The lower slope and rise also contain large submarine fan deposits (e.g. the Laurentian and Hudson Fans off Newfoundland and New York, respectively).

Deeper-water geological processes are reflected in the presence of three seamount (*see* **Seamounts**) chains that impinge upon the margin: the New England Seamounts off Georges Bank, the Fogo Seamounts south of the Grand Banks, and the Newfoundland Seamounts off Newfoundland. A rifted continental fragment forms the submerged circular platform of Flemish Cap.

Several large landslides (*see* **Sedimentary Processes: Landslides**) also modify the margin morphology. The undated Cape Fear slide (Figure 3), south of Cape Hatteras, extends downslope for more than

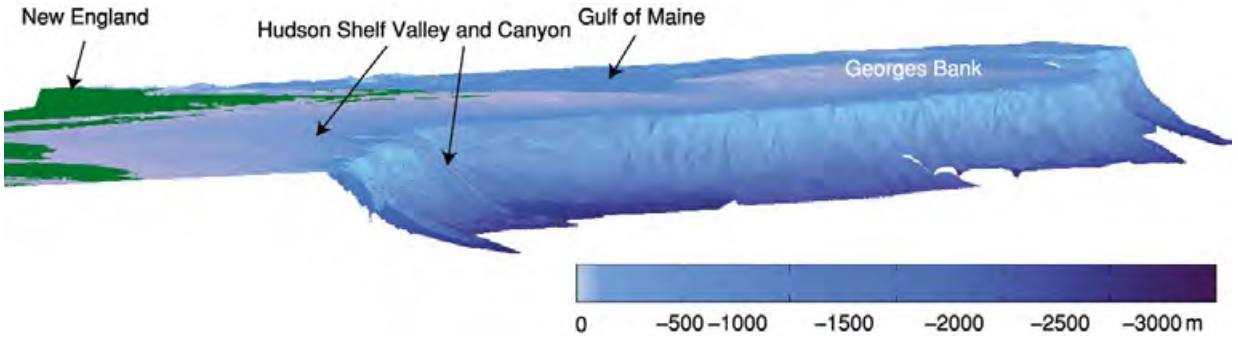


Figure 2 Perspective view of the canyons on the continental margin using bathymetric values from the NOAA coastal relief model (90 m × 90 m grid).

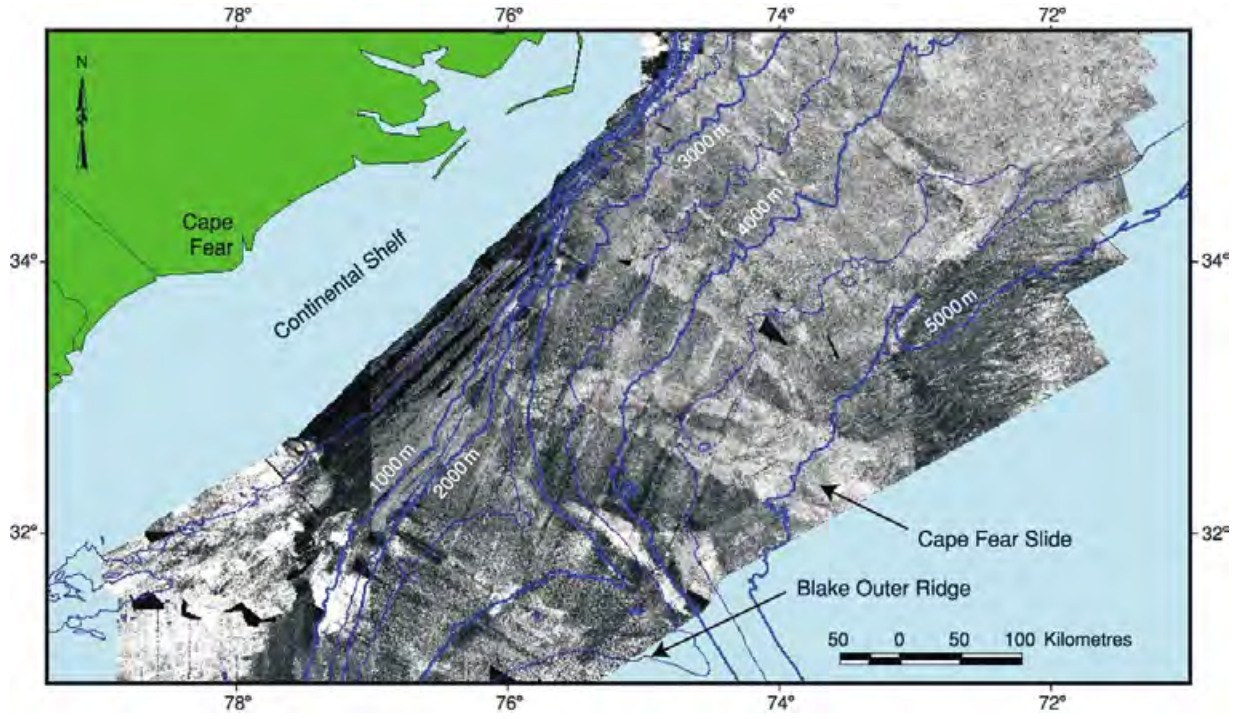


Figure 3 GLORIA side-scan sonar image of the continental slope along a segment of the south-eastern US margin, showing the Cape Fear slide, an undated massive submarine landslide south-east of Cape Hatteras, NC. Dark grey areas are regions of low backscatter signal, commonly from fine-grained deposits; light grey areas are regions of high backscatter signal, commonly from coarse-grained deposits. The Cape Fear slide appears as a light grey region.

300 km between about 1200 m and 5500 m water depth. Off Newfoundland, the slides, turbidites, and other mass-flow bedforms resulting from the 1929 earthquake on the south-eastern Grand Banks can be identified in water depths from about 500 m to 3000 m.

Tectonic Evolution

The North American Atlantic Continental Margin is the product of continental rifting and seafloor spreading (*see* **Plate Tectonics**), which began in the Mesozoic between North America, Africa, and, eventually, Europe. Rifting is the phase of stretching and fracturing of continental crust that culminates in seafloor spreading (drifting), which creates new oceanic crust as the continental blocks move apart. These tectonic processes are directly responsible for the modern configuration of basins, platforms, magnetic anomalies, and oceanic fracture zones along the margin (Figure 4).

Early Triassic rifting began in an extensive interior mountain range in the supercontinent of Pangaea (*see* **Pangaea**) (Figure 5A). A series of rift basins formed between what are now the Gulf of Mexico and the Bay of Biscay. Rifting intensified in the Late Triassic (Norian), and true seafloor spreading progressed from the south at about 200 Ma to the north at about 185 Ma in what is now maritime Canada. Spreading near Newfoundland initially did not continue north of the Gibraltar transform zone at the northern edge of the African block. The Grand Banks were located north of the Gibraltar transform and, although extended during this time, behaved as a unit with Gibraltar and the European Plate.

Rifting and drifting along the entire margin are inferred to have overlapped during the Early Jurassic. A major basaltic extrusion event is recorded in the rift basins at about 200 Ma (Early Jurassic). Basalt flows are exposed throughout the Newark Supergroup. Numerous feeder dykes can be mapped in the exposed rocks surrounding the rifts (e.g. Shelburne Dyke in Nova Scotia, the Great Dyke in the Carolinas). These dykes probably also exist beneath the submerged continental margin, but are difficult to map beneath large thicknesses of postrift sediments. Rifting ceased during the Middle Jurassic.

The East Coast Magnetic Anomaly (ECMA; Figure 4), marks the western edge of the oceanic crust. The oldest oceanic crust near the ECMA was created during the Jurassic magnetic quiet zone and, therefore, is does not contain datable seafloor magnetic anomalies. The oldest seafloor-spreading isochron that can be identified on both sides of the Atlantic is M25 (about 156 Ma, Late Jurassic).

During the Middle and Late Jurassic, the continents dispersed and readjusted to tectonic events that were occurring in other parts of the fragmenting remnants of the Pangaean supercontinent. The initially shallow and restricted ocean basins deepened and widened as North America and Africa separated. In the Late Jurassic, a major plate (*see* **Plate Tectonics**) reorganization between North America and Europe rejuvenated synrift basins that initially formed in the Triassic on the Grand Banks of Canada (e.g. Jeanne d'Arc Basin). This rifting ended when seafloor spreading began between the Grand Banks and Iberia during the early Barremian (126 Ma, Early Cretaceous), initiating the final separation of Europe from North America (Figure 5B). Another plate reorganization in the Early Cretaceous (middle Aptian, about 110 Ma) completed the separation of North America from Europe when the northern part of Flemish Cap (and Orphan Knoll) separated from the Rockall region of the north-west European margin as seafloor spreading began in the Labrador Sea.

Alkaline basaltic magmatism, not obviously associated with breakup, created the New England Seamounts, (*see* **Seamounts**) which extend from the continental slope (Bear Seamount) to the foothills of the Mid-Atlantic Ridge. The age of the seamount chain has been difficult to establish, but it is now generally felt to have formed episodically between the Early Jurassic (125 Ma) and the middle Cretaceous (90–82 Ma). A younger, middle Cretaceous age for some of the White Mountain magma series in New Hampshire indicates that this younger magmatic pulse may have extended inland.

The Cenozoic has been a period of stable plate configurations, in which the Atlantic Ocean has widened and deepened. There are two exceptions to this otherwise stable tectonic regime. First, the Labrador spreading centre was abandoned and the Norwegian and Arctic Oceans began opening in the Late Paleocene. This gave rise to the mid-ocean ridge configuration that persists today between Newfoundland and Florida (Figure 5C). Second, at least two large bolides hit the margin (*see* **Impact Structures**). One excavated a crater 85 km wide beneath Chesapeake Bay in the late Eocene (35 Ma; Figure 6). A second smaller impact structure, 45 km wide, called the Montagnais structure, was formed at about 51 Ma near the shelf–slope break off Nova Scotia.

Passive Margin Structure

During the last 200 Ma, three major tectonic processes have affected the structure of the North American Atlantic Continental Margin: rifting, which broke apart Pangaea, thinned the continental crust by

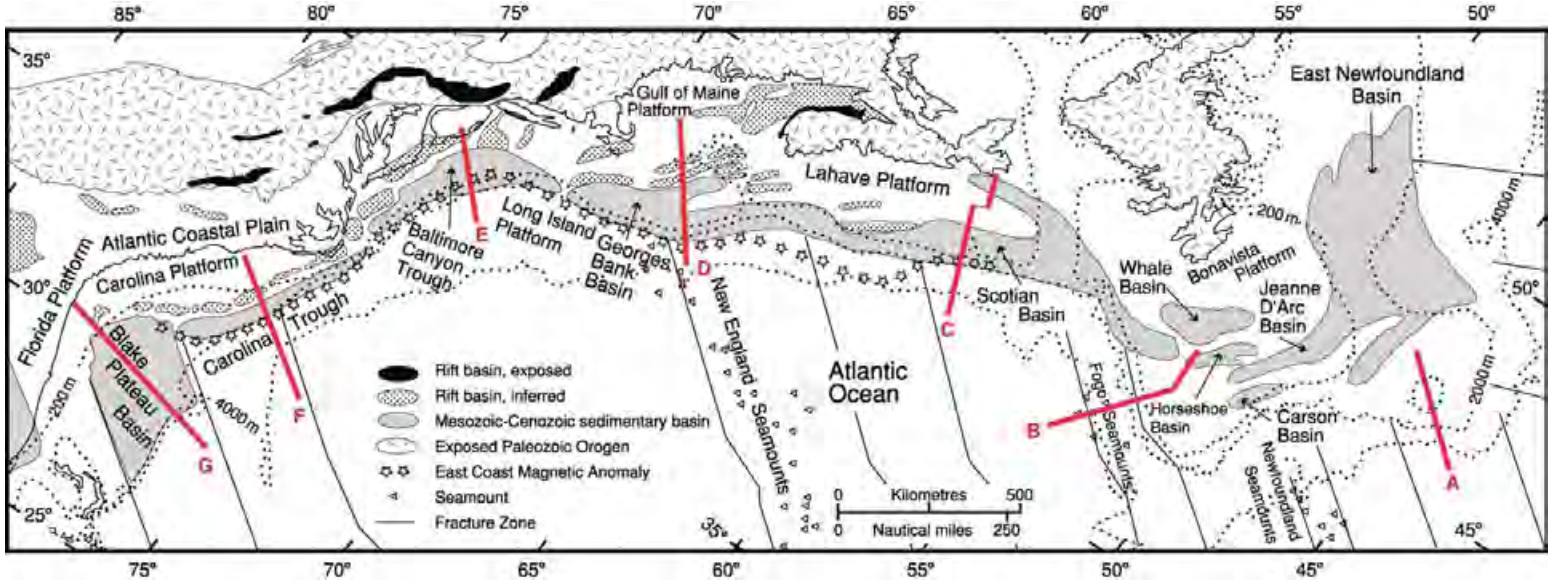


Figure 4 Tectonic map of the North American Atlantic Continental Margin showing the major postrift basins, platforms, and rift basins associated with Mesozoic breakup. The Newark Supergroup consists of the basins labelled as 'Rift basin, exposed.' The red lines show the locations of the crustal cross sections in **Figure 8**. The East Coast Magnetic Anomaly marks the landwards extent of the oceanic crust.

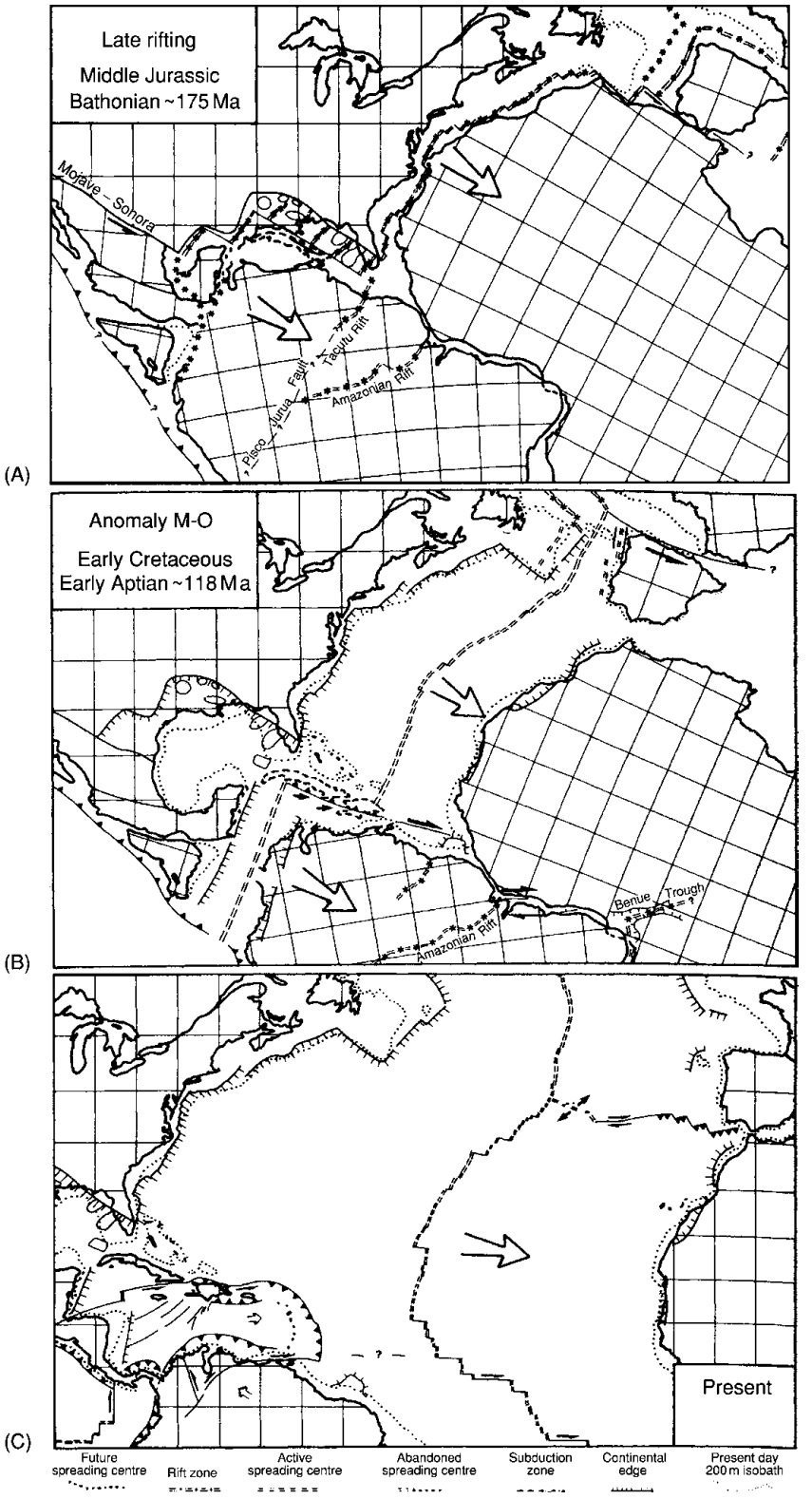


Figure 5 Three plate-tectonic configurations relative to North America. (A) Rifting (Middle Jurassic) and initiation of seafloor spreading when Iberia and Newfoundland were part of a single plate. (B) Midway through the creation of the Atlantic Ocean (Early Cretaceous), shortly after Iberia and Newfoundland separated. (C) The current plate separations. Large arrows indicate the plate separation directions relative to North America. (Modified from Klitgord KD and Schouten H (1986) Plate kinematics of the central Atlantic. In: Vogt PR and Tucholke BE (eds.) (1986) *The Geology of North America, Volume M: The Western North Atlantic Region*, pp. 351–378. Boulder, Colorado: Geological Society of America.)

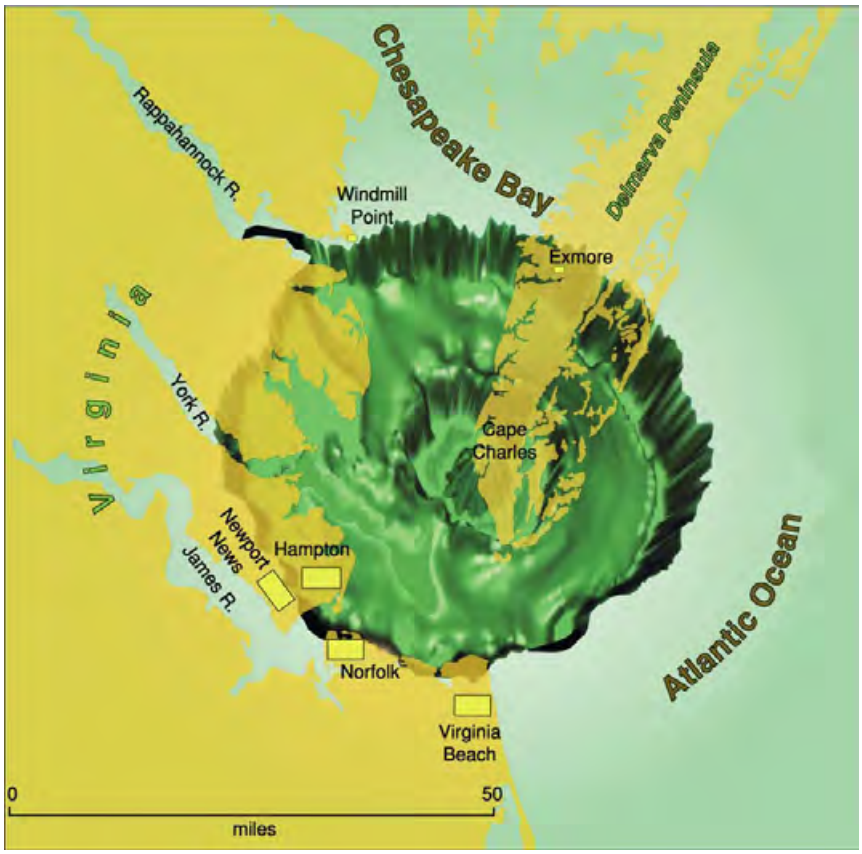


Figure 6 Three-dimensional perspective of the Chesapeake Bay impact crater, which was formed at about 35 Ma on the North American Atlantic Continental Margin. The inner and outer crater walls are shown superimposed on a map of the Virginia part of Chesapeake Bay. (Reproduced from Poag CW, Koeberl C, and Reimold WU (2004) *The Chesapeake Bay Crater: Geology and Geophysics of a Late Eocene Submarine Impact Structure.* New York: Springer-Verlag.)

variable amounts, and created numerous synrift basins; drifting, which created hot new oceanic crust initially adjacent to the rifted continent and now far away along the Mid-Atlantic Ridge; and passive subsidence, caused by cooling of the lithosphere and loading of the crust by sediments, which resulted in the development of large offshore sedimentary basins. Superimposed on these first-order lithospheric events are various oceanographic, magmatic, fluvial, eustatic, sedimentary, diagenetic, and biological processes that have also affected the final shape of the margin. The resulting structure of the margin is an alternating sequence of platforms, which have relatively little sedimentary cover, and basins, which contain enormous thicknesses of sediment (Figure 4).

Mesozoic rifting created the exposed basins of the Newark Supergroup (Figure 4). These fault-bounded basins and/or half-grabens form a linear chain along the exposed part of the margin. The basins contain deposits up to several kilometres thick. Similar synrift basins have been detected seismically and drilled near the coast and beneath the continental shelf. The best-known of these offshore basins is the petroliferous Jeanne d'Arc basin on the Grand Banks, in which the thickness of sediment exceeds 14 km. Synrift deposits are inferred to underlie the deepest portions of the margin, but have not yet been sampled. Synrift basins are also known on the conjugate north-west African continental margin.

A prominent margin-wide unconformity separates the synrift basins from the overlying postrift sedimentary units and is called the postrift or the breakup unconformity (Figure 7). Consequently, sedimentary-thickness maps of the margin often include only those units above the postrift unconformity (i.e. postrift deposits). The postrift unconformity may become a conformable surface within the deepest sediments in the centres of the postrift basins, but evidence for this is ambiguous, because seismic imaging in these deeper regions is often of poor quality and drilling data are lacking. Beneath the Grand Banks, the postrift unconformity is termed the Avalon unconformity.

Above the postrift unconformity, five major sedimentary basins have been documented. From north to south, they are the Scotian basin, Georges Bank basin,

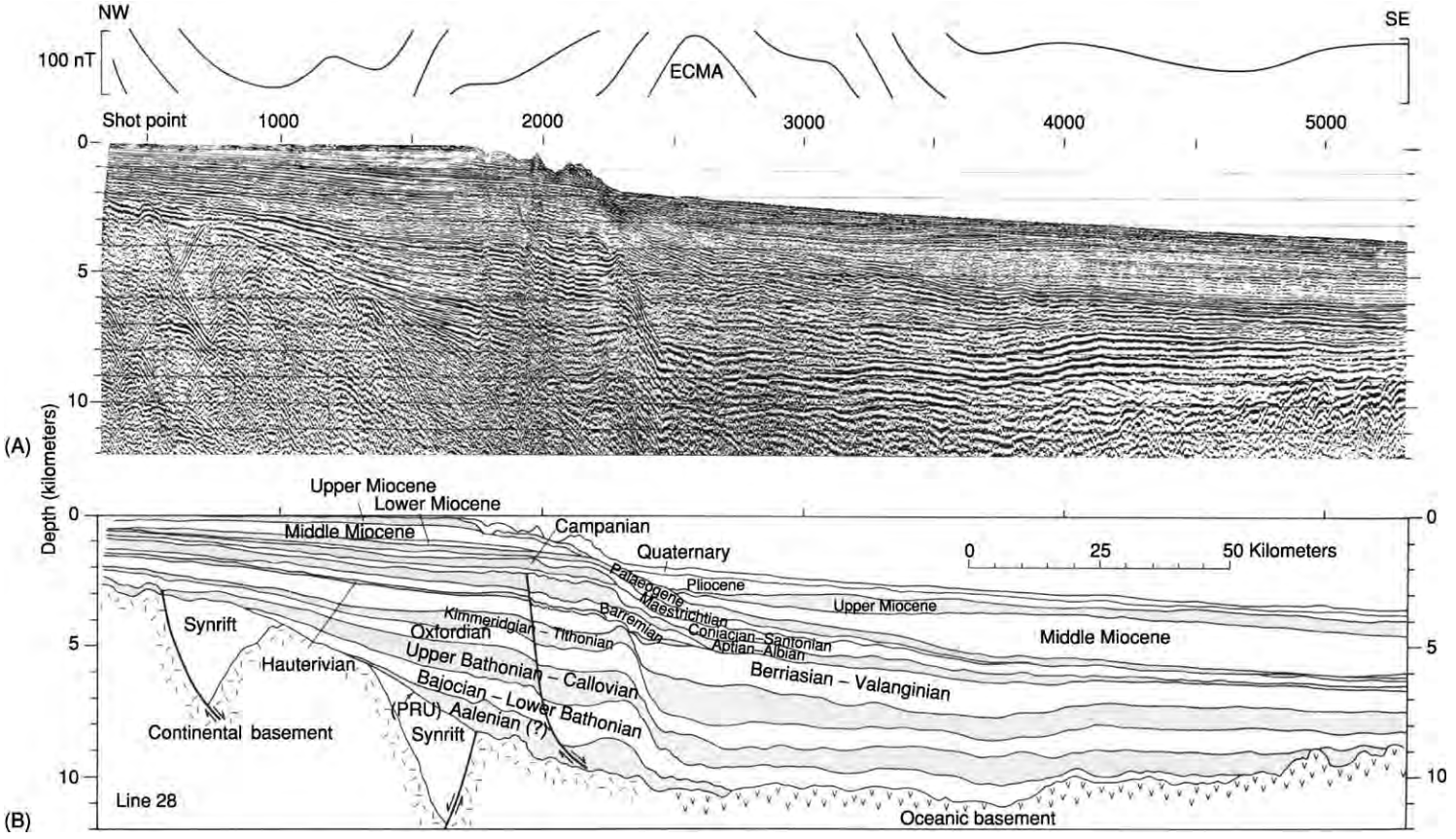


Figure 7 Stratigraphic cross section across the southern Baltimore Canyon trough. (A) Uninterpreted multichannel seismic-reflection data. The magnetic anomaly is plotted across the top of the section and shows the location of the East Coast Magnetic Anomaly (ECMA). (B) Interpretation, giving ages and geometries of units. PRU, postrift unconformity. (Reproduced from Poag CW (1992) US Atlantic continental rise: provenance, dispersal, and deposition of Jurassic to Quaternary sediments. In: Poag CW and de Graciansky PC (eds.) *Geologic Evolution of Atlantic Continental Rises*, pp. 100–156. New York: Van Nostrand Reinhold.)

Baltimore Canyon trough, Carolina trough, and Blake Plateau basin (Figure 4). In plan view, they range in shape from nearly circular (Blake Plateau basin) to extremely elongated (Carolina trough). In cross section (Figure 7), the basin fill is roughly lens shaped, thinning towards both the continent and the ocean. Maximum sediment thickness also varies, from more than 15 km in the Baltimore Canyon trough and Scotian basin to less than 9 km in the Georges Bank basin. The composition of the basin fill is variable, and the sediment is often distributed among several sub-depocentres that together make up a basin complex.

The basins are bounded on their sides by structural highs, which often coincide with major faults. These faults both segmented the margin during rifting and coincided with (and perhaps localized) fracture zones in the newly forming oceanic crust. Most of the fracture zones can be mapped continuously from magnetic-anomaly offsets to the present Mid-Atlantic Ridge and are used as flow-lines of plate movement in plate-tectonic and palaeoenvironmental reconstructions of the ocean basin. Fracture-zone faults near the margin are mostly deeply buried, but they can be identified in the seafloor morphology near the mid-ocean ridge (e.g. Kane Fracture Zone; Figure 1).

Landwards of the large postrift basins are structural platforms. These are, from north to south, the Bonavista, LeHave, Long Island, Carolina, and Florida platforms. These regions may contain locally thick synrift deposits (e.g. the Jeanne d'Arc basin) but generally consist of prerift crystalline basement at shallow depths. Postrift deposits are thin (less than 4 km) and represent the onlapping edges of much thicker offshore units. Coastal-plain deposits are the exposed portions of these thinner younger Mesozoic and Cenozoic units. A hinge zone frequently marks the transition from a platform to a basin, i.e. where sediment thickness changes abruptly. These platforms are also affected by eustatic changes; hence, many parts of the sedimentary section are missing. The platforms are underlain by crystalline rocks from older Palaeozoic orogenies that built the Appalachian, Caledonian–Variscan, and Mauritanian mountains of North America, Europe, and Africa, respectively. The basin-bounding faults of the synrift basins frequently reactivate deeper basement faults and structures.

Variations in the amount of crustal stretching can be seen in cross sections through different parts of the North American Atlantic Continental Margin (Figure 8). Three crustal types are shown: oceanic crust, which is 5–6 km thick and similar to normal oceanic crust; transitional crust, which is generally less than 20–25 km thick, changes thickness rapidly, and has seismic velocities intermediate between those of oceanic and continental crust; and thinned continental crust, which is 30–35 km thick and may contain synrift basins but otherwise appears to be continental in seismic-velocity character (i.e. the platforms). ('Seismic' is used throughout this article to include coincident reflection and refraction techniques, unless otherwise noted.) A fourth type of crust, which is not shown on these cross sections, is normal continental crust, which would be 40–45 km thick, as found in continental interiors. The change from one crustal type to another may be abrupt or gradational and depends on the geological, tectonic, magmatic, and thermal history of the margin prior to and during breakup.

The transform margin of the southern Grand Banks basin (Figure 8) changes abruptly from continental crust to oceanic crust across a 40 km wide zone, whereas most of the basins have transitional crust that is 100–200 km wide (Scotian basin, Georges Bank basin, Baltimore Canyon trough, Carolina trough). The Blake Plateau basin is the most unusual, because the transition from continental crust to oceanic crust extends over nearly 400 km. Deep seismic data do not exist for the Blake Plateau basin, so the interpretation of crustal type is based primarily on poorly constrained gravity models. The basement and crustal structure in this southernmost basin are the least understood of any part of the margin.

The nature of the transitional crust also varies along the margin. For example, the southern part of the margin, towards Florida, is considered to be strongly volcanic, whereas the northern part, near Flemish Cap, is considered to be non-volcanic. The EDGE seismic experiment showed that a thick wedge of high-velocity presumably mafic material exists and coincides with rifting beneath the Carolina trough. Thus, this part of the margin is classified as a volcanic margin. Similar velocities have been interpreted from beneath the Baltimore Canyon trough. In contrast, a recent seismic experiment across Flemish Cap revealed that both serpentinized mantle and unusually thin oceanic crust lie adjacent to thinned continental crust. These anomalous features are interpreted as evidence of ultraslow seafloor spreading associated with a non-volcanic margin. The conjugate Iberian margin also has serpentinized rocks, which are thought to indicate non-volcanic processes. These have been drilled in the Ocean Drilling Program. The deeper structures of the Scotian and Georges Bank basins have not been well imaged using modern seismic techniques. Therefore, the nature of the deeper crust in these regions is uncertain, hindering our understanding of both the transition from continent to ocean and the transition from volcanic to non-volcanic parts of this margin.

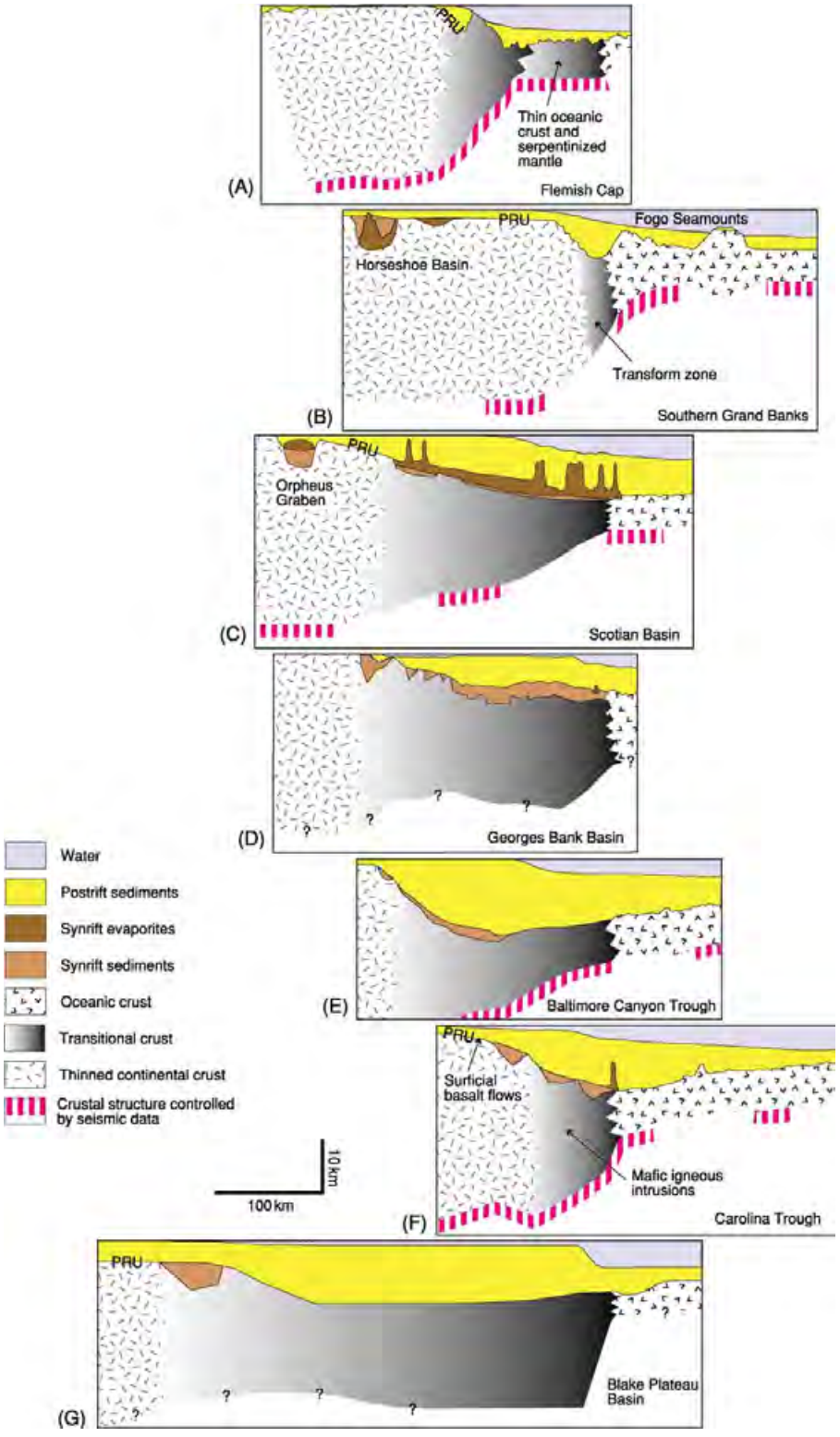


Figure 8 Simplified crustal cross sections through the major postrift basins of the North American Atlantic Continental Margin, aligned on the western edge of the oceanic crust. The transitional zone between continental crust and oceanic crust varies in thickness and width. PRU, postrift unconformity. (A) Flemish Cap, interpreted as a non-volcanic transition (the two transitional-crust patterns differentiate highly thinned continental crust on the landwards side from anomalous oceanic or serpentinized crust adjacent to the oceanic crust). (B) Southern Grand Banks, showing a narrow transition at the position of the fossil transform fault zone. (C) Scotian Basin. (D) Georges Bank Basin. (E) Baltimore Canyon Trough. (F) Carolina Trough, interpreted as a volcanic transition (the transition region is thought to contain magmatically emplaced crust). (G) Blake Plateau Basin.

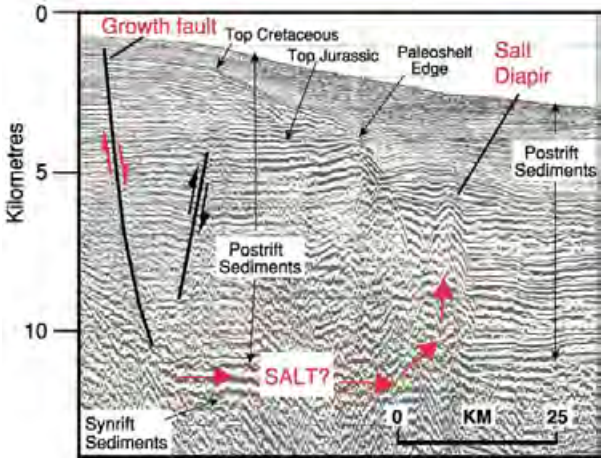


Figure 9 Seismic-reflection profile and interpretation across the Carolina trough, showing the relationship between salt diapirism and faulting. Salt movement, shown by the arrows, is inferred to have caused growth faulting. PRU, postrift unconformity.

Numerous faults that offset postrift sediments are secondary structures on the continental margin. Beneath the Scotian basin and Carolina trough, faults are attributed to salt tectonism (Figure 9). Numerous small Cretaceous and younger north-east-trending reverse faults (e.g. the Block Island and New York Bight faults on the continental margin south of New England) are part of growing evidence for regional compression of the margin from Early Cretaceous through to at least Pleistocene times. The stress field east of the Rocky Mountains remains compressional today. Most of the faulting associated with the Chesapeake Bay bolide impact appears to be localized to the region surrounding the crater.

Sedimentary History and Palaeoenvironments

Interpretations of sedimentary history and palaeoenvironments are based on a combination of seismostratigraphic interpretation and direct samples from dredges, cores, drill cuttings, and submersible observations.

Triassic rifting began in central Pangaea (*see* **Pangaea**), where average elevations were probably more than 1 km above Mesozoic sea-level. Continental clastic redbed sequences filled large lakes that occupied subsiding rift basins, not unlike the large lakes of the modern East African Rift valley. The region from Florida to the Grand Banks spanned a latitudinal range from about 5° S to 20° N and encompassed equatorial rainforests to tropical savannahs. Evaporites (*see* **Sedimentary Rocks:** Evaporites) accumulated in shallower settings. Rift topography, and especially uplifted rift shoulders, exerted a local control on climatic processes (*see* **Tectonics:** Rift Valleys).

By the Late Triassic, shallow-marine conditions had developed around the site of the future continental margin. The initial marine incursion began in the north, when hypersaline waters transgressed the Grand Banks region, probably from across the Newfoundland–Gibraltar fracture zone. A 2000 m thick section of halite in the Osprey evaporites was deposited above redbeds of the Carson sub-basin. More extensive flooding proceeded southwards, and additional evaporites accumulated along the Nova Scotian margin (in a region now called the slope diapiric province). The youngest salt was deposited in the Carolina trough at the southern end of the future continental margin. The marine flooding marked a time when the sedimentary basins along the future margin became interconnected. Isolated rift basins still existed, however, on the distal edges of the zone of extension (i.e. exposed basins of the Newark Supergroup).

Evaporite deposition ended in the Early Jurassic when normal marine conditions became established. The ocean basin may have been as much as 350 km wide, 900 km long, and less than 1 km deep, although seafloor spreading did not become fully established until the Middle Jurassic. Carbonate platforms developed around the edges of the subsiding seaway, and calcareous mudstones were deposited in the deeper central portions. The developing North Atlantic Ocean was connected to the open ocean north of the Grand Banks (via Greenland and Europe) and to the developing Gulf of Mexico to the south. Uplift and basaltic volcanism interrupted and contributed to the Middle Mesozoic sedimentary record.

Between the Late Triassic and Middle Jurassic, the North American continent drifted northwards by about 10–15°, which created more arid conditions in the basins of the Grand Banks, while maintaining humid conditions in the south. The transition to more grey siliciclastic sequences during this interval indicates more humid palaeoenvironments.

The Middle Jurassic was a time of relative sealevel rise, which flooded the continental margins and started the transition from semi-closed to partially open circulation. Final plate separation occurred during the Middle Jurassic, and the ocean basin grew significantly wider (600 km) and deeper (2000 m, perhaps as much as 3000 m). All the large offshore sedimentary basins had formed by this time (Figure 10) and were accumulating enormous volumes of siliciclastics. This influx of siliceous material may have contributed to extensive radiolarian blooms. By the middle Middle Jurassic, local carbonate tracts were developing along the margins; these prograded rapidly seawards

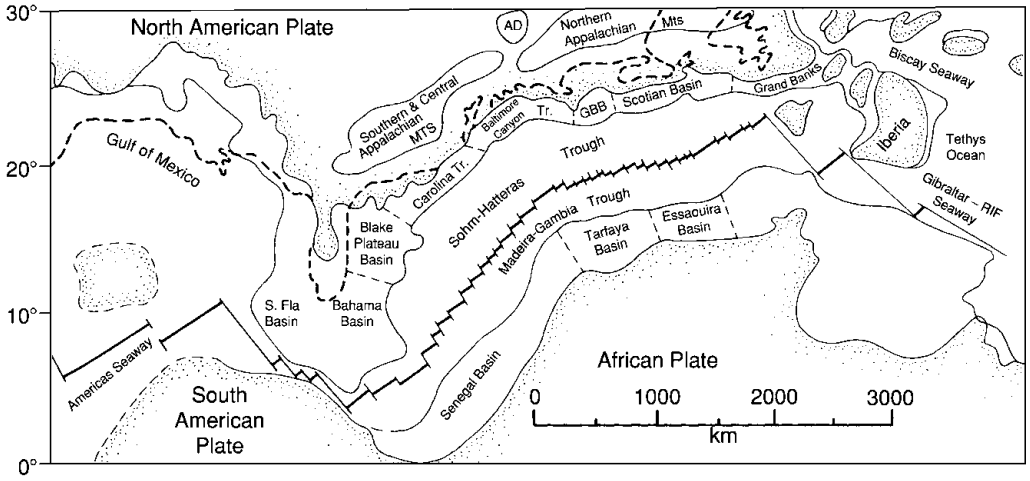


Figure 10 Structural, depositional, and physiographic features of the proto-Atlantic seaway, reconstructed at the end of the Jurassic. Landmass boundaries and palaeolatitudes are approximate. GBB, Georges Bank Basin, AD, Adirondack Mountains. (Reproduced from Poag CW (1991) Rise and demise of the Bahama–Grand Banks gigaplatform, northern margin of the Jurassic proto-Atlantic seaway. *Marine Geology* 102: 63–130.)

in the latest Middle Jurassic and established carbonates as the dominant sediment type.

By the Late Jurassic, a giant carbonate platform extended for more than 5000 km from south of Florida to north of the Grand Banks (Figure 11). The Atlantic Ocean had opened to about 1400 km wide, and maximum seafloor depths reached nearly 4 km. The margin stretched from the equator to about 25° N. A very large carbonate bank formed across the entire Blake Plateau shelf. A more linear semicontinuous carbonate bank–reef system developed in the Baltimore Canyon trough. Farther north, in the Georges Bank and Scotian basins, the reef was more discontinuous. The morphology of the North American Atlantic Continental Margin at this time resembled that of the modern Great Barrier Reef of Australia, in which steep reef fronts mark the edge of the continental shelf and carbonate-rich aprons form on the continental rise seawards of the reef front. Gaps in the reef system channelled siliciclastics into large submarine fan complexes, which carried terrigenous material as far as 300–400 km from the shelf edge. A series of coalescing siliciclastic deltas also deposited terrigenous material shorewards of the carbonate platforms.

Except around the Blake Plateau, the giant carbonate platform ceased to exist in the Early Cretaceous because of rising sea-level, which stressed the carbonate-producing communities, and burial by siliciclastic sediments, which finally filled the back-reef basins and overflowed the bank–reef system. Overall, siliciclastic sedimentation slowed in the major basins, and the margins began to assume shapes similar to their present morphology. Globally, the Cretaceous was a period of poorly oxygenated marine waters,

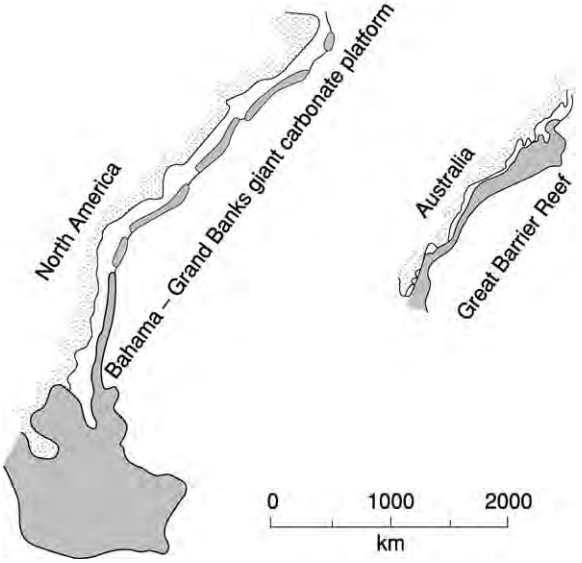


Figure 11 Carbonate tracts (stippled areas) associated with the Late Jurassic Bahama–Grand Banks giant carbonate platform along the North American Atlantic Continental Margin. The modern Great Barrier Reef system of Australia (rotated through 180°) is shown for comparison. (Reproduced from Poag CW (1991) Rise and demise of the Bahama–Grand Banks gigaplatform, northern margin of the Jurassic proto-Atlantic seaway. *Marine Geology* 102: 63–130.)

and studies of the dark carbon-rich Cretaceous deposits of the Hatteras Formation, on the rise and abyssal plain off the coast of North America, have contributed substantially to understanding black-shale sequences. The Cretaceous was also a time when the carbonate compensation depth cycled

between shallow (2000 m) and deep (more than 4000 m).

During the Cenozoic, continental-margin sedimentation was increasingly affected by currents, especially contour-parallel currents (*see* **Sedimentary Environments:** Contourites). Depocentres shifted seawards, causing the continental rise to be built up (Figure 12). The Paleocene and Eocene probably saw the onset of a current-dominated regime (e.g. the ancestral Gulf Stream) and the initiation of small canyons, but much of this early record is thin or missing because of subsequent erosion during the Oligocene. Global temperatures began falling as the Earth completed its transition from a 'greenhouse' in the Cretaceous and Paleocene to an 'icehouse' in the Oligocene. During the Middle Oligocene sea-level lowstand, the Baltimore Canyon shelf retreated by about 30 km. Cold water in the North Atlantic contributed to deep thermohaline circulation, and contour currents became more vigorous in the Early Miocene when cold Norwegian bottom waters overflowed the Norwegian–Greenland ridge and entered the Atlantic, initiating the formation of both the elongate Chesapeake and the Blake Outer Ridge drift deposits. Glaciation of North America in the Late Tertiary increased siliciclastic deposition, intensified down-slope sedimentary processes (e.g. turbidity currents, slumps, slides, and debris flows), and

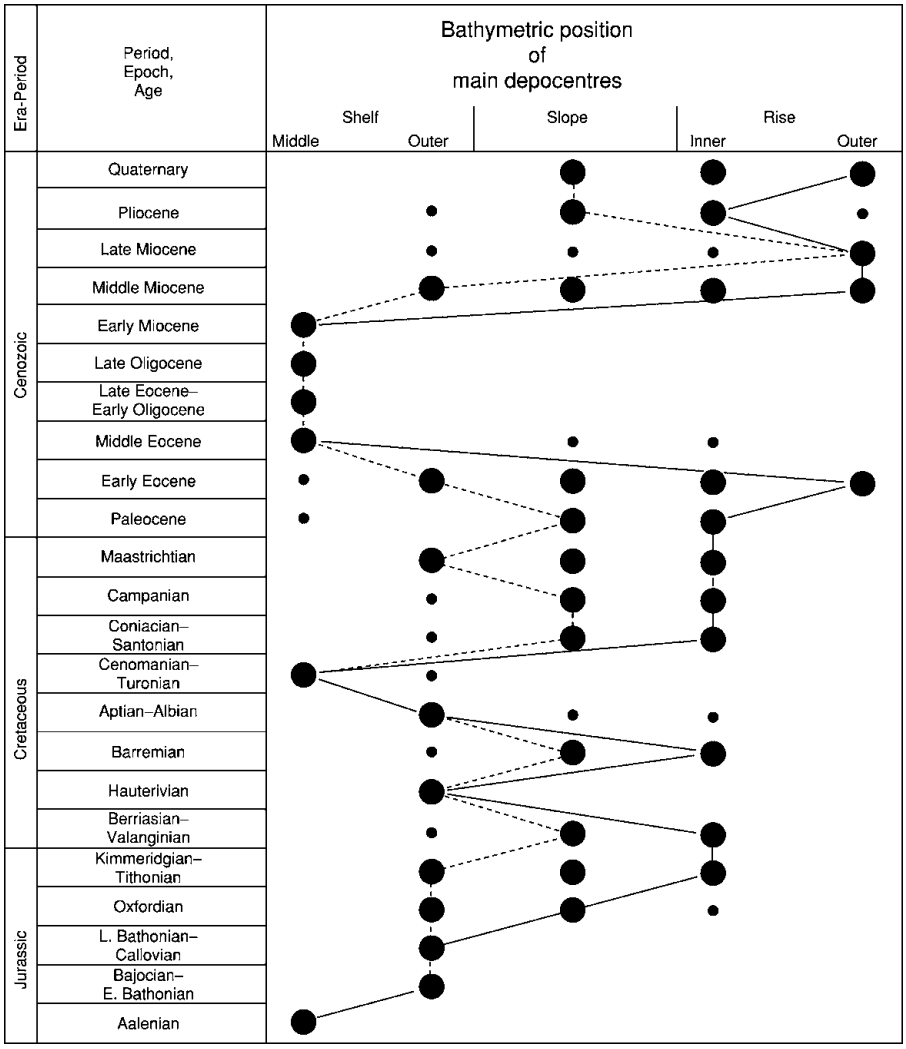


Figure 12 The bathymetric positions of the main depocentres in the Baltimore Canyon trough through the last 187 Ma. Large dots indicate major depocentres (i.e. thickest sediments). Small dots indicate secondary depocentres. Solid line shows the shift in the primary deep-water depocentres through time. Dashed line shows the shift in the primary shallow-water depocentres. This chart shows the migration of the trough seawards and landwards during the history of the margin. Note the marked period of limited shelf and slope deposition in the Late Miocene. (Reproduced from Poag CW and Sevon WD (1989) A record of Appalachian denudation in postrift Mesozoic and Cenozoic sedimentary deposits of the US Middle Atlantic Continental Margin. *Geomorphology* 2: 119–157.)

shifted sedimentary depocentres to deeper water on the rise and abyssal plain. Many new canyons also developed.

Energy, Mineral, and Water Resources

The knowledge base about energy resources on the North American continental margin varies by country: drilling off the US East Coast in the late 1970s and early 1980s produced disappointing shows for both oil and gas in the Baltimore Canyon trough and Georges Bank basin. Neither the Carolina trough nor the Blake Plateau basin have been drilled or tested for petroleum maturation, although models suggest that maturation in the Carolina trough should have reached the oil window only in its deepest Early and Middle Jurassic rocks. The petroleum potential in the rift basins buried beneath the margin remains unknown. Currently the USA has a moratorium on leasing and drilling on its offshore Atlantic margin.

In contrast, Canada has had an active exploration and production programme on the Scotian Shelf and Grand Banks since the late 1970s. The Mesozoic Jeanne d'Arc rift basin contains the most significant discoveries on the Grand Banks (Hibernia, Terra Nova, Hebron-Ben Nevis, and White Rose fields). Both oil and gas have also been produced from Upper Jurassic and Lower Cretaceous sandstones on the Scotian Shelf in the vicinity of Sable Island. In the 25 years since exploration began, a robust offshore technology and infrastructure has grown to support the Canadian petroleum industry.

One of the best-known deposits of an unconventional hydrocarbon, gas hydrate, exists on the Blake Ridge off the coast of the south-eastern USA (Figure 13). Gas hydrate is an ice-like solid that holds high concentrations of methane in a crystalline water matrix, which forms under pressures, temperatures, and geochemical conditions that are often found in sediments near the seafloor deeper than about 500 m water depth (*see* **Petroleum Geology:** Gas Hydrates). The first dedicated gas-hydrate drilling leg of the Ocean Drilling Program occurred on the Blake Ridge in 1995 (Leg 164). Estimates of the total volume of methane in both hydrate and associated free-gas deposits on the Blake Ridge vary from 60 to 100 trillion cubic metres.

Occurrences of offshore minerals are known along the North American continental margin, but they have not been as extensively developed or exploited as their onshore equivalents, primarily because recovery and transportation are more expensive in an offshore environment. Phosphates, used in agricultural fertilizers, are successfully mined in coastal-plain (i.e. postrift) deposits in Florida and North Carolina, but phosphate deposits on the continental shelf between Georgia and North Carolina are largely unexploited. Manganese nodules and ferromanganese crusts, potential sources of manganese, nickel, cobalt, copper, and platinum, have been sampled on the Newfoundland and New England Seamounts, but their distributions are largely unmapped. More detailed studies of similar nodules and crusts have been carried out on samples from the Blake Plateau.

Placer deposits also exist offshore, the most famous being gold-bearing sands and alluvium along the Nova Scotian inner shelf, which are presumed to have a source in pre-Mesozoic basement rocks. Mining for these intriguing finds has been largely uneconomical. Placer deposits containing titanium and iron have been mapped off Florida and are inferred to exist on other parts of the margin. The most plentiful mineral deposits are quartz sands, which are dredged mainly from the inner shelf for beach nourishment (for example, off the coasts of most of the states between New Jersey and Florida). Other resources that have been identified but not developed offshore include high-purity silica sand in the Gulf of St Lawrence and on the Grand Banks (which could be used for glass making), calcium carbonate sand (which is a key ingredient of Portland cement), clay deposits (which supply material for bricks, sewer pipes, and other construction materials), and coarse sand and gravel (which could be used in road building and maintenance). Quaternary glaciation (*see* **Sedimentary Processes:** Glaciers) and modern depositional processes have controlled much of the distribution of unconsolidated sands and gravels on the margin from New England to Newfoundland.

One offshore resource that has received little attention is fresh groundwater within sediments of the continental margin. Freshwater springs are known along the Florida shelf. One of the most spectacular is the Crescent Beach spring, about 3 km off the coast of north-east Florida, in about 20 m water depth. This spring produces a boil with a hydrogen sulphide odour on the sea surface during times of high discharge. Within the fractured and porous limestone (*see* **Sedimentary Rocks:** Limestones) aquifer system of the south-eastern USA, this offshore fresh groundwater is derived from rainfall on land, and its pressure is considered to be in equilibrium with sea-level. In contrast, shallow wells drilled on the continental shelf off New Jersey show a thin wedge of almost-fresh water that extends more than 100 km across the margin at relatively shallow depths (less than 100 m beneath the seafloor, Figure 14). This fossil freshwater accumulated during the Pleistocene falls in sea-level and has been protected from saltwater contamination

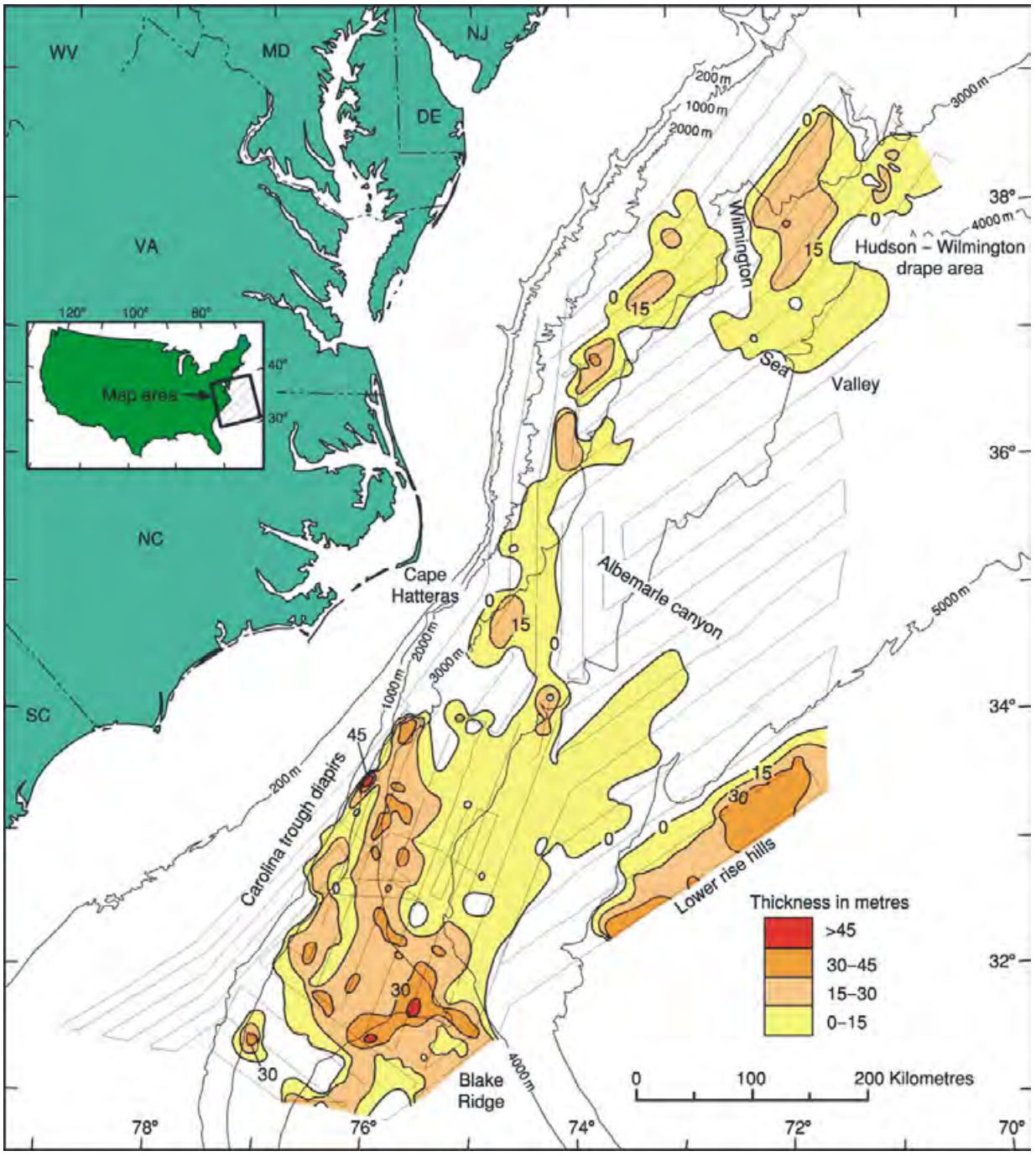


Figure 13 Distribution of gas hydrate inferred from seismic-reflection data on the US Atlantic margin. Contours show the volume of pure gas hydrate in terms of thickness in metres. Volumes were estimated by mapping the extent of blanking in seismic profiles and then using a model to transform blanking to proportion of gas hydrate in the sediment. (Reproduced from Dillon WP, Fehlhaber K, Coleman DF, Lee MW, and Hutchinson DR (1995) *Maps Showing Gas-Hydrate Distribution off the East Coast of the United States*. Miscellaneous Field Studies Map MF-2268, 1 : 1 000 000, 2 plates. US Geological Survey.)

by a low-permeability confining clay unit, which has prevented it from reaching equilibrium with the present saltwater hydrologic system. Freshwater is unknown in the continental-margin sediments off the coast of Canada.

Current and Future Societal Issues

As population growth has intensified in coastal regions, interactions between humans and nature have created an urgent need to understand surficial

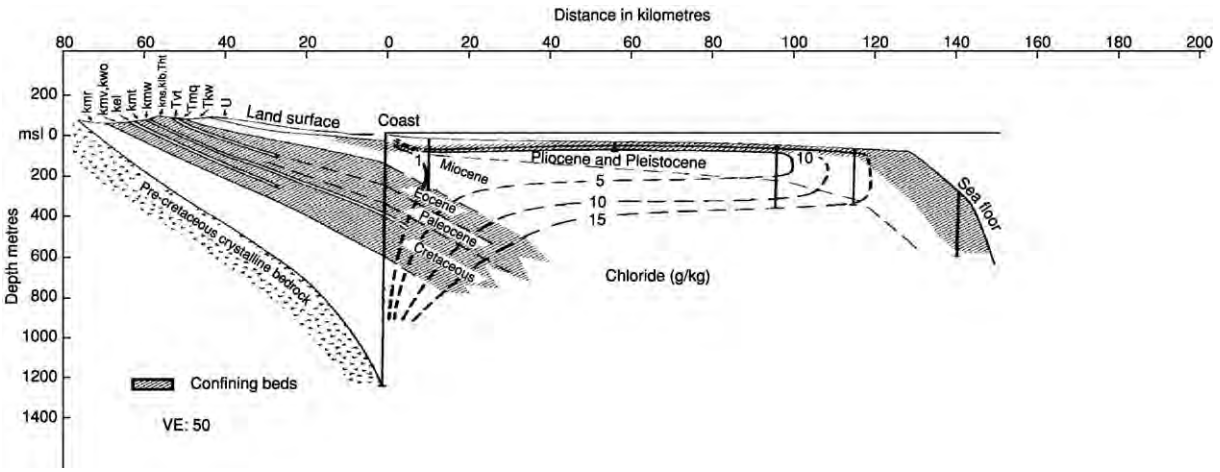


Figure 14 Schematic hydrological section through postrift sediments of the central Baltimore Canyon trough. Dashed contours show the chlorinity of the pore waters. The chlorinity of typical ocean water is about 19‰. (Modified from Kohout FA, Meisler H, Meyer FW, *et al.* (1988) Hydrogeology of the Atlantic continental margin. In: Sheridan RE and Grow JA (eds.) *The Geology of North America, Volume I-2: The Atlantic Continental Margin: U.S.*, pp. 463–480. Boulder, Colorado: Geological Society of America.)

geological processes that affect, for example, coastal erosion, sediment contamination, and habitat occurrence. Erosion from large storms results in loss of and damage to property and sometimes life. Beaches of the USA are among the most heavily artificially replenished of any country in the world: sand is often mined from the inner shelf. Three-dimensional bathymetric and geological mapping offshore has re-emphasized the importance of the geological framework of the shallow continental margin in understanding sediment transport, defining fragile habitats, and predicting the ultimate fate of industrial waste products in the oceans. Mapping techniques that integrate bathymetry, shallow subsurface profiles, video, and sampling have helped to determine which fishing grounds to open on Georges Bank and to manage other offshore fisheries. Many of these societal issues will increase in importance as population growth continues in coastal regions.

New technologies and high-speed computers are yielding new ways to view the continental margin and revealing new opportunities for fundamental scientific research. High-resolution bathymetric maps of the Hudson Shelf Valley reveal major bedforms and debris-flow deposits on the outer shelf, which appear to have been formed by a previously unrecognized enormous flood event. Sampling, dating, and interpretation of these deposits have not yet been achieved but are likely to modify interpretations of the late glacial and postglacial Quaternary history of the northern Atlantic region. Likewise, three-dimensional mapping of gas hydrates continues to challenge researchers to understand the importance of geology in controlling the occurrence of these highly dynamic potential resource reservoirs.

Recent ratification of the United Nations Convention on the Law of the Sea by Canada and the imminent ratification expected by the USA will create the need to identify geodetically accurate positions of the foot of the continental slope and to map accurately the sediment thickness in deep water. Such data can be used by nations to extend the legal boundary of their continental margin. These circumstances provide opportunities in which fundamental science can be integrated with practical needs.

Conclusions

The North American Atlantic Continental Margin is the type example of an Atlantic passive continental margin. For more than 200 million years, this margin has evolved from its initial equatorial rift configuration, accompanied by shallow oceanic flooding, to its present mature mid-latitude position bounding a deep ocean. The margin contains a rich record of diverse geological and tectonic processes and features, which include evaporite basins, extensive carbonate banks and reefs, and the modern sand-dominated partly glaciated sediment system. This margin provides sources of energy and minerals for human use and may hold a potential reserve of freshwater. The twentieth century was a time of great exploration, discovery, and characterization of the margin. At the start of the twenty-first century, new challenges confront researchers, in particular the need to understand how to balance human demands (e.g. for resources, stable coastlines, and waste disposal) with natural

processes (e.g. coastal erosion, habitat maintenance, and sediment and contaminant transport). Much of the continental margin remains an unexplored frontier area with regard to these new challenges.

See Also

Earth: Crust. **Impact Structures**. **Minerals:** Carbonates. **Pangaea**. **Petroleum Geology:** Gas Hydrates. **Plate Tectonics**. **Seamounts**. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Contourites. **Sedimentary Processes:** Glaciers; Landslides; Particle-Driven Subaqueous Gravity Processes. **Sedimentary Rocks:** Evaporites; Limestones. **Tectonics:** Rift Valleys.

Further Reading

Dennison JM and Ettensohn FR (1994) *Tectonic and Eustatic Controls on Sedimentary Cycles*. Concepts in Sedimentology and Paleontology volume 4. Society of Economic Paleontologists and Mineralogists. Tulsa: Oklahoma.

Ford D and Golonka J (2003) Phanerozoic paleogeography, paleoenvironment, and lithofacies maps of the circum-Atlantic margins. *Marine and Petroleum Geology* 20: 249–285.

Gardner JV, Field ME, and Twichell DC (eds.) (1996) *Geology of the United States' Seafloor: The View from GLORIA*. New York: Cambridge University Press.

Keen MJ and Williams GL (eds.) (1990) *The Geology of North America, Volume I-1: Geology of the Continental Margin of Eastern Canada*. Geological Society of America. Ottawa, Ontario, Canada.

Manspeizer W (ed.) (1988) *Triassic–Jurassic Rifting: Continental Breakup and the Origin of the Atlantic Ocean and Passive Margins*. New York: Elsevier.

Poag CW (ed.) (1985) *Geologic Evolution of the United States Atlantic Margin*. New York: Van Nostrand Reinhold.

Poag CW (1991) Rise and demise of the Bahama–Grand Banks gigaplatform, northern margin of the Jurassic proto-Atlantic seaway. *Marine Geology* 102: 63–130.

Poag CW and de Graciansky PC (eds.) (1992) *Geologic Evolution of Atlantic Continental Rises*. New York: Van Nostrand Reinhold.

Scrutton RA, Stoker MS, Shimmield GB, and Tudhope AW (eds.) (1995) *The Tectonics, Sedimentation, and Palaeoceanography of the North Atlantic Region*. Special Publication 90. London: Geological Society.

Sheridan RE and Grow JA (eds.) (1988) *The Geology of North America, Volume I-2: The Atlantic Continental Margin: US*. Boulder, Colorado: Geological Society of America.

Tankard AJ and Balkwill HR (eds.) (1989) *Extensional Tectonics and Stratigraphy of the North Atlantic Margins*. Memoir 46. Tulsa: American Association of Petroleum Geologists.

Vogt PR and Tucholke BE (eds.) (1986) *The Geology of North America, Volume M: The Western North Atlantic Region*. Boulder, Colorado: Geological Society of America.

Withjack MO, Schlische RW, and Olsen PO (1998) Diachronous rifting, drifting, and inversion on the passive margin of Central Eastern North America: an analog for other passive margins. *American Association of Petroleum Geologists Bulletin* 82: 817–835.

OCEANIA (INCLUDING FIJI, PNG AND SOLOMONS)

H Davies, University of Papua New Guinea, Port Moresby, Papua New Guinea
P Bani, Institut de la Recherche pour le Développement, Nouméa, New Caledonia
P Black and I Smith, Auckland University, Auckland, New Zealand
E Garaebiti, Department of Geology and Mines, Port Vila, Vanuatu
P Rodda, Mineral Resources Department, Suva, Fiji

© 2005, Published by Elsevier Ltd.

Introduction

The islands of the south-west Pacific extend for 5000 km from New Guinea in the west to Fiji, Tonga, and Samoa in the east, and include the Solomon Islands, Vanuatu, and New Caledonia (Figure 1). The early European explorers called the area Melanesia because of the generally dark skin colour of the people. Subsequently, the parallel island chains have sometimes been referred to as the inner and outer Melanesian arcs. New Guinea was colonized from the west more than 40 000 years ago; New Britain, at least 35 000 years ago; New Ireland, at least 20 000 years ago; and the northern Solomon Islands, 28 000 years ago. The other Melanesian islands, New Caledonia, Vanuatu, and Fiji, were settled only 3000 years ago.

The islands have many similarities in their geology, in that, for the most part, they have been constructed on oceanic crust by one or more cycles of volcanic activity. Exceptions are the island of New Guinea, which is part of the Australian continent; the older rocks of New Caledonia, which first formed at the margin of the Australian continent; and the outermost Solomon Islands, which are elevated oceanic crust. There are deep-sea trenches on either side of the island chains. On the outer (north-eastern) side are the Manus, Kilinailau, Solomons, and Vitiaz trenches, which are only weakly active or are inactive, and the Tonga–Kermadec Trench east of Tonga (Figure 1), which is active. On the inner (south-western) side is an active system of trenches, the New Britain–Makira–Vanuatu–Hunter–Kadavu trenches. Other seafloor features include the massive Ontong Java submarine plateau of thickened oceanic crust (Figure 1), smaller oceanic plateaus of rifted continental lithosphere in the Coral Sea, and a number of small ocean basins, some of which are actively spreading.

The islands occupy the boundary zone between the north-moving Australian plate and the west–north-west-moving Pacific Plate, and for this reason are tectonically and volcanically active (Figure 2). The plate boundary is not a simple one but, rather comprises a number of smaller plates that exhibit a variety of plate interactions, from seafloor spreading to subduction, transform faulting, and rotation. The lines of earthquakes on the map (Figure 2) coincide with the plate boundaries, and give an idea of the high level of seismic activity in the region. The plot of depth-coded hypocentres shows the trace of a subducted slab dipping westward from the Tonga–Kermadec Trench, in the east, and of a subducted slab dipping northward from the New Britain Trench, in the north-west. The area has evolved since the Cretaceous by repeated episodes of crustal extension and crustal shortening. Extension produced small ocean basins, and shortening resulted in subduction and thrust faulting. The volcanic islands have grown through this time, initially (in the Eocene to Early Miocene) by volcanism related to subduction of the Pacific Plate on the outer trench system, and more recently (the past 10 million years) by volcanism related to subduction of the Australian Plate on the inner trench system.

Papua New Guinea

Papua New Guinea (PNG) comprises the eastern half of the island of New Guinea, the adjacent islands of the Bismarck Archipelago, and the northernmost of the Solomon Islands (Bougainville and Buka). The geology of PNG can be described in terms of four geological provinces (Figure 3): a stable platform in the south-west, where the Australian craton extends beneath the island of New Guinea; a collisional zone in the centre and north of the island, where terranes have been sutured to the Australian craton; a province built by volcanic arc activity in the north-east; and the Ontong Java submarine plateau.

Platform and Foldbelt

The stable platform comprises a continental basement of Palaeozoic and partly Precambrian metamorphic and igneous rocks overlain by a 4-km-thick paraconformable sequence of shelf facies Mesozoic siliciclastic, and Cenozoic carbonate and siliciclastic, sedimentary rocks (Figure 3). Rocks along the northern margin of the platform are buckled and faulted to form a basement-involved foreland fold-and-thrust belt. Northward of the fold-and-thrust belt, the Mesozoic sediments are transitional into a kilometres-thick sequence of black mudstones. These

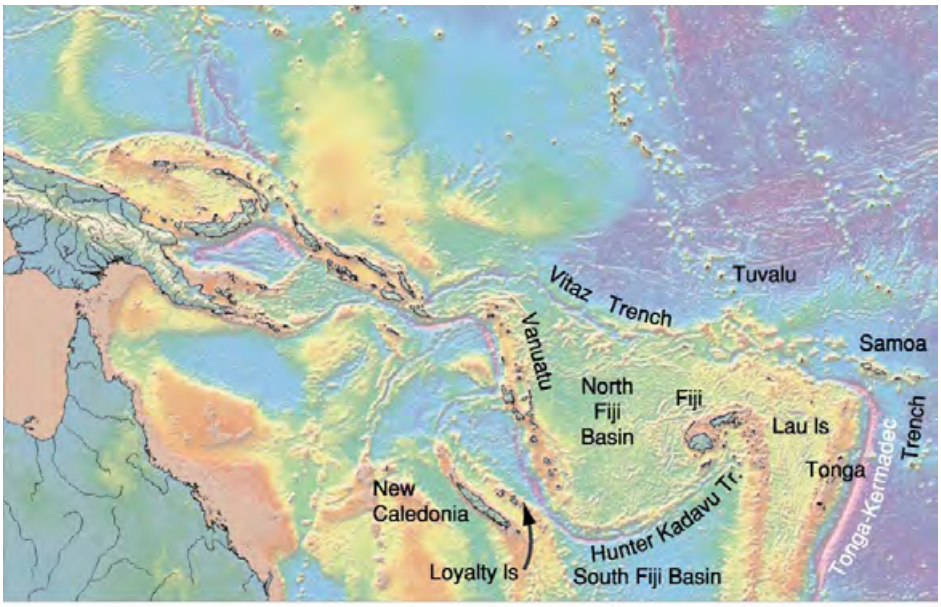


Figure 1 Physiographic map of the south-west Pacific. Reproduced with permission from Smith WHF and Sandwell DT (1997) Global seafloor topography from satellite altimetry and ship depth soundings. *Science* 277: 1957–1962. Available on the Internet at http://topex.ucsd.edu/marine_topo/mar_topo.html.

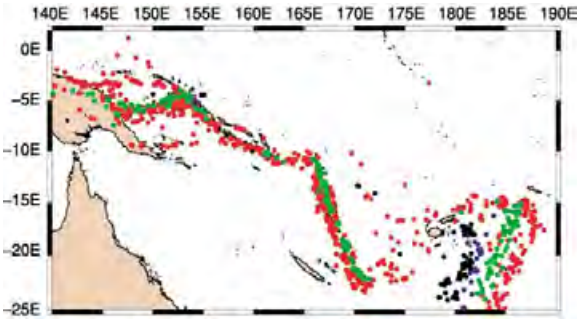


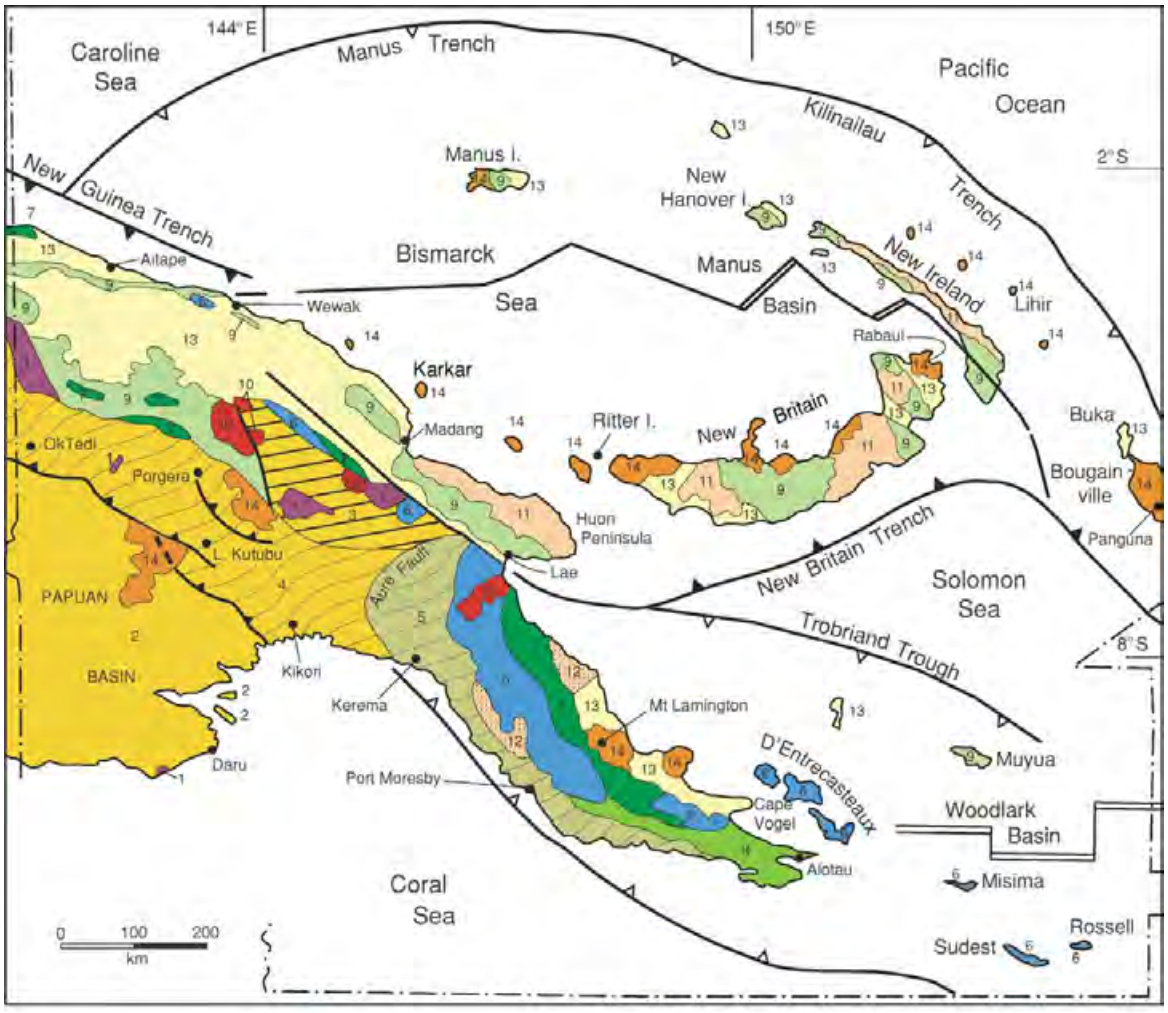
Figure 2 The map shows earthquakes with magnitude 6 or greater, recorded during 1963–2002. Red indicates hypocentre depth of 0–70 km, green 70–300 km, blue 300–500 km and black >500 km. Features of the map are the west-dipping seismic zone beneath Tonga in the east, the north-dipping seismic zone beneath New Britain in the north-west, and the steeply dipping seismic zone beneath Vanuatu. Shallow seismic zones (red dots) mark the spreading ridges and transforms in the Bismarck and Solomon seas and the North Fiji Basin. Map prepared by Emile Okal.

in turn are transitional northwards into greenschist facies graphitic schists. The sediments of the stable platform and foldbelt comprise the Papuan Basin.

The foldbelt changes character east of the Aure Fault (Figure 3). Beyond this point, the foldbelt comprises fault-bounded segments of Late Cretaceous, Paleocene, and Eocene fine siliceous siliciclastic and calcareous sedimentary rocks, associated with faulted slivers of ultramafic rock; Oligocene and younger coarser, partly volcanogenic, dominantly clastic sediments; and locally extensive Oligocene gabbro. There is no indication of a continental or cratonic basement, though such may exist. The Paleocene and much of the Eocene sediments were deposited in a deep marine environment and probably have been added to the mainland by accretion above a north-dipping subduction system.

Collisional Zone

The western part of the central collisional zone comprises fault-bounded terranes of Jurassic to Eocene sedimentary, igneous, metamorphic, and ultramafic rock, and younger (Eocene to Miocene) dioritic intrusive bodies. The association of ultramafic rocks with eclogite, blueschists, submarine volcanic rocks, and volcanogenic sediments indicates a former volcanic arc environment. Sediments associated with the volcanic rocks are of two ages, mid-Eocene and Oligocene. This suggests a history of successive development of volcanic arcs and successive collisions; the Oligocene volcanic rocks are found only on the north side of the northern ranges. Also included in the western part of the central collisional zone are two fault-bounded blocks of Palaeozoic basement. One of these lies on the border with Indonesian Papua and the other extends northward from the Kubor Range, to underlie the Jimi valley (Figure 3). These Palaeozoic terranes lie outward (northward) of younger accreted terranes and thus appear to be 'out of sequence'.



Key

14	Quaternary volcanics
13	Miocene, Pliocene, and Quaternary sediments
12	Miocene volcanics and sediments (mainland)
11	Miocene limestone (islands and Huon Peninsula)
10	Neogene dioritic intrusions
9	Paleogene island-arc-type volcanics
8	Cretaceous and mid-Eocene ocean floor basalts
7	Ophiolite
6	Mesozoic and Paleocene metamorphic rocks
5	Eastern foldbelt (accretionary prism?)
4	Papuan foldbelt
3	Jimi-Kubor: Mesozoic sediments on Palaeozoic basement
2	Papuan Basin Mesozoic and Cainozoic sediments
1	Palaeozoic basement

Figure 3 Geological map of Papua New Guinea, showing active (solid triangles) and inactive or slow-acting trenches (open triangles), transforms, and spreading ridges; major thrust faults on land (solid triangles) are also indicated. There are four geologic provinces: Platform in the south-west, collisional in the centre, volcanic arcs in the north-east, and Ontong Java Plateau. Map drawn by L Cotterall, Geology Department, Auckland University.

Possibly they are indications that the Palaeozoic craton extends northward as a substrate beneath the terranes of the collisional zone. Mesozoic amphibolite facies metamorphics near the north coast (Prince Alexander Range) and Oligocene metamorphics associated with ultramafic rocks just beyond the border on the north coast (near Jayapura) also are 'out of sequence' and may represent material ramped up from a subjacent older stratum.

The geology of the eastern part of the collisional zone, from Lae south-eastward, is much simpler and essentially comprises two rock units: an inner core of metamorphic rocks, flanked on the north-east side by a major ophiolite complex. The protolith of the metamorphics was felsic and psammitic sediments and some submarine basalts, partly of Cretaceous age, and these were metamorphosed in the Paleocene. The ophiolite comprises a 4- to 8-km thickness of tectonite harzburgite, overlain by lenses of cumulate-textured ultramafic rocks, which are in turn overlain by 4 km of gabbro and 4 km of basalt. Paleocene and Eocene volcanic rocks that overlie the ophiolite may be arc related. Ocean floor basalts that form the eastern end of the island are Late Cretaceous and Middle Eocene and may be related to the ophiolite. Miocene and younger, partly volcanogenic, dominantly siliciclastic sediments of the North New Guinea Basin unconformably overlie basement in the western part of the collision zone, and Miocene and younger volcanics and sediments of the Cape Vogel Basin unconformably overlie basement in the eastern part of the collision zone.

North-Eastern Province

The north-eastern province, constructed by volcanic arc activity, embraces the islands of the Bismarck Archipelago and the mainland mountain ranges of the Huon Peninsula. The oldest rocks are Eocene and Oligocene volcanics with some dioritic intrusions. These are partly overlain by Miocene limestone, which is in turn partly overlain by volcanic rocks of the present-day (Plio-Quaternary) volcanic cycle. Bougainville, in the Solomon Islands chain (see later), has similar geology. The land areas are separated by several small ocean basins: the Bismarck Sea, Solomon Sea, and Coral Sea (Figure 3). There are active spreading ridges and transforms in the eastern Bismarck Sea (Manus Basin) and southern Solomon Sea (Woodlark Basin) and active trenches on the inner side of New Britain and the Solomon Islands, on the outer side of the Solomon Islands and extending westward to Manus, and off the New Guinea coast west of Wewak. A trench on the south side of the Solomon Sea, at the Trobriand Trough, is masked by thick sediment. Volcanic activity is associated with the New Britain Trench but not with the trench west of Wewak (New Guinea Trench). The Coral Sea opened by seafloor spreading in the Paleocene (65–55 Ma), the Solomon Sea in the Oligocene (40–30 Ma), and the eastern Bismarck Sea and Woodlark Basin in the Plio-Quaternary (since 6 Ma).

Economic Minerals

The main mineral deposits of Papua New Guinea are porphyry copper–gold deposits, disseminated and vein gold deposits, and alluvial gold. Other less common deposits include lateritic nickel–cobalt ore over ultramafic rocks, minor manganese ore, and the small copper–gold massive sulphide orebodies that have been mined near Port Moresby. The Panguna mine (now closed) on Bougainville Island and the Ok Tedi mine in Western Province are classic porphyry copper–gold deposits. In both locations, copper and gold occur in Plio-Quaternary dioritic to granitic rocks and in the adjacent country rocks. Similar mineralization is known at Frieda River (West Sepik), Yandera (north-west of Goroka), and Wafi, 50 km south-west of Lae.

Disseminated and vein gold deposits are mined at Porgera; Ladolam on Lihir Island; Tolukuma, north of Port Moresby; and, until recently, Misima Island. Production is planned from Irumafimpa near Kainantu and from Hidden Valley, near Wau. Production from these mines, together with Ok Tedi, is sufficient to place Papua New Guinea in the top five of gold-producing nations, worldwide. The Panguna mine also was a major producer of gold. At Porgera, gold mineralization occurs in intrusive dioritic rocks and in the adjacent, baked sedimentary rocks. At Ladolam, the gold is in the floor of the caldera of an eroded volcano, with hot springs nearby. At Tolukuma, Irumafimpa, and Hidden Valley, the gold is in quartz veins. Alluvial gold is mined on a small scale in many parts of Papua New Guinea, notably near Amanab and Maprik in the Sepik, in the Western and Eastern Highlands, around Wau, and in Milne Bay Province.

From 1930 to 1965, with a break for the Second World War, alluvial gold was mined by dredges in the Bulolo valley. A total of about 65 tonnes of gold was produced. Within the last few years, a rich alluvial field was found at Mount Kare, south-west of Porgera. Where earlier gold rushes had been almost exclusively managed by overseas miners, Mount Kare was the first entirely Papua New Guinean gold rush. As many as 8000 miners were on the field at any one time and an estimated 15 tonnes of gold was taken out in 2 years. The only other fields to have produced more than a tonne of gold are the Lakekamu and Sepik alluvial fields, and Misima and Muyua (Woodlark), both of which were mixed alluvial and lode mining. Lateritic soils over ultramafic rocks south of

Madang contain economic concentrations of nickel and cobalt and await development.

Energy Resources

Oil currently is produced from several thrust-bounded anticlines containing the Kutubu, Gobe, and Moran oil fields, all of which are in the vicinity of Lake Kutubu in the Southern Highlands Province. Production from the South-east Mananda oil field within the giant Mananda structure between the Kutubu oil fields and the Hides gas field is planned. The oil was derived from Jurassic black shales and has accumulated in clean, shallow marine quartz sandstones of Late Jurassic and Early Cretaceous age. At the time of the initial discovery, in the Iagifu anticline, recoverable reserves were estimated at around 200 million barrels of oil. By the beginning of 2004, 354 million barrels had been produced from the various fields and remaining recoverable oil reserves stood at 222 million barrels. Production, which began in June 1992, has declined from a peak of almost 150 000 barrels day^{-1} in 1993 to around 50 000 barrels day^{-1} in 2004 as the fields have been depleted.

A small part of the large volume of gas in the Hides anticline, south-west of Tari, provides electrical power to the Porgera mine. The gas at Hides and in other structures in the foldbelt is sufficient to be developed commercially, provided that a market can be found. A pipeline to Queensland, Australia is planned, but the project awaits confirmed markets before detailed engineering and design work and construction can proceed. Large volumes of gas also have been discovered offshore in the Gulf of Papua at Pandora on the north-western corner of the Eastern Plateau, about 150 km south-west of Kerema. This gas occurs in a deeply buried cavernous Miocene limestone reef. Oil and gas seeps are known in the North New Guinea Basin and there are indications of petroleum in the Cape Vogel Basin. A pilot program to utilize geothermal energy at the Ladolam Mine on Lihir Island has proved successful and is being expanded. This is the first use of geothermal energy in PNG.

Natural Hazards

Papua New Guinea has 14 active and 22 dormant volcanoes and these pose a potential danger to 204 000 people. Ten of the fourteen active volcanoes and 12 of the 22 dormant volcanoes are in the Bismarck volcanic arc, from Bam, Manam, Karkar, and Long Island in the west to Langila, Garbuna, Makalia, Pago, Ulawun, and Rabaul in the east. Other groupings of active volcanoes are in the Admiralty Islands, on Bougainville Island, and in eastern Papua. The agency charged with monitoring the PNG volcanoes is the Rabaul Volcanological Observatory.

Rabaul volcano erupted in 1937, killing more than 500 people. When it erupted again in 1994, only five people were killed, but much of Rabaul town was devastated, leading to a decision to relocate government and most business activity to Kokopo. The worst volcanic disaster in historical time was the explosive eruption of Lamington volcano in Oro Province in 1951; 3000 people were killed. A major tsunami generated by the collapse of the west flank of Ritter Island volcano in 1888 caused untold loss of life on New Britain and adjacent islands. More recently, in 1998, the Aitape tsunami destroyed two villages, damaged other villages, and caused more than 1650 deaths.

Solomon Islands

The central and western Solomon Islands form a double chain on either side of a central trough and are bounded on both flanks by deep-sea trenches (Figure 4). The islands comprise three geological provinces: a north-eastern province of Cretaceous–Paleocene oceanic basalts and pelagic sediments, a central province of more complex geology that includes metamorphic and ultramafic rocks and Oligocene–Miocene arc-related volcanics, and a south-western province of Late Miocene to Quaternary volcanic activity. The islands of the north-eastern province are Malaita, the small island of Ulawa, and the north-eastern side of Santa Isabel. On these islands, Cretaceous and Paleocene submarine basalts with intercalated pelagic limestone and mudstone are overlain by younger Cenozoic marine sediments that contain terrigenous material. The Cretaceous and Paleocene rocks originated as part of the Ontong Java submarine volcanic plateau. On Malaita, an intrusion of alnoite, a silica-poor mafic rock, dating to 34 Ma, has attracted interest in the past because of the occurrence of unusual xenoliths, including garnet peridotite.

The islands of the central province are Choiseul, the south-western side of Santa Isabel, the Florida islands, Guadalcanal, and Makira (San Cristobal). Basement on these islands comprises Cretaceous basalt, greenschist and amphibolite facies, mafic schists, and ultramafic rocks; the metamorphic rocks have radiometric ages around 50 Ma, Early to mid-Eocene. Volcanic rocks of Late Eocene(?), Oligocene, and Early Miocene age unconformably overlie the basement rocks, and volcano-related dioritic stocks intrude basement. A thick sequence of Miocene to Pliocene clastic sediments covers much of Guadalcanal Island but is not strongly developed on the other islands.

The south-west province comprises Late Miocene to Holocene arc volcanic rocks and associated

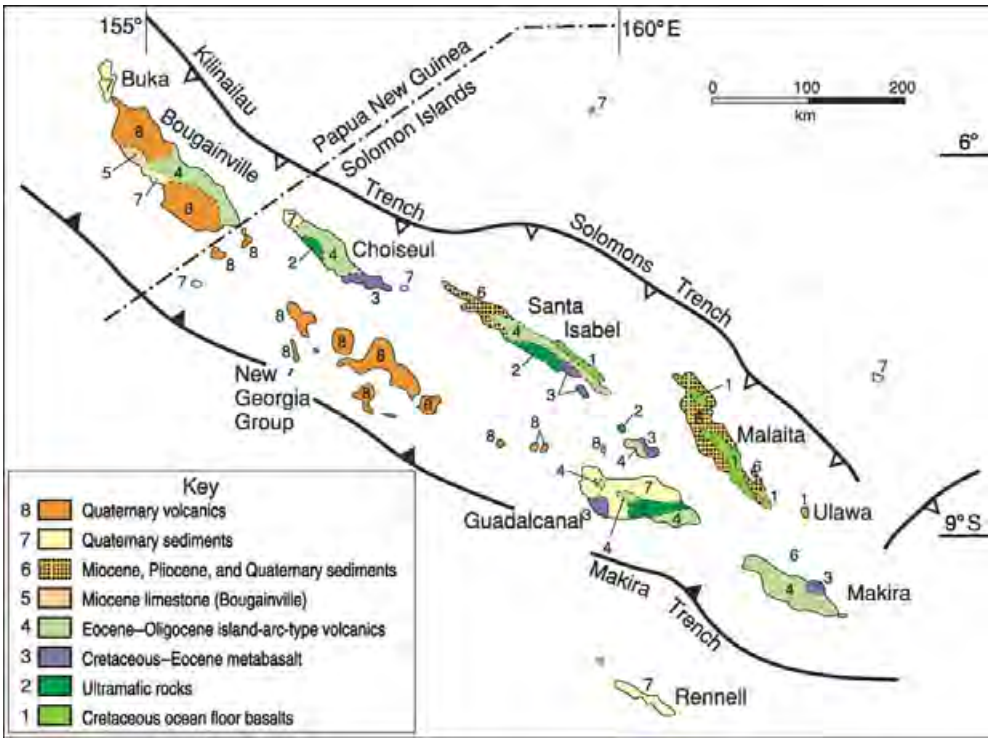


Figure 4 Geological map of the Solomon Islands, including Bougainville and Buka Islands (Papua New Guinea), showing the Makira (San Cristobal) active trench, the Kilinailau inactive trench, and the North Solomons slow-acting trench. The dashed line marks the international border. Map drawn by L Cotterall, Geology Department, Auckland University.

sediments and intrusive rocks. These extend from Bougainville (PNG) and the Shortland Islands in the north-west, to the near end of Guadalcanal in the south-east. There is an outlier of Pliocene volcanics and intrusive rocks in central Guadalcanal, at Gold Ridge. Bougainville Island, politically a part of Papua New Guinea, is the north-westernmost of the Solomon Island chain and combines characteristics of both the central and south-western geological provinces. A basement of Eocene and Oligocene arc-related volcanic rocks is partly overlain by Miocene limestone and Plio-Quaternary volcanoes and is intruded by Pliocene dioritic stocks. The Solomon outer eastern islands are far removed from the central and western islands and are shown on the Vanuatu geological map (Figure 5). The islands of the Santa Cruz Group comprise Miocene to Holocene volcanic rocks and sediments. The Duff Islands, and the islands of Anuta, Fatutaka, and Tikopia (not shown on Figure 5) comprise Quaternary volcanic rocks and sediments. Tinakula in the Santa Cruz Group is the only active volcano.

Geological Evolution

The islands of the north-east province developed as a result of continued convergence across the line of the Kilinailau Trench (Figure 4), and resulting delamination and eastward overthrusting of an upper part of the Ontong Java volcanic plateau. The rocks of the lower part of the Ontong Java plateau continue to be subducted steeply beneath the Solomon Islands. The central province association of ultramafic rocks, Cretaceous basalt, and Eocene metamorphic rocks is thought to have developed as a result of subduction of Pacific plate at the Kilinailau Trench in the Eocene, but the apparently rapid exhumation of these rocks (exhumed before the Oligocene) is not readily explained. The same cycle of subduction yielded the Oligocene to Early Miocene arc-type volcanic rocks that overlie basement on the islands of the central province. The Late Miocene to Holocene volcanic rocks of the south-west province were generated by subduction of the Australian plate at the Makira (San Cristobal) Trench.

Economic Geology

Gold associated with Pliocene intrusives has been mined at Gold Ridge on Guadalcanal Island, and soils on ultramafic rocks on Santa Isabel Island have been tested for lateritic concentrations of nickel and cobalt.

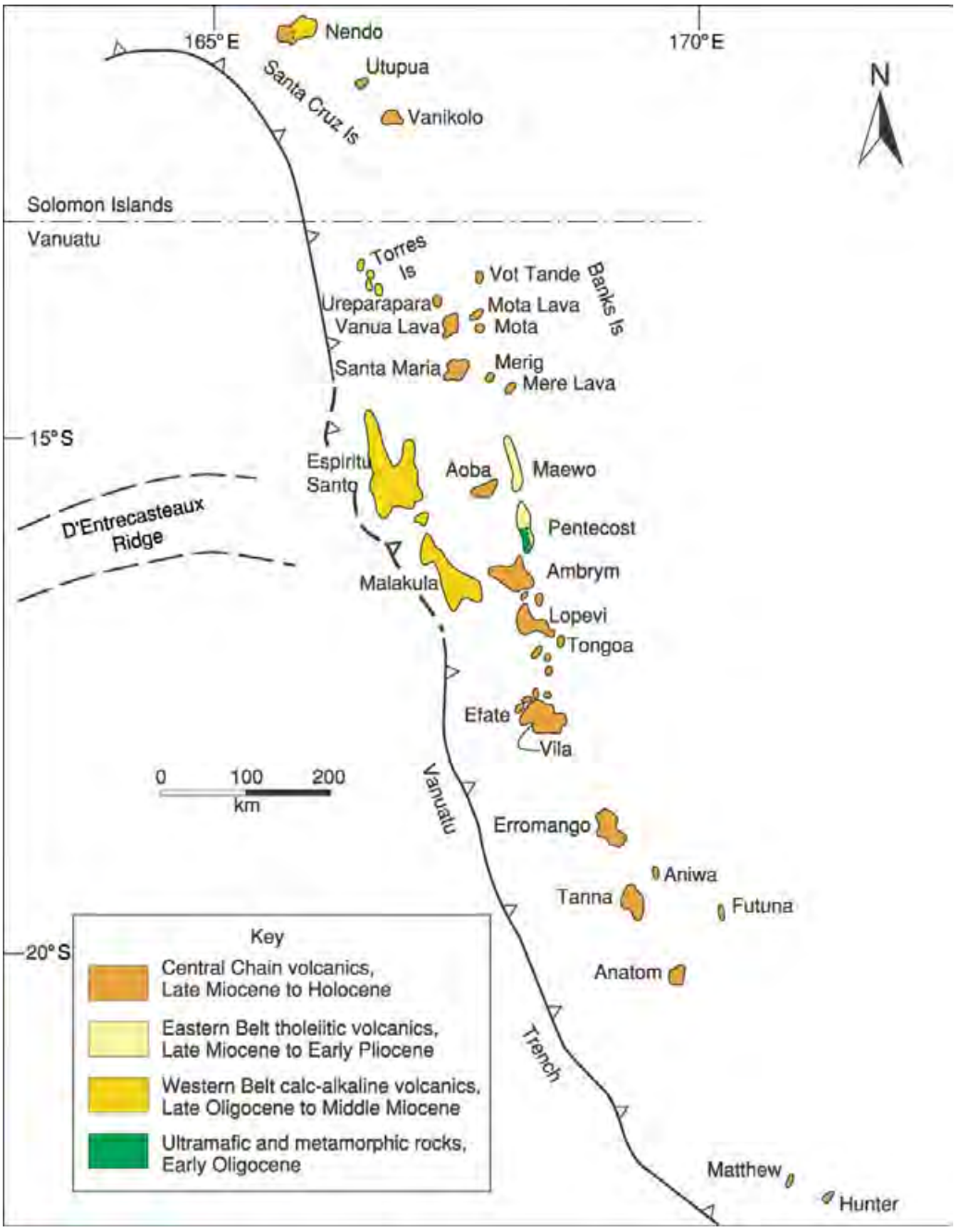


Figure 5 Geological map of Vanuatu, including the Santa Cruz Group in the Solomon Islands (the dashed line marks the international border). The map shows the active Vanuatu trench, the western belt Islands of Late Oligocene to Middle Miocene volcanics, the eastern belt (Maewo and Pentecost Islands) of Late Miocene to Early Miocene volcanic rocks, and the central chain of Late Miocene to Holocene volcanoes. Ultramafic and metamorphic rocks (dating to 35 Ma) are exposed on southern Pentecost Island. Map drawn by L Cotterall, Geology Department, Auckland University.

Vanuatu

The islands of the Vanuatu archipelago (Figure 5) form a narrow chain that extends for more than 1200 km, from the Torres Islands in the north to Matthew and Hunter islands in the south. The island chain is bounded eastward by the North Fiji Basin and westward by the Vanuatu Trench. The islands are, with one exception, entirely volcanic in origin, with associated sediments and intrusive rocks. The exception is on southern Pentecost Island, where ultramafic rocks associated with blocks of schistose amphibolite, metamorphosed pillow lava, and metagabbro are in

fault contact with Pliocene sediments. The metamorphic minerals have a radiometric age of around 35 Ma, and thus are older than any of the volcanic rocks.

The geology of Vanuatu has been described in terms of three volcanic provinces: a western province of Late Oligocene to Middle Miocene age (27–14 Ma; calc-alkaline volcanics), an eastern province of Late Miocene to Pliocene age (7–4 Ma; tholeiitic volcanics), and a central province of Miocene and Pliocene to Holocene volcanic rocks (6 Ma to the present day; Figure 5). The volcanic rocks of the western province developed in response to westward subduction of the Pacific plate at the Vitiaz Trench. The Vitiaz Trench, at that time, lay immediately east of Vanuatu. The trench moved away from Vanuatu (see Figure 1) with the opening of the North Fiji Basin in the past 6 million years.

The volcanic rocks of the eastern province developed in response to eastward subduction of Australian plate at the Vanuatu Trench. Magmatism in the central chain began at 6 Ma and was initially focused on Erromango, Tanna, and Anatom in the south. Volcanism subsequently developed along the entire length of the arc and shifted closer to the trench because of a steepening of the subduction zone to its present inclination of 70°.

The D'Entrecasteaux aseismic ridge, an Eocene–Oligocene island-arc complex, collided with central Vanuatu at 3–1.5 Ma, near Epi, and has migrated northwards as a result of the obliquity of the ridge relative to the trench. The collision has caused a slowing of subduction that may in turn have caused the development of back-arc thrust faults east of Maewo and Pentecost Islands, and the elevation of the ultramafic and metamorphic rocks on Pentecost. Five of the volcanoes of the central province are potentially dangerous. These are Ambae, Gaua, Ambrym, Tanna, and Lopevi. Only Yasur volcano on Tanna Island is monitored.

Economic Geology

Manganese oxide has been mined on Efate, where it typically is developed at contacts between tuff and overlying limestone. Occurrences of copper, zinc, molybdenum, and gold are known.

New Caledonia

The island of New Caledonia and the adjacent islands (Belep Islands to the north and Isle of Pines to the south) lie on the Norfolk Ridge. To the west, beyond the New Caledonia Basin, is Lord Howe Rise, and to the east are the Loyalty Ridge and Loyalty Islands (Figure 1). The rock units of New Caledonia (Figure 6) do not form a simple stratified sequence but, rather, comprise a series of terranes, juxtaposed during a complex history of convergent and extensional tectonic activity, and successor basin sediments. Four terranes make up a pre-Cretaceous basement. This is unconformably and tectonically overlain by Late Cretaceous and younger successor basin sediments and terranes.

Basement Terranes

The pre-Cretaceous basement is a complex amalgam of ophiolite, volcanics, and sediments that developed along the East Gondwana margin during the Permian to Jurassic. Four distinct terranes are recognized. The oldest, the Late Carboniferous Koh ophiolite, is a dismembered fore-arc ophiolite that is probably the basement on which two of the other three terranes, the Teremba and Central Chain terranes, were deposited. The Teremba Terrane, of mid-Permian to Late Jurassic age, has arc-related volcanics and volcaniclastics at its base, but is dominantly composed of a simply deformed thick sequence of diagenetically altered, fossiliferous shallow water sediments. The Central Chain Terrane, considered to be the distal equivalent of the Teremba Terrane, is composed of volcaniclastic sediments and minor arc-tholeiitic basalts that all have been weakly metamorphosed. The Boghen Terrane is a thick sequence of metagreywackes and schists, with sedimentary compositions ranging from pelitic to mafic, on a basal sequence of tholeiitic pillow lavas and red and green radiolarian cherts; detrital zircons in the sedimentary rocks indicate the sediments are of Jurassic age and derived from a south-east-Gondwana arc system and the Antarctic continent. The Boghen Terrane is considered to be a Jurassic accretionary complex, metamorphosed in the Late Jurassic.

Successor Basin Sediments

A transgressive sequence of Late Cretaceous coal measures and conglomerates rests unconformably on all of the older terranes, and forms thick sequences in the Nouméa–Bourail region and in the northern part of New Caledonia. Coarse carbonaceous sediments at the base of this sequence become finer upsection and are overlain by Paleocene to mid-Eocene siliceous shales and a pelagic chert and micrite sequence (pthanites). The coarser grained lower parts of the Late Cretaceous–Paleocene sequence locally contain shoshonitic basic volcanics, and there are also felsic horizons of rhyolitic flows and pyroclastics that appear to be younger than the basic volcanics. Later Eocene sequences, unconformably overlying the pelagic sediments, are carbonates. In the Bourail region, these are overlain by an alternating sandstone and siltstone flysch sequence. These are topped by an

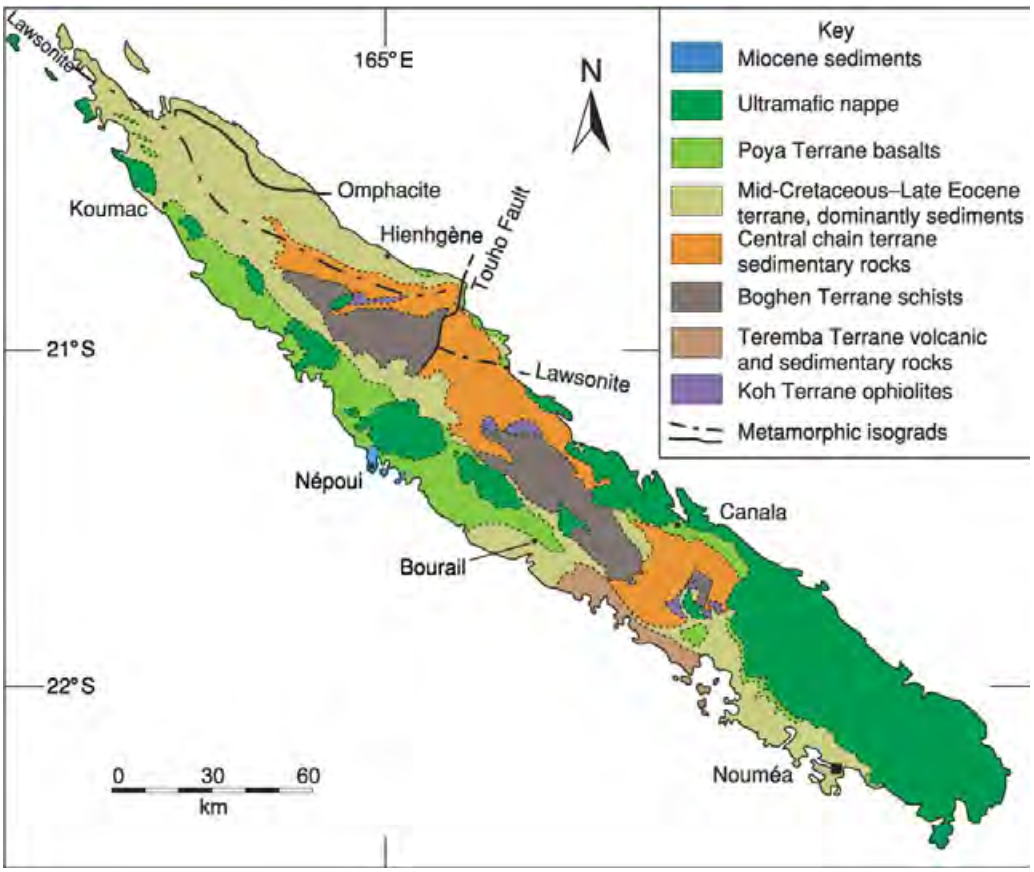


Figure 6 Geological map of New Caledonia (excluding the Loyalty Islands, Belep Islands, and Isle de Pines). The pre-Cretaceous basement comprises the Koh ophiolite and the Teremba, Boghen, and Central Chain terranes. These are unconformably overlain by mid-Cretaceous sediments forming a mid-Cretaceous to Late Eocene terrane or successor basin sequence. Great thrust sheets of Poya Terrane basalts and ultramafic rocks (the ultramafic nappe), emplaced in the Late Eocene, overlie the older terranes. Miocene marine sediments are preserved on the south-west coast. Map drawn by L Cotterall, Geology Department, Auckland University.

olistostrome that contains reworked components of the successor basin sediments, including upper Eocene bioclastic limestone and the flysch.

The Younger Terranes

The New Caledonian ophiolite, emplaced in the Late Eocene, overlies all the older rocks on the island. It consists of two units that appear to be geochemically unrelated: an overlying mainly harzburgitic ultramafic sheet (the ultramafic nappe) and the dominantly basic volcanic Poya terrane. The Poya terrane is composed of slices of geochemically variable submarine mafic volcanic rocks, mainly pillow lavas, together with dolerites and minor gabbros, and associated abyssal sediments. The ophiolite was emplaced as a result of Eocene convergence along the eastern margin of New Caledonia. Convergence was driven by the opening of the North Loyalty and South Fiji basins. Eastward-dipping subduction associated with the convergence consumed most of the South Loyalty Basin and caused high-pressure metamorphism of the Cretaceous–Eocene sediments of northern New Caledonia. The Eocene Loyalty Island volcanic arc formed above the subduction zone.

Miocene

Early Miocene coral reef and shallow-water marine sediments containing material derived from the ultramafic sheet are exposed at Nepoui. Small intrusions of granodiorite of Miocene age intrude the ultramafic sheet east of Nouméa, and north-east of Canala. Buoyancy-driven uplift in the Miocene has exhumed and exposed the high-pressure metamorphosed Cretaceous–Eocene sediments on the north-eastern coast of New Caledonia.

Economic Minerals

Since the Miocene, New Caledonia has been emergent and subjected to intensive tropical weathering, with the result that thick lateritic soils developed on the ultramafic rocks. The laterites have been mined for nickel, cobalt, and iron since the late nineteenth

century and still comprise the largest single reserve of nickel worldwide. Chromite has been mined directly from the ultramafic rock. Manganese deposits are associated with the Poya terrane basalts. Copper, lead, and zinc deposits have been mined from volcano-sedimentary horizons in the metamorphic rocks of the Diahot valley in northern New Caledonia. Gold-bearing quartz veins have been mined in the Diahot valley and gold and molybdenum deposits are associated with the Miocene granodiorite intrusions.

Fiji

The Fiji archipelago (Figure 7) comprises 800 islands, most of which lie between 16° and 19° south, 177° east, and 178° west. The Lau Group of islands in the east extends to 21° south (Figure 1). The largest islands are Viti Levu and Vanua Levu. The island of Rotuma, 650 km north of Viti Levu, is also a part of the Republic of Fiji. The islands are mainly volcanic and of island-arc origin, of basaltic, basaltic andesite, andesitic, and dacitic composition, and tholeiitic, calc-alkaline, or shoshonitic affinities. The younger volcanic rocks are mostly more alkaline and mainly belong to the alkali–basalt suite, and are intraplate rather than subduction related.

Viti Levu

The oldest rocks in the Fiji archipelago are exposed on Viti Levu. These are the volcanic and plutonic rocks of the Late Eocene to Early Oligocene Yavuna arc, and of the Late Oligocene to Middle Miocene (and possibly Late Miocene) Wainimala arc. The two volcanic suites are separated by an unconformity. The Yavuna arc rocks are exposed in western Viti Levu and include basaltic pillow lavas and volcaniclastic rocks, dacite, tonalite, and limestones of both Eocene and Oligocene ages. The basalts are geochemically primitive to arc like.

The Wainimala arc rocks are exposed throughout southern and central Viti Levu and on the islands of Malolo and Waya, immediately to the west. The rocks include mainly dacites in the centre of the island and basalts overlain by pelagic limestone and dacitic and andesitic lavas and volcaniclastics in the south-west. On Malolo and Waya, there are basaltic pillow lavas and breccia, and swarms of basalt dykes, indicating the local arc axis. Gabbro and tonalite of the Colo Plutonic Suite are thought to be part of the Wainimala arc activity and have mid- to Late Miocene radiometric ages (12–7 Ma). They were exposed to erosion in the Late Miocene. The Colo, Yavuna and Wainimala rocks show the effects of some deformation and weak metamorphism. They are unconformably overlain by latest Miocene to Early Pliocene basaltic and andesitic volcanic rocks and associated sediments (radiometric ages, 5.6–4 Ma). These cover much of the northern half of the island. The centre of the island was elevated at about 4 Ma. In the south-east, sediments were deposited in relatively deep water during the latest Miocene and through the Early Pliocene, the Messinian sea-level low being represented by a subaerially eroded coral reef. Shallow-water sediments were laid down in the latest Pliocene. In the south-west, sandstone and siltstone, overlain by limestone, were deposited in the Early Pliocene.

Vanua Levu

Vanua Levu comprises Late Miocene to Pliocene volcanic rocks (7–2.8 Ma). Volcanic activity appears to have progressed with time from east–north-east to west–south-west. The volcanic rocks are generally tholeiitic to calc-alkaline and include basalt, andesite, and some dacite, including a strong development of dacite and rhyolite in the north-east. Boninitic basalts were erupted on Cikobia to the north-east. Vanua Levu was elevated at about 3.5 Ma and Seatura Volcano began to form in western Vanua Levu not later than 3.3 Ma. Clastic sediments were deposited along the south-east coast.

Yasawa Group

The Yasawa and other islands north-west of Viti Levu, other than those of the Wainimala arc, were formed by volcanic arc activity (Yasawa arc) in the Late Miocene, from around 8 Ma. Early submarine dacites and basalts were succeeded by subaerial basalts in the central Yasawa islands, and by varied andesites. Strata generally dip towards Viti Levu, but at the north-eastern and southern ends of the chain have been strongly deformed with the development of overturned folds and thrust faults.

Islands Immediately South of Viti Levu

Vatulele, Beqa, and Yanuca (Serua) are Early Pliocene volcanoes south of Viti Levu. On Vatulele, shoshonitic rocks were erupted through limestone.

Islands of the Koro Sea

The islands of the Koro Sea (Lomaiviti and the Moala Group) mostly formed by volcanic activity in the Early Pliocene. Exceptions are the island of Koro, which formed in the Late Pliocene, and a Late Pleistocene cone on Nairai. The rock types range from tholeiitic through calc-alkaline to alkali basaltic and shoshonitic, and are mostly basaltic, with some hornblende andesite, monzonite, olivine monzonite, and tonalite. Olivine nodules occur in basalts on southern Moala and the western Koro islands.

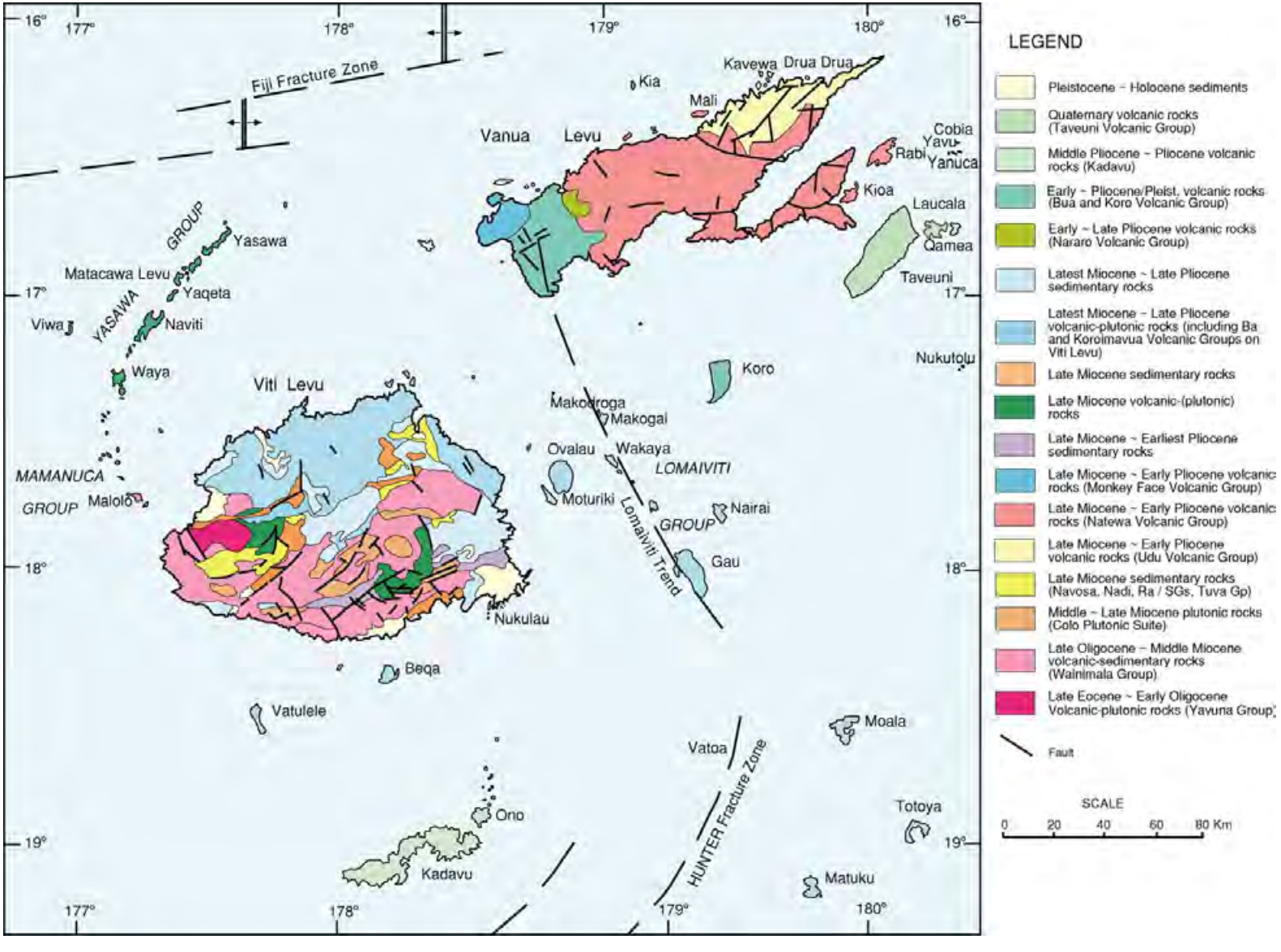


Figure 7 Geological map of Fiji, excluding Lau Islands and smaller Islands. On Viti Levu, the main rock units are Late Eocene–Early Oligocene Yavuna volcanic and plutonic rocks, Late Oligocene–Middle Miocene Wainimala volcanic rocks, Middle to Late Miocene Colo plutonic rocks, Late Miocene sediments, and Late Miocene to Pliocene volcanics and sediments. On Vanua Levu, the main rock units are Late Miocene–Early Pliocene Udu volcanics, Late Miocene–Early Pliocene Natewa volcanics, Early to Late Pliocene Nararo volcanic rocks, and Early Pliocene–Pleistocene volcanics. On Koro Island are Early Pliocene–Pleistocene volcanics; and on the Kadavu Islands are Middle Pliocene volcanic rocks. Map reproduced with permission from Mineral Resources Department, Fiji.

Kadavu Islands

The Kadavu islands, south of Viti Levu, are a volcanic island arc related to subduction at the Kadavu Trench. Ages range from 3.2 Ma to Holocene. Compositions range from basaltic to dacitic and are generally potassic.

Lau Islands

The islands on the Lau Ridge (Figure 1) formed in several periods of volcanic activity, probably beginning in the Eocene. The oldest exposed rocks, on Yacata island, are Middle Miocene (14 Ma). Further south, Late Oligocene rocks have been recovered by dredging. The exposed volcanic rocks are of three ages: Middle and Late Miocene, Middle Pliocene, and Late Pliocene to Pleistocene. Eruptive rocks in the two earlier episodes were tholeiitic to calc-alkaline, and in the third episode were alkali basalt. Limestone was deposited in the Middle to Late Miocene, Late Miocene to Early Pliocene, and Late Pliocene to Quaternary.

Quaternary Volcanic Activity

Quaternary volcanic activity of mainly alkali basalt composition formed the island of Taveuni immediately south-east of Vanua Levu (where the most recent dated eruption was at 340 ± 70 years), and the island of Rotuma, 650 km north of Viti Levu. Many small centres of Quaternary alkali-basalt volcanism are scattered around the archipelago in the north-east, the north-west, in Lomaiviti, adjoining Kadavu and in the Lau Group, but none is known on the Viti Levu platform. High-K andesite to dacite erupted at the western end of Kadavu island.

Economic Geology

Gold has been mined at Vatukoula on Viti Levu since 1933 and has been mined at other locations on Viti Levu and at Mt Kasi on Vanua Levu at different times. Manganese ore was mined from 1950 to 1971, most production being from Viti Levu in the 1950s. Magnetite was mined at Tuveriki in south-west Viti Levu from 1957 to 1963. Deposits of gold, polymetallic sulphides, and disseminated deposits are being explored, notably the Waisoi and other porphyry–copper deposits at Namosi on Viti Levu. Bauxite deposits occur in western Vanua Levu, mainly around Wainunu, and some islands of the Lau Group have deposits of phosphatic clay.

Plate Tectonic Setting

The early volcanic arcs that now form the basement of the Fiji platform developed above subducted Pacific plate in the Eocene, Oligocene, and Miocene. Initially the Yavuna arc and Lau–Tonga Ridge probably were parallel to, and close to, New Caledonia, but were moved to the north-east by the opening of the South Fiji Basin (36–25 Ma). By the Middle Miocene, the Lau–Tonga arc, Fiji, and Vanuatu arcs formed a single east-facing arc above subducted Pacific plate. This simple arc, sometimes referred to as the Vitiaz Arc, was disrupted in the Late Miocene with the onset of back-arc extension in the North Fiji Basin. This extension caused the clockwise rotation of Vanuatu and the anticlockwise rotation of the Fiji platform. In the Pliocene, spreading extended east and then south, causing separation of the Lau and Tonga ridges and the oblique subduction of South Fiji Basin lithosphere at the Kadavu Trench. Rotation of the Fiji platform caused local compression and extension. As Fiji moved away from sites of active subduction, volcanism became mainly alkali basalt in composition.

Tonga

The Tongan archipelago is a north–north-east-trending double-island chain that straddles the northern end of the active convergent plate boundary that separates the Pacific and Australian plates. A well-defined Benioff zone dips westward at 43–45° beneath Tonga and extends to a depth of 700 km. The larger islands of the archipelago form the eastern chain, 130–150 km west of the Tonga–Kermadec Trench. These are low-lying islands made up mainly of coraline limestone. On Eua, the southernmost of this eastern chain, the stratigraphic sequence has been described as consisting of coraline limestone, Miocene tuffs and tuffaceous limestones, Late Eocene tuffaceous marlstone (thickness, 60–80 m), Late Eocene foraminiferal and algal limestone, and Middle Eocene (46 Ma) basaltic and intermediate lava, volcanic breccia, and tuffs cut by dykes. Although the Eocene and Miocene rocks are not exposed elsewhere in Tonga, seismic profiles indicate that they form the basement of the Tonga Ridge. The crust beneath the ridge is 12–18 km thick. Development of the Tonga Trench and uplift of the frontal arc occurred in Middle to Late Oligocene. The Miocene tuffs are thought to represent the initial stages of arc volcanism. The islands of the western chain, which lie 150–200 km west of the trench, comprise the Tonga or Tofua volcanic arc. The volcanoes have a history of intense modern activity, with more than 35 eruptions occurring during the past 200 years, including both subaerial and submarine events. The volcanic products are mainly basalt and basaltic andesite with subordinate dacite, as is to be expected from an intra-oceanic arc system. West of the Tonga or Tofua arc is

the Lau basin, an area of attenuated crust and high heat flow that is interpreted as an extensional back-arc basin.

Samoa

The Samoan archipelago lies between latitudes 13° and 15° S and longitudes 169° and 173° W and extends for 500 km on azimuth 288°, from Rose Island in the east–south-east to Savaii Island in the west–north-west. The archipelago comprises a chain of hotspot-related intraplate volcanoes made up of the alkaline and tholeiitic basalts that are typical of Pacific island chains. The two largest islands are Savaii (1700 km^2) and Upolu (1115 km^2). These, together with the small islands of Manono and Apolima, comprise Western Samoa.

Savaii is the summit of a large basaltic volcano. Many small parasitic cones occur about its flanks and the lavas of the historical eruptions of 1760 and 1905–11 are widely distributed across the northern part of the island. On Upolo Island, young cones and lava flows overlie older, deeply eroded volcanic sequences that form spectacular topography on the eastern end of the island. Similar geology is seen on Tutuila, the third largest island, where post-erosional lavas overlie an older group of volcanics. Further east, the volcano chain is reduced to subaerial remnants in the Manu'a islands and the easternmost Rose Atoll.

Pacific island chains typically increase in age westward. Dating of Samoan lavas and dredged tholeiitic samples from seamounts west of Savaii confirm that this is the case in the older volcanic rocks of the Samoan chain. For example, shield volcanics on Upolu range in age from 2.8 Ma to 1.7 Ma and on Tutuila, to the east, range from 1.54 to 1.28 Ma. The post-erosional volcanic activity on Savaii, Upolu, and Tutuila, from 700 000 years ago to the present, does not fit the 'older westward' pattern. Possibly this is a side effect of the sharp change in the orientation of the Tonga–Kermadec Trench which, at a point about 100 km south of the Samoan islands, swings from a near north–south azimuth to near east–west (Figure 1). Flexural bending of the Pacific plate at this point presumably could cause rifting in the Samoan archipelago and thus provide a conduit for renewed volcanic activity.

See Also

Gondwanaland and Gondwana. **Plate Tectonics**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Hydrothermal Vents At Mid-Ocean Ridges.

Further Reading

Coleman PJ (ed.) (1973) *The Western Pacific: Island Arcs, Marginal Basins and Geochemistry*. Perth: University of Western Australia Press.

Hillis RR and Muller RD (2003) *Evolution and Dynamics of the Australian Plate. Geological Society of Australia Special Publication 22*. Sydney: Geological Society of Australia.

Kroenke LW (1984) *Cenozoic Development of the Southwest Pacific. SOPAC Technical Bulletin 6*. Suva: South Pacific Applied Geoscience Commission.

Veevers JJ (ed.) (2000) *Billion-Year Earth History of Australia and Neighbours in Gondwanaland*. Sydney: GEMOC Press, Macquarie University.

General

The following Earth Science Series, publications on *The Geology and Offshore Resources of Pacific Island Arcs*, are published by the Circum-Pacific Council for Energy and Mineral Resources (Menlo Park, CA) and are available from the AAPG Bookshop, PO Box 979, Tulsa, OK, USA.

Series 1: Howell DG (ed.) (1985) *Tectonostratigraphic Terranes of the Circum-Pacific Region*.

Series 2: Scholl DW and Vallier TL (eds.) (1985) *Tonga Region*.

Series 3: Brocher TM (ed.) (1985) *Geological Investigations of the Northern Melanesian Borderland*.

Series 4: Vedder JG, Pound KS, and Boundy SQ (eds.) (1986) *Central and Western Solomon Islands*.

Series 7: Taylor B and Exon NF (eds.) (1987) *Marine Geology, Geophysics and Geochemistry of the Woodlark Basin – Solomon Islands*.

Series 8: Greene HG and Wong FL (eds.) (1988) *Vanuatu Region*.

Series 9: Marlow MS, Dadisman SV, and Exon NF (eds.) (1988) *New Ireland and Manus Region, Papua New Guinea*.

Series 12: Vedder JG and Bruns TR (eds.) (1989) *Solomon Islands and Bougainville*.

Fiji

Colley H and Flint DJ (1995) *Metallic Mineral Deposits of Fiji. Fiji Mineral Resources Department Memoir 4*. Suva: Fiji Mineral Resources Department.

Hathway B and Colley H (1994) Eocene to Miocene geology of southwest Viti Levu, Fiji. In: Stevenson AJ, Herzer RH, and Ballance PF (eds.) *Geology and Submarine Resources of the Tonga–Lau–Fiji Region. SOPAC Technical Bulletin 8*, pp. 153–169. Suva: South Pacific Applied Geoscience Commission.

Rodda P (1994) Geology of Fiji. In: Stevenson AJ, Herzer RH, and Ballance PF (eds.) *Geology and Submarine Resources of the Tonga–Lau–Fiji Region. SOPAC Technical*

Bulletin 8, pp. 131–151. Suva: South Pacific Applied Geoscience Commission.

New Caledonia

Aitchison JC, Ireland TR, Clarke GL, Cluzel D, Davis AM, and Meffre S (1998) Regional implications of U/Pb SHRIMP age constraints on the tectonic evolution of New Caledonia. *Tectonophysics* 299: 333–343.

Black PM (1995) High-Si rhyolites and shoshonitic volcanics; a late Cretaceous bimodal association, Nouméa basin, New Caledonia. In: Mauk JL and St George JD (eds.) *Proceedings of the Pacrim Congress, September 1995, Auckland, New Zealand*, pp. 55–58.

Cluzel D and Meffre S (2002) L'unité de la Boghen (Nouvelle-Calédonie, Pacifique sud-ouest): un complexe d'accrétion jurassique. Données radiochronologiques preliminaries U-Pb sur les zircons détritiques. *CR Geoscience* 334: 867–874.

Cluzel D, Aitchison JC, and Picard C (2001) Tectonic accretion and underplating of mafic terranes in the Late Eocene intraoceanic fore-arc of New Caledonia (Southwest Pacific): geodynamic implications. *Tectonophysics* 340: 23–59.

Paris JP (1981) *Géologie de la Nouvelle-Calédonie. Un Essai de Synthèse. Memoirs du BRGM, No. 113*. Orleans: BRGM.

Yokoyama K, Brothers RN, and Black PM (1986) Regional facies in the high-pressure metamorphic belt of New Caledonia. In: Veevers JJ and Powell CM (eds.) *Permian–Triassic Pangean Basins and Foldbelts along the Panthalassan Margin of Gondwanaland. Geological Society of America Memoir 184*, pp. 407–571. Boulder, CO: Geological Society of America.

Papua New Guinea

Corbett GJ and Leach TM (1998) *Southwest Pacific Rim Gold–Copper Systems: Structure, Alteration and Mineralisation. Economic Geology Special Publication 6*. Littleton, CO: Society of Economic Geologists.

Davies HL, Perembo RCB, Winn RD, and KenGemar P (1997) Terranes of the New Guinea orogen. In: Hancock G (ed.) *Proceedings of the Geology Exploration and Mining Conference, Madang*, pp. 61–66. Melbourne: Australasian Institute of Mining & Metallurgy.

Dow DB (1977) *A Geological Synthesis of Papua New Guinea, Bulletin 201*. Australia: Bureau of Mineral Resources.

Pigram CJ and Davies HL (1987) Terranes and accretionary history of the New Guinea orogen. *BMR Journal of Australian Geology and Geophysics* 10: 193–211.

Solomon Islands

Coulson FI and Vedder JG (1986) Geology of the Central and Western Solomon Islands. In: Vedder JG, Pound KS, and Boundy SQ (eds.) *Geology and Offshore Resources of Pacific Island Arcs – Central and Western Solomon Islands, Earth Science Series 4*, pp. 59–87. Menlo Park, CA: Circum-Pacific Council for Energy and Mineral Resources.

Hughes GW, Craig PM, and Dennis RA (1981) *Geology of the Outer Eastern Islands. Geological Survey Division Solomon Islands Bulletin 4*. Honiara, Solomon Islands: Mines and Minerals Division.

Vanuatu

Auzende JM, Pelletier B, and Eissen J-P (1995) The North Fiji Basin: Geology, structure and geodynamic evolution. In: Taylor B (ed.) *Backarc Basins: Tectonics and Magmatism*, pp. 139–175. New York: Plenum.

Calmant S, Pelletier B, Lebellegard P, Bevis M, Taylor FW, and Phillips DA (2003) New insights on the tectonics along the New Hebrides subduction zone based on GPS results. *Journal of Geophysical Research* 108(B6): 2319.

Greene HG, Collot J-Y, Fisher MA, and Crawford A (1994) Neogene tectonic evolution of the New Hebrides island arc: a review incorporating ODP drilling results. In: Greene HG, Collot J-Y, Stokking LB, *et al.* (eds.) *Proceeding of the Ocean Drilling Program, Scientific Results 134*, pp. 19–46. College Station, TX: Ocean Drilling Program.

Macfarlane A and Carney JN (1987) *Geology and Mineralisation of North and Central Malekula. General Report*. Republic of Vanuatu: Department of Geology, Mines & Rural Water Supplies.

Macfarlane A, Carney JN, Crawford AJ, and Greene HG (1988) Vanuatu – a review of the onshore geology. In: Greene HG and Wong FL (eds.) *Geology and Offshore Resources of Pacific Island Arcs – Vanuatu Region, Earth Science Series 8*, pp. 45–91. Menlo Park, CA: Circum-Pacific Council for Energy and Mineral Resources.

Maillet P, Ruellan E, Gérard M, Person A, Bellon H, Cotten J, Joron JL, Nakata S, and Price RC (1995) Tectonics, magmatism, and evolution of the New Hebrides back-arc (south-west Pacific). In: Taylor B (ed.) *Backarc Basins: Tectonics and Magmatism*, pp. 177–235. New York: Plenum.

Mitchell AHG and Warden AJ (1971) Geological evolution of the New Hebrides island arc. *Journal of the Geological Society of London* 129: 501–542.

Warden AJ (1971) *Manganese Mineralisation in the New Hebrides. New Hebrides Condominium Geological Survey.* New Hebrides: British Service.

ORIGIN OF LIFE

J Bailey, Anglo-Australian Observatory and Australian Centre for Astrobiology, Sydney, Australia

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Establishing the nature of the processes by which life originated is one of the most fundamental unsolved scientific problems. Fifty years of study have enabled us to assemble many pieces of the puzzle but have also left a number of critical gaps in our understanding. Even the issues of when and where life originated are still poorly constrained, but life probably appeared on Earth between about 4.2 Ga and 3.5 Ga ago. There is good evidence that the current DNA–RNA–protein basis for life developed from a stage known as the 'RNA world' in which RNA performed the information-storage and catalytic roles currently performed by DNA and proteins. Direct formation of the RNA world from prebiotic chemistry seems unlikely, so much current research focuses on possible precursors to the RNA world. Various suggestions for pre-RNA worlds have been made, but none has yet proved entirely satisfactory. A number of terrestrial and extraterrestrial processes can produce prebiotic organic molecules, but there are still problems in obtaining the right molecules in sufficient quantities.

Development of Ideas on the Origin of Life

The idea that all life on Earth has a common origin became well established only in the twentieth century. Early ideas on the evolution of life were most clearly expressed by Lamarck, who described the process as one of progression from simpler to more complex and advanced forms (*see* **Evolution**). The observation that both simple and complex organisms are currently present therefore required that simple organisms are appearing even today from inorganic matter by a process of 'spontaneous generation' (a concept that has its origins in antiquity). Thus, the ideas of transformism (evolution) and spontaneous generation were closely linked in the early nineteenth century and put Lamarck and his followers in conflict with the religious and scientific establishment, who argued for the immutability of species, which were divinely created in their current forms. Lamarck's ideas were criticized by scientists such as Cuvier (*see* **Famous Geologists:** Cuvier) (in a celebrated debate with Geoffroy Saint-Hilaire in 1830) and later by Pasteur, who carried out his famous 'swan-necked flask' experiments to discredit the idea of spontaneous generation.

Darwin's theory of evolution by natural selection opened the way for the modern view of the origin of life (*see* **Famous Geologists:** Darwin). In Darwin's theory there is not necessarily a ladder of progress from simple to more complex forms. Simple organisms can be as evolutionarily successful as complex ones. This allowed the idea that all life, simple and complex, had a single origin in the distant past. While the subject of the origin of life is hardly mentioned in Darwin's published writings, the following quote (from a letter to his botanist friend Joseph Hooker) gives us an inkling of his thoughts on the subject at that time:

> If (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity present, that a protein compound was chemically formed, ready to undergo still more complex changes... .

Oparin and Haldane in the early twentieth century developed the idea that chemical reactions on the early Earth could have led to the production of a range of organic compounds, forming a 'primordial soup' in which the required building blocks for life would have been present.

But it was not until 1953 that these ideas received experimental support. Stanley Miller, then a graduate student working in the laboratory of Harold Urey, set up his famous experiment in which electrical discharges were passed through a mixture of gases (methane, ammonia, hydrogen, and water vapour) simulating a thunderstorm on the primitive Earth. The experiment produced a mixture of several amino acids, the building blocks of proteins. Miller speculated that this was how organic compounds had been made on the early Earth. In the same year, Crick and Watson published their structure for DNA, the first step in elucidating the fundamental molecular basis of life. These two discoveries meant that we had both a plausible way of generating simple organic building blocks and an understanding of the macromolecules on which

life depends. A detailed experimental and theoretical study of the origin of life was now possible.

The Earliest Life

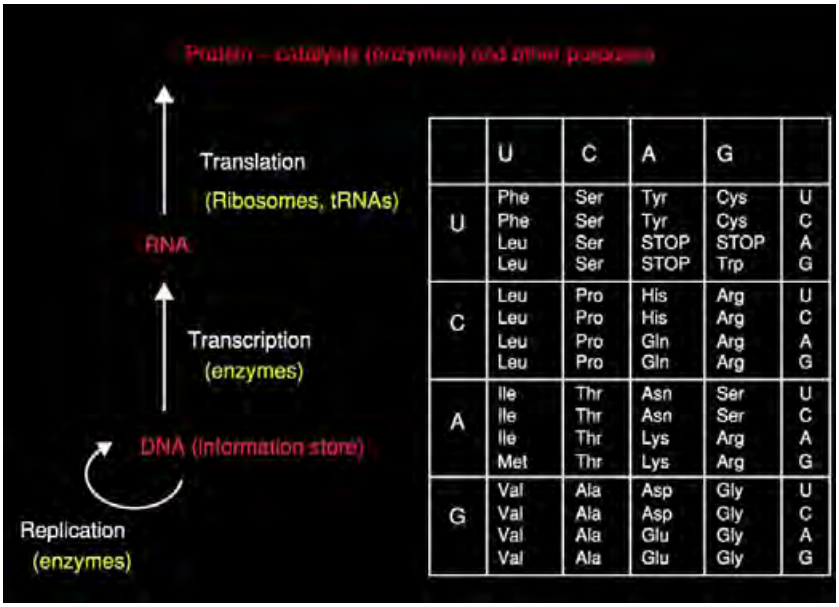
Sedimentary rocks of about 3.5 Ga old are found in the Pilbara region of Western Australia and in the Barberton Mountain Land in South Africa; these are the oldest such rocks that are sufficiently well preserved to contain fossils. These rocks contain structures that are interpreted as fossil stromatolites (layered structures built by colonies of micro-organisms) as well as filamentous microstructures interpreted to be microfossils of bacteria (*see* **Biosediments and Biofilms**). The biogenicity of many of these early fossils is controversial, but, if accepted, their presence indicates that life was well established on Earth at 3.5 Ga. It is even more difficult to say much about the nature of this early life. Some of the Pilbara microfossils have been suggested to resemble present-day cyanobacteria, but recent reinterpretation of the stratigraphy of the region suggests that the rocks originated in a hydrothermal environment.

Evidence for life has also been claimed in rocks of about 3.8 Ga from Greenland. These rocks are too metamorphosed to preserve physical fossils, but they contain carbon depleted in ^{13}C, which is consistent with the occurrence of biochemical reactions. These claims are also controversial.

The Tree of Life

The early history of life can also be studied through the techniques of molecular biology. It has been found that all life on Earth shares the same fundamental chemical processes. All life uses DNA as its basic genetic material. The DNA message is read by copying it onto RNA (transcription), and this RNA message is used to synthesize proteins (translation) by means of an RNA–protein complex called the ribosome (Figure 1). The resulting proteins are most frequently used as catalysts (enzymes) to drive the many chemical processes in the cell. The genetic code – which translates three-'letter' sequences of the four DNA bases (A, C, G, and T) into the corresponding amino acid to be added to a protein – is almost universal (with a few minor variations).

The common chemical basis for life strongly supports the idea that all life on Earth is related and descended from a single universal ancestor, sometimes known as LUCA (last universal common ancestor). Moreover, by comparing the sequences of genes involved in these fundamental processes common to all life, we can construct a 'family tree' showing the relatedness of all groups of life. Such molecular phylogenetic analysis has led to the 'three domain' taxonomy, in which life is divided into the domains Bacteria, Archaea, and Eucarya (Figure 2).



	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Figure 1 Summary of the key molecular processes common to all life. Genetic information is carried by DNA and can be copied (replication) onto new DNA molecules. The genetic message is read by first copying onto RNA (transcription) and then using the RNA message to synthesize a protein (translation). Three-letter sequences of the RNA bases (uracil (U), cytosine (C), adenine (A), and guanine (G)) are translated into a corresponding amino acid to include in a protein according to the genetic code shown on the right.

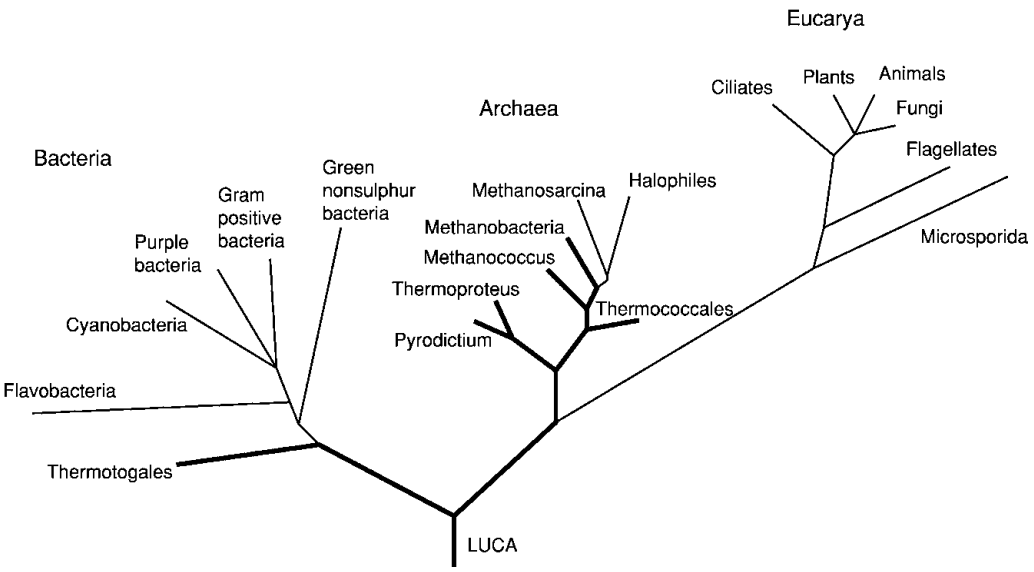


Figure 2 The tree of life derived from analysis of the small subunit ribosomal RNA gene, showing the division of life into the three domains Bacteria, Archaea, and Eucarya. Heavy lines indicate the branches containing hyperthermophilic organisms and illustrate the hypothesis that LUCA (the last universal common ancestor) was a hyperthermophile.

One feature of this phylogenetic tree is that many of the deepest branches closest to the 'root' of the tree are found to contain the hyperthermophiles (organisms adapted to very high temperatures). This has led to the suggestions that LUCA itself was probably a hyperthermophile and that perhaps life originated in a high-temperature environment. However, the interpretation of the tree is complicated by the likely occurrence of horizontal gene transfer (transfer of genes between different species), and there are possible interpretations of the tree that do not require LUCA to be hyperthermophilic.

Even if LUCA was a hyperthermophile, this does not necessarily tell us much about the environment in which life originated, since we do not know how much evolution occurred before LUCA. For example, one scenario is that life had already adapted to a wide range of environments before LUCA, but a catastrophic event such as a major impact wiped out everything except the hyperthermophiles.

LUCA is likely to have been similar to present-day bacteria. While these are usually regarded as simple organisms, even the simplest bacteria still contain substantial complexity, having genome sequences of several hundred thousand base pairs and using several hundred enzymes to catalyse a wide range of metabolic processes. The basic problem of the origin of life is to understand how such complexity could arise from a system simple enough to have assembled by chance in a prebiotic environment.

The key process is likely to be evolution by Darwinian natural selection. This process is capable of developing increased complexity. We therefore need to identify simpler systems that preceded the current DNA–RNA–protein basis for life but retain the basic requirements for natural selection to operate. While we have no fossil record to guide us in studying this stage of life's history, there are clues in the molecular mechanisms of modern organisms.

The RNA World

The DNA–RNA–protein basis for life presents a chicken-and-egg problem. DNA requires protein enzymes to catalyse its replication and transcription. However, these proteins can be made only by using the DNA genes that code for them. Hence, DNA is dependent on proteins and vice versa. A possible solution to the problem of how such a system originated was provided by the discovery by Cech and Altman that RNA can exhibit catalytic activity. This means that RNA can perform the roles of both a replicating information store (like DNA) and a catalyst (like protein enzymes).

This discovery led to the concept known as the RNA world – the idea that there was a stage in the development of life in which RNA acted as both the genome and the catalyst. These roles were subsequently taken over by DNA and proteins, leaving RNA in its modern role as an intermediary between DNA and proteins. The RNA-world concept is supported by a number of features of the molecular structure and chemistry of modern cells.

First, the structure of the ribosome, which catalyses protein synthesis, has recently been determined in detail. Although the ribosome contains both RNA and protein components, it has been found that the important catalytic activity comes from the RNA components, with the protein playing a peripheral role. This shows that protein synthesis (translation) is carried out almost exclusively by RNA and therefore probably originated at a time when only RNA catalysts were available. By contrast, the replication and transcription of DNA is carried out mostly by enzymes. Second, the synthesis of the nucleotides of DNA is carried out by building the corresponding RNA nucleotides first and then converting them to DNA nucleotides. Third, there are a number of important biochemical processes that use RNA (but not DNA) nucleotides or closely related molecules. These include the use of ATP (adenosine triphosphate) as the main energy carrier, and the cofactors used to assist a number of enzymes.

All these lines of evidence point to RNA being more ancient than either DNA or proteins and suggest that a pure RNA world 'invented' protein synthesis, giving rise to an RNA–protein world; subsequently, DNA was adopted as the primary genetic material.

For a pure RNA world (i.e. one in which proteins play no role) there must have been an RNA catalyst (or ribozyme) capable of catalysing the replication of RNA. No such molecule exists in modern organisms (although there are protein enzymes, known as RNA-dependent RNA polymerases, that do this). However, attempts have been made to synthesize such an RNA polymerase ribozyme. This is done by a process of 'in-vitro evolution', in which many copies of an RNA strand are made with random variations and then selected for appropriate catalytic activity. After many cycles of selection, ribozymes have been made that are capable of copying short sections of RNA (up to 14 bases) but are too inefficient to copy larger molecules.

The Origin of the RNA World

An RNA world is now widely accepted as being a stage in the development of life. The RNA world was either the first stage of life or preceded by an earlier stage of life, the pre-RNA world.

If the RNA world was the first stage of life then there must be a way of making a viable RNA world through the random processes of prebiotic chemistry. However, there are a number of difficulties with such a process. RNA consists of a sugar–phosphate backbone carrying four possible bases – guanine, adenine, cytosine, and uracil. Prebiotic syntheses have been suggested for the bases, but some require contrived conditions. The ribose sugar can be produced by the formose reaction but only in small quantities as part of a complex mixture of other sugars. The assembly of a complete ribonucleotide from these components presents further difficulties.

The nucleotides would then need to be assembled into RNA strands of sufficient length to show useful catalytic activity. Experiments using montmorillonite clay as a catalyst have led to RNA molecules of up to about 40 nucleotides. However, for this to be possible the nucleotides must be provided in their activated, or energy-rich, triphosphate form.

A further difficulty with an RNA world is imposed by the chirality of RNA. RNA replication (like many processes in biological chemistry) depends on the RNA nucleotides being based on the right-handed version of the two mirror-image structures for the sugar (i.e. D-ribose rather than L-ribose). However, any prebiotic process would be expected to make the sugar as a racemic mixture (i.e. equal quantities of D-ribose and L-ribose).

Pre-RNA Worlds

The difficulties with the prebiotic synthesis of RNA lead many researchers to conclude that direct production of an RNA world is highly improbable, and there must have been a preceding stage, or pre-RNA world, that used some other genetic material. Many different suggestions have been made.

The most obvious precursor to RNA is some other organic polymer that carries information in the same way as DNA and RNA. One suggestion is the peptide nucleic acids (PNA) (Figure 3). These molecules use the same set of bases as DNA and RNA but have a much simpler backbone linked by peptide bonds. The prebiotic synthesis of PNA seems to be more straightforward than that of RNA. Furthermore, PNA is achiral, thereby avoiding the chirality problems associated with RNA.

Another possible RNA precursor is TNA. This is a nucleic acid in which the four-carbon sugar threose replaces the five-carbon sugar ribose used in RNA. TNA can base pair with RNA or DNA and is the structurally simplest member of the RNA family. The four-carbon sugar should be easier to make than ribose in prebiotic conditions.

Some other suggestions for pre-RNA worlds involve very different ideas. For example, it has been suggested that inorganic clay crystals, rather than organic polymers, provided the first genetic material. Another suggestion is the 'compositional genome', in which the genetic information is carried by a set of molecules

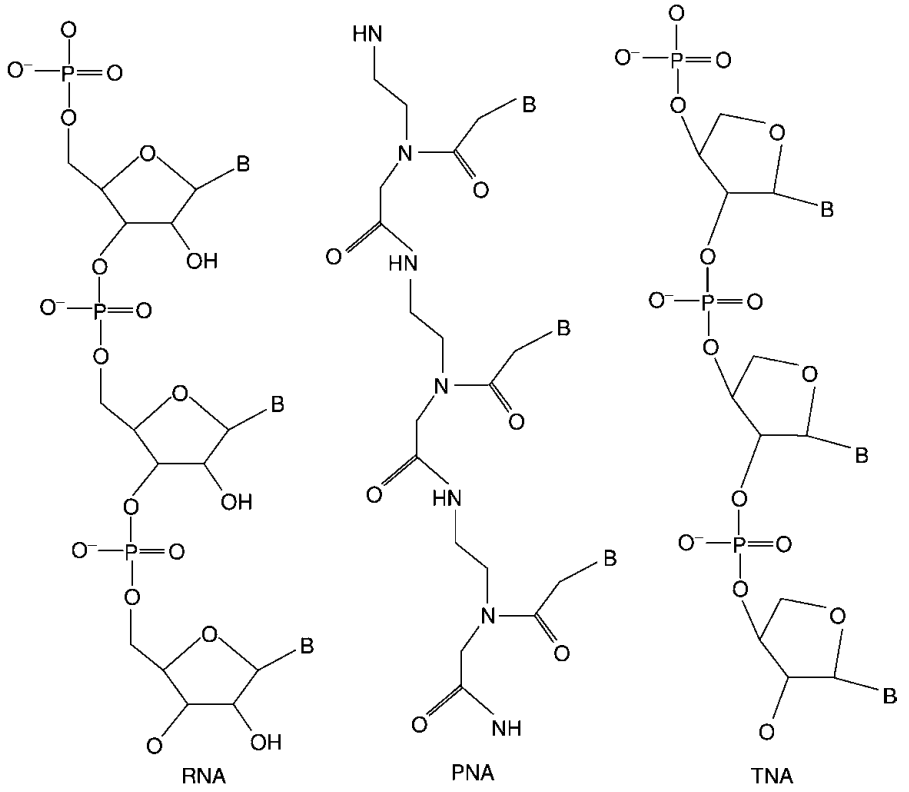


Figure 3 The molecular structures of RNA and two molecules that have been suggested as possible precursors of RNA. PNA (peptide nucleic acid) uses a simpler backbone based on peptide bonds, and TNA replaces the five-carbon sugar ribose with the four-carbon sugar threose. All three molecules can use the same set of four bases (attached at the positions marked B) and can base-pair with RNA or DNA.

enclosed by a membrane. Other models propose that networks of reactions that are the forerunners of modern metabolic processes preceded the development of genes. All such models present a number of problems, not least in explaining how the transition to an RNA world could occur (Figure 4).

Suffice it to say that no model of a pre-RNA world has yet achieved any degree of consensus, and the problem of how to build an RNA world out of the available prebiotic material remains unsolved.

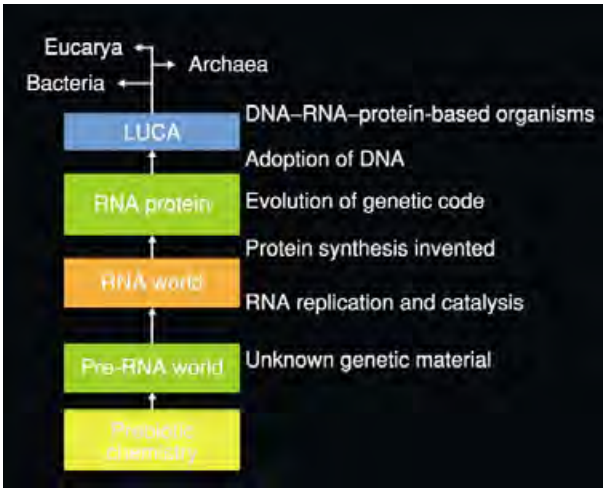


Figure 4 Outline of a possible series of steps in the development of life from prebiotic molecules up to the modern DNA–RNA–protein basis for biological chemistry.

Sources of Prebiotic Organic Molecules

The Miller experiment, described above, provided the first demonstration of how organic molecules of biological relevance might be produced in the atmosphere of the primitive Earth. Similar experiments have produced a wide range of organic molecules in this way. However, to achieve a substantial yield, a strongly reducing atmosphere (i.e. one containing gases such as hydrogen, methane, and ammonia) is required. Current ideas suggest that the atmosphere of the early Earth was instead composed mostly of carbon dioxide and nitrogen (*see* **Atmosphere Evolu-**

tion). There may have been a small amount of methane (up to about 100 ppm), making the atmosphere, at best, weakly reducing and allowing the production of small amounts of organic material in this way.

A possible alternative source of organic molecules is extraterrestrial delivery. We know from astronomical observations that interstellar clouds contain at least 100 species of simple organic molecule, while experiments simulating interstellar conditions have shown that more complex molecules, including amino acids, might be produced by photochemistry in ice-coated dust grains. The Murchison meteorite, which fell in Murchison, Australia in 1969, has been found to contain a huge range of organic compounds including more than 70 different amino acids. During the phase of heavy bombardment in the first few hundred million years after the solar system was formed, the Earth would have been subject to frequent impacts of comets and asteroids (*see* **Earth Structure and Origins**). While the heat produced by the impacts of these large bodies may destroy much of the organic material, recent experiments and simulations suggest that at least some organic material may be able to survive these major impacts. Organic material can also be delivered to the Earth's surface by interplanetary dust particles. However, current estimates of the amount of organic material that might have been delivered to the early Earth vary widely.

One intriguing discovery is that at least some of the amino acids in the Murchison meteorite show a small excess of the L or left-handed form. This excess is found in amino acids not known to be present in the biosphere, so it cannot be due to contamination, and suggests that a chiral imbalance existed in the organic material of the early Solar System. This could be relevant to the origin of biological chirality. A possible cause of the original excess of L-amino acids is photolysis by circularly polarized light, which could have occurred in the star-formation region in which the Sun originated.

Another source that has been suggested for organic compounds is reactions occurring in the superheated water around deep-sea hydrothermal vents (*see* **Tectonics:** Hydrothermal Vents At Mid-Ocean Ridges). Reactions involving mineral catalysts could build organic molecules from simple molecules such as carbon dioxide. However, some workers argue that the high temperatures in hydrothermal vents are more likely to destroy organic molecules than create them. It has been calculated that virtually all the water in the oceans circulates through hydrothermal vents on a timescale of about 10^7 years, and the temperatures would be high enough to destroy molecules such as amino acids. This limits the amount of organic material that can build up in the oceans from any of the above sources.

Thus, while there are a number of possible sources of prebiotic organic compounds, we are still left with two significant problems. First, the total amount of organic material produced by any of these mechanisms seems to be enough to lead to only very weak concentrations in the ocean. Second, the molecules that are most easily made do not necessarily appear to be the most relevant. For example, amino acids are readily produced and are certainly important in modern biology, but the adoption of proteins seems to be a relatively late step in the development of life. The building blocks needed for RNA (such as ribose) are more difficult to make, although RNA appears to be essential for early life.

Where did Life Originate?

Given that we do not yet understand the chemical processes required for the early stages of life, it is hard to say much about the likely location. However, there are some general considerations that can be applied. The early Earth would have been subject to bombardment by asteroids at a high but rapidly reducing rate, and some of the largest impacts would have been 'sterilizing impacts' capable of vaporizing the oceans and destroying any life. Clearly life must have appeared after the last sterilizing impact, which might have been around 4–4.2 Ga ago. (There could, of course, have been one or more earlier origins of life that were wiped out in the impacts.) However, somewhat smaller impacts might have destroyed life near the surface while allowing it to be preserved deep in the oceans or within the Earth's crust. This suggests that life might have originated in a deep-ocean location or even deep within the crust. Another important consideration is that life requires an energy source (for modern life this is normally solar energy via photosynthesis). For a deep location the likely energy source is geothermal energy. A location on or near the surface would make available other energy sources such as ultraviolet radiation and lightning. A surface origin would, however, be much more exposed to impacts, and life would therefore have to have arisen somewhat later, towards the end of the heavy bombardment. This leaves a rather small time window between the end of the bombardment and the first evidence of life on Earth. However, the timings of both these events are somewhat uncertain.

Finally, the possibility must be considered that life did not originate on Earth at all. More than 20 meteorites have been identified as coming from Mars. The most famous is ALH84001, which has been sug-

gested to contain fossil evidence of Martian life. While many are unconvinced by these claims, the meteorite does provide evidence that material can be transferred from Mars to Earth with relatively good preservation, and suggests that something like a bacterial spore could seed life from one planet to another. There is evidence that early Mars had liquid water and may have been at least as favourable as Earth as a site for the origin of life. During the heavy-bombardment phase transfer of material from Mars to Earth would have been much more frequent than it is today.

Glossary

ALH84001 Martian meteorite discovered in the Allan Hills ice-field in Antarctica in 1984. It has been suggested that this meteorite contains evidence of Martian microbial life.

Amino acid Simple organic molecule containing an amino group (NH_2) and a carboxyl group (COOH). Amino acids link together in chains to form proteins.

Archaea One of three domains of life, the other two being the Bacteria and the Eucarya. The archaea are prokaryotic organisms, which were shown to be distinct from the bacteria by molecular phylogeny using 16S ribosomal RNA. The archaeal domain includes many extremophilic organisms.

Bacteria One of three domains of life, the other two being the Archaea and the Eucarya. The bacterial domain includes all prokaryotic organisms not classified as archaea.

Base pairing (Watson–Crick base pairing) The pairing of the DNA and RNA bases through chemical bonds, which link the double-helix structure of DNA and allow the copying of information. Adenine always pairs with thymine (or uracil), and cytosine pairs with guanine.

Bases The components of DNA and RNA that carry the genetic information in their sequence. DNA uses adenine, guanine, cytosine, and thymine. In RNA, thymine is replaced by uracil.

Chirality A chiral molecule is a molecule with an asymmetric structure that can exist in two mirror-image forms (or enantiomers). In living organisms such molecules are usually found in only one of the two possible enantiomers (homochirality). Thus, amino acids are normally in the left-handed or L-enantiomer, while sugars are in the right-handed or D-enantiomer.

Cyanobacteria A class of bacteria that make use of oxygen-producing photosynthesis; commonly referred to as blue-green algae.

DNA Deoxyribonucleic acid; the molecule that constitutes the genome of living cells. DNA molecules have a double-stranded helical structure built from a sugar–phosphate backbone and a set of four bases (adenine, guanine, cytosine, and thymine). The sequence of bases specifies the genetic information.

Domain The highest taxonomic division in the classification of living organisms. The three domains are the Archaea, the Bacteria, and the Eucarya. Domains are subdivided into kingdoms. While the three-domain model is widely used in astrobiology, some biologists prefer other schemes, such as the five-kingdom system.

Enzyme A protein that acts as a catalyst. Most chemical processes in living cells are enzyme catalysed.

Genetic code The set of rules by which three-letter 'words' in a DNA or RNA sequence describe an amino acid to be incorporated in a protein.

Genome The complete set of genetic information for a particular organism.

Heavy bombardment During the first few hundred million years after the formation of the solar system, the Earth and the other planets were subject to an intense bombardment by the debris left over from the formation of the solar system. It is during this phase of heavy bombardment that most of the craters on the Moon were formed. The emergence of life on Earth appears to coincide roughly with the end of the heavy bombardment.

Hyperthermophile An organism adapted to life at very high temperatures. Hyperthermophiles have optimum growth temperatures above 80°C, and a number can grow at temperatures above 100°C.

Interstellar molecule Molecules in interstellar space are most commonly detected by means of radio-frequency emission lines coming from the gas in molecular clouds. Molecules can also be detected from the infrared spectra emitted from dust. Well over a hundred molecular species have been detected by these methods. Some of the more complex molecules found include acetic acid, acetone, and ethanol.

Lateral gene transfer The transfer of genes between different species. Lateral gene transfer may have been widespread in the early stages of life on Earth, and this complicates the interpretation of the tree of life.

LUCA Last universal common ancestor; the last common ancestor of all living organisms.

Martian meteorites Meteorites that originate from Mars, also known as SNC meteorites. Their Martian origin is demonstrated by bubbles of gas trapped within them, which have identical composition to the atmosphere of Mars.

Miller–Urey experiment (or Miller experiment) An experiment carried out by Stanley Miller and Harold Urey in 1952, which demonstrated the synthesis of amino acids in conditions simulating a thunderstorm on the early Earth. The experiment can be seen as marking the beginning of the experimental study of the chemistry of life's origins.

Nucleic acid The molecules that carry genetic information – DNA and RNA.

Phylogeny The evolutionary relationships between different species of organism, represented in the form of a phylogenetic tree. In molecular phylogeny these relationships are determined by analysing the differences in the sequences of genes common to the various species.

Pre-RNA world A hypothetical early stage in the development of life, which preceded the RNA world and used some other genetic material in place of RNA or DNA.

Ribosomal RNA The RNA components of a ribosome. One of these components, the small subunit ribosomal RNA (also known as 16S ribosomal RNA in prokaryotes or 18S ribosomal RNA in eukaryotes), has been widely used to determine the tree of life.

Ribosome A structure composed of protein and RNA molecules that reads genetic information from messenger RNA and synthesizes the corresponding protein.

Ribozyme An RNA molecule that acts as a catalyst. The discovery of RNA catalysis led to a Nobel prize for Sidney Altman and Thomas Cech, and to the concept of the RNA world.

RNA Ribonucleic acid; a molecule that can carry genetic information in a similar way to DNA. In modern cells, RNA molecules are important in the process of protein synthesis, in the form of messenger RNA, ribosomal RNA, and transfer RNA.

RNA world A hypothetical early stage in the development of life, in which RNA molecules provided both the genome and the catalysts, roles that were subsequently taken over by DNA and proteins.

Stromatolites Layered structures built by colonies of micro-organisms that are commonly found in the Archaean and Proterozoic fossil records. Modern examples can be found in sites such as Shark Bay in Western Australia.

Tree of life A phylogenetic tree covering all groups of life on Earth. The term is commonly used for the tree derived by molecular phylogeny using small subunit ribosomal RNA, as pioneered by Carl Woese in the 1970s.

See Also

Atmosphere Evolution. **Biosediments and Biofilms**. **Earth Structure and Origins**. **Evolution**. **Famous Geologists:** Cuvier; Darwin. **Precambrian:** Eukaryote Fossils; Prokaryote Fossils. **Tectonics:** Hydrothermal Vents At Mid-Ocean Ridges.

Further Reading

Bailey J (2001) Astronomical sources of circularly polarized light and the origin of homochirality. *Origins of Life and Evolution of the Biosphere* 31: 167–183.

Davies P (1998) *The Fifth Miracle*. London: Penguin Books.

De Duve C (1995) *Vital Dust*. New York: Basic Books.

Ferris JP (1998) Catalyzed RNA synthesis for the RNA world. In: Brack A (ed.) *The Molecular Origins of Life*, pp. 255–268. Cambridge: Cambridge University Press.

Kasting JF and Brown LL (1998) The early atmosphere as a source of biogenic compounds. In: Brack A (ed.) *The Molecular Origins of Life*, pp. 35–56. Cambridge: Cambridge University Press.

Miller S (1953) A production of amino acids under possible primitive Earth conditions. *Science* 117: 528–529.

Miller SL and Lazcano A (2002) Formation of the building blocks of life. In: Schopf JW (ed.) *Life's Origin*, pp. 78–112. Berkeley: University of California Press.

Orgel LE (2002) The origin of biological information. In: Schopf JW (ed.) *Life's Origin*, pp. 140–157. Berkeley: University of California Press.

Stetter K (1998) Hyperthermophiles and their possible role as ancestors of modern life. In: Brack A (ed.) *The Molecular Origins of Life*, pp. 315–335. Cambridge: Cambridge University Press.

Wills C and Bada J (2000) *The Spark of Life*. New York: Perseus Publishing.

Woese CR, Kandler O, and Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proceedings of the National Academy of Sciences USA* 87:4576–4579.

Yusupov MM, Yusupova GZ, and Baucom A (2001) Crystal structure of the ribosome at 5.5Å resolution. *Science* 292: 883–896.

PALAEOCLIMATES

B W Sellwood, University of Reading, Reading, UK
P J Valdes, University of Bristol, Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

An evaluation of the climatic regime under which sedimentary successions accumulated has often been integral to their overall interpretation. It is an essentially uniformitarian approach. Palaeoclimatology, the understanding of past climatic regimes (palaeoclimates), has only relatively recently become a subject in its own right and is currently undergoing a major phase of expansion. One reason for this is the light thrown by palaeoclimatic studies on the natural variability that exists in the Earth's climate in the context of future, anthropogenically influenced, climate changes. Another is the realization that many of the Earth's sedimentary resources (e.g. coal, hydrocarbons, and other extractable minerals) formed under particular types of climatic regime.

Through time, the climate of the Earth has undergone many changes – the most recent ice age ended a mere 500 human generations ago. Astronomical data suggest that the Sun is of a type that steadily increases its heat output as it ages, increasing by about 1% per 100 Ma (*see* **Solar System:** The Sun). This presents a problem (the 'faint young sun paradox'). Simply stated, in its early days the Sun would have emitted about 45% less than its current heat output, so Earth should have been frozen, but there is ample evidence from Earth's most ancient rocks (e.g. in Greenland, South Africa, and Australia) to suggest that aqueous processes were operating. This paradox can be overcome if it is accepted that Earth's atmosphere was much denser and much richer in greenhouse gases (such as carbon dioxide and methane) during its early history. The rock record makes it clear that the Earth was periodically affected by major freezes during its early history, and controversy exists over the possibility that the Earth periodically froze from the poles to the equator (the 'snowball Earth hypothesis') during Proterozoic and older times (*see* **Atmosphere Evolution**).

Despite the possibility that the Sun was significantly cooler 500 Ma ago, abundant and widespread geological evidence suggests that through much of the subsequent time the Earth has been significantly warmer than at present. Since the Cambrian there appear to have been only three episodes that were cooler than now. One was in the Ordovician, when glaciation was centred on the present Sahara (*see* **Palaeozoic:** Ordovician), and one was in the Late Carboniferous (*see* **Palaeozoic:** Carboniferous) and Early Permian, with glaciation centred on southern Gondwana (*see* **Gondwanaland and Gondwana**). The latest 'ice ages' took place during the Quaternary (*see* **Tertiary To Present:** Pleistocene and The Ice Age), from about 1.67 Ma, but a growing body of evidence suggests that this latest phase had its origins at around the Eocene–Oligocene boundary (36 Ma), when ice began to accumulate over eastern Antarctica. For around 250 Ma, since the Late Permian, there appears to have been a period of extreme warmth, often referred to as the 'greenhouse world'. The Earth's past climate regimes can now be simulated using general circulation models (GCMs) similar to those used in long-term weather forecasting.

Orbital Forcing of Climate

In 1875, James Croll was the first to describe the cyclic way in which the Earth's orbit behaves. The eccentricity varies from near-circular to elliptical over a cycle of about 110 Ka and affects the total amount of solar radiation reaching the Earth. Changes in the tilt of the Earth's axis of rotation (obliquity; currently about 23.4° but varying between 21.8° and 24.4°), which affect the latitudinal distribution of received solar radiation, take around 40 Ka. In addition, a rotational wobble in the Earth's axis of rotation (as if the poles describe a circle against the background of the stars) causes the precession of the equinoxes and takes around 22 Ka to complete a cycle. In the 1930s and 1940s, Milutin Milankovitch, improving on Croll's work, realized that each of these cycles interacts with the others to cause changes in the amount of insolation received by the Earth that are sufficient to promote climate changes (*see* **Earth:** Orbital Variation (Including Milankovitch Cycles)). These cycles are now named after Milankovitch and have been used to predict the time-frame of glacial cycles in the Quaternary. Changes in palaeoclimate have been recognized in Neogene oceanic sediments, which provide a metronomic time signal, synchronized with the Milankovitch timeframe, and hence a means of short-term correlation that was unavailable until recently.

Geological Proxies of Palaeoclimate (Figures 1A and 1B)

Sedimentary facies and palaeontological data generally provide qualitative evidence of palaeoclimates

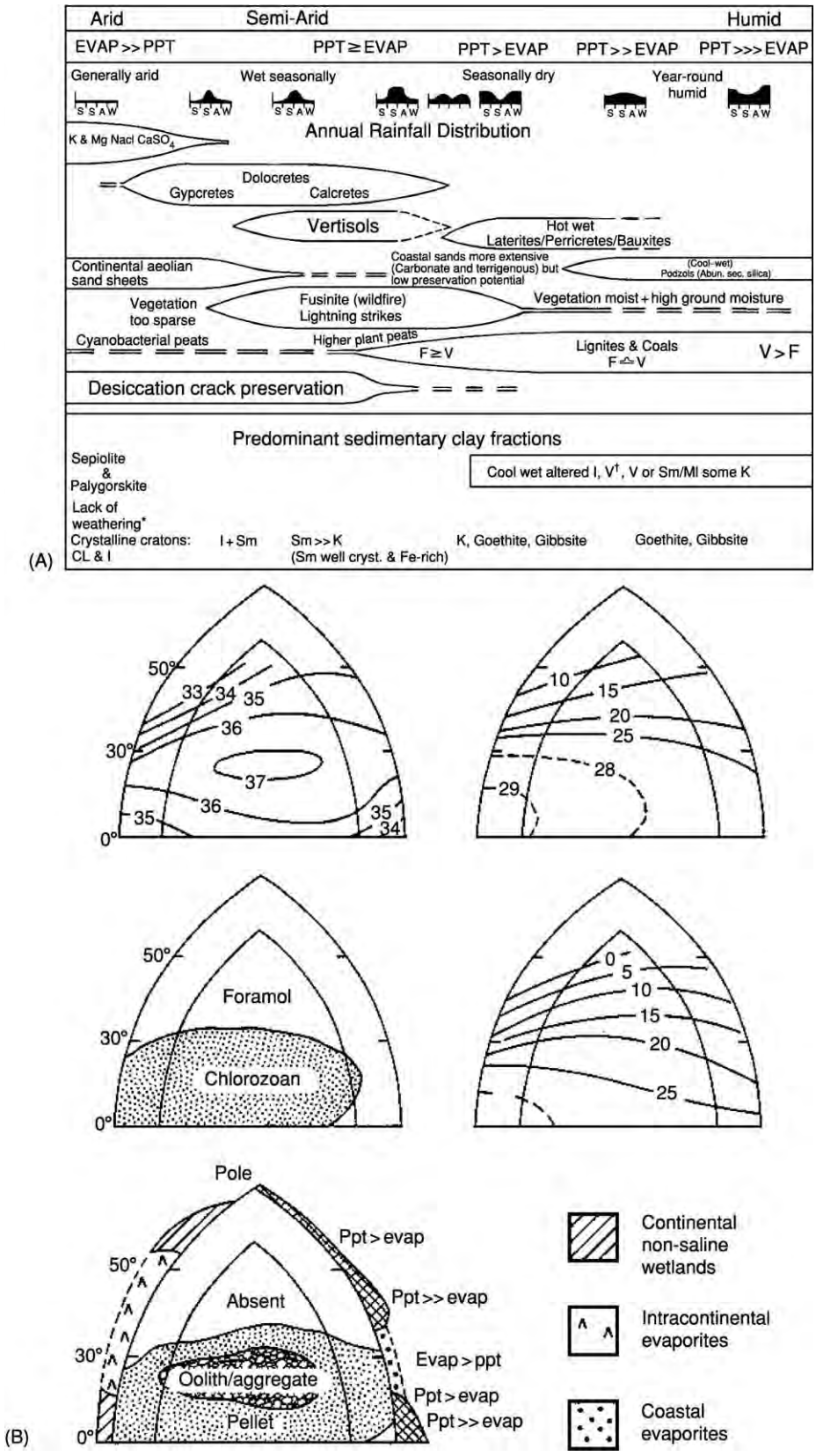


Figure 1 Geological palaeoclimate proxies for (A) terrestrial environments and (B) marine carbonates and evaporites (top left, salinity (‰); top right, maximum temperature (°C); middle left, skeletal associations; middle right, minimum temperature (°C) Lower left, predicted modern carbonate and evaporite distributions. Cross-hatching shows maritime latitudes where precipitation exceeds

and climate changes. These are palaeoclimate proxies. The most climatically informative structures are glaciogenic features (striated pavements, tills, and tillites), distinctive soil types (e.g. laterites and bauxites), and indicators of aridity (e.g. evaporites and aeolianites). Supplementary evidence of climate is provided by a range of other features, such as other types of palaeosol (e.g. calcretes, gypcretes, and vertisols), clay mineralogy (reflecting weathering regime), storm deposits, reefs and oolites, glendonites, and fusain (from wildfires). Coals, often considered to require moist tropical climates, are now known to accumulate equally well in temperate mires. Critical information on past climates (particularly quantitative information on palaeotemperature) comes from oxygen isotopic ($\delta^{18}O$) data derived from carbonate fossils, but such data have to be treated with care because of problems associated with diagenetic resetting. Palaeobotanical data, particularly regarding plant appearance (physiognomy), are generally qualitative, and have also been claimed to provide quantitative information. The density of leaf stomata may be related to the concentration of atmospheric carbon dioxide, and there is, today, a striking relationship between mean annual temperature and leaf shape (percentage of entire-margined, as opposed to toothed-margined, species) in a biome. Entire-margined leaves dominate warm biomes.

Fossil reptiles such as crocodilians, like their modern descendants, have global distributions that are principally controlled by temperature. A coldest-month mean temperature of 5.5°C marks the minimum thermal limit for the group (corresponding to a modern mean annual temperature of *ca*. 14.2°C). Such evaluations of temperature have been successfully applied globally for the Mesozoic and Cenozoic and compared with GCM output on climate for those times.

Marine Carbonates

Modern shelf seas are dominated by carbonate facies in areas where organic productivity is high, generally in low latitudes (between about 30°N and 30°S) with strong insolation and warm normal-salinity seawaters (>20°C and 35‰ salinity). Carbonate shelves also occur in temperate waters, particularly where rates of terrigenous run-off are low. However, rates of productivity in these areas are much lower than in warmer waters, and the biota is dominated by different species. Nonetheless, dominance of carbonates in a particular rock succession may not necessarily indicate warm waters, particularly for more ancient rocks in which the biota have only tenuous links with modern forms (*see* **Sedimentary Environments:** Carbonate Shorelines and Shelves).

Present-day warm-water shelves unpolluted by fine-grained terrigenous run-off have diverse communities with hermatypic (reef-building) corals and codiacian calcareous algae and may include ooids, aggregates, and pellets; these latter grain types reflect carbonate precipitation and early cementation. This is known as the chlorozoan skeletal grain association and can be recognized as far back as the Triassic. Modern temperate-water carbonate factories are dominated by benthic foraminiferans, molluscs, bryozoans, barnacles, and calcareous red algae. This is termed the foramol skeletal grain association and can be recognized, albeit equivocally, into the Mesozoic Era. The polewards expansion of reef and carbonate facies (by about 10° latitude) in the Mesozoic and Early Cenozoic, relative to today (30° today, *ca*. 40° in the Mesozoic and Palaeogene), suggests that the Earth was more equable during these times.

The principles of palaeotemperature determination from carbonates using oxygen isotope ratios are well established, and for many years oxygen isotopic ratios have been used to quantify sea-surface temperatures and bottom-water temperatures. The technique has most successfully been applied using the fossil shells of organisms considered to be robust to diagenetic changes (e.g. planktonic forams, coccolithic oozes). Such data have shown that ocean waters have changed their isotopic ratios in concert with the waxing and waning of the ice-caps during Quaternary glacial cycles, the oceans becoming relatively depleted in ^{16}O during glacial episodes (shells therefore showing ^{18}O enrichment) by comparison with interglacials. In the longer term the $\delta^{18}O$ trend for ocean waters suggests that seawater has become progressively depleted in ^{16}O (leading to positive $\delta^{18}O$ values) since about 55 Ma. This has occurred in a series of steps, probably reflecting, first, the cooling of deep waters, followed by the onset of permanent ice on Antarctica at around the Eocene–Oligocene boundary. Then followed a series of significant more positive excursions, through the Oligocene and Miocene, before a major and permanent swing in the Miocene (15 Ma), which is taken to reflect the formation of the Antarctic ice-sheet. The growth of the northern hemisphere ice-sheets is

evaporation). Under lignites and coals V, vitrinite; F, fusinite; I, illite; V, vermiculite; Sm, smectite; K, kaolinite; CL, chlorite; ppt, precipitation; evap, evaporation; *, unaltered parent rock with fines under both arid and frigid terrains; †, vermiculite can form under a wide variety of conditions.

marked by a further more positive excursion in the Quaternary.

Evaporites

Evaporite salts may form anywhere on Earth where evaporation exceeds rainfall (*see* **Sedimentary Rocks:** Evaporites). Their preservation is controlled by the nature of the post-depositional climate, the ingress of groundwaters, and subsequent burial history. Modern coastal evaporites accumulate between 15° and 35° latitude, but intracontinental evaporites may occur beyond 50°, sometimes being frozen for several months in the year. So salts do not necessarily imply permanent warmth. The formation and preservation of the more soluble salts of potassium require low atmospheric humidity (less than 50%). Salts may form very rapidly under ideal conditions (100 m in 1 Ka), so short-lived episodes of evaporation may lead to great thickness of salts.

Glaciogenic and Cold Water Sediments

About 10% of the Earth is covered by ice today, but this increased to 30% during the Quaternary glacials. Ice and permafrost provide a series of erosional and depositional fingerprints (e.g. striated pavements and boulders, dropstones, ice wedges, and glacial tills (diamicton)) (*see* **Sedimentary Processes:** Glaciers) as well as a range of geomorphological features, such as eskers, kames, etc., some of which have been recognized in association with ancient glacial systems (e.g. in the Lower Palaeozoic of the Sahara and the Late Palaeozoic of Gondwana). However, boulders may be dropped into deep water through mechanisms other than ice rafting (e.g. from the roots of floating trees). More cryptic evidence of cold-water precipitation comes from glendonites – carbonate nodules pseudomorphing the mineral ikaite, the low-temperature polymorph of calcium carbonate ($CaCO_3 \cdot 6H_2O$). It is today associated with clathrates in cold deep waters and in shallow polar waters. Glendonites are reported from Cretaceous localities in Australia and Canada and from the Jurassic of Siberia.

Coals and Lignites

Coals and lignites are indicative of an excess of precipitation over evaporation: vegetation accumulates in wetland mires, which, if rain occurs on most days of the year, will grow upwards above the general level of the regional water table. Preservation of peat requires rapid burial and/or anoxic conditions. Frigid conditions favour preservation, as do environments in which rainfall is evenly spread throughout the year. Today such conditions occur in equatorial zones and in high mid-latitudes. In the Mesozoic equatorial zones appear to have been more arid than they are today.

Wildfires and Palaeosols

Fusain, almost pure carbon in sedimentary rocks, is a product of wildfires. Natural fires, mostly triggered by lightning strikes, are favoured in strongly seasonal climates where vigorous plant growth is promoted by a distinct wet season, followed by a dry season creating tinder that becomes fire-prone during thunder storms at the onset of the next wet season. Fusain-rich sediments are frequently associated with other evidence of strong seasonality (distinct wet–dry seasons) such as vertisols (and associated smectite-rich clay mineralogies) and calcretes.

Laterites (or ferricretes) reflect the relative accumulation of haematite, goethite, aluminium hydroxides (e.g. gibbsite), and kaolinite within soil profiles (*see* **Soils:** Palaeosols). Pedogenic and groundwater laterites form best in humid tropical climates with a distinct dry season. Bauxites (aluminium-rich laterites) often occur in association with laterites. Some, like those of the Cretaceous of the Tethyan region, occur on karstic surfaces cut into carbonate platforms (*see* **Sedimentary Processes:** Karst and Palaeokarst). In such cases the detrital materials are likely to have been brought into the system, probably by aeolian activity.

Clay minerals in oceanic sediments reflect the weathering processes in the adjacent continental areas and the climatically controlled distribution of soils. Differential settling causes some depositional segregation of clay species (e.g. near-shore accumulation of kaolinite); nonetheless, recognizable patterns in the ancient can be preserved (*see* **Clay Minerals**).

Aeolianites

Although coastal dune complexes may form across a range of latitudes, major sand seas occur predominantly in arid environments. Such aeolianites provide both local and regional palaeowind data, which may be used to test palaeoclimate models generated by GCMs (see below). (*see* **Sedimentary Processes:** Aeolian Processes)

Palaeoclimate Models

A recent approach to understanding past climate regimes on Earth has been through the application of complex computer models, specifically atmospheric general circulation models (AGCMs), ocean general circulation models (OGCMs), and recently even more complex coupled ocean–atmosphere general circulation models (OAGCMs). There are now many contributions in this field, the palaeoclimatic

approach having been pioneered by John Kutzbach and Eric Barron.

General Circulation Models

GCMs use the laws of physics and an understanding of past geography to simulate climatic responses. They are objective in character. They require powerful computers to handle vast numbers of calculations. Nevertheless, it is now possible to compare the results of different GCMs for a range of times and over a wide range of parameterizations for the past, present, and future (e.g. in terms of predictions of surface air temperature, surface moisture, precipitation, etc.). GCMs are currently producing simulated climate predictions for the Mesozoic and Cenozoic that compare favourably with the distributions of the geological climate proxies discussed above. They can be used effectively to predict sites of oceanic upwelling and the distribution of petroleum source rocks and phosphorites. Models also produce evaluations of parameters that do not leave a geological record (e.g. cloud cover, snow cover) and quasi-parametric phenomena such as storminess. Parameterization is the main weakness of GCMs (e.g. palaeogeography, palaeobathymetry, sea-surface temperature, orography, cloud behaviour), and model output for continental interiors is still colder in winter than indicated by palaeontological data. The sedimentary and palaeontological record provides an important way of evaluating GCMs, and this is important because the same GCMs are currently being used to predict possible changes in future climate.

The outputs discussed below were generated by an AGCM (HadAM3) and an OAGCM (HadCM3L), which is currently state-of-the-art. The model was developed at the Hadley Centre for Climate Prediction and Research, which is part of the UK Meteorological Office. The GCM consists of a linked atmospheric model, ocean model, and sea-ice model. The horizontal resolution of the atmospheric model is 2.5° latitude and 3.75° longitude. This provides a grid spacing at the equator of 278 km north–south and 417 km east–west. The atmospheric model consists of 19 layers. It also includes a radiation scheme that can represent the effects of minor trace gases. Its land-surface scheme includes a representation of the freezing and melting of soil moisture. The representation of evaporation includes the dependence of stomatal resistance on temperature, vapour pressure, and carbon dioxide concentration. There is an adiabatic diffusion scheme, to simulate the horizontal mixing of tracers.

The ocean model has the same spatial resolution as the atmosphere model and 20 vertical layers, with a time step of 30 min. This contrasts with HadCM3 (the standard version of the Hadley centre OAGCM), which uses a horizontal resolution of $1.25° \times 1.25°$.

Palaeoclimate of the Mesozoic – Model Output and Geological Data

The Mesozoic Earth was an alien world, as illustrated here by reference to a Triassic GCM simulation and geological data. Throughout the Mesozoic dense forests grew close to both poles, experiencing months of continual daylight in warm summers and months of continual darkness in cold snowy winters. Neither Triassic nor Jurassic oceanic sediments provide good evidence, but in the Late Cretaceous, from ODP (Ocean Drilling Program)/DSDP (Deep Sea Drilling Program) data, the ocean depths appear to have been warm (8°C or more at the ocean floor), and reefs grew 10° further north and south than at the present time. During the era the whole Earth was warmer than now by at least 6 °C, giving more atmospheric humidity and a greatly enhanced hydrological cycle. However, from modelling studies, it seems that much of the rainfall could have been predominantly convective in character, often focused over the oceans. The model output might help to explain geological data suggesting that major desert expanses extended across the continents in low latitudes. From the model, polar ice sheets are unlikely to have been present because of the high summer temperatures. The model suggests the possibility of extensive sea ice in the nearly enclosed Arctic seaway through parts of the year, but there is as yet no proxy data against which such predictions may be tested. The Triassic world was a predominantly warm world; the model outputs for evaporation and precipitation conform well to the known distributions of evaporites, calcretes, and other climatically sensitive facies.

Triassic: Comparison of Model and Proxy Data – A Case Study

Modelled temperatures (Figure 2) A significant feature of the Triassic Earth is that the landmasses were almost symmetrically distributed in a broad arc about the equator (*see* **Mesozoic:** Triassic). A major aspect of the modelled Earth is its overall warmth. Despite temperatures plunging to −20°C and below over Siberia in the northern-hemisphere winter and to similarly low values over southernmost Gondwana in the southern-hemisphere winter, the annual average temperature is subdued in these high-latitude areas because of the high summer values achieved there (*ca.* 24°C). These high summer values preclude the possibility of year-round ice and snow.

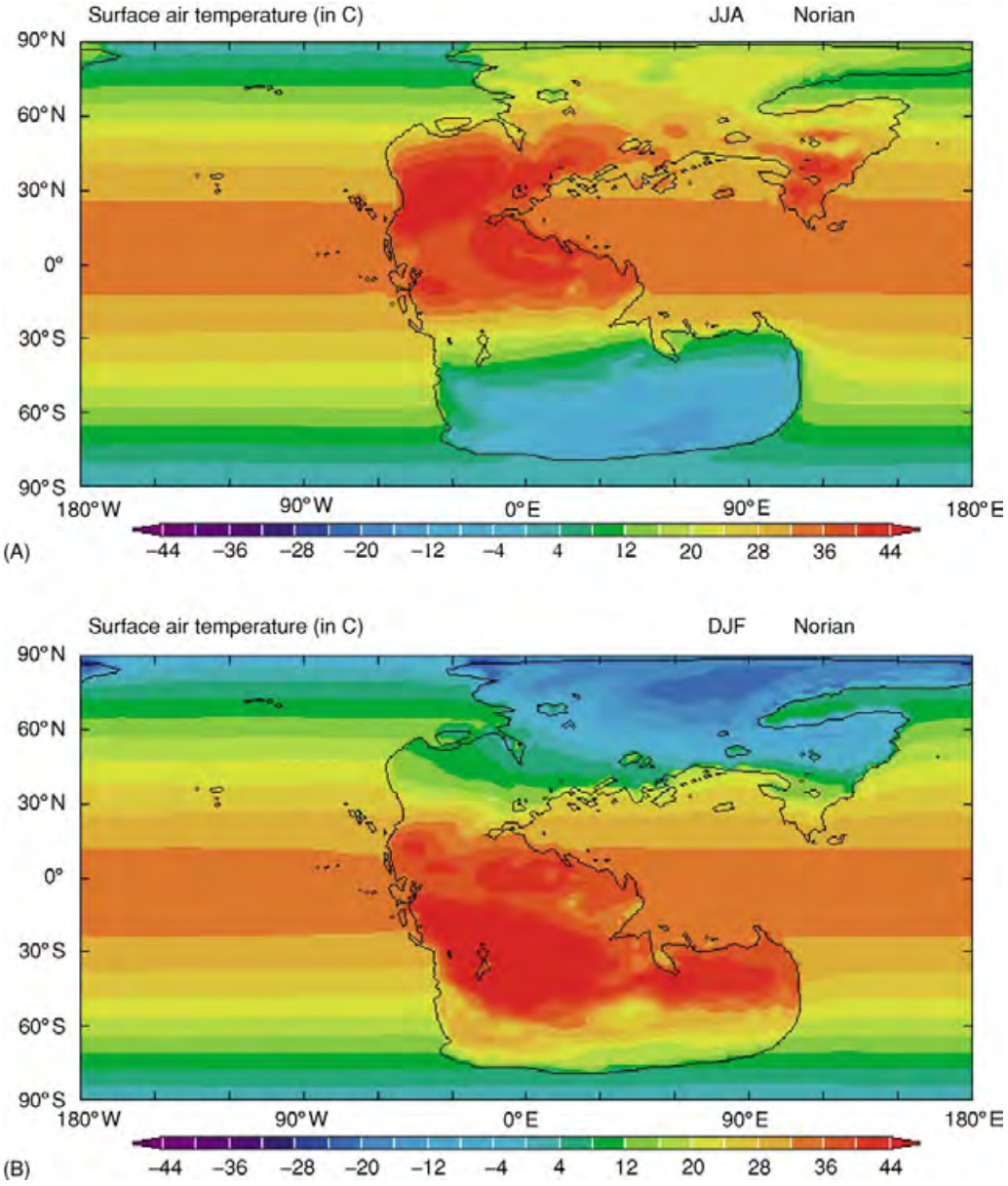


Figure 2 Model-simulated mean seasonal temperatures for the Norian for (A) December–February and (B) June–August. Units are °C and the contour interval is every 4°C. The ocean sea-surface temperatures were prescribed. JJA, June, July, August; DJF, December, January, February.

Temperature-limited facies According to the model, low temperatures would not have been a significant barrier to coral reef development around the Tethys. Reef systems developed and diversified through the Triassic, extending their range from between 2°S and 25°N in the mid-Triassic to a maximum range of 35°S to 33°N in the Late Triassic. Range limits are very close to the 20°C isotherm in the model.

Precipitation (Figure 3) Large tracts of Pangaea between about 40°N and 40°S are modelled to receive very little rainfall. Much of the rainfall is over the oceans, being convective in character, with the main zone of rainfall migrating north and south through the year with the movement of Intertropical convergence zone (ITCZ). The equatorial lands surrounding eastern and southern Tethys are modelled to receive relatively little rainfall during December–April. North-eastern Gondwana is modelled to have a monsoonal climate (winter wet). Northern and eastern Siberia have some rainfall throughout the year, the wettest period generally being through June and the

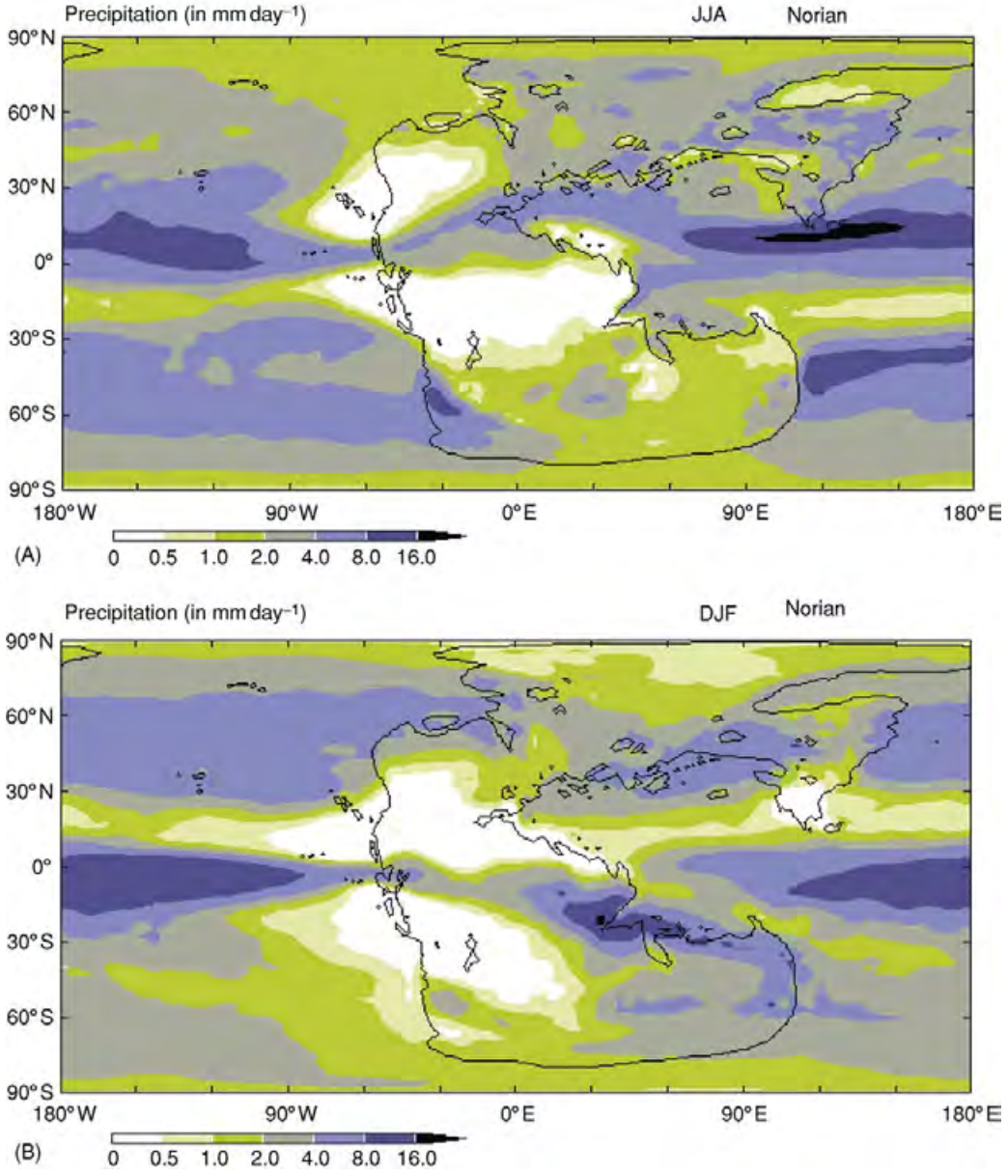


Figure 3 Model-simulated mean seasonal precipitation for the Norian for (A) December–February and (B) June–August. Units are in mm day^{-1} and the contour interval is not regular. JJA, June, July, August; DJF, December, January, February.

driest during the summer months of July and August (also equivalent to a winter-wet biome). South-western Gondwana is also modelled to be winter wet; the eastern parts of Gondwana are moister than the western parts. Nonetheless, large tracts of Pangaea, particularly in the tropics but extending to nearly 50°N and 50°S, have a large excess of evaporation throughout much of the year, as does a large part of western Tethys. Southernmost Gondwana is also winter wet, including the southern polar area during its months of darkness.

In the Triassic (Figure 4), the following climate zones have yet to be recognized from geological data: ice; polar; cold temperate; and cold temperate arid (steppe/desert). Tillites have not been reported. An absence of both ice caps and several of these biomes would be expected from the model simulations, but with some caveats as the biomes are modelled by reference to modern equivalents (e.g. cold temperate).

In the Triassic the highest-latitude floras contain a lycopod that has a worldwide distribution, reflecting a global climate largely devoid of frost. From

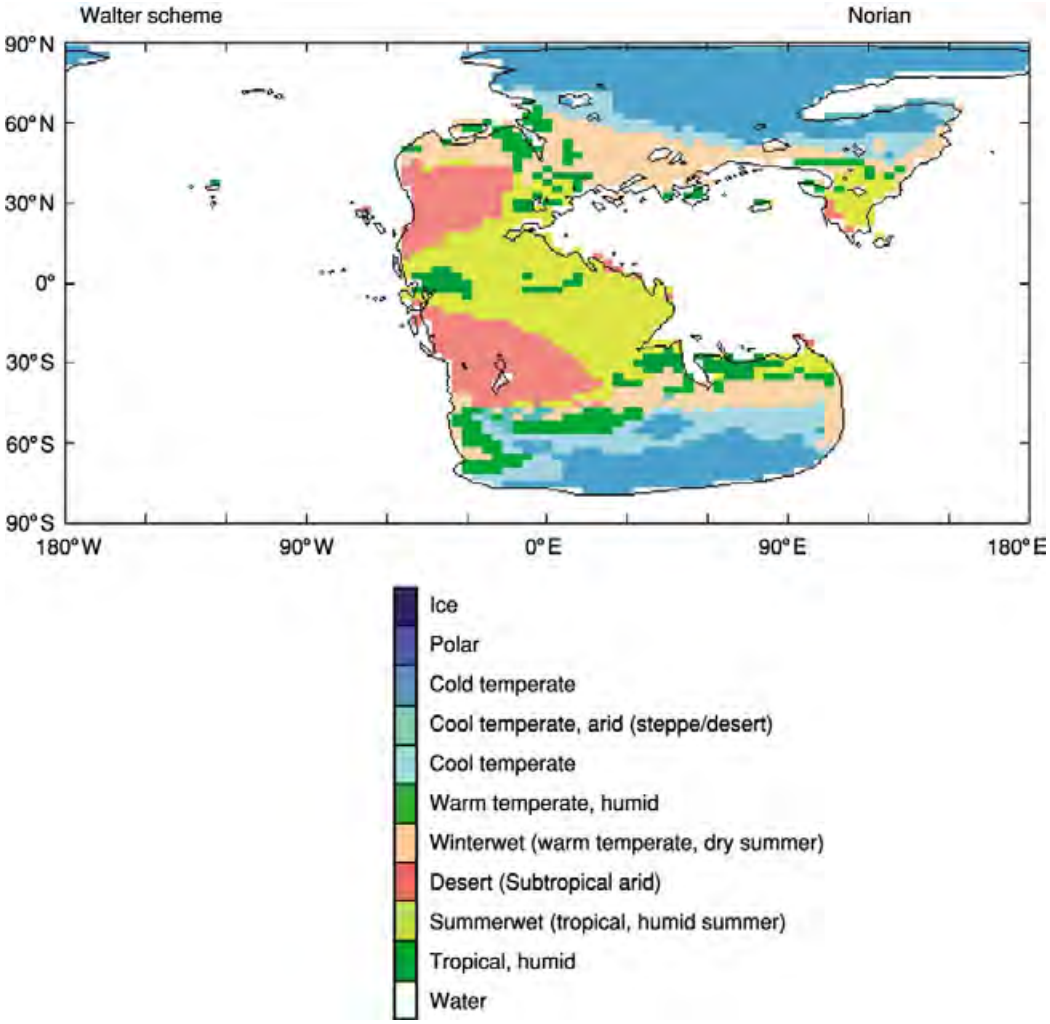


Figure 4 The Walter biome zones for the Triassic based on the model-predicted temperatures and precipitation.

geological proxy data the interior of Pangaea was hot and dry, with warm temperate climates extending to the poles. In Siberia the flora was deciduous and displays well-defined annual growth bands; the climate is interpreted to be cool temperate, with cool winters and warm summers and a growing season of four to eight months. For this biome to extend to 75° in both hemispheres, as is observed, the poles must have been far warmer than at present. This is consistent with modelled output. The distribution of extensive calcretes throughout Europe and the western Tethyan margin is in agreement with the modelled climate: hot and predominantly arid but with a short wet season.

Relatively low-diversity communities of conifers were confined to lower latitudes and are often found in the proximity of evaporites (usually found in both hemispheres between about 10° and 40° latitude). They thus represent the subtropical dry environment, and their distributions are closely in line with model predictions in terms of the ratio between precipitation and evaporation.

On the basis of facies and palaeobotanical data, warm temperate climates existed between 5° and the equator, with high-diversity floras of ferns, cycads, and seed ferns, as well as conifers and ginkgoes. The cool temperate zone extended to central Siberia and is associated with the postulated uplands of the Central Siberian Volcanic Traps.

During the Middle Triassic, abundant evaporites and calcretes were deposited in Europe, whereas in northern and southern China warm temperate and cool temperate floras predominated. These observations are fully in agreement with model output. Warm temperate floras extended to about 70° N in north-eastern Siberia, and the first northern-hemisphere coal swamps since the end-Permian appear in this area. Southern-hemisphere coal swamps are found in

South America and Antarctica, again compatible with modelled zones where precipitation exceeded evaporation for much of the year. Biomes and facies suggest that the equatorial zone is a predominantly arid belt, again lending support to the validity of the model. Exquisitely preserved fossil wood from the Middle Triassic of Antarctica (70–75° S in the Triassic) indicates forests with growth rates that were one to two orders of magnitude greater than those seen at such latitudes today.

In the Late Triassic the warm temperate biome was dominant between 30° and 50° of latitude in both hemispheres. The model generally supports this scenario, but lack of precipitation typifies many regions between these latitudes, especially away from the coastal zones.

Thus, the model output shows a good correlation with globally derived data, particularly for terrestrial plant biomes, terrigenous proxy facies, and the distribution of reefs. Terrestrial reptile distributions are difficult to relate directly to palaeoclimate, but can be interpreted to be broadly in line with model output. This illustrates the sort of model–proxy correlation that is currently possible.

Conclusions

Model simulations for the Mesozoic throw important light on climatic processes and the behaviour of Earth systems under greenhouse conditions. Model output generally compares favourably with climate proxy data, but discrepancies between data and models reveal some serious modelling problems (or possibly incorrect interpretation of the proxies). In continental interiors, model simulations generate winter conditions that are considered to be too cold (by 15°C or more) according to the interpretation of some palaeontological data.

Ocean temperatures could have been much warmer than today, not only at the surface but also at great depths. Such temperatures would have greatly reduced the oceanic storage capacity of both carbon dioxide (in ocean waters) and methane hydrates (in ocean-floor sediments) relative to the present day. Model output is in close agreement with proxy climate data in coastal and open-sea areas.

In its 'greenhouse' mode the Earth's palaeoclimate has greatly enhanced evaporation and precipitation, by comparison with the present, but much of the rainfall is convective in character, falling over the oceans. To replicate climatic conditions similar to those indicated by proxy facies throughout the Mesozoic a very large increase in atmospheric carbon dioxide is required (at least four times the present-day values).

Through the Phanerozoic, the Earth's climate has changed significantly, both on a variety of time-scales and over a range of climatic states, usually baldly referred to as 'greenhouse' and 'icehouse', although these terms disguise more subtle states between these extremes.

See Also

Atmosphere Evolution. **Clay Minerals**. **Earth:** Orbital Variation (Including Milankovitch Cycles). **Gondwanaland and Gondwana**. **Mesozoic:** Triassic. **Palaeozoic:** Ordovician; Carboniferous. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Reefs ('Build-Ups'). **Sedimentary Processes:** Aeolian Processes; Glaciers; Karst and Palaeokarst. **Sedimentary Rocks:** Evaporites. **Soils:** Palaeosols. **Solar System:** The Sun. **Tertiary To Present:** Pleistocene and The Ice Age.

Further Reading

Allen JRL, Hoskins BJ, Sellwood BW, Spicer RS, and Valdes PJ (eds.) (1994) *Palaeoclimates and Their Modelling: With Special Reference to the Mesozoic Era*. London: Chapman and Hall.

Barron EJ (1987) Eocene equator-to-pole surface ocean temperatures: a significant climate problem. *Palaeoclimatology, Palaeoecology, Palaeogeography* 2: 729–740.

Billups K and Schrag DP (2002) Paleotemperatures and ice volume of the past 27 Myr revisited with paired Mg/Ca and $^{18}O/^{16}O$ measurements from benthic foraminifera. *Paleoceanography* 17:10.1029/2000PA000567.

Crowley TJ and North GR (1991) *Palaeoclimatology*. Oxford: Oxford University Press.

D'Hondt S and Arthur MA (2002) Deep water in the late Maastrichtian ocean. *Paleoceanography* 17:10.1029/1999PA000486.

Frakes LA, Francis JE, and Syktus JI (1992) *Climate Modes of the Phanerozoic*. Cambridge: Cambridge University Press.

Haywood AM, Valdes PJ, Sellwood BW, and Kaplan JO (2002) Antarctic climate during the middle Pliocene: model sensitivity to ice sheet variation. *Palaeogeography, Palaeoclimatology, Palaeoecology* 182: 93–115.

Huber BT, MacLeod KG, and Wing ST (eds.) (2000) *Warm Climates in Earth History*. Cambridge: Cambridge University Press.

Kiessling W, Flügel E, and Golonka J (eds.) (2002) *Phanerozoic Reef Patterns*. SEPM Special Publication 40072. Tulsa: Society for Sedimentary Geology.

Kutzbach JE and Gallimore RG (1989) Pangean climates – megamonsoons of the megacontinent. *Journal of Geophysical Research – Atmospheres* 94: 3341–3357.

Markwick PJ (1998) Fossil crocodilians as indicators of Late Cretaceous and Cenozoic climates: implications for using palaeontological data in reconstructing palaeoclimate. *Palaeogeography, Palaeoclimatology, Palaeoecology* 137: 205–271.

Price GD, Sellwood BW, and Valdes PJ (1995) Sedimentological evaluation of general circulation model simulations for the 'greenhouse' Earth: Cretaceous and Jurassic case studies. *Sedimentary Geology* 100: 159–180.

Ruddiman WF (2000) *Earth's Climate: Past and Future*. New York: WH Freeman and Co.

Scotese CR (2000) Paleomap Project. http://www.scotese.com.

Valdes PJ, Spicer RA, Sellwood BW, and Palmer DC (1999) *Understanding Past Climates: Modelling Ancient Weather*. CD-ROM. OPA (Overseas Publishers Association), Amsterdam.

PALAEOECOLOGY

E M Harper, University of Cambridge, Cambridge, UK
S Rigby, University of Edinburgh, Edinburgh, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Ecology is the scientific study of the way in which modern organisms interact with their environment. In practice, ecologists study a very wide range of questions, but there are two main branches of study, autecology and synecology. Autecology is concerned with the ecology of an individual taxon, such as a cheetah, blue mussel, or vestimentiferan worm. Questions tackled might concern mode of life, feeding and breeding habits, and physical and chemical factors that limit distribution. Synecology examines the complex interactions between populations of a single species or between communities of organisms; examples of synecological studies might concern the relationships between a variety of big cats and their prey on the plains of Africa, the organisms associated with a mussel patch on a rocky shore, or the chemosymbiotic relationships within deep-sea vent communities, examining topics such as trophic structure and diversity. Palaeoecologists seek to gain similar understandings about fossilized organisms and ancient communities.

Palaeoecological studies may have relevance outside the sphere of palaeobiology. Organisms are not randomly distributed and are adapted to particular sets of environmental conditions (e.g., salinity, light levels, substrate consistency, and temperature). Once it has been established that a particular fossil taxon had a particular set of habitat requirements, the presence or absence of that taxon may be used in a broader geological context as a palaeoenvironmental indicator. For example, stenohaline organisms such as echinoderms and brachiopods, which have no ability to excrete water gained by osmosis, are limited to normal marine salinity, thus the presence of the body fossils of these higher taxa has been widely used as an indicator of prevailing normal salinity. For other groups, a detailed knowledge of the palaeoecologies of their component genera or species has been invaluable in finely subdividing palaeoenvironments. For example, the relative proportions of milioline, rotaliine, and textularine foraminifera in sediments give important information concerning the salinity of a depositional environment from brackish to hypersalinity. Brackish waters are dominated by rotaliinids and textularinids, whereas miliolinids and rotaliinids make up the dominant species in areas of normal or elevated salinity.

In terrestrial sediments, details of the palaeoecology of plants, such as the relationship between leaf shape and temperature and between concentration of stomata and atmospheric carbon dioxide levels, may be used in palaeoenvironmental reconstruction. Most recently, there has been an added interest in palaeoecology because the study of how the ecologies of different organismal groups and communities have altered during periods of dramatic environmental change (for example, during glacial periods or after mass extinctions) may give useful clues to how life on Earth may react to future climatic perturbations linked to human activities.

Palaeoautecology

The palaeoautecology of an extinct organism may be taken broadly to equate to its mode of life. Of course, there are aspects of fossil palaeoautecologies for which there is very little hope of being able to reconstruct; for example, it is difficult to imagine that anything will ever be known of mating dances in extinct flies, but a remarkable amount can be gleaned. Arguments made from homology with living relatives and analogy from organisms with apparently similar adaptations, from trace fossils or from models of fossils, can be used. It is almost invariably preferable to use a combination of approaches in order to allow cross-referencing.

When fossil organisms belong to taxonomic groups that still have extant representatives, much can be learned by applying knowledge about living taxa, using principles of homology. Clearly this is likely to

be most successful in groups which are most closely related, and as this relationship is stretched, any conclusions that extant and fossil groups may have lived in the same way are tenuous. Interpretation of the modes of life of most fossil bivalve molluscs is relatively straightforward. Several features of bivalves make these organisms ideal for palaeoecological reconstruction. First, modern bivalves are more diverse, both in terms of taxonomic units and in life habits, than ever before in their 500-million-year evolutionary history, meaning that interpretations can be based on a wide range of possible models. Second, a wide variety (for example, oysters, scallops, mussels, and clams) are of great commercial interest and consequently their biologies have been studied intensively. Third, it is well demonstrated that bivalve shell morphology and life habit are intimately related (*see* **Fossil Invertebrates:** Bivalves).

An example of this methodology is illustrated in Figure 1, which shows the right valve of *Eopecten*, a large bivalve known from Middle and Upper Jurassic hardground faunas of Normandy (France) and Portugal. Several aspects of its morphology can be used to interpret the palaeoecology of this animal, based on comparison with a living bivalve of the same family. The oldest part of the shell, close to the umbo, is a regular scallop shape and shows the morphological features associated with attachment to a hard substrate by byssal threads spun by the foot. The surrounding section (later growth) of the shell is highly irregular and its form varies among individuals, because at this size the bivalve began to cement to the seafloor and modified its shape to suit the substrate. Experimental evidence has shown that cementation in bivalves is a good defence against predators.

Clearly some caution is required in using modern organisms to interpret the palaeoecology of extinct fossils, even when they are closely related. Some organisms have undergone a major shift in aspects of their life habits over their evolutionary history. For

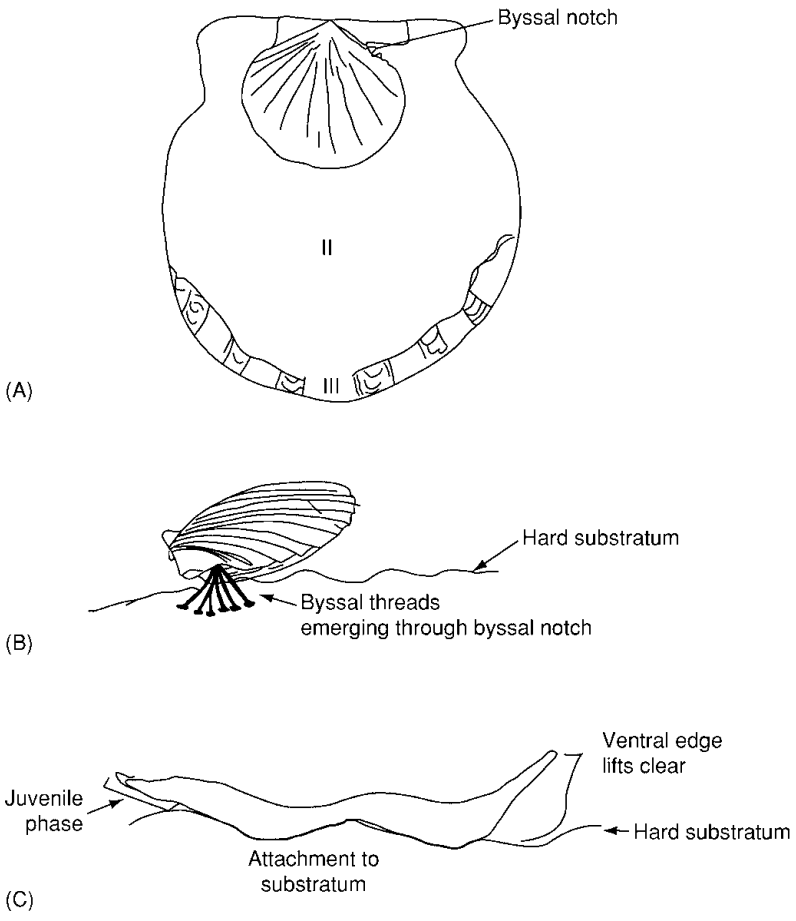


Figure 1 Reconstruction of life habits of Jurassic *Eopecten*. (A) External view of the right valve. I, Juvenile phase (the byssal notch is later filled with shell material); II, later growth (attachment scar, no ribbed ornament); III, valve lifts clear of substrate, coarse-ribbed ornament. (B) Juvenile stage byssally attached to hard substrate. (C) Cross-section through a mature animal, showing extensive cementation scar and raising of ventral margins above substrate.

example, monoplacophorans, a primitive caplike molluscan class (*see* **Fossil Invertebrates:** Molluscs Overview), are known from a variety of shallow marine sediments from the Lower Palaeozoic, apparently living in communities alongside brachiopods and trilobites. They then apparently disappear from the fossil record during the Devonian and were presumed extinct until the celebrated rediscovery of *Neopilina* during dredging of deep water (>3500 m) off the coast of Mexico during the Danish-led Galathea expedition in 1952. However, it would clearly be inappropriate to reconstruct the palaeoecology of the Palaeozoic forms based solely on our understanding of the biology of *Neopilina*. Presumably during the Upper Palaeozoic, monoplacophorans were forced, perhaps by competition or predation pressure, into deep-water refuges that are not represented in the sedimentary record. Other Palaeozoic taxa that were important and widespread in shelf communities, such as brachiopods and stalked crinoids, are postulated by some also to have been driven into refugia. These examples clearly demonstrate the need to use several lines of evidence when reconstructing palaeoecology and, most importantly, not to divorce the study of fossils from their lithological context.

What happens when there are no convenient living homologues? In such cases, the palaeoecologists must use more imaginative approaches. Trilobites (*see* **Fossil Invertebrates:** Trilobites) became extinct by the close of Palaeozoic, although much may be gleaned of their palaeobiology by homology and by analogy with other arthropod groups (such as understanding their moulting and likely internal anatomy). It is believed that trilobites exploited a wide range of life habits; although most were benthic, others were pelagic or nektonic. Some (e.g., *Calymene*) are interpreted as having been voracious predators, whereas others were thought to have been filter feeders or deposit feeders. The lines of evidence used to elucidate these different palaeoecological details are varied. Much is reliant on detailed studies of functional morphology, but evidence from trace fossils (including some that show evidence of trilobites tracking worms) and experimental modelling have also played a role. An interesting example is the interpretation of the Lower Palaeozoic olenellids as possible chemosymbionts in an elegant study by Richard Fortey. Using a combination of evidence based on the functional morphology of the family, principally the large number and wide nature of the thoracic segments and degenerate hypostome, the lithological evidence that olenellids tended to inhabit sulphur-rich unbioturbated black shales, and by analogy with totally unrelated modern chemosymbiotic organisms such as lucinid bivalves, Fortey suggested that these trilobites might well have harboured sulphate-reducing bacteria on their gills, which allowed them to live in 'difficult' oxygen-poor environments.

When living relatives and even good analogues are absent, more unconventional approaches can be taken in the study of the fossil mode of life. Graptoloids (*see* **Fossil Invertebrates:** Graptolites) were a significant group of macrozooplankton living in the Ordovician to mid-Devonian open seas. They have very remote living relatives, the pterobranchs, but their benthic, encrusting mode of life makes direct comparisons unrealistic. Instead, work has concentrated on the response of graptoloid shapes to seawater, which had the same properties in the Lower Palaeozoic as it does now. Physical and computer-derived models show that both the overall colony shape and its detailed morphology would have had important consequences for the behaviour of the species in water. Colonies were designed to rotate during movement, thereby increasing the volume of water that could be sampled. Flow of water was directed into the thecal apertures where the zooids lived by an arrangement of spines, hoods, or hooks in many species. A standard orientation relative to the dominant direction of water flow was maintained by structures such as the virgella and virgula, arrayed at the proximal and distal ends of the colony (**Figure 2**). Understanding the physical effects of adaptations shown by fossils allows inferences about life habits to be made (for example, concerning the likely feeding position of the zooids and the sampling requirements that conditioned their fitness and hence evolution).

Palaeosynecology

Ecologists define a community as a recurring group of taxa (in fixed proportions) that live together in a particular habitat. The focus of interest is in the way in which populations of single species or different taxa interact and the manner in which resources within that habitat are shared between the different groups. Ecological studies can be well founded on quantitative analyses based on direct observations. Palaeoecologists, however, are faced with a number of problems and must use less direct methods. First, a high proportion of any community may be soft-bodied or weakly skeletonized such that they are unlikely to be preserved in the fossil record. It is estimated that more 70% of modern taxa fall into this category and the Middle Cambrian Burgess Shale lagerstätte (*see* **Lagerstätten**) provides a salutary lesson, showing the wealth of poorly or unskeletonized organisms (such as worms and arthropods) not usually preserved in contemporary deposits. Second, relatively few organisms are preserved in life position and many may have been

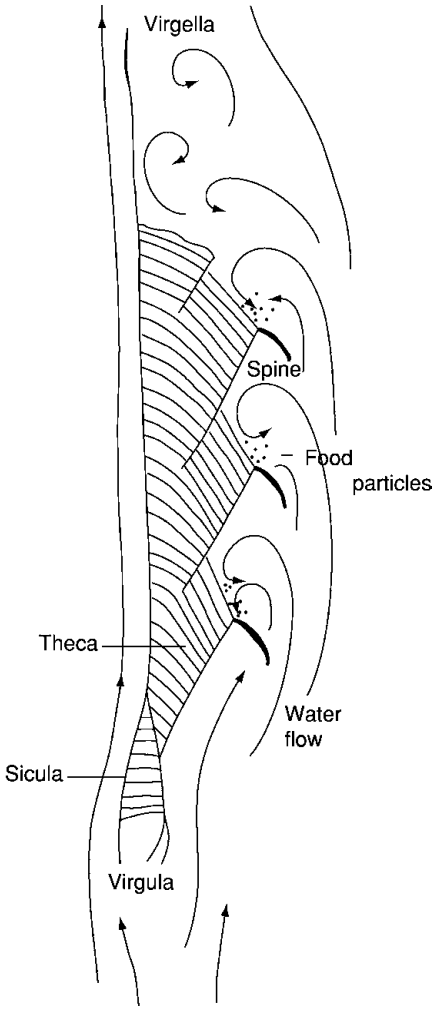


Figure 2 Sketch of flow patterns of water over a monograptid graptoloid model. The virgula and virgella maintain the colony in a stable orientation relative to the main current. Spines at the thecal openings produce local turbulence and direct flow over the apertures, concentrating food particles conveniently for animals living in the thecae.

transported far from their original habitats. As a result, the palaeoecologist must consider whether the group of fossils being studied is autochthonous (i.e., preserved in life position), parautochthonous (i.e., elements are disturbed and disarticulated, but the fossils are not necessarily far removed from their origins), or allochthonous (i.e., organisms from different habitats have been swept together) before drawing any conclusions. Third, even if a group of fossils is considered autochthonous, it has to be recognized that any bed, or even bedding plane assemblage, is going to be time averaged, and it will thus contain a sample of individuals that lived on that site over many centuries or even thousands of years. In this way, individuals of the same species but of very different generations may be found next to one another, or taxa that lived in successive communities on that site may appear to have co-existed. As a result of these complications, palaeoecologists prefer to use the term 'assemblage' rather than 'community'. In addition, the soft-bodied components of the original communities are seldom found fossilized. If similar assemblages are found to occur widely, either geographically or temporally, this provides a good indication that they are a real ecological phenomenon, and these are given the name 'associations'.

Investigating Fossil Populations

Populations of organisms are sensitive to characteristics of the environment and can be studied in two main ways, by the examination of life strategy and through survivorship analysis (Figure 3). Two extreme life strategies, commonly termed r and K, are derived from the logistic equation that states that the number of individuals (N) in a population will be a function of reproductive rate (r) modified by closeness of the number of individuals to the total number that can be supported by an environment (the carrying capacity, or K):

$$dN/dt = rN[(K - N)/K]$$

The r-strategist species depend heavily on high rates of reproduction, often breed early (at small size), and may die after breeding. They generally invest little in the care of their offspring, but have many offspring at once. The strategy is characteristic of unstable environments in which the carrying capacity varies significantly over short time-scales. The K-strategist species breed late, often bearing single young over whom care is expended. Individuals often live for a considerable time, even after breeding has ceased. This strategy is typical of stable environments in which the carrying capacity is highly predictable. Although these two descriptions are extremes, and most organisms show a mixture of r and K features, the generalizations have great use in making first-level interpretations of population dynamics.

Survivorship analysis is a complementary tool for use in the examination of fossil populations. The number of survivors of each age group of a population is plotted on a logarithmic scale as a function of age. In fossil populations, the number of survivors can be inferred if the bedding plane is time averaged, or can be observed if mass mortality of a single group produced the association. Most survivorship curves fall into three distinct categories, though variations exist, and organisms with very different life habits at different stages of their lives may show composite patterns. In cases in which high infant mortality occurs, as might be seen, for example, in fish stocks,

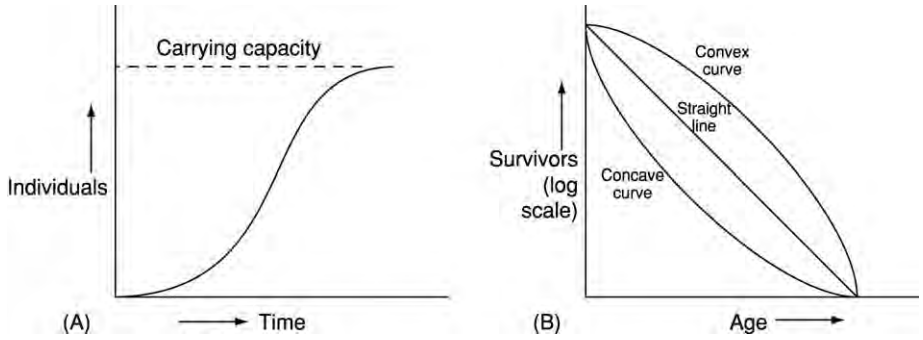


Figure 3 Key graphs for the interpretation of fossil populations. (A) A plot of the logistic equation, showing the rate of population increase slowing as the population size approaches the carrying capacity of the local environment. (B) Survivorship curves for populations with a high juvenile mortality (concave curve), an equal chance of mortality at any age (straight line), and a high gerontic mortality (convex curve).

the survivorship curve will be concave. In environments in which there is an equal risk of mortality at any age, the survivorship curve will be straight. The classic example of this type of curve is that of garden birds. Palaeontological examples are found, especially in extreme environments in which mortality is generated by frequent, unpredictable, and unselective events. In more stable environments, mortality rate is low initially but increases with age after a certain point, producing a convex survivorship curve. In these cases, organisms can be fit enough to survive the variation seen in their environment, but this characteristic is reduced later in life through illness or gerontic effects.

Investigating Interactions between Organisms

Organisms may interact in a number of different ways. Most important among these are competition for resources (e.g., for nutrients, space, light, or mates), either intra- or interspecifically; predator–prey interactions; and symbiosis, which runs through a spectrum, from being beneficial to both partners (commensalism) to parasitism, in which one partner's advantage is deleterious to the other. Ecologists have found that these interactions play an important role in controlling the fitness of modern organisms, and competition and predation in particular are thought to be extremely important driving forces in evolution. At first sight it might seem that the fossil record is likely to be virtually mute on the interactions between organisms; however, this is not the case. Clearly, individual cases must be carefully assessed, but it is possible to generate quantitative data on interactions between organisms from the fossil record.

Competition

The competitive exclusion principle suggests that two species that have exactly the same requirements cannot exist together within a habitat. Where two species do appear to have the same requirements, one will be prevented from maintaining a viable population by the success of the other, or one will be forced to adapt in order to use slightly different resources. The latter strategy may involve character displacement (e.g., modification of filter-feeding apparatus to cope with particles of a different size).

Competition has often been invoked as a cause of macroevolutionary change, although it is difficult to demonstrate. For example, some have suggested that the post-Palaeozoic decline in articulated brachiopods was at least in part influenced by the superior competitive advantages of bivalves that may occupy the same niche, and a similar argument has been made to link the success of dinosaurs with the decline of synapsids towards the end of the Triassic. One situation in which it has been possible to collect high-quality data concerns interactions between borers and skeletonized encrusters (e.g., bryozoans, oyster-like bivalves, and serpulid worms). These organisms have the advantage of having been preserved *in situ* and thus the fossils maintain their spatial relationships to one another and to the substratum. It must be remembered that soft-bodied encrusters (e.g., anemones, sponges) that may also have co-existed will be unrepresented, although evidence of their existence may be revealed by bioimmuration (preservation by the overgrowth of a skeletonized organism). In these cases, it is also possible to examine competitive overgrowth. It is known that space on hard substrates is at a premium in modern marine habitats, from which it may be inferred that taxa that are able rapidly to colonize space will be at an advantage over those that do so less aggressively. Much of this type of research has been centred on two clades of extant bryozoans, the cyclostomes and cheilostomes (*see* **Fossil Invertebrates:** Bryozoans). Both evolved independently from soft-bodied ctenostome ancestors, the cyclostomes in

the Ordovician and the cheilostomes in the Cretaceous. The cheilostomes replaced the cyclostomes as the dominant bryozoan clade and it has been argued that the former were more successful competitors. An analysis of competitive overgrowths between the two groups from a range of localities (Cretaceous to Recent) by McKinney has revealed that cheilostomes consistently 'won' around 66% of all interactions between the two clades, although interestingly this figure appears to have remained constant over geological time.

Predation

All organisms die, and a great many deaths are the result of being eaten by predators, becoming part of the food chain. The need to avoid predation is a very important selection pressure, hence the battery of defensive morphologies (e.g., spines and thick shells), behaviours (such as leaping by cockles or swimming by scallops), and life habits (such as deep burrowing) that may be demonstrated in modern organisms. Very rarely, predators may be fossilized in the act of predation, with prey remains preserved in their guts, for example. But predator–prey interactions may be interpreted from the fossil record in a number of other ways, such that palaeoecologists are not limited to exceptional preservation to chart the evolution of different predatory groups and methods. Various aspects of an organism's functional morphology, such as possession of teeth or claws, may suggest that it was a predator. For example, the large, toothed chelicerae of Palaeozoic eurypterids (large aquatic scorpions) have been used to suggest that they may have been powerful predators of the early fish. Most theropod dinosaurs (*see* **Fossil Vertebrates:** Dinosaurs), including the well-known *Tyrannosaurus rex*, have narrow, sharp-pointed curved teeth, many with serrated edges, which suggests that they were carnivores. The most useful evidence, however, has come from the study of prey remains. Although some predatory groups eat their prey without leaving any evidence of damage (for example, starfish that prise apart bivalve shells, or fish that simply swallow their prey whole), there are two important and widely used predatory methods on shelled prey, crushing and drilling, that can be recognized in the fossil record. A whole variety of arthropod and vertebrate predatory groups feed by crushing open the shells of their victims, many leaving very diagnostic breakage patterns. Similarly, a number of worms, gastropods, and octopods drill neat holes in their victims, for either the extraction of tissue or the injection of toxins (Figure 4). Where such damage is evident, it is possible to undertake quantitative surveys of the frequency of attack by a particular predatory group and the frequency of prey repair (which provides a measure of the effectiveness of their defences), and also some aspects of behaviour, such as prey preference or stereotypic (predictable and so diagnostic of the predator) attack strategy.

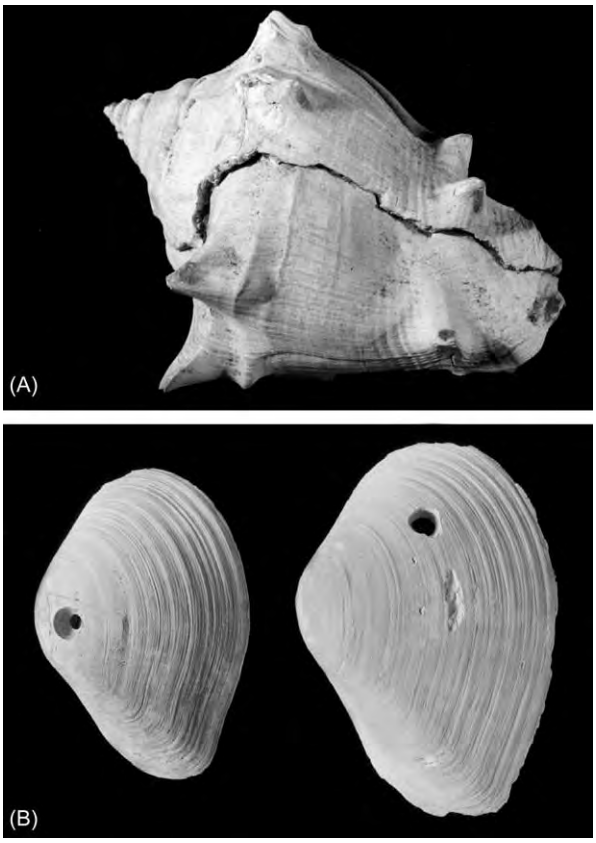


Figure 4 (A) Repaired predator-induced breakage to gastropod prey. Miocene, France (×1). (B) Predatory drill holes produced by naticid gastropods in bivalve prey. Pliocene, United States (×1.5).

There is now a strong body of evidence to suggest that levels of predation pressure have increased during the Phanerozoic. In particular, it is evident that there have been marked increases in the past 270 million years, especially during the Cretaceous and Early Tertiary, during which time the numbers of predator taxa increased and more specialized feeding techniques evolved. This is part of the so-called Mesozoic Marine Revolution, a major restructuring of marine communities following the Permian–Triassic extinctions and leading to ecosystems taking on their 'modern' aspect. It can be argued that this increase in predation pressure had a profound effect on the evolution of individual taxa and changed the ecological properties of communities. It has been effectively shown that post-Palaeozoic gastropods have become better armoured (thickening their shells,

more taxa with slit-like apertures, and fewer taxa with loose coiling) and therefore better able to resist crushing predators, such as crabs. Many bivalve taxa adopted 'new' life habits such as cementation and deep burrowing, which can be shown to have decreased the threat from predators, and stalked crinoids moved away from shallow water communities into deeper habitats where there would have been fewer predators.

Symbioses

If two different fossil taxa are frequently found together, this may be evidence of a symbiotic relationship, but it may not be easy to demonstrate where the benefits and the costs of the relationships lay. Upper Palaeozoic platyceratid gastropods are frequently preserved in close association with the calyces of a variety of crinoids, in particular in a position close to or over the anus. The margins of these snails mirror closely the morphology of the calyx, implying that the animals were in close proximity long-term and that the gastropod literally grew to fit its location. The relationship appears to have been detrimental to the crinoid (i.e., a case of parasitism), because those with platyceratids are usually smaller than those without.

Hermatypic scleractinian corals (*see* **Fossil Invertebrates:** Corals and Other Cnidaria) are involved in an intimate symbiotic relationship with zooxanthellae that live within their soft tissue. Theirs is a commensal relationship of mutual benefit to both partners: the photosynthetic zooxanthellae gain a living site free from sediment cover, held aloft in the water and protected from the stinging nematocysts of the host, and the coral uses carbon fixed by the zooxanthellae to augment dramatically their own calcification processes. It is a strong relationship and many corals cannot survive without their zooxanthellae, which explains the very stringent requirements of scleractinian corals to inhabit clear water within the photic zone.

Examples of Palaeosynecological Studies

The Middle Cambrian Burgess Shale is one of the most important lagerstätten to have been discovered. The vast collection of material from the site has allowed a detailed examination of the palaeosynecology of the deposit. In a survey of more than 65 000 specimens collected from the Phyllopod Bed of the Burgess Shale, Conway Morris was able to examine the life habits of each of the taxonomic groups represented and to make calculations of biomass. It was also possible to undertake a trophic analysis (i.e., examining the feeding methodologies of different elements of the fauna). Largely using evidence from functional morphology, as well as some instances of preserved gut contents, Morris demonstrated the presence of deposit and suspension feeders, predators, and scavengers and allowed the tentative reconstruction of a trophic web.

In another study, Fürsich, Palmer, and Goodyear examined the palaeosynecology of bivalve-dominated patch reefs from the Upper Jurassic of the southern UK. Their studies recognized a number of 'guilds', or species that exploit a particular way of life or resource. Primary framebuilders (mostly bivalves, solenoporacean algae, and bryozoans) accounted for 55–70% of the volume of the reef, but there were also accessory encrusters (other bivalves, bryozoans, serpulids, and forams); a crevice-dwelling fauna (including byssate bivalves, small terebratulid brachiopods, and worms) that occupied cavities within the reef and borings made by endoliths; borers on a variety of scales, from large borings of bivalves to microscopic borings made by a variety of taxa (including phoronids,

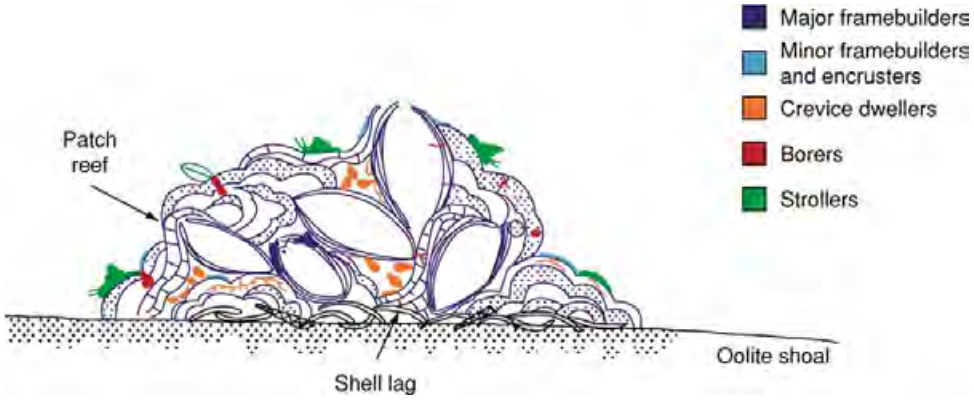


Figure 5 A reconstruction of a patch reef from the Upper Jurassic Portland Beds of South England. The key shows the different guilds of organisms, linking taxa by function within the reef, rather than by family tree.

crustaceans, and bryozoans); and a small number of vagile taxa, or strollers, largely gastropods, that roamed the surfaces of the reefs. It is likely that this latter guild is under-represented. These studies suggest that the reefs built up after initially unstable oolitic shoals had been stabilized by both infaunal and epifaunal bivalves, the shells of which then provided the firm seafloor onto which the reef-forming organisms were recruited (**Figure 5**).

See Also

Fossil Invertebrates: Trilobites; Bryozoans; Corals and Other Cnidaria; Graptolites; Molluscs Overview; Bivalves; Gastropods. **Fossil Vertebrates:** Dinosaurs. **Lagerstätten**. **Microfossils:** Foraminifera. **Palaeopathology**.

Further Reading

Brenchley PJ and Harper DAT (1998) *Palaeoecology: Ecosystems, Environments and Evolution*. London: Chapman and Hall.

Briggs DEG and Crowther PR (1990) *Palaeobiology: A Synthesis*. Oxford: Blackwell.

Briggs DEG and Crowther PR (2001) *Palaeobiology II*. Oxford: Blackwell.

Conway Morris S (1986) The community structure of the Middle Cambrian Phyllopod Bed (Burgess Shale). *Palaeontology* 29: 423–467.

Fortey RA (2000) Olenid trilobites: the oldest known chemautotrophic symbionts? *Proceedings of the National Academy of Sciences* 97: 6574–6578.

Fürsich FT, Palmer TJ, and Goodyear KL (1994) Growth and disintegration of bivalve-dominated patch reefs in the Upper Jurassic of southern England. *Palaeontology* 37: 131–171.

Gould SJ and Calloway CB (1980) Clams and brachiopods – ships that pass in the night. *Paleobiology* 6: 383–396.

Harper EM and Palmer TJ (1993) Middle Jurassic cemented pectinids and the missing right valves of *Eopecten*. *Journal of Molluscan Studies* 59: 63–72.

Kelley PH, Kowalewski M, and Hansen TA (2002) *Predator–Prey Interactions in the Fossil Record. Topics in Geobiology 20*. New York: Kluwer Academic/Plenum Publishers.

McKinney FK (1995) One hundred million years of competitive interactions between bryozoan clades: asymmetrical but not escalating. *Biological Journal of the Linnean Society* 56: 465–481.

Rollins HB and Brezinski DK (1988) Reinterpretation of crinoid-platyceratid interaction. *Lethaia* 21: 207–217.

Taylor PD and Wilson MA (2003) Palaeoecology and evolution of marine hard substrate communities. *Earth Science Reviews* 62: 1–103.

PALAEOMAGNETISM

T H Torsvik, Geological Survey of Norway, Trondheim, Norway

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Palaeomagnetism is the study of the Earth's magnetic field preserved in rocks. The discovery that some minerals, at the time of their formation, can become magnetized parallel to the Earth's magnetic field was made in the nineteenth century. Early in the twentieth century, Bernard Brunhes made the startling discovery that some rocks are magnetized in the opposite orientation to the Earth's present-day magnetic field. This led him to propose that the Earth's magnetic field had reversed its polarity in the past. These reversals have subsequently been shown to be non-periodic and the Earth's magnetic field reversal history is now well known for the past 175 million years and more sketchingly understood to the beginning of the Palaeozoic (*ca*. 545 Ma).

Palaeomagnetism has a range of application potential (*see* **Magnetostratigraphy, Analytical Methods: Geochronological Techniques**), and the focus here is on understanding the importance of palaeomagnetism as an investigative tool in assembling palaeogeographical reconstructions.

Fundamentals

The Earth's magnetic field is believed to originate from the outer fluid core and, at the surface, the field is described by its inclination (angle with respect to the local horizontal plane), declination (angle with respect to the Greenwich meridian), and field strength (**Figure 1**). The inclination of the Earth's field varies systematically with latitude, which is of prime importance for palaeomagnetic reconstructions. At the north magnetic pole, the inclination of the field is $+90^\circ$ (straight down), at the equator the field inclination is zero (horizontal) pointing north, and at the south magnetic pole the inclination is -90° (straight up; **Figure 2**). The magnetic north and south poles currently differ from the geographical north and south poles by 11.5° because the magnetic axis is inclined from the geographical (=rotation) axis. The

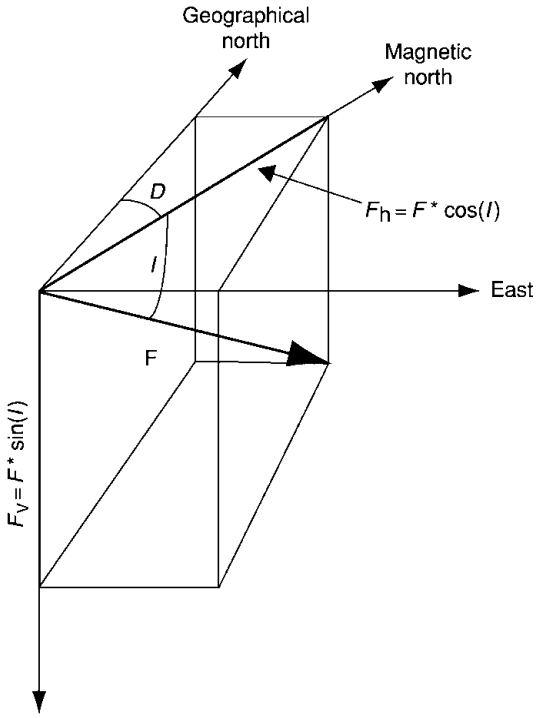


Figure 1 The direction and intensity of the total field vector (**F**) decomposed into declination from geographical north (D) and inclination from the horizontal (I). The equations relate the horizontal (F_h) and vertical (F_v) components of the total field (**F**) with I and D.

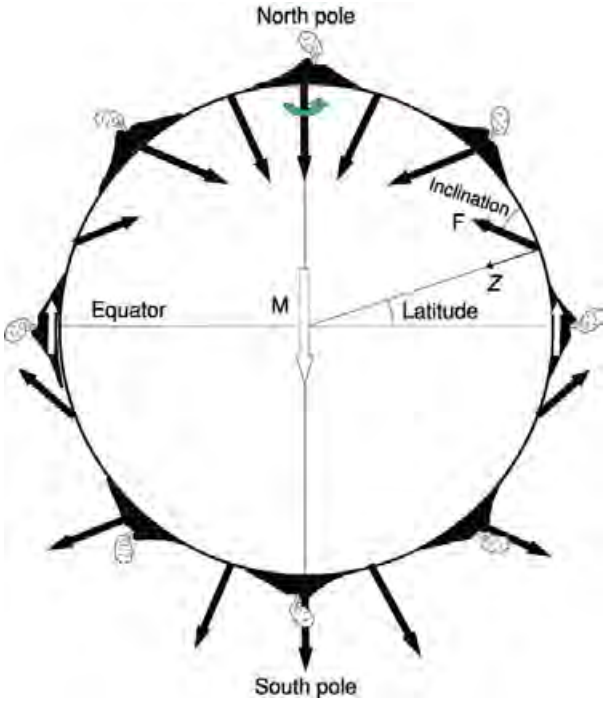


Figure 2 Field lines at the Earth's surface for a geocentric axial dipole. At the equator, the inclination is flat (zero) and, at the north and south poles, the inclination is vertical (+90° and −90°, respectively). The inclination recorded in volcanoes formed on the Earth's surface is dependent on the latitude (see also **Figure 3**). Declinations in a normal polarity field, such as today's, should point to the north.

magnetic axis, however, is slowly rotating/precessing around the geographical axis (known as secular variation) and, over a period of a few thousand years, it is hypothesized that the averaged magnetic poles correspond reasonably well with the geographical poles. This is known as the geocentric axial dipole (GAD) hypothesis. We can therefore imagine that a magnetic dipole is placed at the centre of the Earth and aligned with the Earth's rotation axis (**Figure 2**). In palaeomagnetic studies, it is therefore important to sample rocks whose ages range over more than a few thousand years; a study of a single dyke or basalt flow, for example, represents an instantaneous reading of the Earth's magnetic field (cools within a scale of days to weeks) and will not accurately record the position of the Earth's rotation axis.

When rocks form on an ideal planet, they acquire a remanent (permanent) magnetization parallel to the Earth's magnetic field at that location (**Figures 2 and 3**). There are a number of ways in which a rock can acquire a remanent magnetization, but most rocks are magnetized by one of the following processes: (1) as magma solidifies and cools below the Curie temperature (T_C, i.e., temperature above which a magnetic material loses its magnetism because of thermal agitation), magnetic minerals acquire a thermoremanent magnetization aligned with the Earth's magnetic field at the time of cooling; (2) during the deposition of sediments, magnetic mineral grains settle statistically in the direction of the Earth's magnetic field and a detrital remanent magnetization is acquired; and (3) when magnetic minerals are formed during chemical processes (diagenesis or metamorphism), the magnetic minerals grow as magnetized crystals with their magnetization in the direction of the external magnetic field; this creates a chemical remanent magnetization. Most magnetic minerals are iron–titanium oxides that belong to two solid solution series (**Figure 4**): the titanomagnetites (e.g., the end-member magnetite) and the titanohaematites (e.g., haematite). In the titanomagnetite series ($Fe_{3-x}Ti_xO_4$), there is an approximately linear variation of spontaneous magnetization ($\mathbf{M}_s$) and T_C with composition (x). Other common magnetic minerals include goethite and pyrrhotite (**Table 1**).

As an example, we can consider the remanence acquisition of a basaltic lava flow during cooling. The most important magnetic mineral in basaltic rocks is titanomagnetite. Low-titanium phases (e.g., pure

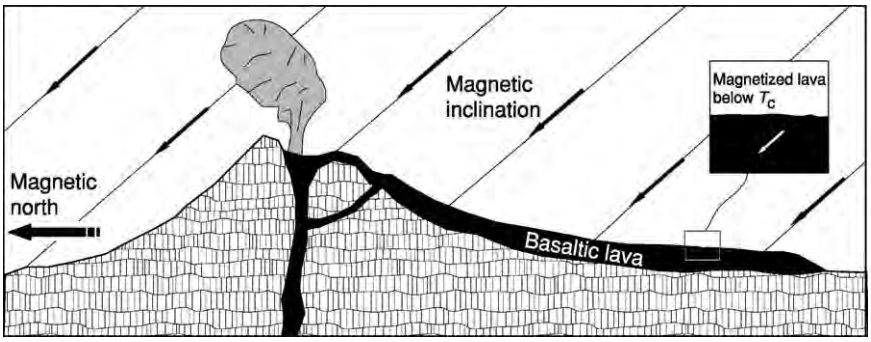


Figure 3 Example of acquisition of a thermoremanent magnetization (TRM) at intermediate northerly latitudes (acquired in a normal polarity field similar to today's). A lava will acquire a TRM upon cooling below the Curie temperature (see text), and the inclination will parallel the inclination of the external field and have declinations due north.

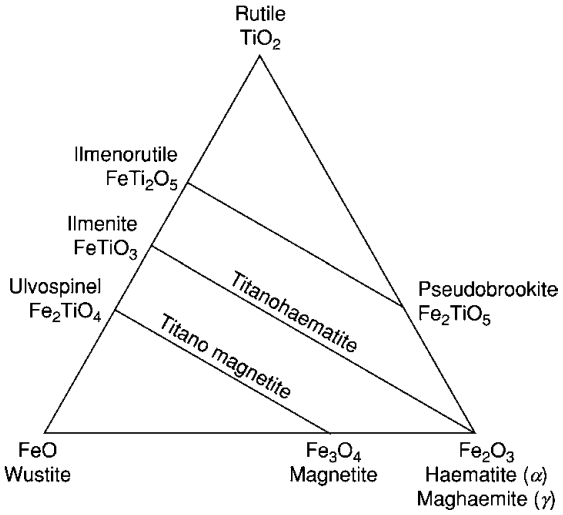


Figure 4 Ternary diagram showing the magnetically important iron oxide minerals in the titanomagnetite and titanohaematite solid solution series.

Table 1 Magnetic properties of some common minerals in rocks

Mineral (composition)	M_s*(kA m^{-1})*	T_C *(°C)*
Titanomagnetite ($Fe_{2.4}Ti_{0.6}O_4$)	125	150
Magnetite (Fe_3O_4)	480	580
Maghaemite (γ-Fe_2O_3)	380	590–675
Haematite (α-Fe_2O_3)	~2.5	675
Goethite (α-FeOOH)	~2	120
Pyrrhotite (Fe_7S_8)	~80	320

magnetite; Table 1) have T_C values close to 580°C, whereas the presence of titanium lowers the T_C value. During a basaltic volcanic eruption, the temperature of a lava is approximately 1200°C; when a lava flow cools below T_C, the Earth's magnetic field is recorded within the lava flow (Figure 3). The declination, inclination, and magnetization intensity, which are proportional to the strength of the field, can today be measured in the laboratory. The magnetization intensity can vary by several orders of magnitude between different rock types, and thus different laboratory instruments are required to measure the magnetization precisely. Volcanic rocks normally have high intensities and the magnetization can be measured on standard spinner magnetometers. Conversely, sedimentary rocks can be extremely weakly magnetized, and highly sensitive superconducting magnetometers (superconducting quantum interference devices, SQUIDs) are required to measure and unravel their magnetization history.

Palaeomagnetic Analysis

In the early days of palaeomagnetic studies, it was common to measure the magnetization in a rock and assume that this magnetization, referred to as the natural remanent magnetization (NRM), represented a primary magnetization that had survived magnetic resetting from subsequent thermal or chemical activity. However, during the 1970s, it became more and more evident that rocks can undergo magnetic resetting, and it is therefore now standard procedure to test the stability of the NRM by thermal, alternating field, or chemical (rare) demagnetization. With the former method, a sample is measured following heating to higher and higher temperatures in a 'zero' field oven. From the 1980s, it became standard procedure to display the demagnetization data in orthogonal vector plots, also referred to as Zijderveld diagrams (Figure 5). These diagrams portray directional and intensity changes on a single diagram – magnetization components are identified as linear segments in both the horizontal ('declination') and vertical ('inclination') planes. Components and the degree of linearity can be computed using least-squares algorithms. A single-component

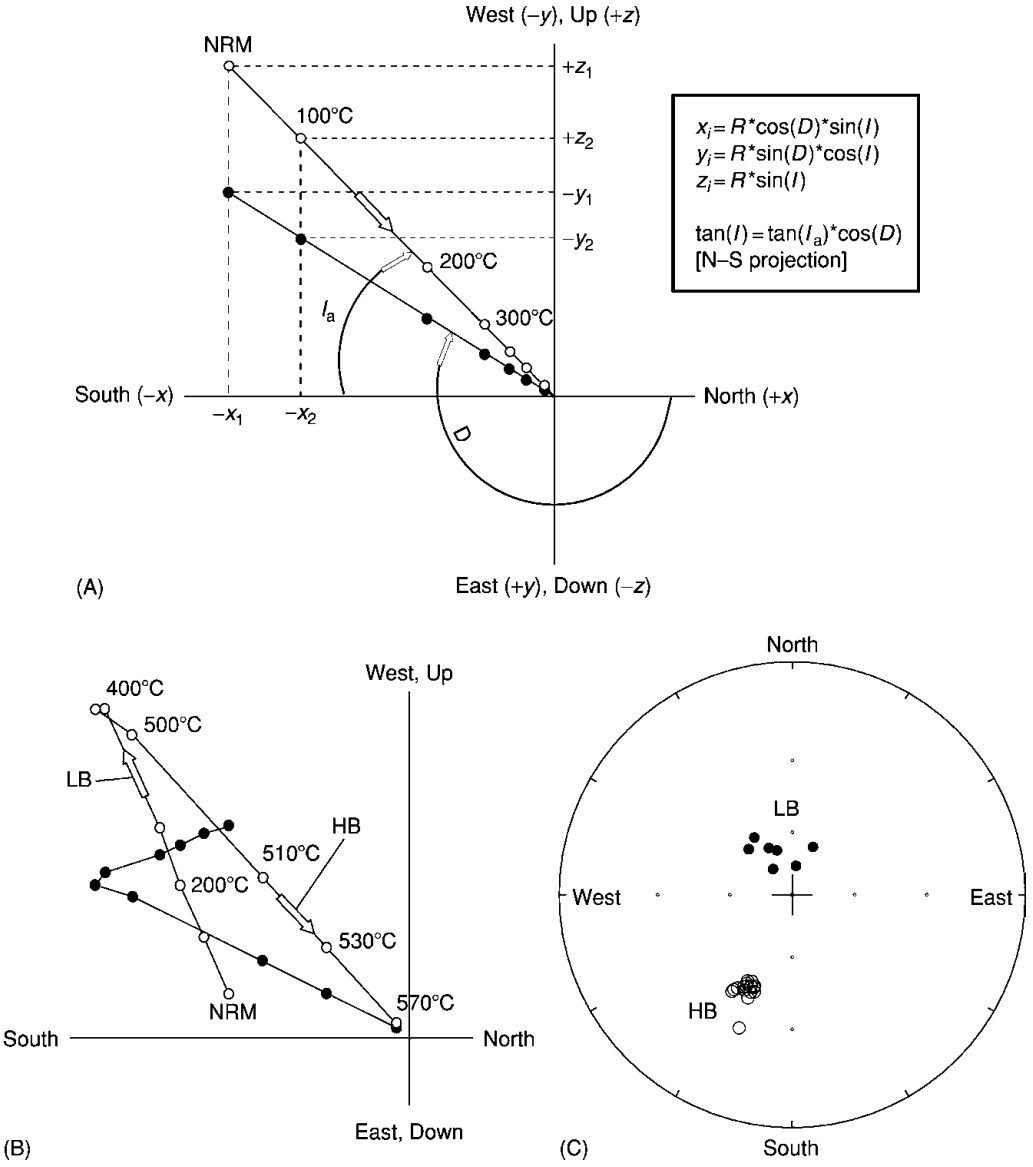


Figure 5 A Zijderveld plot illustrating vector components of magnetization during progressive thermal demagnetization, projected onto orthogonal horizontal (declination; filled symbols) and vertical (inclination; open symbols) planes. For each demagnetization step, the measured declination (D), inclination (I), and magnetization intensity (R) are decomposed to cartesian coordinates, x_i, y_i, and z_i, using the formulae in the box in (a). N–S (+x, −x) is selected as the projection plane as the declination is closer to N–S than E–W. The horizontal component is plotted as $x_i y_i$, whereas the vertical plane is plotted as $x_i z_i$. This procedure is repeated for each demagnetization step (in the example, natural remanent magnetization (NRM): 100°C, 200°C, etc.). Using this procedure, a magnetic component is recognized as a linear vector segment. The declination for a component can be read directly from the diagram, normally numerically computed by least-squares analysis, whereas the inclination is apparent (I_a) and always larger than the real I; the 'distortion' of I_a depends on the projection plane. (A) Single-component magnetization decaying towards the centre of the diagram. (B) Two-component magnetization with a low unblocking component (LB) identified below 400°C and a high unblocking component (HB) decaying towards the centre of the diagram; the stability up to 570°C suggests pure magnetite as the remanence carrier. (C) Mean site compilation of HB and LB components shown in a stereoplot. LB is a recent overprint, whereas HB should be considered to be a Late Permian magnetization from the Oslo area (Baltica). The mean declination/inclination in this hypothetical study is 205°/−41° (95% confidence circle around the mean, α_{95}, is 2.3°). The calculated palaeomagnetic pole for the site (60° N, 10° E) is 49.3° N and 152.3° E. In the stereoplot, open (filled) symbols denote negative (positive) inclinations.

magnetization is identified by single vector decay towards the origin of the diagram as the sample is progressively demagnetized (Figure 5A). Multi-component magnetizations, in which a primary component has been partly overprinted by younger components, can be recognized by the presence of two or more linear segments (Figure 5B). In the latter example, a hypothetical Permian dyke magnetization

from the Oslo region (Norway), with SSW declinations and negative (upward-pointing) inclinations (high unblocking component), is partially overprinted by a younger magnetization of Holocene origin, with NNW declinations and positive (downward-pointing) inclinations (low unblocking component). When the blocking temperatures overlap, curved segments are observed.

In a typical palaeomagnetic study, 5–10 samples are analysed from each site. Magnetization components are computed by least-squares analysis, a mean direction is computed for each site using Fisher statistics, and, finally, a mean direction is calculated from all the sites. In our example, we can imagine that the stereoplot in Figure 5C represents the mean directions from numerous individual dykes in the Oslo region.

Palaeomagnetic Stability Tests

Magnetic overprinting or resetting presents a problem. However, there are four fundamental tests used to check the stability and the potentially primary character of magnetizations: (1) the fold test; (2) the reversal test; (3) the conglomerate test; and (4) the contact test (Figure 6). The fold test determines whether a magnetization was formed prior to or after folding of a rock, and is therefore a relative test that does not prove a primary origin of the remanence. If a remanence is pre-fold, the site vectors should be dispersed after folding (Figure 6A); conversely, if a magnetization is post-fold, the site vectors should be similar throughout the fold structure (Figure 6B). The presence of antipodal stratigraphically linked reversals in a sedimentary or basaltic sequence is the best evidence for primary remanence (Figure 6A). The conglomerate test is also a powerful test. If magnetizations are random between individual boulders (Figure 6A), this is an indication for a primary magnetization in the host rock. Conversely, if boulder magnetizations concur with those in the surrounding rocks, magnetic overprinting is indicated (Figure 6B). A contact test is employed to check whether an intrusion or a dyke carries a primary magnetization. A dyke and its baked margin should coincide, whilst non-baked samples should differ if significantly older

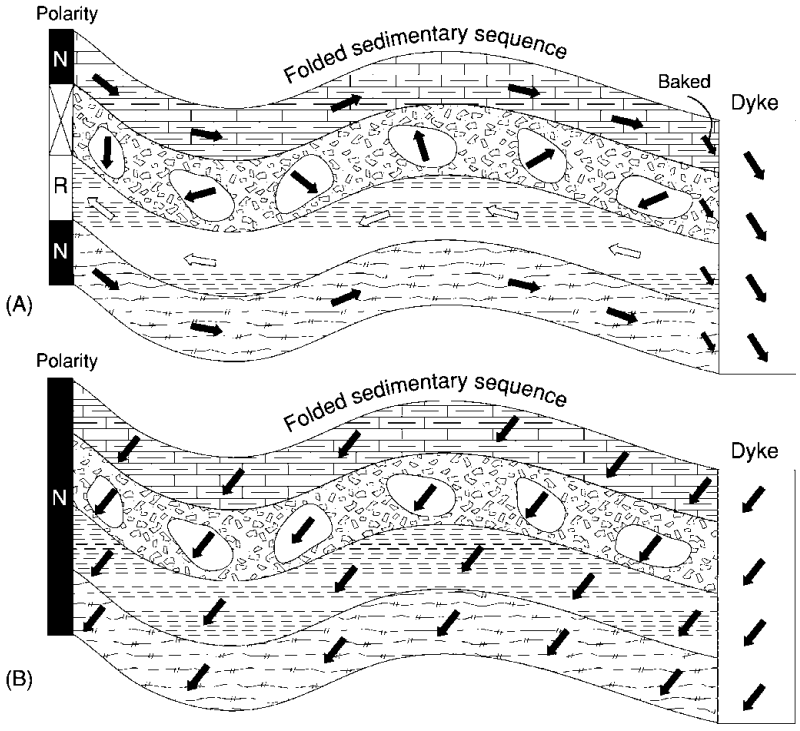


Figure 6 (A) Positive field tests. We assume that a sedimentary sequence was deposited at the equator (flat inclinations) and later folded. Polarity is indicated by the arrows. As the inclinations follow the bedding (except in the conglomerate), this is a pre-fold magnetization. Evidence for a primary magnetization is witnessed by layers with antipodal polarity, and we can therefore establish a magnetostratigraphy (alternating normal and reversed magnetic fields). Further evidence for a primary magnetization is observed from a layer with boulders that show random magnetization vectors. A dyke intruded the folded sequence later, and steeper inclinations (indicating that the continent must have moved to higher latitudes) from the dyke and its baked/chilled margin indicate a primary dyke magnetization. (B) Negative field tests. All layers, conglomerate boulders, and the dyke have similar magnetizations and no reversals are observed. This indicates a regional secondary overprint after the folding and dyke intrusion.

than the dyke (Figure 6A). Conversely, if the dyke and all the surrounding rocks have a similar magnetization, the dyke records a younger overprint (Figure 6B).

Palaeomagnetic Poles and Reconstruction of a Continent

Based on the measurement of the remanent inclination, we can calculate the ancient latitude for a continent when the rock formed from the formula: $\tan(I) = 2 \times \tan(\text{latitude})$. In addition, the remanent declination, which deviates from 0° or 180° (depending on the polarity of the Earth's magnetic field), provides information about the rotation of a continent.

The inclination and declination change with the position of the sampled rock on the globe (Figure 2), but the position of the magnetic pole of a geocentric axial dipole is independent of the locality at which the rock acquired its magnetization. Thus, it is practical to calculate pole positions in order to compare results from various sites or to perform plate tectonic reconstructions.

Ideally, as a time average, a palaeomagnetic pole (calculated from the declination, inclination, and the geographical site location) for a newly formed rock will correspond with the geographical north or south pole. If a continent moves later, the palaeomagnetic pole must move with the continent. To perform a reconstruction with palaeomagnetic poles, we therefore have to calculate the rotation (Euler) pole and angle which will bring the palaeomagnetic pole back to the geographical north or south pole, and then rotate the continent by the same amount (Figure 7A). In our example, a palaeomagnetic pole (latitude, 49.3° N; longitude, 152.3° E), calculated from the situation depicted in Figure 5C, will position the Baltica continent (most of northern Europe eastward to the Urals) at latitudes between 15 and 50° N, causing the city of Oslo to have been located

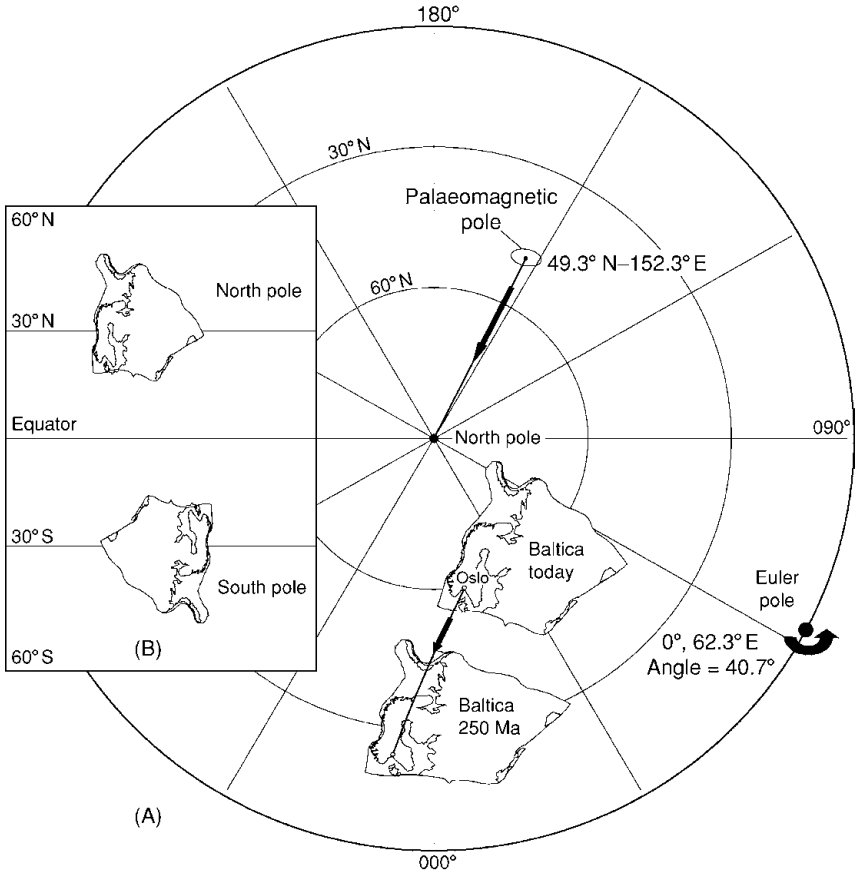


Figure 7 (A) The reconstruction of a continent, for example Baltica, is performed as follows. Determine the Euler pole needed to rotate a palaeomagnetic pole (in our case 49.3° N, 152.3° E) to the geographical north pole (we calculate 0°, 62.3° E and a rotation of 40.7°). This Euler pole is then used to rotate the continent by the same amount. Thus, Baltica today is rotated back about this pole to the position it occupied in Permian times. (B) In (A), we assumed that the palaeomagnetic pole was a north pole. If we assume a south pole, then the continent will be placed in the opposite hemisphere and geographically inverted.

at 24° N in the Late Permian (Figure 7A). Because the current latitude of Oslo is 60° N, Baltica must have drifted northwards since the Permian.

Palaeomagnetic data can only constrain latitude (based on inclination) and the amount of angular rotation (based on declination). Because the palaeolongitude is unknown, we can position Baltica at any longitude we wish, subject to other geological constraints. In addition to that uncertainty, we cannot tell in old rocks whether a palaeomagnetic pole is a south or north pole. In Figure 7A, we assumed that the pole was a north pole, but if we used a south pole, Baltica would plot in the southern hemisphere but in a geographically inverted orientation (Figure 7B). Hence, there is freedom to select north or south poles when producing reconstructions, placing the continent in an opposite hemisphere and rotated by 180°.

Apparent Polar Wander Paths

Apparent polar wander (APW) paths represent a convenient way of summarizing palaeomagnetic data for a continent or terrane, instead of producing palaeogeographical maps at each geological period. APW paths represent the apparent motion of the rotation axis relative to the continent, depending on whether one plots the movement of the north or south pole. APW paths can therefore be constructed as north or south paths. To construct an APW path, a set of palaeomagnetic poles of varying geological age

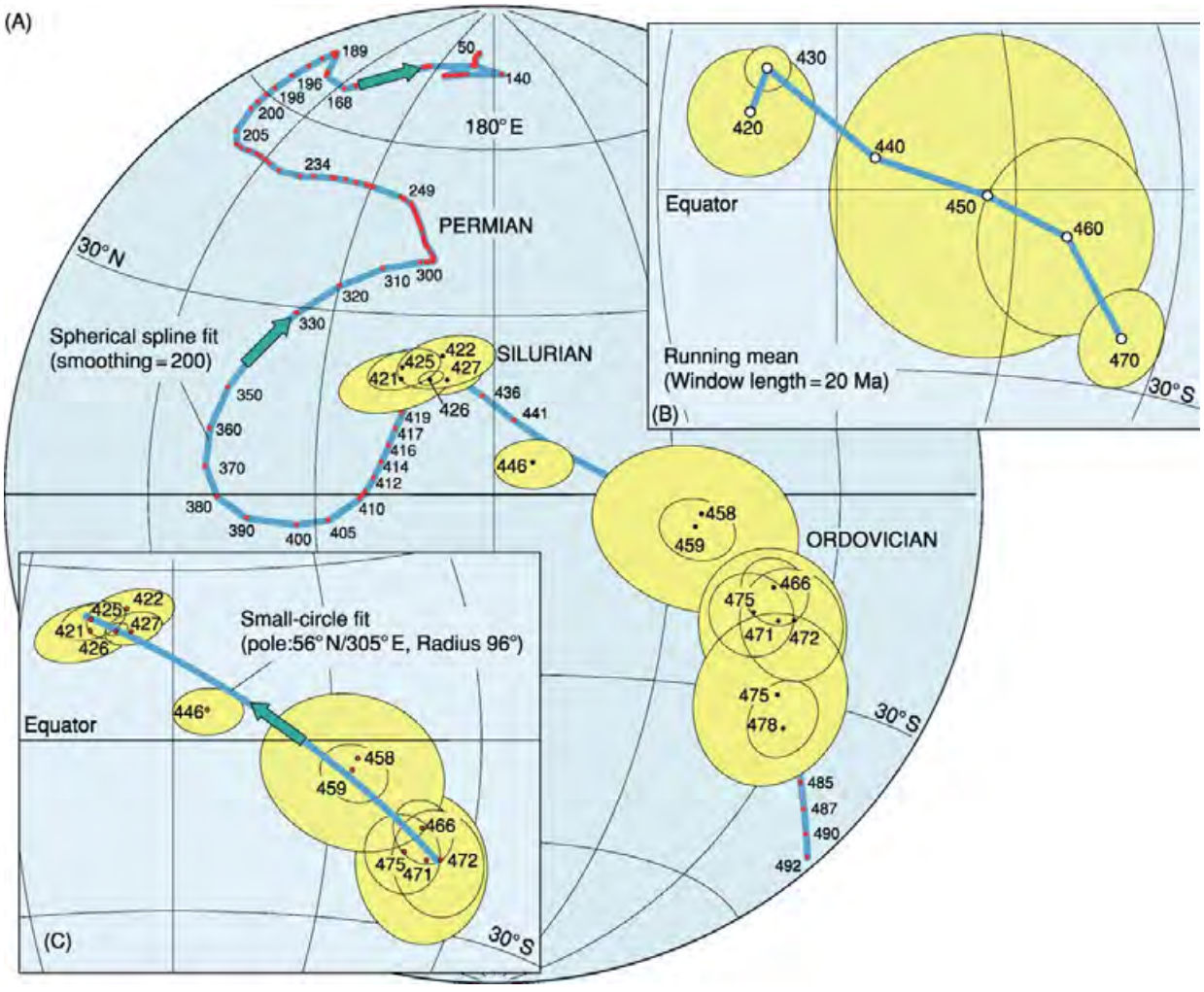


Figure 8 Examples of apparent polar wander (APW) paths for Baltica (include stable European data from Permo-Carboniferous times). (A) Moderately smoothed spherical spline APW path from Early Ordovician to Early Tertiary times. Only Ordovician through Silurian input poles are shown. (B) Running mean path using the same input poles as in (A) and a 20 million year window. Mean poles are shown with 95% confidence ellipses (A95), except for the 440 Ma mean pole for which there was only one pole entry. The running mean path is only shown for the Ordovician–Silurian section of the APW path. (C) Small circle path fitted to Ordovician–Silurian poles as in (A). Input poles in (A) and (C) are shown with 95% confidence ovals (known as dp/dm).

are presented in a single diagram, and a synthetic path is fitted to the incrementing poles (Figure 8A). There are three common methods for generating APW paths: (1) spherical splines; (2) running mean (sliding time window); and (3) the small circle method.

The spherical spline method of modelling APW paths has been employed since the late 1980s. In brief, a spline constrained to lie on the surface of a sphere is fitted to the palaeomagnetic poles (Figure 8A), themselves weighted according to the precisions of the input palaeopoles. In the running mean method, palaeomagnetic poles from a continent are assigned absolute ages, a time window is selected (e.g., 20 million years), and all palaeomagnetic poles with ages falling within the time window are averaged. Using Fisher statistics, 95% confidence ellipses (known as A95 when averaging poles) can be calculated for each mean pole (Figure 8B). Both the spline method and the running mean technique are effective in averaging out random noise and allowing the basic pattern of APW paths to be determined.

The small circle method is based on the fact that movements of continents, APW paths, hotspot trails, ocean fracture zones, etc., must describe small circular paths if the Euler pole is kept constant. It is reasonable to assume that continents may drift around Euler poles that are kept constant for, say, some tens of millions of years. One can therefore fit APW segments along an APW path. This is demonstrated in Figure 8C where we can fit a small circle to Baltica poles from 475 to 421 Ma. However, after 421 Ma, the path changed direction markedly and this resulted from the collision of Baltica with Laurentia (North America, Greenland, and the British Isles north of the Iapetus Suture), which radically changed the plate tectonic boundary conditions and the APW path for Baltica.

Palaeolatitudes and Drift Rates – Links to Facies

Based on APW paths, we can calculate palaeolatitudes and plate velocities for a specific geographical location. Plate velocities are minimum velocities as the longitude is unconstrained; we only calculate latitudinal velocities. Figure 9 shows an example of such calculations based on the APW path in Figure 8A. In this diagram, we have also separated the northward and southward drift of Baltica. Drift velocities are typically below 8 cm per year, but peak velocities of around 14 cm per year are seen after collision of Baltica with Laurentia in Late Silurian–Early Devonian times.

The calculation of latitudinal velocities is important in order to check whether drift rates are

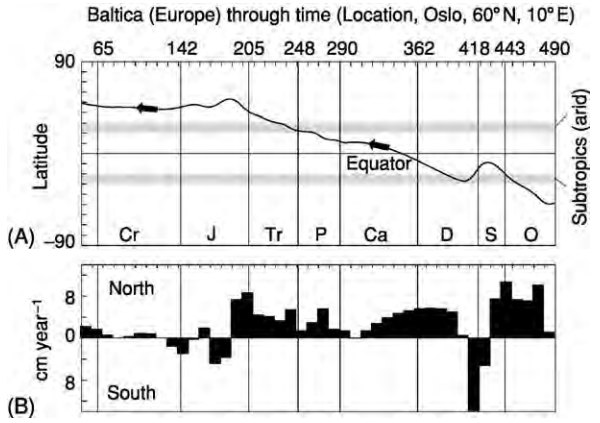


Figure 9 Latitude motion (A) and velocities (B) for Baltica (city of Oslo) from Ordovician to present times based on palaeomagnetic data (Figure 8A). Baltica was in the southern hemisphere during Ordovician through Devonian times, crossed into the northern hemisphere in the Early Carboniferous, and continued a general northward drift throughout the Mesozoic and Cenozoic. Northward and southward latitudinal translations throughout the Early Palaeozoic were accompanied by velocity peaks in the Early Silurian (northward) and the earliest Devonian (southward). Cr, Cretaceous; J, Jurassic; Tr, Triassic; P, Permian; Ca, Carboniferous; D, Devonian; S, Silurian; O, Ordovician.

compatible with 'modern' plate tectonic velocities. A rate of 18 cm per year (India) is the highest reliable value reported for the last 65 million years. When values appear unrealistically high (e.g., more than 20–30 cm per year), some authors have appealed to true polar wander (TPW) as a plausible explanation. TPW is a highly controversial subject that implies rapid tilting of the Earth's rotation axis, and is not generally accepted.

The distribution of climatically sensitive sediments, such as glacial deposits, coal, carbonates, and evaporites, is useful to check the palaeolatitudes derived from palaeomagnetic data. Glacial deposits are usually confined to polar latitudes and, except during the recent ice ages, there is no evidence for such deposits in Southern Baltica, as predicted by the palaeomagnetic data (maximum 60° S in the Early Ordovician). Carbonates, particularly in massive build-ups, such as reefs, are more common in lower latitudes. During Ordovician and Silurian times, Baltica drifted to subtropical and tropical latitudes, as witnessed by the presence of Bahamian-type reefs in Southern Baltica. Evaporites typically record dry climates within the subtropics (20–30°). During the Late Permian, Baltica was located at subtropical northerly latitudes, and the Late Permian coincides with large evaporite deposits in the North Sea area that subsequently became important in hydrocarbon trap development.

Palaeomagnetism and Palaeogeography: the Big Picture

Quantitative reconstructions are most commonly derived from hotspot traces (Cretaceous–Tertiary times) and ocean-floor magnetic anomalies, but prior to the earliest *in situ* ocean floor preserved today (approximately 175 Ma), the positioning of continents can only be quantitatively recognized by palaeomagnetism. As longitude is unknown from palaeomagnetic data, the identification and discrimination of faunas and floras can indicate that continents with similar faunas were in proximity to one another; conversely, different faunas of the same age can indicate the separation of the continents.

In order to construct a global palaeogeographical map, palaeomagnetic data from individual continents or terranes must be compiled and evaluated in terms of reliability. At any given time, palaeomagnetic data may not exist for all continents and additional criteria, such as fauna, flora, facies, and tectonic and

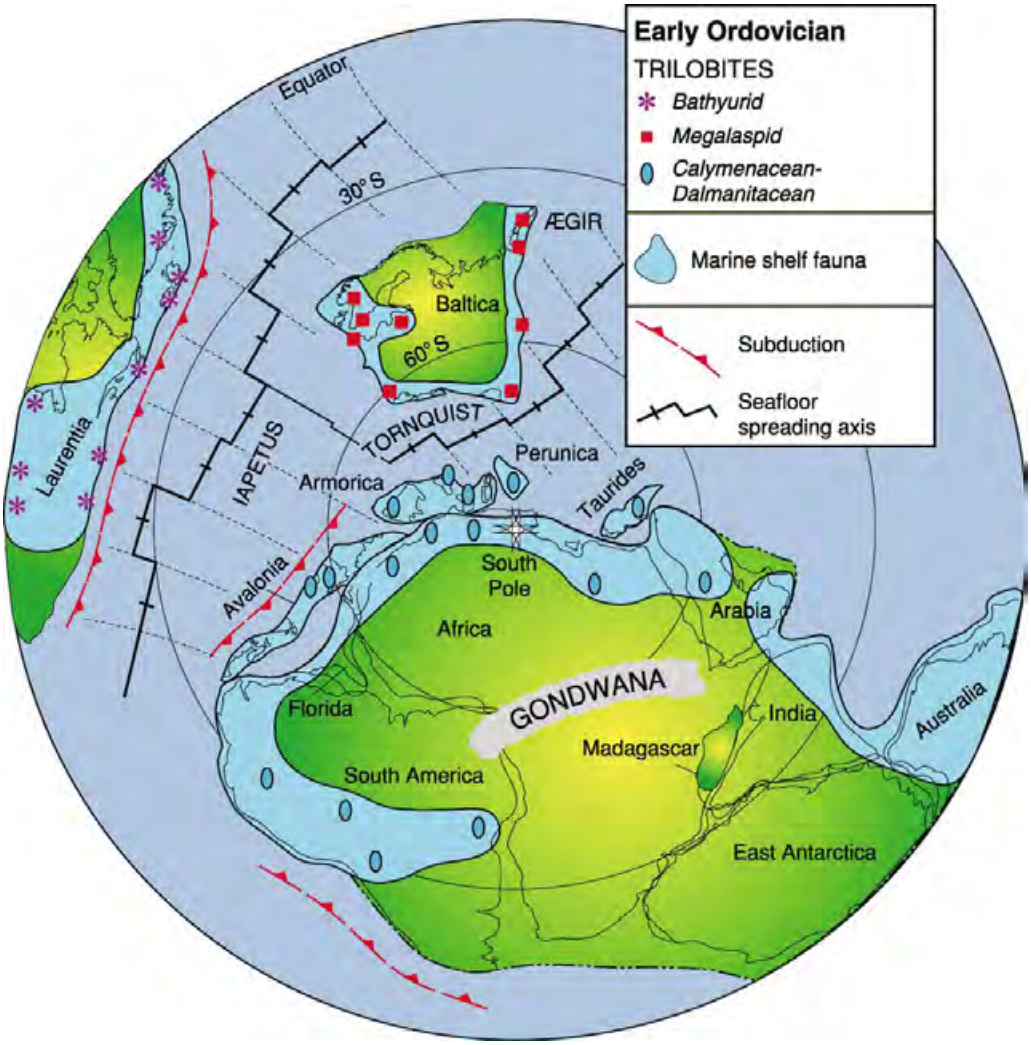


Figure 10 Reconstruction for Early Ordovician times (490–470 Ma) with the major terranes and some key Arenig–Llanvirn trilobite faunas. Laurentia (located at the equator) includes North America, Greenland, and the British Isles north of the Iapetus Suture. Baltica (intermediate southerly latitudes) includes most of Scandinavia and Russia eastwards to the Urals. The core of Gondwana consists of Africa, Arabia, Madagascar, Greater India, most of Antarctica, most of Australia, Florida, and most of South America. Gondwana formed at around 550 Ma and covered more than 90° of latitude in the Early Ordovician. Gondwanan dispersal history commenced with the rifting off of Avalonia at *ca.* 465 Ma. Avalonia includes the British Isles and north-west Europe south of the Iapetus Suture, eastern Newfoundland, most of the Maritime Provinces of Canada, and parts of the eastern USA. Armorica includes the Armorican Massif of Normandy and Brittany, the Massif Central, and the Montagne Noire areas of France, together with parts of the Iberian Peninsula. Perunica comprises the area north of the Barrandian basin of Bohemia. The Taurides comprises most of central and southern Turkey.

magmatic history, must be incorporated in order to construct rational maps. Figure 10 illustrates an Early Ordovician (490–470 Ma) reconstruction; it does not show all the continents or terranes that existed at this time, but some major players, such as Gondwana, Baltica, Laurentia, and some selected peri-Gondwanan terranes with palaeomagnetic and/or faunal data (Avalonia, Armorica, Perunica, and Taurides). Not all the Gondwanan continents have reliable palaeomagnetic data for this time. However, Gondwana was amalgamated at around 550 Ma and continental elements, such as South America and India (no palaeomagnetic data), remained attached to Africa until the breakup of the Pangea supercontinent during the Mesozoic.

An integrated approach of palaeomagnetic and faunal analysis is applicable for the entire Phanerozoic, but works best for the Early Palaeozoic and, notably, the Early Ordovician (Figure 10). At this time, Gondwana stretched from the south pole (Africa) to the equator (Australia and East Antarctica), Baltica occupied intermediate southerly latitudes, separated by the Tornquist Sea, whereas Laurentia straddled the equator. Laurentia was separated from both Baltica and Gondwana by the Iapetus Ocean that had opened in the Late Precambrian. The Iapetus Ocean (approximately 5000 km across the British sector) and the Tornquist Sea (approximately 1100 km between southern Baltica and Armorica–Perunica) were at their widest. This is probably why benthic trilobites from Laurentia (bathyurid) and north-west Gondwana (calymenacean–dalmanitacean) are so markedly different from those of Baltica (megalaspid).

In summary, palaeomagnetism can be seen to be the best and only quantitative method to establish the positions of old terranes and continents as they drifted across the globe over geological time.

See Also

Analytical Methods: Geochronological Techniques. **Gondwanaland and Gondwana**. **Magnetostratigraphy**. **Mantle Plumes and Hot Spots**. **Palaeozoic:** Ordovician. **Plate Tectonics**.

Further Reading

Butler RF (1992) *Paleomagnetism – Magnetic Domains to Geologic Terranes*. Oxford: Blackwell Scientific Publications.

Cocks LRM and Torsvik TH (2002) Earth geography from 500 to 400 million years ago: a faunal and palaeomagnetic review. *Journal of the Geological Society, London* 159: 631–644.

Cox A and Hart RB (1986) *Plate Tectonics – How it Works*. Oxford: Blackwell Scientific Publications.

Dunlop DJ (1979) On the use of Zijderveld vector diagrams in multicomponent palaeomagnetic studies. *Physics of the Earth and Planetary Interiors* 20: 12–24.

Fisher NI, Lewis T, and Embleton BJJ (1987) *Statistical Analysis of Spherical Data*. Cambridge: Cambridge University Press.

Gordon RG, Cox A, and O'Hare S (1984) Paleomagnetic Euler poles and the apparent polar wander and absolute motion of North America since the Carboniferous. *Tectonics* 3: 499–537.

Kirschvink JL (1980) The least square line and plane and the analysis of palaeomagnetic data. *Geophysical Journal of the Royal Astronomical Society* 62: 699–718.

McElhinny MW and McFadden PL (2000) *Paleomagnetism, Continents and Oceans*. Cambridge: Academic Press.

Torsvik TH, Smethurst MA, Meert JG, *et al.* (1996) Continental break-up and collision in the Neoproterozoic and Palaeozoic: A tale of Baltica and Laurentia. *Earth Science Reviews* 40: 229–258.

Van der Voo R (1993) *Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans*. Cambridge: Cambridge University Press.

PALAEONTOLOGY

L R M Cocks, The Natural History Museum, London, UK

Copyright 2005, Natural History Museum. All Rights Reserved.

Introduction

The word 'fossil' has no exact definition, but it is loosely taken to mean any organism whose remains or traces of remains are preserved in some kind of sediment. The term is derived from the Latin *fossare*, meaning 'to dig'. The definition implies nothing about age: fossils range from those that are about three billion years old to those that are preserved in lime-saturated water deposited only a few days or even hours ago. The study of fossils is termed palaeontology, which is a word derived from Greek that literally means 'knowledge of ancient things'; however, for more than two centuries the word 'palaeontology' has been restricted to the study of formerly living (organic) not inorganic remains or traces. The word palaeobiology is sometimes used as an alternative. In this encyclopaedia there follow numerous

articles on the various groups of fossils, and this article is designed simply so that the enquiring reader may find his or her own way through the ensuing contributions.

The Preservation and Condition of Fossils

Very few fossils are found that show no changes from the original organism. There is a very approximate correlation in many cases between the age of the fossil and the degree of change, but age is in itself no guide. For example, a fossil may be distorted by tectonic activity within a few hundred thousand years of its deposition, whilst other rocks and their contained fossils undergo little tectonic change over long periods. For example, the Lower Cambrian shale in the St Petersburg area of Russia, which is some 430 Ma old, is so soft that it can be dug with a spade, and the marine shells preserved within it can be washed out and look not dissimilar in preservation from those found on a beach today. The term diagenesis (*see* **Diagenesis, Overview**) describes the changes within both rocks and fossils as the sediments become dewatered and chemical transformations occur. Voids, from microscopic size upwards, present both in the original organism and in the surrounding sediment, are usually filled during diagenesis as a result of mineral-charged fluids circulating through the rock. This may occur at any time from soon after the fossil was deposited to many millions of years later, when the entombing rock is subjected to sedimentary pressure or tectonic events. Replacement of the chemicals in the fossils themselves is widespread; for example, the calcium carbonate ($CaCO_3$) of a brachiopod shell is often replaced by silica (SiO_2) during diagenesis.

Fossils may also be hugely distorted from the original shape of the organisms by tectonic processes. In addition the fossils may dissolve, and their chemical contents may form other substances; for example all 'fossil fuels', such as crude oil (*see* **Petroleum Geology:** The Petroleum System) and natural gas, are made up of the concentrated remains of fossil plankton, and coal is formed from the now usually unrecognizable remains of fossil plants.

Another process is termed disarticulation. Obviously, if an animal or plant has only a single hard part, such as the shell of a snail, then that may become weathered or broken, but there is still only the one shell. However, if an organism has more than one hard part (ranging from the two shells of a bivalved mollusc or brachiopod to the several hundred bones found in a mammal skeleton), then the process of transport from its place of death to its final burial place, where it will eventually form a fossil, will often cause the organism to break up or disarticulate. The degree of disarticulation often depends on how strong or weak the original articulation was; for example, some brachiopods have weak hinge structures, and others, such as the common Silurian genus *Atrypa*, have complex interlocking mechanisms between the two valves and are thus very commonly found fossilized with their two valves still closed together.

Classification of Organisms

Fossils and organisms living today are all included within the same system of classification: whether an organism is living or dead, or the species extinct, is not relevant to its position in the classification. Of course, all the various systems of classification, whilst attempting to reflect truly natural relationships and groupings, are actually man made and in the last analysis subjective, although objective measurements and methods of analysis have in many cases assisted the systematists in their analyses and helped them to reach their taxonomic conclusions. The term 'systematics' is used in a specialized sense by biologists and palaeontologists to describe the study and arrangement of organisms into classifications and hierarchies.

All organisms are classified within the binomial method; in other words, each organism or taxon is defined as a species within a genus. The binomial system was invented by the eighteenth-century Swedish naturalist Carl Linnaeus (or Linné). Linnaeus published many classificatory books and papers, and subsequent systematists have agreed that the 12th edition of his book *Systema Naturae*, published in 1758, should be the technical starting date for animal classification; however, there are several different starting dates for plant names. All generic names start with capital letters, and all species names are written in lower case: for example, our own genus is named *Homo* and our species name is *sapiens*. Although the generic and specific names are the formal minimum for all distinctively named organisms, some genera are divided into subgenera (which are therefore above the species level), and some species are divided into subspecies. For example, the English song thrush is the species *Turdus philomelos*, and those birds in my garden in London are the main subspecies *Turdus philomelos philomelos*, whilst the subspecies occurring only in northwest Scotland is *Turdus philomelos hebridensis*. Generic, subgeneric, specific, and subspecific names are normally printed in italic. Above the genus level, organisms are grouped within families, which may again be divided into subfamilies or grouped together as superfamilies. Above the family are the order, the class, the

phylum, and the kingdom. Plants are one kingdom, animals are another. Within the animal kingdom, the phylum Mollusca (*see* **Fossil Invertebrates:** Molluscs Overview), for example, is divided into the class Bivalvia (*see* **Fossil Invertebrates:** Bivalves) (including oysters), the class Cephalopoda (*see* **Fossil Invertebrates:** Ammonites; Cephalopods (Other Than Ammonites)) (which includes ammonites, nautiloids, and squids), the class Gastropoda (*see* **Fossil Invertebrates:** Gastropods) (snails), and various smaller classes with few included orders and families. Family names and above are not normally italicized in print. Families, genera, and species are subject to the Law of Priority, but orders and the higher taxa are not. This means that if a genus is named but is subsequently recognized as being within the same generic concept as another genus named earlier (but after 1758), then the later-named genus is placed within the synonymy of the earlier-named one.

Organisms are also classified, but less formally, according to their mode of life. They are obviously divided into marine (living in the sea), freshwater, and non-marine. Those that drift in the water without substantial self-propulsion are termed plankton, those that swim are termed nekton, and those that are restricted to the ocean floor are termed benthos. Benthic animals may be anchored to the seafloor (like most corals), move around on it (like most gastropods and trilobites), or burrow within the sediment (like most bivalves).

Trace Fossils

As stated in the introduction, the term 'fossil' includes not only the remains of an organism but also its traces preserved in the rock. These traces fall into two main groups. The first is when, for example, the shell of a bivalve is preserved in a sediment that is subsequently changed through diagenesis into a lithified rock, after which the shell vanishes, usually through the action of circulating fluids, leaving only the mould of the original creature's hard parts (in that case the shell) in the rock. That mould will, in favourable circumstances, faithfully reproduce the image of the shell (Figure 1). The second group of trace fossils are the burrows, tracks, feeding scrapes, and other biological traces preserved in soft sediment that has eventually hardened into rock. These are often found even though all the remains of the organism that caused the traces have entirely disappeared.

Human Understanding of Fossils

Fossils have been noticed by humans from time immemorial. Their use as grave goods dates back at least to the Neolithic, and superstitions abounded; for example, the incurved fossils of the bivalve *Gryphaea* were identified as toenails of the Devil, and the Chinese thought that dinosaur bones were the remains of dragons. The naturalist Gilbert White described the ammonites from the Cretaceous chalk of Selbourne in

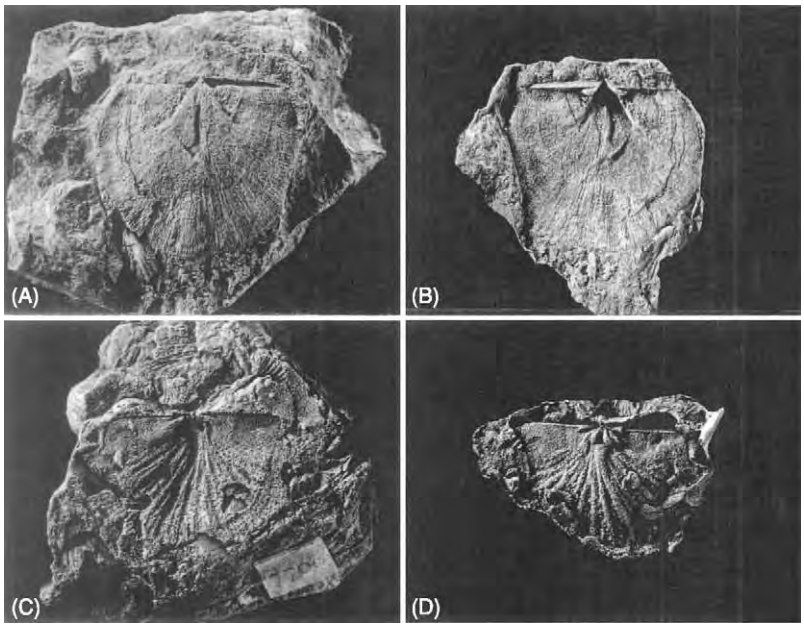


Figure 1 The (A,B) ventral (×1.5) and (C,D) dorsal (×2.0) valves of the Late Silurian brachiopod *Castelloarina fascifer* from Argentina. The views in (A) and (C) are natural internal moulds, as found in the geological outcrop, with the original carbonate shell entirely dissolved by the action of circulating groundwater. The views in (B) and (D) are artificial latex casts of the moulds, faithfully reproducing the topology of the original shell.

Hampshire, England, and retold the legend of St Keyna, of Keynsham near Bristol, who lived alone in a wood full of snakes, which she turned into stone (ammonites) by the power of prayer. It was usually thought that fossils represented dead specimens of living species of animals and plants, but it was not until the eighteenth century that it was realized that in many cases those animals and plants were now extinct. It was William Smith (*see* **Famous Geologists:** Smith) who realized that rocks could be dated in a relative way because different suites of fossils were found in rocks of successive ages. In the past 200 years all Phanerozoic (Cambrian to Holocene) rocks have been dated using fossils, and it is only since about 1960 that relatively reliable radioisotopic ages have been assigned to those zones. Even today, isotopic dates are not precise to within a million or so years over most of the Phanerozoic, whilst by interpolation many fossil zones are understood to have lasted for as little as 100 000 years. Thus, fossils have been extensively used to correlate rocks over considerable distances, and this knowledge has been of vast economic benefit in the exploration and discovery of coal and hydrocarbons, as well as being used to understand past Earth geographies.

Fossils in this Encyclopaedia

Animals are divided into those without backbones (invertebrates) and those with backbones (vertebrates). The chief Encyclopedia entries on fossils are arranged under the major headings of "Fossil Invertebrates", "Fossil Vertebrates", "Fossil Plants" and "Microfossils". Invertebrates are the most varied and commonest fossils. They are arranged here in order of increasing biological complexity. For about 50 years, each genus of fossil invertebrate has been illustrated and described in the 40 or so volumes of *The Treatise on Invertebrate Paleontology* (published jointly by the Geological Society of America and Kansas University Press between 1953 and 2004, and which is continuing), which is to be found in most geological libraries and in which much more detail on each group may be found; however, publication intervals between successive editions of the various parts of the *Treatise* are usually substantial. Despite the fact that fossils with original hard parts only occur in Cambrian and later rocks, a very substantial amount is now known about Precambrian fossils. In this Encyclopedia those are treated under **Origin of Life, Biosediments and Biofilms,** and also in the four articles within the Precambrian section, namely **Precambrian:** Eukaryote Fossils; Prokaryote Fossils; Vendian and Ediacaran.

There are also, at appropriate and separate places, articles on more general palaeontological and biological topics, namely **Biodiversity, Biological Radiations and Speciation, Creationism, Evolution, Fake Fossils, Lagerstätten, Palaeoecology, Palaeopathology, Pseudofossils** and **Trace Fossils.** In addition, the treatment of fossil specimens is considered in **Conservation of Geological Specimens** and **Micropalaeontological Techniques.** Of course the names of many fossil biozones are given in every one of the main stratigraphical articles on systems, which are grouped into Palaeozoic, Mesozoic, and Tertiary to Present, and also in other articles, particularly those on regional geology.

See Also

Biodiversity. **Biological Radiations and Speciation**. **Biosediments and Biofilms**. **Conservation of Geological Specimens**. **Creationism**. **Diagenesis, Overview**. **Evolution**. **Fake Fossils**. **Famous Geologists:** Smith. **Fossil Invertebrates:** Molluscs Overview; Bivalves; Cephalopods (Other Than Ammonites); Gastropods; Ammonites. **Lagerstätten**. **Micropalaeontological Techniques**. **Origin of Life**. **Palaeoecology**. **Palaeopathology**. **Petroleum Geology:** The Petroleum System. **Precambrian:** Eukaryote Fossils; Prokaryote Fossils; Vendian and Ediacaran. **Pseudofossils**.

Further Reading

Benton MJ (ed.) (1993) *The Fossil Record 2*. London: Chapman and Hall.

Benton MJ and Harper DAT (1997) *Basic Palaeontology*. Harlow: Addison Wesley Longman.

Briggs DEG and Crowther PR (eds.) (1990) *Palaeobiology: A Synthesis*. Cambridge: Blackwell Science.

Clarkson ENK (1998) *Invertebrate Palaeontology and Evolution*, 4th edn. Cambridge: Blackwell Science.

Fortey RA (2002) *Fossils: The Key to the Past*, 2nd edn. London: The Natural History Museum.

McKerrow WS (ed.) (1978) *The Ecology of Fossils: An Illustrated Guide*. London: Duckworth. MIT: USA.

Murray JW (ed.) (1985) *Atlas of Invertebrate Macrofossils*. London and New York: Longman and John Wiley.

Rudwick MJS (1992) *Scenes from Deep Time*. Chicago: University of Chicago Press.

Stewart WN and Rothwell AC (1993) *Paleobotany and the Evolution of Plants*, 2nd edn. Cambridge: Cambridge University Press.

Willis KJ and McElwain JC (2002) *The Evolution of Plants*. Oxford: Oxford University Press.

PALAEOPATHOLOGY

S G Lucas, New Mexico Museum of Natural History, Albuquerque, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In 1913, the British scientist, Sir Armand Ruffer, originally defined palaeopathology as "the science of diseases which can be demonstrated in human and animal remains of ancient times". Palaeopathology thus encompasses two distinct areas, one palaeoanthropological, in which ancient human remains are examined, and the other more strictly palaeontological, in which non-human animal fossils are examined. Furthermore, not only disease, but also evidence of injuries, is normally studied by palaeopathologists. Moreover, pathological conditions are not only known in fossil animal remains, but have also been reported in fossil plants. Here, discussion focuses on the strictly palaeontological aspects of palaeopathology. Palaeopathology is also used here as a noun to refer to any evidence of disease or injury in a fossil.

Assumptions and Methods

Palaeopathology relies on a straightforward uniformitarian assumption, namely that the manifestations of diseases have been relatively stable through time. Thus, for example, a disease that produces a particular skeletal abnormality at present would have done just the same millions of years ago. If this is the case, then it will be possible to diagnose medically (determine the underlying cause of) a palaeopathology.

Also critical to palaeopathology is the need to distinguish an abnormal structure in a fossil, caused by disease or injury, from the effects of post-mortem damage, the processes of rock formation and fossilization, and of erosion and weathering. Abnormal structures in fossils are sometimes caused by these effects, and have been mistakenly identified as manifestations of pathology.

Once a pathological condition is identified in a fossil, diagnosis is strictly by analogy. The palaeopathology is compared with similar pathologies in extant plants and animals, and a match in pathologies suggests identical underlying causes in the living organism (for which we know the cause) and the fossil (for which we infer the cause). Bacteria and viruses that cause disease are almost never preserved in fossils, and so the 'diagnosis by analogy' method is the principal way to determine the underlying causes of palaeopathologies.

Palaeopathologies in Fossil Plants

Pathologies caused by diseases have been identified in fossil leaves and wood. These include evidence of bacterial and fungal infections of ancient plants (*see* **Fossil Plants:** Fungi and Lichens). Most important, however, in the study of fossil plant pathologies is the fossil record of predation and infestation of plants by insects.

This record of predation and infestation is seen in damage of the fossil plant, and specific types of damage include:

- leaf mining – the extraction of living leaf tissue by larvae or mites;
- galling – when an insect or mite inserts an egg into or burrows into plant tissue, so that an abnormal, three-dimensional expansion of the plant tissue (a gall) is formed;
- wood boring – when insects burrow into wood;
- piercing and sucking – when insects penetrate plant tissue to remove fluid;
- sporivory, pollinivory, and nectarivory – consumption by insects of spores, pollen, and nectar, respectively.

The most significant aspect of these palaeopathologies is that they provide direct evidence of plant–insect interaction in the fossil record. This enables palaeontologists to determine the history and intensity of insect herbivory in the fossil record, and allows inferences to be made about the evolution of plant defences against such herbivory.

Palaeopathologies in Fossil Invertebrates

Evidence of deformities and other abnormalities in fossil shells is attributed to disease or genetic defects. Evidence of parasitism on the shells of stationary Palaeozoic animals, such as crinoids (sea lilies), is also well documented. Amongst the fossil invertebrates, there has been extensive study of healed injuries due to predation, especially in fossil clams (bivalves) and snails (gastropods). This evidence of predation (Figure 1) is particularly significant as it allows palaeontologists to understand the history and evolution of attack and defence behaviours in fossil invertebrates.

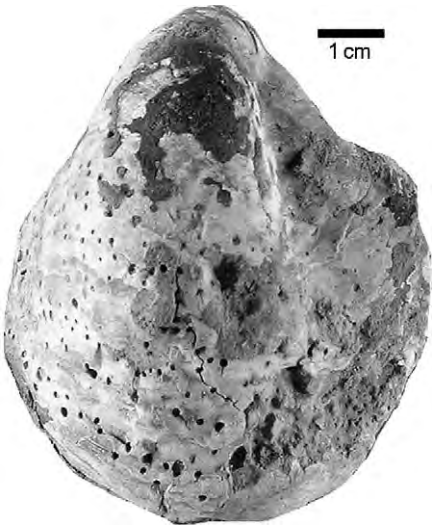


Figure 1 Shell of a 100-million-year-old clam (*Texigryphaea*) from eastern New Mexico covered with tiny holes, which are the drill marks left by snails (gastropods) that attempted to prey on the clam.

Palaeopathologies in Fossil Vertebrates

The majority of palaeopathologies (and the vast majority of the literature on them) are found in the teeth and bones of extinct vertebrates, especially reptiles and mammals. The most commonly identified palaeopathologies are: (1) so-called arthritic conditions, although many are not demonstrably arthritis; (2) mechanical injuries (Figure 2) and related infections, and the healing of these injuries if the animal survived them; (3) wounds due to predation or combat between animals; (4) dental anomalies, including caries (tooth decay); and (5) bone lesions caused by infections and/or parasites.

A list of the most common skeletal pathologies of fossil vertebrates includes:

- arthritis – joint diseases that can be caused by infections or mechanical injuries;
- dental anomalies – rotated, misplaced, malformed, or extra teeth (Figure 3);
- dental caries – cavities (decay) in the teeth;
- exostosis – swellings on bone surfaces;
- fracture – broken bones which, when naturally healed, leave a bone callus (a patch of healed bone) (Figure 2);
- hypertrophy – increase in size of the bone or portions of the bone;
- necrosis – pitting and disintegration of dead bone;
- osteomyelitis – bone infections;
- osteoma – bone tumours;
- osteoporosis – loss of bone density;

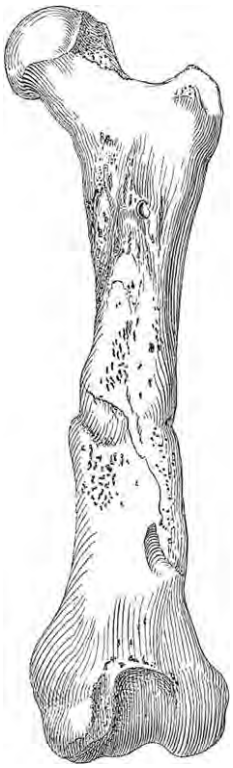


Figure 2 Anterior view of the left femur (thigh bone) of a Pleistocene cave bear from Europe, showing a healed fracture of the bone shaft. The pits above and below the fracture may also indicate necrosis. From Moodie RL (1930), Fig. 5.

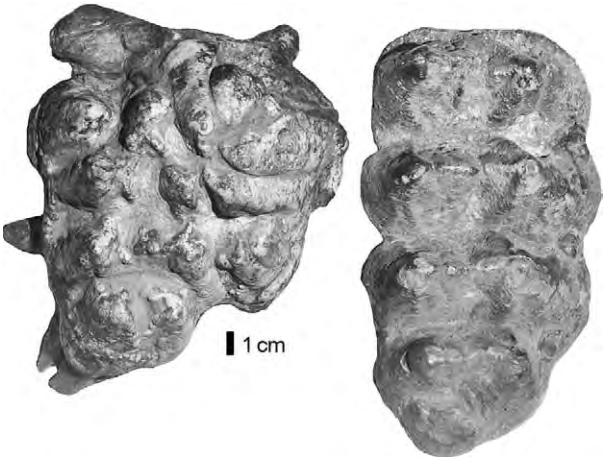


Figure 3 Left: crown view of a pathological molar of an American mastodon (*Mammut americanum*) from the Pleistocene of Oklahoma. Right: normal mastodon molar for comparison. On the pathological molar, there is a normal-shaped molar (left side) covered by a second (right side), highly deformed molar. This secondary (or supernumerary) tooth probably resulted from the fusion of two tooth buds where, normally, only one tooth bud would have been present.

- synostosis – co-ossification (fusion) of normally separate bones.

Dinosaur Palaeopathology

The study of palaeopathology in fossil vertebrates is very active, with many new cases documented in the technical literature each year. This is especially true of dinosaur palaeopathology, which, during the last two decades, has become the largest field of palaeopathological research on non-human fossils. Prior to this interest, reports of palaeopathologies in dinosaurs go back to the 1800s, but they were mostly of isolated curiosities that often led to unfounded speculation. A good example of this is the unsupported suggestion that the frills (head shields) of horned dinosaurs (ceratopsians) grew so large because the dinosaur had too much growth hormone.

More recent studies of dinosaur palaeopathology have examined large samples (populations) using stringent standards for disease recognition. They have been accompanied by attempts to identify blood cells, DNA, and disease pathogens in dinosaur bone. These studies have identified a variety of common palaeopathologies in dinosaur skeletons. Most obvious are broken and healed bones (Figure 4), fused bones, and certain types of bone tumour. These palaeopathologies are well known in horned dinosaurs (ceratopsians), duck-billed dinosaurs (hadrosaurs), and tyrannosaurs. Indeed, the famous skeleton of *Tyrannosaurus rex* nicknamed 'Sue', has many healed injuries in its bones. The new studies of dinosaur palaeopathology demonstrate that dinosaur skeletons show much evidence of injury and trauma, but little evidence of disease.

Figure 4 Toe bone (metatarsal) of a small, meat-eating dinosaur from the Upper Cretaceous of New Mexico (about 73 million years old) with a healed fracture (f) visible in the photograph (left) and the X-ray (right). Courtesy of R. M. Sullivan.

Application of Palaeopathology

As the above paragraphs make clear, the study of palaeopathology provides important information about more than just the history of disease. It also provides direct evidence of interaction between plants and animals, and between animals, mostly between predators and prey. It is thus important to the understanding of the evolution of behaviour. Palaeopathology can also contribute to an understanding of palaeoecology. For example, leaf mining by insects is extremely intense in today's tropics, and the intensity of fossil leaf mining has been used to infer the presence of ancient tropical conditions. Palaeopathology is thus an important subject in its own right, and a valuable tool for inferring ancient behaviour and ecology from the fossil record.

Although observations of palaeopathologies have been made for nearly two centuries, it is fair to say that the study of palaeopathology remains a relatively understudied field. Disease-causing agents generally do not fossilize, and so their history, which would be gleaned directly from the fossils of bacteria and viruses, is still very poorly understood. New technologies promise the ability, at least in some cases, to identify fossil bacteria and viruses associated with palaeopathologies.

Most diseases do not affect the fossilizable parts of an animal (its skeleton), and so much about ancient disease may never be extracted from the fossil record. Most published records of palaeopathologies are of isolated occurrences, often questionably diagnosed. Surveys that provide an idea of the frequency of palaeopathologies are almost unknown. Moreover, sweeping claims about the role of disease in the past – such as the idea that disease caused dinosaur extinction – lack supporting evidence. What is needed, and what is being pursued today, are more extensive efforts to identify palaeopathologies in all kinds of fossils, coupled with more rigorous attempts at diagnosis.

See Also

Fossil Plants: Fungi and Lichens. **Fossil Vertebrates:** Dinosaurs.

Further Reading

Baker J and Brothwell D (1980) *Animal Diseases in Archaeology*. London: Academic Press.

Brothwell DR (1972) *Digging up Bones*. British Museum (Natural History), London.

Jarcho S (ed.) (1966) *Human Palaeopathology*. New Haven: Yale University Press.
Labandeira C (1998) The role of insects in Late Jurassic to Middle Cretaceous ecosystems. *New Mexico Museum of Natural History and Science Bulletin* 14: 105–124.
Moodie RL (1923) *Paleopathology: An Introduction to the Study of Ancient Evidences of Disease*. Urbana: University of Illinois Press.
Pindborg JJ (1970) *Pathology of the Dental Hard Tissues*. Philadelphia: W. B. Saunders.
Rothschild BM and Martin LD (1993) *Paleopathology: Disease in the Fossil Record*. Boca Raton, FL: CRC Press.
Tanke DH and Rothschild BM (2002) Dinosaurs: an annotated bibliography of dinosaur paleopathology and related topics – 1838–2001. *New Mexico Museum of Natural History and Science Bulletin* 20: 1–96.
Vermeij GJ (1987) *Evolution and Escalation: An Ecological History of Life*. Princeton: Princeton University Press.
Wilf P, Labandeira CC, Johnson KR, Coley PD, and Cutter AD (2001) Insect herbivory, plant defense, and early Cenozoic climate change. *Proceedings of the National Academy of Sciences of the USA* 98: 6221–6226.

PALAEOZOIC

Contents

Cambrian
Ordovician
Silurian
Devonian
Carboniferous
Permian
End Permian Extinctions

Cambrian

N C Hughes, University of California, Riverside, CA, USA
N A Heim, University of Georgia, Athens, GA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Cambrian period, the oldest division of the Phanerozoic time-scale, encompasses several important global-scale environmental and evolutionary transitions. Recent advances in the understanding of Cambrian environments, faunas, and chronostratigraphy have all provided an enriched view of this critical interval in the history of the Earth – the time during which life's direct impact upon the rock record first became explicit. The term 'Cambrian' was first applied by Sedgwick (*see* **Famous Geologists:** Sedgwick) in 1835 to a series of sedimentary rocks well exposed in north Wales. The Cambrian system, as defined today, equates to only a portion of what Sedgwick originally considered to be 'Lower Cambrian'. Cambrian deposits are distributed widely, and a basic stratigraphical and palaeogeographical framework has been established for almost all regions. Nevertheless, a great deal of additional research will be necessary to secure our understanding of the geological framework in which the events of the Cambrian were structured. This article presents an overview of the defining temporal, geographical, and environmental attributes of the Cambrian, and then considers these in relation to the most striking aspect of the Cambrian system – the diversification of abundant metazoan life.

Chronostratigraphical Framework

The ratification of formal boundaries for the Cambrian system and improvements in the precision of isotope geochronology have recently resulted in substantial adjustments to views concerning the absolute age and duration of the Cambrian period. These results have fundamentally changed perspectives on the historical events encompassed within the Cambrian.

The Bounds of the Cambrian System

Formal definitions for both the base and the top of the Cambrian system have been ratified within Global Standard Stratotype Sections and Points (GSSPs) in recent years, both with type sections in Newfoundland, Canada. The base of the Cambrian system is currently placed at the occurrence of the trace fossil, *Treptichnus pedum*, 2.4 m above the base of Member 2 of the Chapel Island Formation at Fortune Head, Newfoundland. The base of the subsequent geological system, the Ordovician system, is placed at the first occurrence of the conodont, *Iapetognathus fluctivagus*, at a level of 101.8 m, within Bed 23, in the measured section of the Lower Broom Point Member, Green Point Formation at Green Point, western Newfoundland.

Stratotype points are located at the occurrence of diagnostic markers within particular stratotype sections, and both the top and the base of the Cambrian system rely on the occurrence of fossils. Such boundaries commonly relate to the stratigraphically lowest occurrence of the diagnostic marker fossil in the stratotype section. In the case of the base of the Cambrian, this is not so, because unequivocal specimens of *Treptichnus pedum* have recently been described some 4 m below the GSSP at the stratotype section. The occurrence of other *Treptichnus* specimens still lower in this faulted section could compromise the utility of this section as a stratotype. Nevertheless, the diversification of trace fossils during the Precambrian–Cambrian boundary interval was a significant evolutionary event, and offers the possibility of global correlations based on the impressions of homologous morphological characteristics.

The absolute ages of both the base and the top of the Cambrian system rely on recent U–Pb zircon dating of volcanic deposits intercalated within sedimentary successions. Suitable volcanic deposits are not known from the critical intervals of either of the stratotype sections. Hence, all absolute age estimates of the boundaries of the Cambrian rely on radioisotopic dates obtained from other sections that have been correlated with the stratotypes using alternative criteria, such as fossils or geochemical signatures. The base of the Cambrian system is commonly considered to be at about 542 Ma. This view is based on a minimum date for the first appearance of *Treptichnus pedum* of 539.4 ± 1.0 Ma in the Swartkloofberg section of Namibia, and a maximum age of 543.3 ± 1.0 Ma based on the occurrence of an assemblage of Ediacaran fossils in the same section. An age of 542 Ma also corresponds with a sharp negative excursion in carbon isotope ratios found in several parts of the world. Nevertheless, Ediacaran fossils have recently been recovered from rocks bearing *Treptichnus pedum* in South Australia, and the exact correlation of the Namibia section with the Newfoundland stratotype awaits resolution. The maximum age of the Cambrian–Ordovician boundary has recently been estimated from latest Cambrian volcaniclastic sandstone beds in north Wales to be 489 ± 0.6 Ma. The minimum age is 483 ± 1.0 Ma based on a Tremadocian ash from Nova Scotia.

Stratigraphical Subdivision of the Cambrian System

Formal stratigraphical stages have been established for Cambrian sequences around much of the globe. These sequences are correlated using cosmopolitan fossils, isotopic signatures, and correlative stratigraphical surfaces. Certain key horizons, characterized by distinctive and widespread markers, can be correlated with some accuracy around the globe. Other portions of the Cambrian are less securely correlated due to such factors as faunal endemism (partly real, partly the result of provincial biases in taxonomic descriptions) or questionable alignments of isotopic profiles. All global correlation schemes for Cambrian stratigraphical sequences (e.g., **Figures 1** and **2**) should be viewed as provisional.

The Cambrian period is commonly divided into three portions (**Figure 3**), although the global utility and bounds of these divisions have yet to be formalized. The perception of the nature of the Cambrian period has changed dramatically recently. The decision to place the base of the system near the first occurrence of *Treptichnus pedum* has meant that the first appearance of trilobites (and some other skeletonized macrofossils) occurs after about one-third of Cambrian time had passed. Furthermore, the discovery that the oldest trilobites date from around 520 Ma has dramatically reduced the apparent duration of the trilobite-bearing Cambrian to an interval of approximately 35 million years in duration. This suggests that the biostratigraphical zones of trilobite-bearing portions of the Cambrian are of a time resolution comparable to those of Mesozoic ammonites, on average as short as 1 million years.

Cambrian Palaeogeography

Palaeogeographical reconstructions for the Cambrian, constructed by different earth scientists, generally agree in terms of the overall distribution of the major landmasses, although doubts are commonly expressed about the quality of the palaeomagnetic database used to underpin such reconstructions. The assembly of the core Gondwana components is

generally considered to have been completed prior to the onset of the Cambrian. The positions and movements of many of the 'outboard' terranes that jostled about the margins of Gondwana, however, remain poorly constrained. This is particularly true of the equatorial sector of Gondwana, where faunal provinciality – commonly useful for regional demarcation – was low. Overall, Cambrian palaeogeography was characterized by open water in the northern polar region, and dominated by the major landmass, Gondwana, that stretched from the lowest latitudes in the south, across the equator, and well into the northern hemisphere. Other major landmasses, such as Laurentia, Siberia, and Baltica, were equatorial or lay within the southern hemisphere (Figure 4).

With the exception of the movements of minor terranes, global palaeogeography remained broadly consistent throughout the Cambrian period, with most of the changes apparently related to the rotation and dispersion of Laurentia, Siberia, and perhaps Baltica relative to a stable Gondwana. These movements, coupled with the patterns of sedimentation, are consistent with the breakup of a late Precambrian supercontinent that has been named Pannotia. Details of late Neoproterozoic palaeogeography remain hotly debated and these debates have important implications for interpreting secular changes during Cambrian time.

Cambrian Environments and Climates

The deposits of most major palaeocontinents record a substantial transgression during much of the Cambrian. Such a pattern is clearly displayed in Laurentia, where deposits encroached progressively further into the interior of the continent. Cambrian sedimentation patterns are consistent with relatively high rates of thermal subsidence, perhaps related to the fragmentation of Pannotia. In equatorial settings, particularly within Gondwana, latest Neoproterozoic carbonate platforms were drowned by relatively rapid flooding events that were followed by the deposition of phosphorites and black shales. The shoreward spread of oxygen-depleted and nutrient-enriched waters onto the shelf was associated with the replacement of fossils by secondary phosphates and within cherts. Substantial evaporite deposits developed within interior basins. In the absence of evidence of persistent glaciation, the Cambrian transgression is inferred to have been related to high rates of seafloor spreading.

The analysis of secular changes in the stable isotopes of strontium and carbon provides supplementary tools for correlation amongst sections, together with insights into the evolution of Cambrian environmental conditions (Figure 5). The striking rise in Cambrian stable isotopic ratios is thought to reflect enriched ^{87}Sr values due to the uplift and erosion of cratonic material. This rise in $^{87}Sr/^{86}Sr$ ratios is comparable with that seen in the Neogene, associated with the uplift of the Himalaya. Events associated with the latest stages of the prolonged 'Pan-African' orogeny (*see* **Africa:** Pan-African Orogeny) and the final consolidation of Gondwana may have been responsible, but the sequence and timing of these events remain incompletely resolved.

Investigations of Cambrian seawater chemistry, both directly from fluid inclusions and indirectly from inorganic and organic precipitates, suggest that Cambrian seawaters were generally characterized by relatively low Mg/Ca ratios. This is consistent with the deposition of calcitic, as opposed to aragonitic, biomineralized metazoan skeletons. The late Proterozoic to Cambrian interval witnessed the transition from an 'aragonite sea' to a 'calcite sea', corresponding to a transition from 'icehouse' to 'greenhouse' climatic conditions, and a change from $MgSO_4$ to KCl late-stage salts in evaporites. When viewed in the context of trends seen throughout the Phanerozoic, such transitions correspond with long-term changes in plate tectonic regime. The rapid seafloor spreading associated with the breakup of Pannotia probably drove changes in mid-ocean ridge brine fluxes, caused a global sea-level rise, and increased global volcanism.

Marked fluctuations in Cambrian carbon isotopic ratios over short time-scales have been noted in association with the base of the Cambrian and the Early–Middle Cambrian boundary. In these cases, the excursions were towards lighter carbon isotopes. A more prolonged, positive excursion towards the heavier isotope, ^{13}C, has been defined within the Late Cambrian Steptoean stage within Laurentia, Kazakhstan, and portions of core and outboard Gondwana. The determinants of such fluctuations are complex because carbon isotopic ratios are influenced by multiple factors. Explanations commonly invoked include alternations between periods of well-oxygenated, warmer waters on continental shelves and periods in which the shelves were flooded by cooler, oxygen-poor waters. These explanations receive additional support from marked periodic fluctuations in macrofaunal diversity that roughly parallel the excursions, at least in some cases. Cambrian oceans were similar to oceans throughout the Phanerozoic in that oxygen penetrated the sediment–water interface even within deep ocean basins. Recent models of atmospheric oxygen levels throughout the Phanerozoic suggest that O_2 constituted between 10% and 20% of the composition of the atmosphere. Although these estimates are subject to considerable uncertainty, they suggest that Cambrian

Siberian platform			Australia		China		Spain
Stages	Trilobite, Archaeocyath, and Small Shelly Fossil Zones		Archaeocyath Zones	Trilobite Zones (Stages)	Stages	Trilobite and Small Shelly Fossil Zones	Stages
Amgan	*Schistocephalus*	1		*Xystridura templetonensis/ Redlichia chinensis*	Maochuangian	*Yaojiayella*	Leonian
Toyonian	*Anabaraspis splendens*	3			Lungwangmiaoan	*Redlichia nobilis* / *Redlichia chinensis* / *Hoffetella*	
	Lermontovia grandis/ Irinaecyathus shabanovi-Archaeocyathus okulitchi beds	2	*Archaeocyathus abacus beds*	*(Ordian/ Early Templetonian)*	Tsanglangpuian	*Megapalaeolenus/ Palaeolenus*	Bilbilian
	Bergeroniellus ketemensis	1		*523 Ma			
Botomian	*Bergeroniellus ornata*	4	*Syringocnema favus beds* EB	*Pararaia janeae* *525 Ma		*Drepanuroides*	Marianian
	Bergeroniellus asiaticus	3	Unnamed beds	*Pararaia bunyerooensis*		*Yunnanaspis/ Yiliangella*	
	Bergeroniellus SB *gurarii*	2				*Malungia*	Ovetian
	Bergeroniellus micmacciformis/ Erviella	1		*Pararaia tatei* / *Abadiella huoi*	Chiungchussuan CG	*Eoredlichia/Wutingaspis* / *Parabadiella/ Mianxidiscus*	
Atdabanian	*Fansycyathus lermontovae*	4	*Jugalicyathus tardus* / *Spirillicyathus tenuis*		Meishucunian	*Lapworthella/ Tannuolina/Sinosachites*	
	Nochoroicyathus kokoulini	3	*Warriootacyathus wilkawillinensis*				
	Carinacyathus pinus	2					
	Tetecoscinus zegebarti	1					Cordubian
Tommotian *535 Ma	*Dokidocyathus lenaicus/ Tumuliolynthus primigenlus*	4					
	Dokidocyathus regularis	2/3					
	Nochoroicyathus sunnaginicus	1				*Siphogonuchites/ Paracarinachites*	
Nemakit-Daldynian *545 Ma	*Purella antiqua*	2					Alcudian
	Anabarites trisulcatus	1				*Anabarites/ Protohertzina/ Arthrochites*	

Figure 1 Correlation chart for the major regions of Lower Cambrian rocks. CG, Chengjiang fauna; EB, Emu Bay Shale; MC, Mount Cap Formation; SB, Sinsk fauna; SP, Sirius Passet fauna; RW, Ruin Wash fauna. Modified with permission from Zhuravlev AY and Riding R (2000) *The Ecology of the Cambrian Radiation*. New York: Columbia University Press. © Columbia University Press, 2000.

oxygen profiles and levels in both ocean and atmosphere were essentially modern in aspect.

As noted above, the Cambrian marked a transition interval between 'icehouse' and 'greenhouse' climatic conditions. The details of this transition are poorly resolved, however. Direct evidence of warm conditions in the later Cambrian, derived from oxygen stable isotope ratios, is questioned by some authorities. Yet, the presence of widespread evaporites suggests warm conditions, at least in the equatorial Gondwana region. Broad, but indirect, climatic indicators, such as the distributions of evaporites, ironstones, and carbonates, are consistent with climatic gradients similar to those that characterized much of the Palaeozoic. Recent suggestions of Early Cambrian glacial deposits in West Africa are yet to be accepted generally, but it is likely that the overall climatic conditions warmed through the Cambrian.

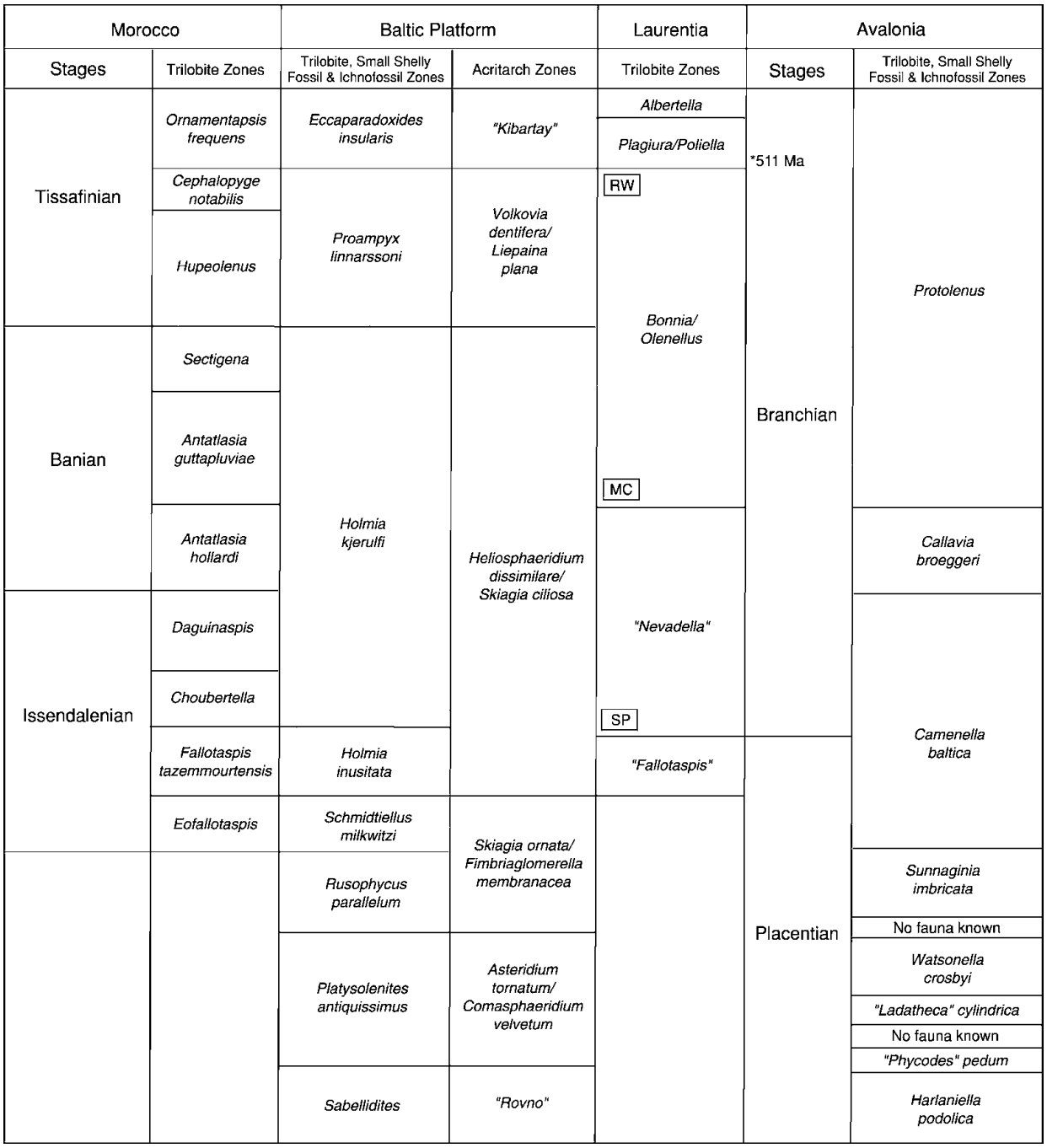


Figure 1 Continued

Information on Cambrian terrestrial conditions is sparse, and the transition between marine and terrestrial sedimentary rocks is difficult to determine in many Cambrian sequences, partly because fine-grained sedimentary particles were easily winnowed in the absence of binding by macroflora. Definitive non-marine sedimentary structures are relatively rare within Cambrian sandstones. Nevertheless, the common tacit assumption of an abiotic terrestrial realm during the Cambrian is probably unrealistic. A diverse microbiota may have existed in some settings. Furthermore, there is abundant trace fossil and sparse body fossil evidence of both diploblastic and triploblastic metazoans active in marginal marine environments, together with spores that are apparently non-marine in origin.

Due to the decreasing rotation speed of the Earth, the Cambrian year would have contained more days

Kazakhstan & Siberia			Australia		China
Ungurian	*Dilelokephalina*	1	Warendian	*Cordylodus lindstromi*	Xingchangian
Batyrbayan	*Euloma limitaris/ Batyraspis*	1	Datsonian	*Cordylodus prolindstromi*	
				Hirsutodontus simplex	
				Cordylodus proavus	
	Lotagnostus hedini	3	Payntonian	*Mictosaukia perplexa*	Fengshanian
	Harpidoides/Troedsonia				
	Lophosaukia			*Neoagnostus quasibilobus/ Shergoldia nomas*	
	Trisclcagnostus trisulcus			*Sinosaukia impages*	
Aksayan	*Eolotagnostus scrobicularis*	2	Iverian	*Lophosukia*	Changshanian
				Rhaptagnostus clarki prolatus/ Cazbnaia sectarix	
	Neoagnostus quadratiformis	1		*Rhaptagnostus clarki patulus/ Caznaia squamosa/ Hapsidocare lilyensis*	
	Oncagnostus ovaliformis			*Peichiashania tertia/ Peichiashania quarta*	
	Oncagnostus kazachstanicus			*Peichiashania secunda/ Peichiashania glabella*	
	Pseudagnostus pseudangustilobus			*Wentsua iota/ Rhaptagnostus apsis*	
Sakian	*Ivshinagnostus ivshini*	3		*Irvingella tropica*	
	Pseudagnostus "curtare"	2	Idamean	*Stigmatoa diloma*	
	Oncagnostus longifrons	1		*Erixanium sentum*	
				Proceratopyge cryptica	
	Glyptagnostus reticulatus			*Glyptagnostus reticulatus*	
Aysokkanian	*Glyptagnostus stolidotus*	6	Mindyallan	*Glyptagnostus stolidotus*	Kushanian
	Agnostus pisiformis	5		*Acmarhachis quasivespa*	
Mayan	*Leiopyge laevigata/ Aldanaspis truncata*	4	Boomerangian *495 Ma	*Erediaspis eretis*	Changhian
				Damesella torosa/ Ascionepea jantrix	
	Anomocarioides limbataeformis	3		*Holteria arepo*	
				Proampyx agra	
				Ptychagnostus cassis	
	Anopolenus henrici/ Corynexochus perforatus	2	Undillan	*Goniagnostus nathorsti*	
				Doryagnostus notalibrae	
				Ptychagnostus punctuosus	
			Late Templetonian/ Floran	*Euagnostus opimus*	
Amgan	*Pseudanomocarina*	3		*Acidusus atavus*	Hsuchuangian
				Triplagnostus gibbus	
	Kounamkites	2		*Xystridura templetonensis/ Redlichia chinensis*	Maochuangian KF
	Schistocephalus	1			

Figure 2 Correlation chart for the major regions of Middle and Upper Cambrian rocks. Global correlations for the top of the Cambrian System are not finalized, but the top of the Sunwaptan Stage of North America approximates the top of the Cambrian in these figures. Rocks above this belong to the Lower Ordovician. BS, Burgess Shale; KF, Kali fauna; MF, Marjum Formation; OR, *orsten*; WF, Wheeler Formation. Modified with permission from Zhuravlev AY and Riding R (2000) *The Ecology of the Cambrian Radiation*. New York: Columbia University Press. © Columbia University Press, 2000.

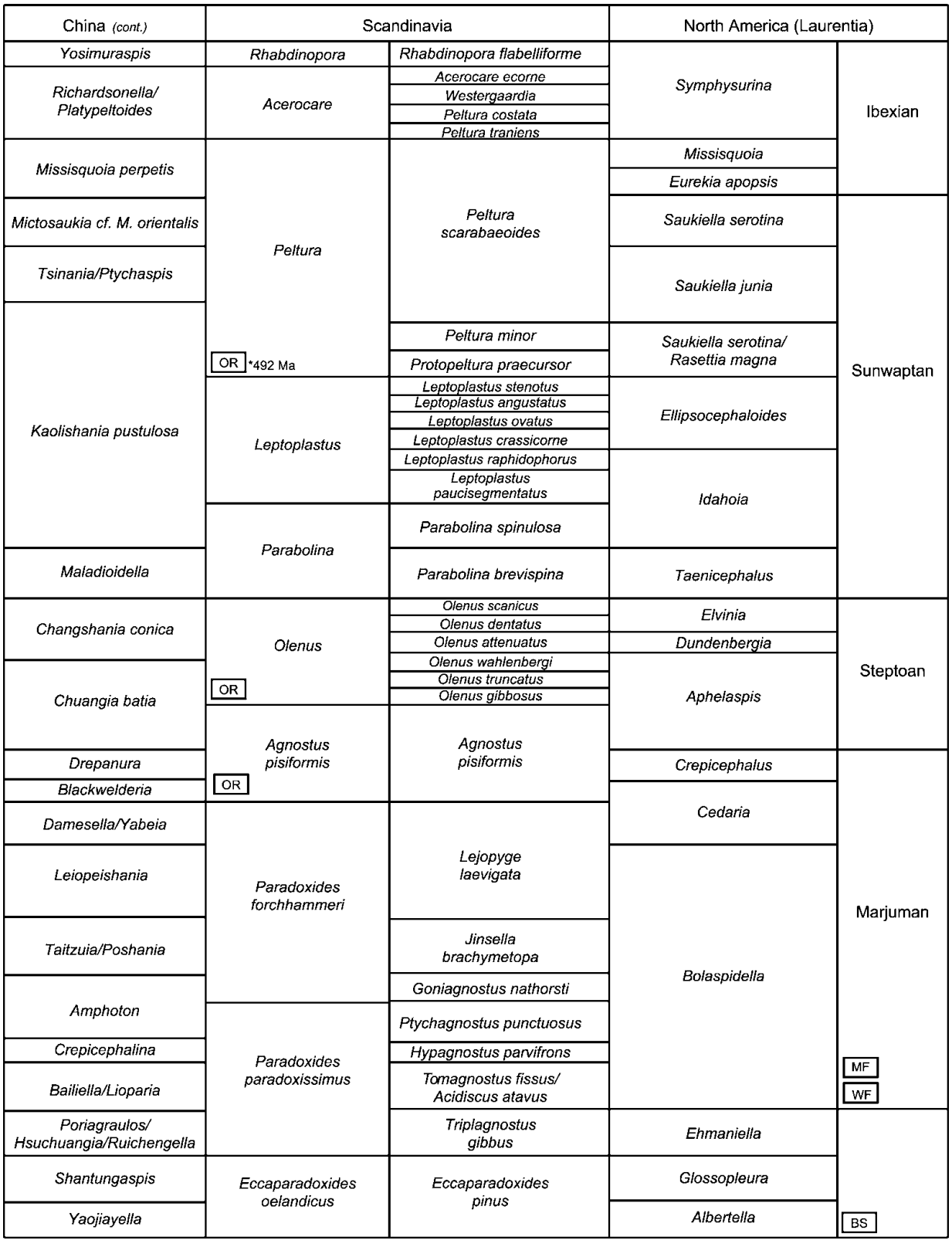


Figure 2 Continued

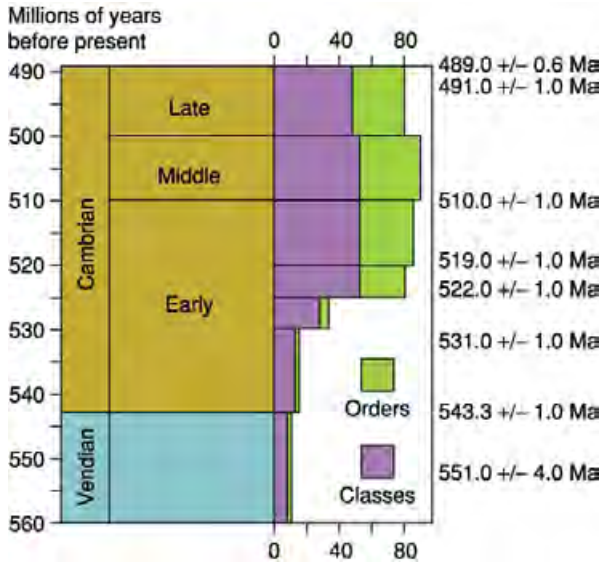


Figure 3 A broad chronology of the Cambrian, and part of the underlying Precambrian Vendian period, showing the major stratigraphical divisions of the system that are commonly used, an approximation of the diversity of metazoan order and classes, and some radiometric dates. Modified with permission from Bowring SA and Erwin DH (1998) A new look at evolutionary rates in deep time: uniting paleontology and high-precision geochronology. *GSA Today* 8: 1–7.

than at present, and each day would have been shorter than 24 h. The Cambrian year was likely to have been about 420 days long. If the shorter distance between the Earth and its moon resulted in a stronger tidal action during the Cambrian, as might also be expected, Cambrian epeiric seas may have been dominated by tides. However, there is no observational support for increased tidal dominance in the Cambrian relative to comparable recent settings. Overall, Cambrian environmental conditions were apparently primarily controlled by the global tectonic setting. The dispersion of the continents, rates of thermal subsidence, and seafloor spreading strongly influenced both the accumulation of sedimentary rocks and their preservation potential. Cambrian environmental conditions were similar to those seen at other parts of the Phanerozoic and essentially modern in aspect. In marked contrast, several aspects of ocean chemistry and stratification during the Neoproterozoic apparently differed significantly from the Phanerozoic conditions. Hence, the Neoproterozoic–Cambrian interval remains a fascinating and important time of environmental transition, in addition to encompassing critical biotic transitions.

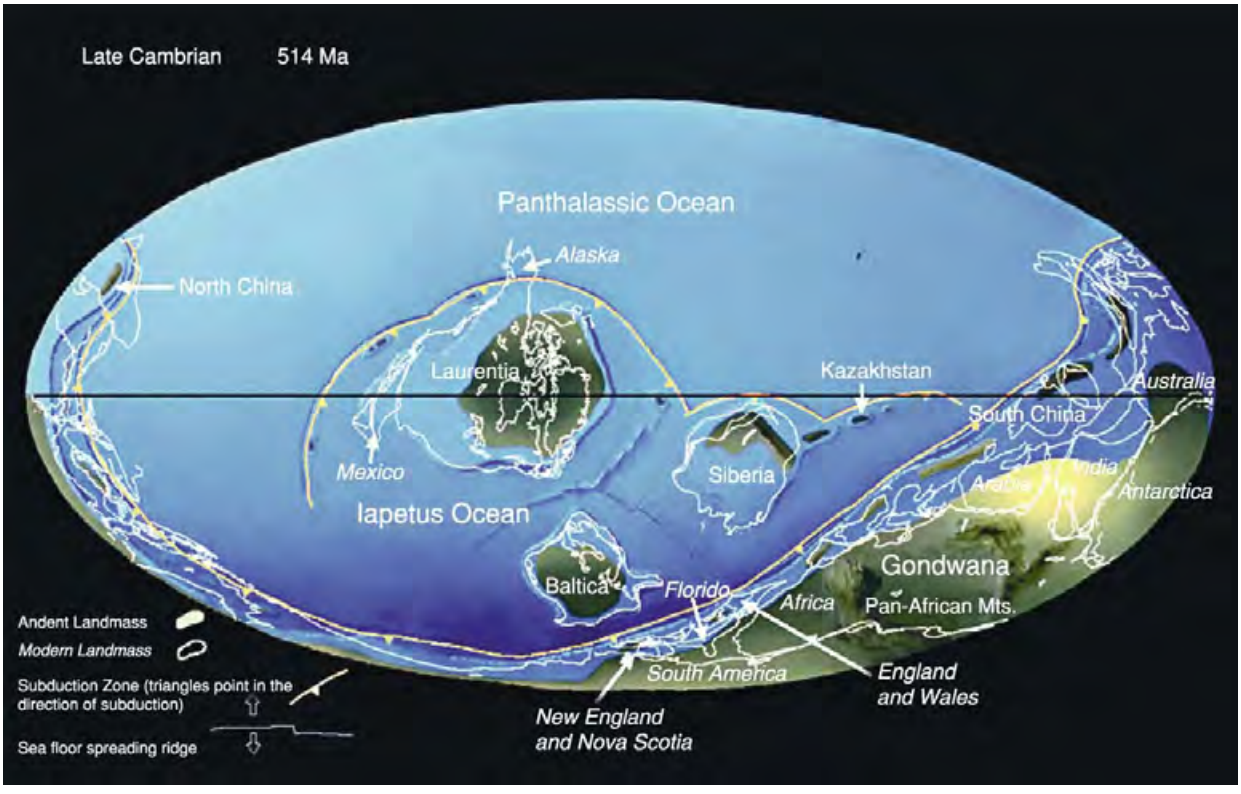


Figure 4 A reconstruction of Cambrian palaeogeography at approximately 514 Ma. Reproduced with permission from Chris Scotese and the Paleomap project.

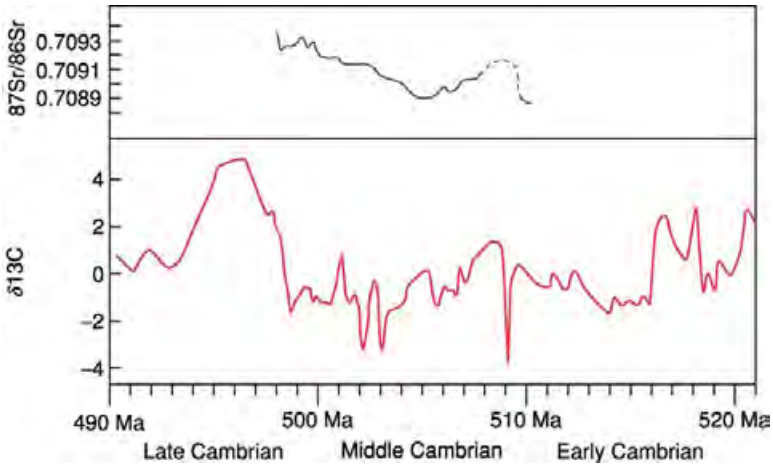


Figure 5 Secular changes in carbon and strontium isotope values during the Cambrian. Modified with permission from Montañez IP, Oselger DA, Banner JL, Mack LE, and Musgrove ML (2000) Evolution of the Sr and C isotope composition of Cambrian oceans. *GSA Today* 10: 1–7.

The Neoproterozoic–Cambrian Evolutionary Radiation

Cambrian Life

The morphological and taxic diversity of trace fossils expanded markedly in association with the basal Cambrian boundary. Surficial trails of simple morphology present in Late Neoproterozoic Ediacaran assemblages were joined by varied morphologies, suggesting motile, triploblastic organisms that mined the sediment infaunally and possessed complex anterior–posterior body patterning. The first traces that can be attributed to the scleritized appendages of arthropods also occurred at about 530 Ma, significantly before the appearance of arthropod body fossils, but after the base of the Cambrian system.

Biomineralized skeletons were present in the latest Neoproterozoic, but were uncommon there relative to their abundance in the Early Cambrian. The pre-trilobitic Cambrian contains a myriad of relatively small biomineralized fossils, collectively known as 'small shelly fossils'. These were commonly composed of calcium carbonate or calcium phosphate and displayed shapes that ranged from simple conical tubes to highly sculptured plates. The diversity of these forms increased through the lowermost Cambrian. Although the phylogenetic relationships of many small shelly fossils remain obscure, in some instances considerable progress has been made in assessing their affinities. Whilst some small shelly fossils represented protective skeletons surrounding the bodies of a variety of diploblastic organisms, others are known to be components of a sclerite meshwork containing hundreds or thousands of separate units that covered portions of the external surface of triploblastic organisms (Figure 6). The synchronous biomineralization of numerous separate metazoan and protist lineages, together with the appearance of Foraminifera that agglutinated inorganic minerals to their external surfaces, suggests that the possession of a reinforced external surface conferred a selective advantage in the Early Cambrian. This pattern, together with evidence from body fossil size and mouthpart anatomy of some Early Cambrian arthropods, suggests that Cambrian marine ecologies were structured with significant feedback between higher and lower food chain levels.

The appearance of trilobites marked the start of the traditional Cambrian fauna, as it has come to be known following Jack Sepkoski's static modelling of Phanerozoic marine diversity. This assemblage of fossils, based on the record of biomineralized fauna, was dominated by certain basal trilobite, linguliform brachiopod, echinoderm, and poriferan clades, and persisted into the lower Ordovician. Although the designations 'Cambrian fauna' and 'Palaeozoic fauna' aptly describe the marked contrast in the composition of biomineralized faunas, dynamic modelling of diversity does not suggest that these two faunas behaved as distinct ecological entities. Moreover, it is clear that basal representatives of many of the clades that rose to dominance in the Ordovician were also present in the Cambrian.

Cambrian benthic macrofossils were also characterized by a periodic series of major evolutionary radiations and extinctions. Several of these extinctions (e.g., those associated with the Early to Middle Cambrian transition) had a cosmopolitan effect,

Figure 6 The Early Cambrian triploblastic metazoan *Halkieria evangelista*. Note the external surface consisting of multiple skeletal elements. Each element individually constitutes a 'small shelly fossil', many of which were combined on a single individual. Photograph: John S. Peel.

during which important groups, such as the archaeocyathid sponges and olenelloid trilobites, were exterminated. Marked episodes of diversity expansion were typically followed by stratigraphically sharp extinctions. These striking cycles of diversity fluctuation occurred repeatedly during the Cambrian, and have been documented most clearly in the Late Cambrian 'biomeres' of Laurentia. Although the mechanisms responsible for the biomere pattern are still debated, it appears that large-scale eustatic changes of sea-level, probably coupled with fluctuations in seawater temperature and oxygen levels, were driving factors.

Cambrian Faunal Provinces

The distributions of Cambrian taxa can be used as tools for both biostratigraphical correlation and palaeogeographical reconstruction. Geologically short-lived, abundant, and cosmopolitan species (e.g., agnostoid trilobites) have great utility for global correlations. Other forms had more restricted distributions, with their geographical ranges limited by barriers (e.g., temperature tolerances, the inability to cross deep ocean basins; Figure 7). Distributions of fossils can be used as independent criteria to constrain palaeogeographical reconstructions based on other data. Laurentian shelf faunas were apparently the most distinctive, which is consistent with the notion that Laurentia was geographically isolated during Cambrian times. A wide-ranging shelf fauna also occurred about the peri-Gondwanan margin, although restriction of some elements to specific regions suggests some palaeolatitudinal limits on faunal distribution. In general, species adapted to cooler waters had more widespread occurrences than those restricted to equatorial shelf environments. Additionally, the occurrence of Cambrian faunas diagnostic of one province surrounded by rocks of another province has proved key to the recognition of Cambrian microcontinents.

The Neoproterozoic–Cambrian Biotic Transition

The traditional criterion for recognizing the base of the Cambrian is the first appearance of macrofossils in the stratigraphical record – the change from 'hidden' to 'obvious' life. Although it is certain that Cambrian global tectonics favoured the accumulation and preservation of fossils, and may even have fuelled evolutionary radiation, the extent to which Cambrian tectonic settings triggered the metazoan radiation is less clear. Clarification of this depends on a knowledge of the nature of biotic diversity and structure prior to the base of the Cambrian. If the clades that first appeared in the Cambrian had phylogenetic origins significantly prior to the base of the Cambrian and had also assumed their 'modern' ecological roles within Neoproterozoic ecosystems, the appearance of abundant fossils in the Cambrian may be related to conditions that favoured both biomineralization and the preservation of shelfal sediments. Alternatively, the Cambrian might chronicle the fundamental steps in metazoan phylogenetic and ecological diversification, in which case clues for the trigger might be sought within Cambrian deposits.

It is now apparent that the base of the Cambrian did not coincide with the phylogenetic appearance of most of the clades, both metazoan and protist, that became common in Cambrian rocks. Although metazoan divergence patterns and times based on molecular data are fuzzy, they are consistent in several major aspects, placing the origins of diploblastic organisms before the triploblasts at least tens of millions of years before the base of the Cambrian. Furthermore,

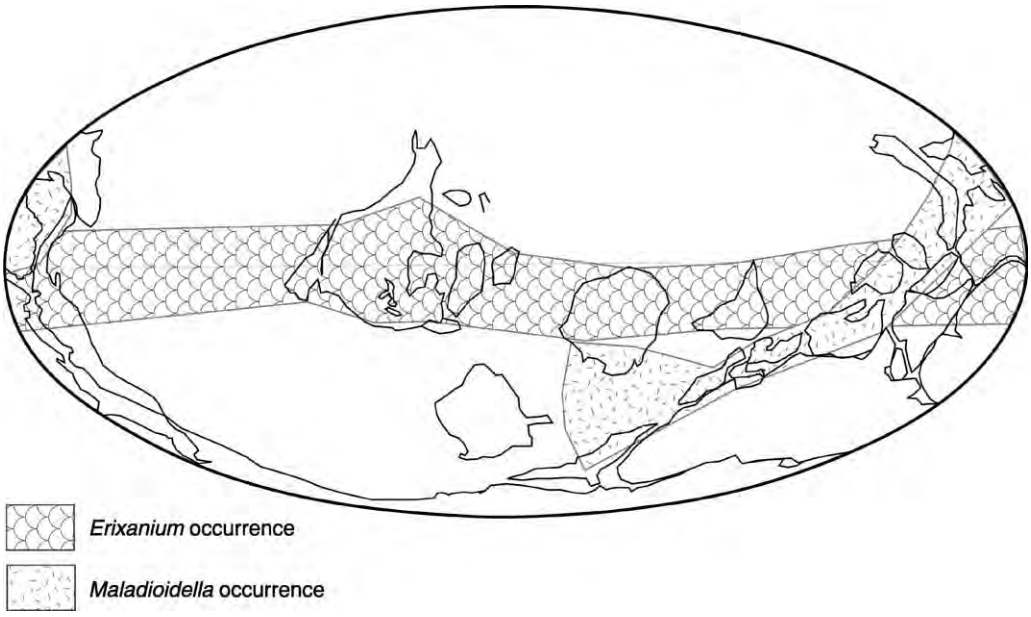


Figure 7 Contrasting biogeographical distributions of two widespread Late Cambrian trilobite genera. The *Erixanium* distribution was apparently constrained by factors related to latitude. In contrast, the *Maladioidella* distribution was apparently limited by open ocean basins. Modified with permission from Zhuravlev AY and Riding R (2000) *The Ecology of the Cambrian Radiation*. New York: Columbia University Press. © Columbia University Press, 2000.

direct fossil evidence of Neoproterozoic metazoans is provided by phosphatized embryos aged about 575 Ma from southern China, in addition to other various late Neoproterozoic fossils such as definitive sponges and possible cnidarians. Evidence for Neoproterozoic triploblasts is still being debated. The phylogenetic affinities of the famous Ediacaran fossil assemblages (*see* **Precambrian:** Vendian and Ediacaran), consisting of large bodies commonly constructed metamerically, also remain hotly contested. Even for some of the best known forms, such as *Dickinsonia*, recent interpretations range from it being a member of an extinct kingdom, a multicellular prokaryote, a lichen, a cnidarian, and a basal triploblastic metazoan capable of movement. Unequivocal Precambrian triploblasts pursuing ecologies closely comparable to their roles in the Phanerozoic have yet to be described, although the Ediacaran *Kimberella* – which may be a mollusc with an associated grazing trail – provides the most plausible candidate.

The presence of likely Late Precambrian metazoans that differed in gross morphology and ecology from their Cambrian counterparts suggests that the transition into 'obvious life' was not merely an artefact related to the advent of biomineralized skeletons with high preservation potential (although biomineralization clearly did appear synchronously and independently during Early Cambrian time in various metazoan and protist clades). Such a suspicion is confirmed by two other data sets that are arguably free of strong preservational biases across the Neoproterozoic–Cambrian boundary. The first of these is the record of trace fossils. Ediacaran fossils are commonly found preserved in sandstones with evidence of mat-like bedding surfaces bound by algae. Finer grained substrates, such as mudstones, were apparently quite firm throughout the Neoproterozoic to Cambrian transition interval. Such substrates would have provided a suitable medium for the preservation of surficial movement traces and shallow, infaunal burrowing. Examination of the trace fossil record shows that the diversity, complexity, depth, and extent of bioturbation increased progressively across the Neoproterozoic–Cambrian boundary, with trends begun in the Neoproterozoic continued into the Early Cambrian. These observations suggest that this interval witnessed important evolutionary innovations in metazoan size, morphology, and behaviour coincident with the establishment of benthic ecologies of essentially modern aspect. The second critical data set is the burst in morphological diversity of pelagic microflora at the onset of the Cambrian following marked conservatism during the late Neoproterozoic. New morphologies found in the Cambrian are consistent with the presence of active pelagic herbivores (mesozooplankton) of which there is direct evidence from Lower

Cambrian rocks. If mesozooplankton were essential to the establishment and maintenance of the Cambrian benthic communities, as they are in modern marine communities, it could be that innovations in the zooplankton drove diversification of the Cambrian macrobenthos.

It seems that the phylogenetic origins of the clades that first appeared in the Cambrian probably predated the base of the Cambrian by at least 50 Ma (and possibly far longer), as did most of the genetic regulatory architecture required to organize large, complex, body plans. This is indicated by the remarkable extent to which body patterning mechanisms are shared amongst disparate metazoan groups. For this reason, it seems unlikely that the Cambrian radiation was directly coupled to basic biological innovations that made metazoan diversity mechanistically possible. Nonetheless, the Cambrian radiation was more than a simple expansion of biotic diversity and ecological structure present in the terminal Neoproterozoic triggered by favourable tectonics. Cambrian ecologies, both pelagic and benthic, were different from those of the Neoproterozoic in ways that suggest enhanced trophic interactions and ecospace utilization. Biotic innovations coincident with the Neoproterozoic–Cambrian transition, both morphological and ecological, appear to have been critical to the adaptive radiation that clearly took place during Cambrian time.

Although there were marked differences between late Neoproterozoic and Cambrian biotas, this does not imply that the transition between the two was sharp, even on a geological time-scale. Both the sequence and apparent spacing of events are consistent with evolutionary processes and mechanisms operative today, although the rates of anatomical and behavioural innovations may have been elevated in some lineages relative to later times. This pattern, and the persistent compatibility of new data with the expectations of the evolutionary model, belies any basis for the idea that the Cambrian radiation requires alternative, non-evolutionary explanations. Furthermore, it can be argued that the types of evolutionary transformation that took place during the Precambrian–Cambrian boundary interval persisted into the Cambrian.

Many Early Cambrian body fossils arguably belong to stem group lineages of clades whose derived crown group representatives only appeared later in the Cambrian or thereafter. According to this view, many varieties of early arthropods present in Cambrian Lagerstätten (*see* **Lagerstätten**) (e.g., Chengjiang in Yunnan, China; the Burgess Shale in British Columbia, Canada) represent steps *en route* to derived euarthropods, and scenarios for the evolution of basic features of arthropod morphology, such as rigid, metameric sclerites, can be postulated on the basis of character state transformations seen amongst Cambrian fossils. Although the specifics of such scenarios are subject to debate, the fact that they can be reasonably postulated affirms that major aspects of arthropod evolution were probably accomplished subsequent to the appearance of the earliest arthropod body fossils. Furthermore, the opportunity to explore the connection between the ontogenetic and phylogenetic development of body patterning in early Palaeozoic arthropods (e.g., Figure 8) suggests that aspects of the basic control of body plan (e.g., the numbers and varieties of segments) evolved during the Early Palaeozoic. Although such opportunities may offer a fascinating glimpse of the evolution of developmental processes, they apparently concern relatively minor adjustment of developmental systems established prior to the base of the Cambrian.

Figure 8 The Middle Cambrian trilobite, *Xystridura templetonensis*, from Australia, a typical representative of the 'Cambrian fauna'. Photograph: Nigel C. Hughes.

See Also

Africa: Pan-African Orogeny. **Analytical Methods:** Geochronological Techniques. **Evolution**. **Famous Geologists:** Sedgwick. **Fossil Invertebrates:** Trilobites. **Lagerstätten**. **Palaeozoic:** Ordovician. **Precambrian:** Vendian and Ediacaran. **Trace Fossils**.

Further Reading

Berner RA (2001) Modeling atmospheric O_2 over Phanerozoic time. *Geochimica et Cosmochimica Acta* 65: 685–694.

Bowring SA and Erwin DH (1998) A new look at evolutionary rates in deep time: uniting paleontology and high-precision geochronology. *GSA Today* 8: 1–7.

Budd GE and Jensen S (2000) A critical reappraisal of the fossil record of the bilaterian phyla. *Biological Reviews* 75: 253–295.

Conway Morris S (1998) *Crucible of Creation*. Oxford: Oxford University Press.

Cooper RA, Nowlan GS, and Williams SH (2001) Global Stratotype Section and Point for base of the Ordovician System. *Episodes* 24: 19–28.

Davidek K, Landing E, Bowring SA, *et al.* (1998) New uppermost Cambrian U–Pb date for Avalonian Wales and age of the Cambrian–Ordovician boundary. *Geological Magazine* 135: 305–309.

Dickson JAD (2002) Fossil echinoderms as monitor of the Mg/Ca ratio of Phanerozoic oceans. *Science* 298: 1222–1223.

Droser ML, Jensen S, and Gehling JG (2002) Trace fossils and substrates of the terminal Proterozoic–Cambrian transition: implications for the record of early bilaterians and sediment mixing. *Proceedings of the National Academy of Sciences, USA* 99: 12 572–12 576.

Gehling JG, Jensen S, Droser ML, Myrow PM, and Narbonne GM (2001) Burrowing below the basal Cambrian GSSP, Fortune Head, Newfoundland. *Geological Magazine* 138: 213–218.

Grotzinger JP, Bowring SA, Saylor BZ, and Kaufman AJ (1995) Biostratigraphic and geochronologic constraints of early animal evolution. *Science* 270: 598–604.

Hughes NC (2003) Trilobite body patterning and the evolution of arthropod tagmosis. *BioEssays* 25: 386–395.

Johnson WJ and Goldstein RH (1993) Cambrian sea water preserved as inclusions in marine low-magnesium calcite cement. *Nature* 362: 335–337.

Landing E, Bowring SA, Davidek KL, *et al.* (2000) Cambrian–Ordovician boundary age and duration of the lowest Ordovician Tremadoc Series based on U–Pb zircon dates from Avalonian Wales. *Geological Magazine* 137: 485–494.

Landing E, Bowring SA, Fortey RA, and Davidek KL (1997) U–Pb zircon date from Avalonian Cape Breton Island and geochronologic calibration of the Early Ordovician. *Canadian Journal of Earth Sciences* 35: 724–730.

Montañez IP, Oselger DA, Banner JL, Mack LE, and Musgrove ML (2000) Evolution of the Sr and C isotope composition of Cambrian oceans. *GSA Today* 10: 1–7.

Narbonne GM, Myrow PM, Landing E, and Anderson MM (1987) A candidate stratotype for the Precambrian–Cambrian boundary, Fortune Head, Bruin Peninsula, southeast Newfoundland. *Canadian Journal of Earth Sciences* 24: 1277–1293.

Sepkoski JJ Jr (1979) A kinetic model of Phanerozoic taxonomic diversity. II. Early Phanerozoic families and multiple equilibria. *Paleobiology* 5: 222–251.

Zhuravlev AY and Riding R (2000) *The Ecology of the Cambrian Radiation*. New York: Columbia University Press.

Ordovician

R A Fortey, The Natural History Museum, London, UK

Copyright 2005, Natural History Museum. All Rights Reserved.

Introduction

The Ordovician System, which is radioisotopically dated to have lasted from 495 Ma to 443 Ma, follows the Cambrian in the geological time-scale and is overlain by the Silurian. The Ordovician comprises the middle part of the Lower Palaeozoic, and has been traditionally divided into six series, of which two (Llanvirn and Llandeilo) have been combined to give a five-fold subdivision. These series are currently in the process of revision, largely to bring them into line with the threefold division (Lower, Middle, and Upper) used in other geological systems. It is still, however, convenient to describe Ordovician rocks in terms of five classical subdivisions, which are in ascending order: Tremadocian (489–483 Ma), Arenigian (483–463 Ma), Llanvirnian (463–*ca*. 455 Ma), Caradocian (*ca*. 455–*ca*. 448 Ma), and Ashgillian (*ca*. 448–443 Ma). Of these, the Tremadocian was previously included in the topmost part of the Cambrian by some authorities, but has now been assigned to the Ordovician by international agreement. The top of the Ordovician is marked by a major ice age and an important mass extinction of marine organisms, which forms a natural boundary with the base of the Silurian. Ordovician rocks are found on every continent except the Indian Peninsula. During the Ordovician the continental masses were widely dispersed, other than the Gondwana supercontinent. Ordovician geology, therefore, varies richly from one area to another. The Ordovician saw the inception of several marine ecosystems that are still with us today.

Several important animal groups (e.g. corals and nautiloid cephalopods) also underwent evolutionary radiations during the Ordovician.

History

The Ordovician System was a comparative latecomer to the geological column. It was born out of a historical compromise between competing claims of what strata should be included in the underlying Cambrian and overlying Silurian systems. The seminal studies of this part of the stratigraphical succession were carried out in the British Isles. Cambrian rocks had been studied in Wales by the Woodwardian Professor at the University of Cambridge, the Reverend Adam Sedgwick (*see* **Famous Geologists:** Sedgwick) – a pioneering field geologist whose classes had been attended by the young Charles Darwin. Rocks comprising the Silurian System were described in considerable detail by Sir Roderick Murchison (*see* **Famous Geologists:** Murchison) in one of the classics of geology, *The Silurian System* (first published in 1839). Murchison was a gentleman of private means and a dominant personality, who had studied rock successions in Shropshire and across southern Wales, where inevitably he overlapped stratigraphically with Sedgwick's ground. Murchison's attempts to 'appropriate' some of Sedgwick's Cambrian rocks into the Silurian met with stern opposition from the Cambridge cleric, and a vigorous controversy ensued, which was still not resolved on the death of both protagonists. A compromise solution to the problem was devised by Charles Lapworth, a Professor in the Department of Geology at the new University of Birmingham. This compromise involved hiving off the disputed middle ground into a separate System, the Ordovician. The name was derived – like that of the Silurian – from the name of an ancient British tribe. Although Lapworth's proposal was made in 1879, it was a long time before it was universally accepted. Indeed, one can find Murchison's 'Silurian' still in use in Russia more than 50 years later. However, as with many of Lapworth's innovations, it came to be seen as a wise solution, and the Ordovician has since been accepted as one of the more natural divisions of geological time.

An outstanding confusion remained, however, about whether to include the rocks around the north Welsh town of Tremadoc in the Cambrian or in the Ordovician. Lapworth had originally stated that the base of the Ordovician should be drawn at the base of the Arenigian rocks, thereby assigning the underlying Tremadocian rocks to the Cambrian. To many observers at the close of the nineteenth century, the Tremadocian seemed to show 'intermediate' features between the two stratigraphical contenders. In Scandinavia, the Tremadocian was incorporated in the Ordovician. The situation was further complicated because in much of North America – where fossil faunas are very different from those in Europe – Tremadocian rocks were not recognized as such, and a different set of names (e.g. 'Ozarkian' and 'Canadian') were applied to rocks of Late Cambrian and Early Ordovician age. Only when correlation between these separate continents improved did it become practicable to apply a standard series of names to the subdivisions of Ordovician time. As late as 1972, the Tremadoc 'series' continued to be allocated to the Cambrian by British scientists, although by then a majority of other nations had settled on a horizon for the base of the Ordovician that approximated the base of the Tremadocian. Such problems lend themselves to arbitration by the International Union of Geological Sciences, and the Tremadocian was eventually deemed to be the basal Ordovician subdivision, by general consensus, in 1988. The horizon chosen to mark the base of the Ordovician also marks the appearance of planktonic graptolites of *Rhabdinopora* (formerly *Dictyonema*) *flabelliformis* type, and thus respects the spirit of Lapworth, who recognized the stratigraphical utility of these ubiquitous fossils well over a century earlier.

Type Areas and Sections

Since Ordovician rocks are so varied, it has not proved possible to define a single 'type area' for subdivision of the System, although the Anglo-Welsh area remains historically important. The base of the Ordovician System has recently been defined by international agreement within the Cow Head Group at Green Point, on the western coast of Newfoundland, Canada. Here, graptolite and conodont fossils occur together in a series of limestones and shales, permitting ready international correlation. In some areas (e.g. the western USA) conodonts are found in the absence of graptolites, whereas the situation is reversed in Spain and northern Africa. The index fossil is the first appearance of the conodont *Iapetognathus fluctivagus*. The base of the Silurian is, by definition, the top of the Ordovician System and has been similarly defined by international agreement in one of Lapworth's classical sections at Dob's Linn, near Moffat, in Scotland. This horizon lies at the base of the *Akidograptus acuminatus* biozone, which is characterized by the eponymous graptolite.

Attempts to define Lower, Middle, and Upper Ordovician Series are in progress. The base of the Middle Ordovician has been mooted, on the basis of conodonts, to lie at a horizon within the Arenig 'Series' of traditional usage. Correlation remains a problem,

however, and to date a type section has yet to be chosen. A standard time division close to the Middle Ordovician, the Darriwilian, has been recognized on the basis of graptolites. The base of the Upper Ordovician has been agreed to be the base of the widespread *Nemagraptus gracilis* biozone. This horizon coincides with the base of the Caradoc 'Series' as now defined. A type section in the Fågelsang section, Sweden, has been selected for its definition. Because this threefold series subdivision has not yet achieved wide currency, it is expedient to continue to use the fivefold subdivision in this chapter. Strictly speaking, these should no longer be referred to formally as 'Series', as they are effectively regional stages. The suffix '-ian' is therefore employed below. Because of the high degree of regional differences in Ordovician strata, several widely separated areas have become important in defining Ordovician temporal subdivisions on a regional to global scale. Biozonal standards have been based on graptolites, conodonts, and trilobites, but these are found together in only a minority of rock sections and they, too, show provinciality. Much effort has accordingly been expended in tying together graptolitic and conodont-based zonal schemes.

Tremadocian

The small town of Tremadoc lies in North Wales at the northern end of Cardigan Bay. A series of dark shales and mudstones of the Dol cyn Afon Formation – often cleaved and weakly metamorphosed – overlie rocks of the Upper Cambrian. There is often a stratigraphical break between the two, but in at least one section a complete and conformable transition between the two systems has been demonstrated. The Tremadocian is recognized by the incoming of net-like rhabdosomes of the widespread graptolite *Rhabdinopora flabelliformis* subspecies, the oldest planktonic graptoloid species. These rocks were deposited in comparatively open shelf environmental conditions, and the graptolites are accompanied by a variety of trilobites, of which members of the Family Olenidae are a prominent component, some species of which are also very widespread. The genus *Jujuyaspis*, for example, is known from North and South America, Scandinavia, and China in earliest Ordovician strata. In North America, the stage name Ibexian (approximately the same as 'Canadian' used by earlier authors) is employed for Early Ordovician strata, based on well-exposed sections in Utah in the Great Basin. The Ibexian embraces all of the Tremadocian and part of the Arenigian. Both the Tremadocian and Ibexian have been further subdivided into chronostratigraphical subdivisions of regional application. The graptolite and conodont succession is particularly well-known in the Baltoscandian area and indicates the presence there of an upper part of the Tremadocian (Hunnebergian) that is not well-developed in Britain.

Arenigian

Arenigian rocks overlie the Tremadocian rocks in North Wales, and were named after a mountain, Arennig Fawr, where they are well exposed. However, in North Wales there is an unconformity at the base of the Arenigian, which is marked by a transgressive sandstone, and so it is inadequate as a type area (Figure 1). The Arenigian succession is better developed in South Wales and Shropshire, where thick sequences of mudstones, shales, sandstones, and turbidites have yielded diverse trilobite faunas of shallow-to-deep shelf aspect with, more locally, graptolites and brachiopods. Because of the disparate biogeography of the Early Ordovician, most of the species are of relatively local occurrence, and correlation of these strata internationally can be difficult. For some years a *de facto* base of the Arenigian has been recognized at the base of the widespread graptolite *Tetragraptus* (*Etagraptus*) *approximatus* biozone, which can be identified in North and South America,

Figure 1 Ordovician (Arenigian) strata lying unconformably over Cambrian rocks on the Llyn Peninsula, North Wales.

Scandinavia, Australia, and New Zealand. The most varied succession of graptolite faunas is found in the shales of the region of Bendigo, Victoria, Australia; this has achieved the status of a global standard, and a fine-scale division into biozones (and stages) has been established, which has been widely applied elsewhere. In south-western China sequences span facies from inshore to deep-water, as do sequences in Scandinavia and the Russian Platform. In North America, strata equivalent to the Arenigian embrace the upper part of the Ibexian and the lower part of the Whiterockian, and have been documented in largely calcareous successions in the Great Basin. In many limestone sequences, conodonts have become the biostratigraphical standard. This disparity of stratigraphical criteria together with regional differences have meant that a plethora of stratigraphical subdivisions of Arenigian strata has grown up over the years, and different regional names are used in, among others, China, Scandinavia, Britain, Australia/New Zealand, and North America. Correlation between these schemes is difficult. However, certain horizons have proved to have international utility. Near the top of the Arenigian, for example, the appearance of biserial graptolites marking the *austrodentatus* biozone has been recognized on most palaeocontinents, and this will form one basis of future standardization.

Llanvirnian

The Llanvirn 'Series' was recognized in the nineteenth century in black slates exposed on the coast in south-western Wales, where it takes its name from an insignificant farmhouse. It is typified by a multitude of 'tuning fork' graptolites (e.g. *Didymograptus artus*, *D. murchisoni*, and related forms), appearing in the Aber Mawr Formation, which have been recognized widely in continental Europe and Scandinavia. The type area is complicated by volcanics and difficult structure, as is the area around Fishguard, and the succession of strata in Britain is better inspected in Shropshire and central South Wales. In Scandinavia, the Upper *Didymograptus* Shales and their equivalents in limestone strata have enabled comparisons to be made with a rich conodont fauna, which provides an international basis for correlation. The Llandeilan (the lower part of the original Llandeilo 'Series') has recently been incorporated into the upper part of the Llanvirnian. The flaggy limestones around the town of Llandilo in central South Wales were already well known by Murchison's time and are among the most fossiliferous rocks of the British Ordovician. Similar problems to those of the Arenigian apply to the international correlation of subdivisions of the Llanvirnian. The Whiterockian of the North American standard includes the Llanvirnian but extends downwards into the Arenigian. Across Europe and Asia the base of the Llanvirnian is marked by a transgressive event, indicated by a deepening in the biofacies and the appearance of graptolites.

Caradocian

The type area of the Caradocian is in Shropshire, England, and it takes its name from a prominent hill, Caer Caradoc, in the vicinity of Church Stretton. In that area the Caradoc is transgressive and its base is defined by an unconformity at the base of the Hoar Edge Grit. As the sequence deepens upwards into a varied succession of mudstones, sandstones, and siltstones (from which several formations are mapped), a great variety of trilobites and brachiopods appear, which were originally studied in detail by B. B. Bancroft. He divided the Caradocian into fine subdivisions ('stages') based on faunal turnover, which, while useful at a local level, are of limited service internationally. In the Shelve Inlier, in South Wales, and near Builth the facies represent deeper water and the sections are without unconformities. However, the effect of the Caradocian transgression is to bring in many graptolite species of stratigraphical utility, including *Dicellograptus* and *Nemagraptus*, replacing the restricted fauna of the Llandeilan *teretiusculus* biozone beneath. This event is of worldwide significance, and the graptolite species concerned can be widely employed for correlation purposes.

The base of the Caradocian (and of the Upper Ordovician) is defined as the base of the *Nemagraptus gracilis* biozone, which can be traced into many areas including the Laurentian platform edge, Bohemia, continental Europe, and the standard graptolitic sequences of Australia and New Zealand. *Nemagraptus gracilis* itself is probably associated with relatively 'oceanic' conditions, and its appearance in basinal successions is probably not entirely synchronous, but the co-occurrence of some of its attendant species of *Dicellograptus* usually places correlation on a sound footing. This, together with the beginnings of a breakdown of the faunal provinciality typifying the earlier Ordovician, means that international correlation of Caradocian strata is less of a problem than is the case with Arenigian and Llanvirnian strata. Nonetheless, there are still distinctive fossil faunas and concomitant separate stratigraphical schemes in the platform limestone successions (Trentonian and Blackriverian) of North America and the Baltic areas. In China, a distinctive diachronous red often nautiloid-rich deep-water formation, the Pagoda Limestone, appears in the Caradocian and continues

into the Ashgillian. North African Gondwanan Caradocian successions include endemic genera adapted to cool-water conditions close to the pole. In deep-water sites, graptolitic successions are well-studied, particularly in the Glenkiln and Hartfell Shales of the Southern Uplands of Scotland, where Lapworth was a pioneer in establishing the succession of faunas. In Wales, the Caradocian was a time of intense volcanic activity in the area around Snowdon, with several kilometres of varied pyroclastic and eruptive rocks, which are classics in the study of 'fossil' volcanoes.

Ashgillian

The youngest, and shortest, subdivision of the Ordovician System (Figure 2) has a type section in the English Lake District, at Ash Gill. A correlative section at Foggy Gill shows continuous passage downwards into the Caradocian. The base of the Ashgillian at the base of the Pusgillian local 'stage' is recognized entirely on a shelly fauna of brachiopods and trilobites. Conodonts and, until recently, graptolites were unrecognized from this locality, and this has made it difficult to locate the base of the Ashgillian in the graptolitic (and conodont) schemes, although a horizon within the *Pleurograptus linearis* biozone has often been accepted. Fossil faunas are not strikingly different from those of the underlying beds and are diverse, but higher in the succession there is a decline in the Rawtheyan local 'stage', presaging a much more drastic reduction as the climate deteriorated at the end of the Ordovician. Shallow-water tropical strata in North America correlative with the Ashgillian, the Cincinnatian limestones, are prodigiously fossiliferous. The stratigraphy of Ashgillian strata in Bohemia, Scandinavia, and China is known in equal detail. There is, in general, a more cosmopolitan fauna in this part of the Ordovician, but cool-water high-latitude faunas still endured close to the pole in northern Africa at the core of the Gondwana supercontinent. Graptolitic deposition continued in deep-water sites distant from continental influences, as in southern Scotland and Taimyr. The youngest subdivision of the Ashgillian, the Hirnantian, marked the onset of the Ordovician glaciation, which is accompanied by drastic lithological and faunal changes in most stratigraphical sections caused by a worldwide regression. In many places there is an unconformity at this horizon, while in all sections spanning the interval there is a reduction in faunal diversity, often accompanied by shallow-water clastic sedimentation.

Figure 2 Late Ordovician (Ashgillian) strata underlying Conway Castle, North Wales.

Life in the Ordovician

The Ordovician is arguably the period of greatest marine diversification in the fossil record. Groups of animals that appeared in the Cambrian (e.g. bivalved molluscs and gastropods) became much more diverse and widespread. Other groups appeared and diversified for the first time. These included corals and bryozoans, which, after a rare occurrence in the Tremadocian, formed diverse bioherms and even reefs before the Caradocian. Cephalopod molluscs – the nautiloids – achieved great size and variety of form, and presumably were among the most effective predators that served to structure the ecosystem. The free-floating graptolites (*see* **Fossil Invertebrates:** Graptolites) became widespread in the world's oceans and are often abundant macrofossils in off-shelf sites where few Cambrian animals can be found. Trilobites (*see* **Fossil Invertebrates:** Trilobites) continued to diversify from the Cambrian and included free-swimming as well as benthic forms, which did not survive into the Silurian. Filter-feeding trilobites were highly characteristic of soft substrates; for example, trinucleid trilobites are abundant in, and confined to, Ordovician strata. Among predatory trilobites, the largest example known from the whole group (*Isotelus*) was discovered in Ordovician strata in Canada.

Articulated brachiopods became numerous enough to form shell beds and evolved into the orders that would dominate the Palaeozoic shallower seas. Echinoderms saw radiations in the starfish, sea lilies, and the earliest sea urchins. By the mid-Ordovician vertebrates with phosphatic external skeletons had become quite varied. Conodonts (now considered to be related to vertebrates) (*see* **Microfossils:** Conodonts) became abundant, and showed a variety of feeding apparatuses. It is likely that the first land plants, allied to liverworts, had appeared by Llanvirnian times, although their fossil record is dominated by spores alone. Moreover, because of the diverse palaeogeography of the time, various groups of animals evolved differently on different palaeocontinents, adding to the total biodiversity. Climatic zonation aided this differentiation. If there was a watershed between Cambrian-style faunas and those characterizing later strata, it occurred at the base of the Whiterockian, when taxa typical of younger strata increased at the expense of those with an archaic history extending back into the Cambrian. It has also been shown that there was an increase in the complexity of trace fossils and bioturbation through the period, with younger Ordovician animals burrowing deeper in the sediment and making 'galleries'. The kind of tiering characteristic of later infaunal assemblages had its origins in the Ordovician. It is possible that increased predation drove these innovations forwards. Allowing for fluctuations in the fossil record, there was an increase in global biodiversity through the Ordovician until the Caradocian (Figure 3). The enrichment process was interrupted by the extinction event at the end of the Ordovician, but communities were able to re-establish comparable complexity in the Silurian.

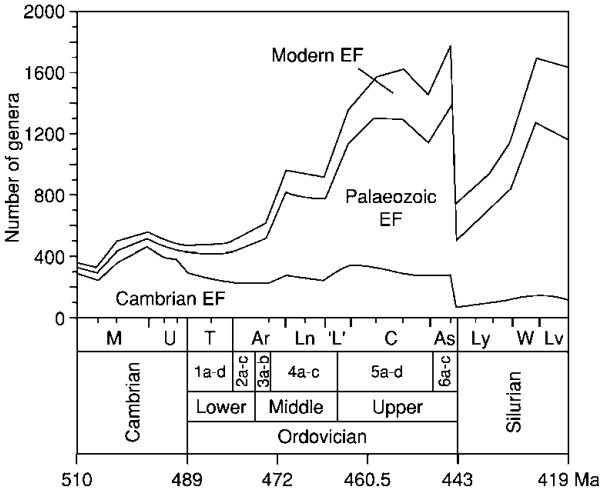


Figure 3 Changes in biodiversity through the Ordovician, showing changes in the diversity of taxa belonging to the different 'evolutionary faunas' (EF) identified by JJ Sepkoski. Maximum diversity is reached late in the Ordovician. The numbers 1a-d, 2a-c etc., refer to finer scale temporal subdivisions used in calibrating this diagram. Reproduced from Webby BDE, Paris F, Droser ML, and Percival IG (eds.) (2004) *The Great Ordovician Biodiversification Event*. New York: Columbia University Press.

The End-Ordovician Extinction and Glaciation

The mass extinction at the end of the Ordovician is one of the major events in the fossil record, being exceeded only by those at the end of the Palaeozoic and at the Cretaceous–Tertiary boundary. While the ultimate controls on the latter events are still a subject of dispute, the end-Ordovician extinction coincided closely with a short-lived but very extensive glaciation, and it is natural to link the biological event with the physical event. Exactly how the secondary effects associated with the glaciation caused the extinction is still the subject of research.

The ice age climaxed in the final subdivision of the Ordovician, the Hirnantian 'stage'. It was evidently short-lived, and current evidence indicates that the Hirnantian may have been as brief as 0.5 Ma. There is good reason to suppose that the climate was deteriorating before the glacial climax. Evidence for glacial deposits is widespread near the Ordovician south pole in North Africa, and they have also been discovered in South America and the Arabian Peninsula. All these areas were part of the Gondwanan continent in the Ordovician, and the scenario of an extensive ice-sheet expanding outwards from a Gondwanan core is supported by evidence on the ground. In Morocco, there are spectacular glacial pavements, which are associated with tillites and finer-grained diamictites. Evidence of ice movement away from the polar centre has been obtained from striae, roches moutonées, and the fining directions of glacial deposits (*see* **Sedimentary Processes:** Glaciers). The glacial event does not seem to have involved multiple advances and retreats, as in the Pleistocene, although a double ice-advance pulse has been claimed. Correlation within non-marine sediments is difficult at this early time, however.

The effects of the glaciation extended well beyond Gondwana. Oceanic water became tied up in the ice-sheet, resulting in a general marine regression. This event was accompanied by climatic cooling, such that the tropical carbonate belt became extremely restricted. This obviously affected the many organisms adapted to warm carbonate environments, which either died out or were pushed into small refugia. Clastic deposits are typical of the Hirnantian, but in many localities the interval was one of general scouring and erosion. A striking faunal change occurs

in the same interval. A low-diversity set of fossil genera known as the *Hirnantia* Fauna is found almost ubiquitously in strata deposited during the regression. *Hirnantia* itself is a brachiopod, and brachiopods are probably most characteristic, but a dalmanitoid trilobite, *Mucronaspis*, is almost invariably found alongside the other shelly fossils. Nearer the palaeoequator a somewhat more diverse fossil fauna can be collected, but there is little sign of the extraordinarily diverse life of the Early Ashgillian. The fact that most members of the *Hirnantia* fauna had a previous Gondwanan pedigree supports the idea that it represents an invasion of cool-water taxa into lower latitudes – a 'crisis' fauna. It even appears in the midst of graptolitic sequences in China, where the sea-level must have dropped sufficiently to allow the colonization of the seafloor by benthos.

This extinction event was much more profound than one just affecting the shelf benthos. The diversity of those animals inhabiting the open ocean was severely reduced. The fossil record is probably better for the planktonic graptolites than for anything else, and they were extinguished to the extent that only two or three species survived through the crisis. These are thought to have given rise to all the Silurian faunas. Oceanic trilobites such as agnostids and olenids, which had endured for tens of millions of years, disappeared completely, along with many of the typical Ordovician families of shelf trilobites (asaphids and trinucleids among them). It has been suggested that these trilobites had planktonic larvae, in which case an oceanic event affecting graptolites and larvae alike is implied. Such an event could have been extensive anoxia. Some, at least, of the shelf faunas were able to lurk in refugia through the crisis, and in some cases they reappeared as Lazarus taxa as late as the Wenlock (Silurian).

Ordovician Geography and Tectonics

The Ordovician was a period when major continents were separated by wide oceans. At the edges of these continents, microcontinents and islands arcs – many of which were subsequently accreted to one continent or another – produced a complex local geography, which was an additional factor in the rich biotic composition of the Ordovician.

The largest continent was Gondwana, which comprised Africa and the southern part of Europe, South America, the Indian and Arabian Peninsulas, Antarctica, and Australia (Figure 4). It extended from the south pole in Africa to north of the palaeoequator in Australia. Much of it, especially central Africa and India, was land throughout the period and largely lacks Ordovician strata. But shallow seas immersed northern Africa, Arabia, and large parts of Australia, and these areas preserve strata rich in fossils. South China was close, if not actually attached, to Gondwana through much of the Ordovician. The kinds of rocks that accumulated reflect the Ordovician palaeolatitudes: mostly clastic rocks near the pole, and predominantly carbonate rocks as the palaeoequator is approached. Spanning the palaeoequator, but well removed from Gondwana, the Laurentian palaeocontinent included what is now North America and Greenland, plus some fragments that lie today on the opposite side of the Atlantic Ocean. Throughout most of the Ordovician, thick deposits of limestones accumulated over platform Laurentia (Figure 5). The Laurentian continent appears to have been stable in

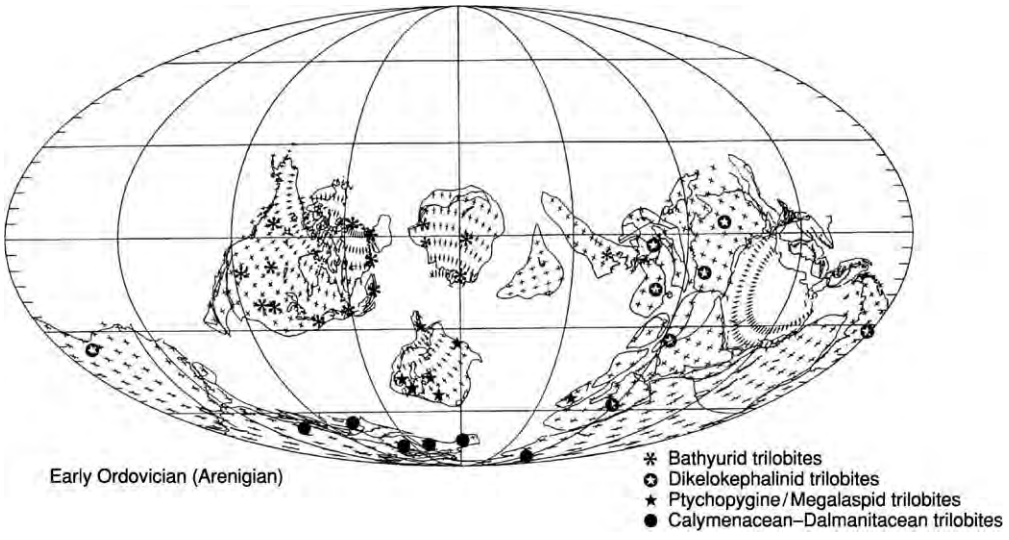


Figure 4 World geography in the Early Ordovician. The distributions of some of the characteristic trilobites are shown.

Figure 5 Limestone strata typical of Ordovician tropical palaeolatitudes: the white limestones of the Meiklejohn Bioherm (Middle Ordovician) near Beatty, Nevada, USA.

Figure 6 A Caradocian world reconstruction showing the migration of the Avalonian Terrane across Iapetus in the middle of the map.

its palaeoequatorial position throughout the Ordovician. The Siberian continent lay between Laurentia and Gondwana. Lying at an intermediate latitude through much of the Ordovician, the continent of Baltica comprised most of present-day Scandinavia and the Russian platform as far east as the Urals. Palaeomagnetic evidence suggests that it rotated between the Middle Cambrian and the Middle Ordovician to attain its current orientation. The palaeocontinents were surrounded by continental shelf and slope deposits or marginal basins related to offshore island arcs. In these, graptolite-bearing shales and mudstones accumulated, and benthic animals different from those inhabiting the shallow shelves were able to prosper. Further seawards again there were oceans, direct evidence for which is provided by obducted ophiolites preserved in mountain belts. These vanished oceans have been given several names: Iapetus separated Laurentia from Baltica and Gondwana; the Tornquist Sea separated Baltica from Gondwana. The oceans separated the major continents sufficiently for each to evolve its own endemic faunas in the earlier Ordovician.

Geography changed continually during the Ordovician, as oceans waned. For example, the Tornquist Sea closed, and Iapetus became much narrower. As this happened, former ocean islands became added to the edges of active continental margins along the subduction zones. The changing geography also facilitated the migration of faunas from one continent to another. Microcontinents migrated across the oceans: Avalonia – comprising the Anglo-Welsh area plus part of what is now eastern North America – moved from Gondwana towards Laurentia, and a new ocean (the Rheic Ocean) opened up behind it (Figure 6). Simultaneously, part of Laurentia drifted away to become accreted to what is now South America before the end of the period. This crustal mobility makes the Ordovician one of the more difficult geological periods for which to reconstruct the geography in detail.

Mountain belts grew along the lines of the subduction zones. The Ordovician saw the early phases of the growth of the Appalachian–Caledonian chain along the western margin of Iapetus. Another mountain belt (now partly buried beneath younger strata) marks the suture between Baltica and continental Europe. Another belt runs along eastern Australia. Volcanic activity along island arcs and in connection with subduction produced varied local rock sequences in different marginal terrains, in which volcanogenic and sedimentary rocks are interbedded. Palaeogeographical (latitudinal) coordinates for these contentious fragments can be determined from palaeomagnetic data, which, combined with evidence from lithology and faunas, contribute to a kinematic description of the shifting palaeogeography. Some mineralization of economic significance is associated with the Ordovician mountain belts. Extensive lead and silver deposits were mined in Wales and elsewhere along the Caledonian chain. Gold is known from many sites close to the main Iapetus oceanic suture in, for example, Ireland and Newfoundland.

Vulcanicity and Geochronology

The high level of tectonic activity during Ordovician times means that there is much evidence of explosive volcanic eruptions that spread ash deposits widely over the seafloor. These are recognized today as bentonites, which usually appear as very soft pale clay beds contrasting with the stratified deposits with which they are associated. In some successions, such as that of the eastern USA, bentonites occur more or

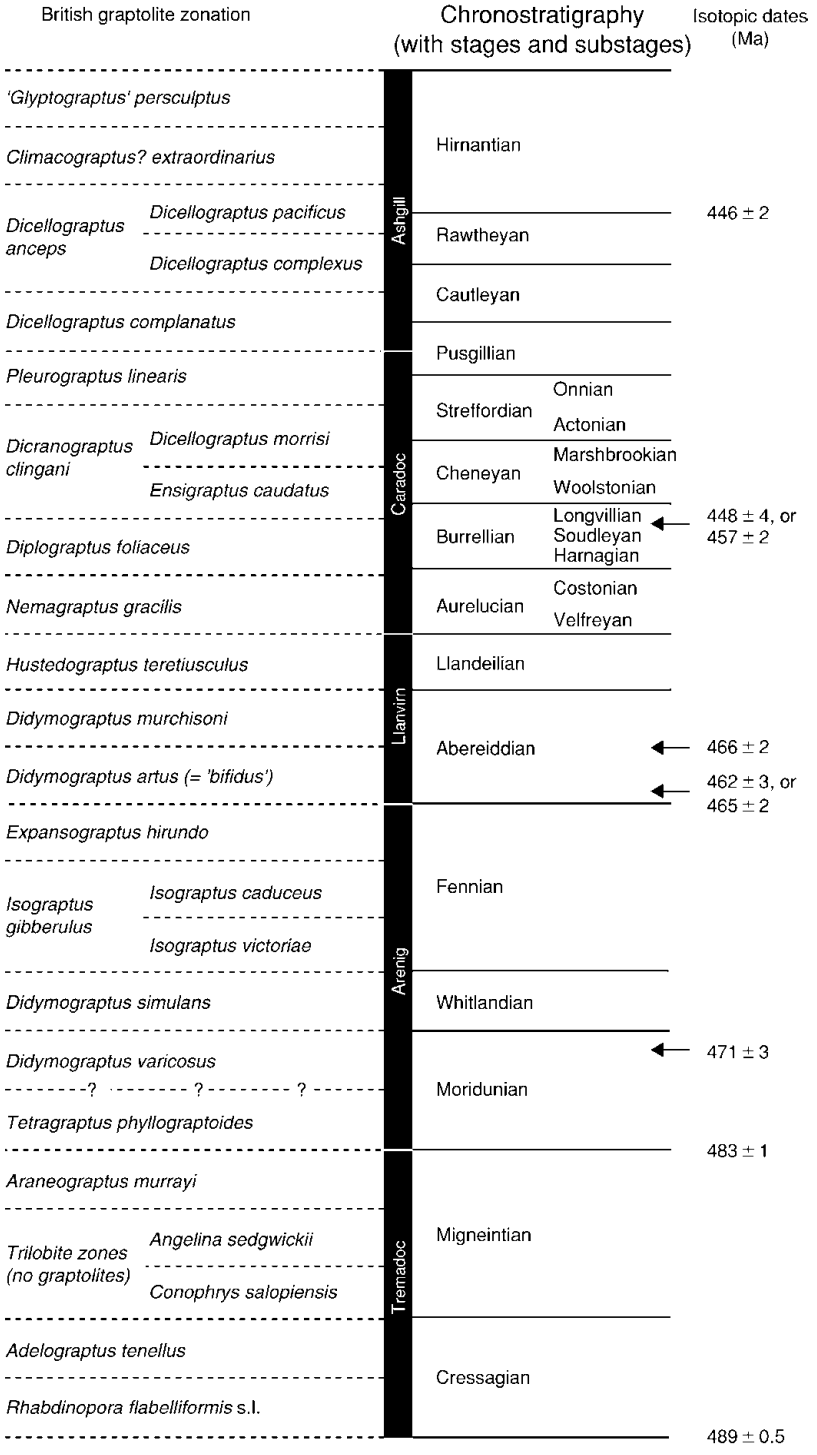


Figure 7 Chronostratigraphy of the Ordovician, showing some of the radioisotopic dates on the right. The traditional series names are employed without the '-ian' suffix in this diagram. Taken from Fortey RA, Harper PAT, Ingham JK, Owen AW, Parks MA, Rushton AWA and Woodcock NH (2000). A revised correlation of Ordovician Rocks in the British Isles. Geological Society. Special Report No. 24, pp. 38.

less regularly through hundreds of metres of strata. The Caradocian 'Big Bentonite' has been claimed to be the product of the explosion of an enormous 'super-volcano', an eruption greater than any that has been recorded in historical times. Bentonites useful in geochronology because they yield datable zircons and are often interbedded with graptolitic deposits, enabling a precise biostratigraphical tie-in (Figure 7). Other

sources of zircons include ash beds and tuffs. Subdivisions within the Ordovician are well-dated with zircon ages, often to an accuracy of ±1.0 Ma.

Biostratigraphy

Historically, biostratigraphical zonation using graptolites was the first scheme to be applied to Ordovician rocks for the purposes of international correlation of strata. Graptolites underwent a successive series of easily recognizable evolutionary changes through the Ordovician, which, even today, give a rapid method of estimating geological age. The 'stipe reduction series' – in which the graptolite fauna as a whole changes from species with many (eight or more) stipes to species with just two – still affords a 'ready reckoner' to distinguish Tremadocian to Llanvirnian strata in the field. Caradocian strata often include numerous Y-shaped colonies of *Dicranograptus*. However, finer zonal subdivision requires the accurate determination of species. Currently, about 25 biozones have been recognized. The solution of limestones to yield conodont fossils provided an additional and refined biostratigraphical system. This was especially useful because it could be applied to the bulk sampling of otherwise unpromising carbonate rocks. Since the 1950s conodont stratigraphy has complemented the graptolite-based schemes, with only slightly lesser refinement. However, such work requires dedicated laboratories. Recently, the stratigraphic utility of organic-walled microfossils – acritarchs and chitinozoa – has been appreciated, especially for dating clastic rocks from which conodonts are difficult to extract. Laboratory facilities capable of handling hydrofluoric acid are essential for this kind of work. These microfossils have proved exceptionally useful in dating rocks from high palaeolatitudes in the Ordovician. Trilobites and brachiopods have been used in local stratigraphy, but individual species are not usually of wide intercontinental geographical distribution, the notable exception being the *Hirnantia* Fauna. For international correlation, sections that include more than one index group are considered ideal: a combination of graptolites and conodonts is especially useful.

See Also

Biozones. **Famous Geologists:** Murchison; Sedgwick. **Fossil Invertebrates:** Trilobites; Graptolites. **Microfossils:** Conodonts. **Palaeomagnetism**. **Palaeozoic:** Cambrian; Silurian. **Sedimentary Processes:** Glaciers.

Further Reading

Bassett MG (ed.) (1976) *The Ordovician System*. Cardiff: University of Wales Press and National Museum of Wales.

Brenchley PJ (1984) Late Ordovician extinctions and their relationship to the Gondwana glaciation. In: Brenchley PJ (ed.) *Fossils and Climate*, pp. 291–316. New York: John Wiley and Sons.

Bruton DL (ed.) (1984) *Aspects of the Ordovician System*. Oslo: Universitetsforlaget.

Cocks LRM and Fortey RA (1982) Faunal evidence for oceanic separations in the Palaeozoic of Britain. *Journal of the Geological Society of London* 139: 467–478.

Fortey RA (1995) Trilobite life habits and Lower Palaeozoic continents. *Geoscientist* 5: 16–17.

Ross RJ (1982) The Ordovician System in the United States: correlation chart and explanatory notes. *Monograph 12*. International Union of Geological Sciences. Denver, Colorado.

Webby BDE, Paris F, Droser ML, and Percival IG (eds.) (2004) *The Great Ordovician Biodiversification Event*. New York: Columbia University Press.

Webby BDE and Laurie J (eds.) (1992) *Global Perspectives on Ordovician Geology*. Rotterdam: Balkema.

Silurian

L R M Cocks, The Natural History Museum, London, UK

Copyright 2005, Natural History Museum. All Rights Reserved.

Introduction

The Silurian System, which is radioisotopically dated to have extended from 443 Ma to 418 Ma (*see* **Time Scale**), follows the Ordovician (*see* **Palaeozoic:** Ordovician) in the geological time-scale and is immediately overlain by the Devonian (*see* **Palaeozoic:** Devonian). The Silurian forms the latest part of the Lower Palaeozoic and is divided into four standard series: the Llandovery (from 443 Ma to 429 Ma), the Wenlock (from 429 Ma to 424 Ma), the Ludlow (from 424 Ma to 420 Ma), and the Pridoli (from 420 Ma to 418 Ma). The type sections for the first three are in Great Britain and the last in the Czech Republic. The base of the Silurian is marked by an unconformity in many parts of the world, since the glaciation that occurred near

the Ordovician–Silurian boundary was responsible for a global lowering of sea-level. The system is well represented in most regions today, apart from Antarctica and peninsular India, which both formed interior parts of the immense supercontinent of Gondwana.

This article describes the history, type areas, other key areas, palaeontology, tectonic activity, climate, and palaeogeography of the Silurian.

History

The Silures were an ancient British tribe who occupied southern Wales before the Roman occupation in the first century AD. The name 'Silurian' was coined by Roderick Impey Murchison (1792–1871) (*see* **Famous Geologists:** Murchison) to denote a geological system in a short paper in 1835. That paper was followed in 1839 by *The Silurian System*, a substantial book that not only described the geology of the Silurian, particularly of Wales and the Welsh Borderland of England, but also contained many illustrations and descriptions of characteristic Silurian strata and fossils. Successive editions, with the shortened title *Siluria*, followed until just after Murchison's death in 1872, by which time the scope of both the book and the system had been extended to cover the whole world (Figures 1 and 2).

The Silurian System was originally envisaged as encompassing the rocks between the Cambrian below and the Devonian above. However, it became increasingly clear that the 'Lower Silurian' of Murchison included the same rocks and geological time as the upper part of Adam Sedgwick's underlying Cambrian System. Sedgwick (*see* **Famous Geologists:** Sedgwick) was originally a friend of Murchison's (they named the Devonian together), but as time progressed fierce arguments broke out between the two men over the scope and definition of both the Cambrian and the Silurian. The matter was not resolved until after both Sedgwick and Murchison had died, when, in 1879, Charles Lapworth erected the Ordovician System to encompass both the upper part of Sedgwick's Cambrian and the lower part of Murchison's Silurian, which had at its base the Llandeilo Flags.

For many years the restricted Silurian was divided into three parts: the lowest was termed the Llandovery (or sometimes the Valentian from the graptolite-dominated faunas in the southern Uplands of Scotland), and this was followed by the Wenlock and the Ludlow. At the type area in the Ludlow area of Shropshire, marine Silurian rocks are conformably overlain by the Old Red Sandstone. At the boundary between the Ludlow and the non-marine Old Red Sandstone is a remanié deposit rich in disarticulated fish remains, known as the Ludlow Bone Bed. This thin bed was, for many years, regarded by British geologists as defining the top of the Silurian. However, in continental Europe, beds stratigraphically higher than the Ludlow Bone Bed equivalent were regarded as Silurian in age, largely because they contained graptolites. It was not until the 1960s that the views of British and continental geologists were reconciled by erecting the Pridoli Series (type area in the Bohemian part of the Czech Republic), which is now internationally accepted as a separate series within the Silurian, above the Ludlow Series and below the Devonian System.

Type Areas and Sections

The base of the Silurian System, which is also the top of the Ordovician System, the base of the Llandovery Series, and the base of the Rhuddanian Stage, has been set by international agreement within the mudstones of the Birkhill Shale Formation at the base of the *Parakidograptus acuminatus* graptolite Biozone (*sensu lato*) at Dob's Linn, near Moffat, Scotland. Some biostratigraphers, however, now divide that zone into a lower *P. ascensus* Biozone and an upper *P. acuminatus* Biozone (*sensu stricto*). The Birkhill Shales, and most of the southern Scottish Uplands in which they occur, were deposited as deep-water ocean-floor deposits and thus represent fairly continuous deposition of sediments, and faunas occurred without substantial break over the Ordovician–Silurian boundary time interval. At Dob's Linn, the underlying latest Ordovician *complexus, extraordinarius*, and *persculptus* graptolite biozones are all represented by diagnostic graptolite faunas, as are the Early Llandovery *atavus, acinaces*, and *cyphus* graptolite biozones above the *acuminatus* Biozone. Fossils other than the abundant graptolites are less common in the Dob's Linn section, but include some inarticulated brachiopods, the blind dalmanitid trilobite *Songxites*, and conodont, acritarch, and chitinozoan microfossils, the last of which are helpful in international correlation.

The Llandovery Series

Llandovery is a town standing on late Ordovician rocks in the Twyi Valley, central Wales, but the low hills to the east of the town are formed from rocks of Early Silurian age. These form the type area of the Llandovery Series (Figure 3). This series is formally divided into three stages, which are, in ascending order, the Rhuddanian, Aeronian, and Telychian. The base of the Rhuddanian, which is synonymous with the base of the Silurian, is formally defined at Dob's Linn, Scotland, some 300 km from Llandovery. Nevertheless, Rhuddanian rocks are well exposed in the type Llandovery area, consisting of up to 250 m of

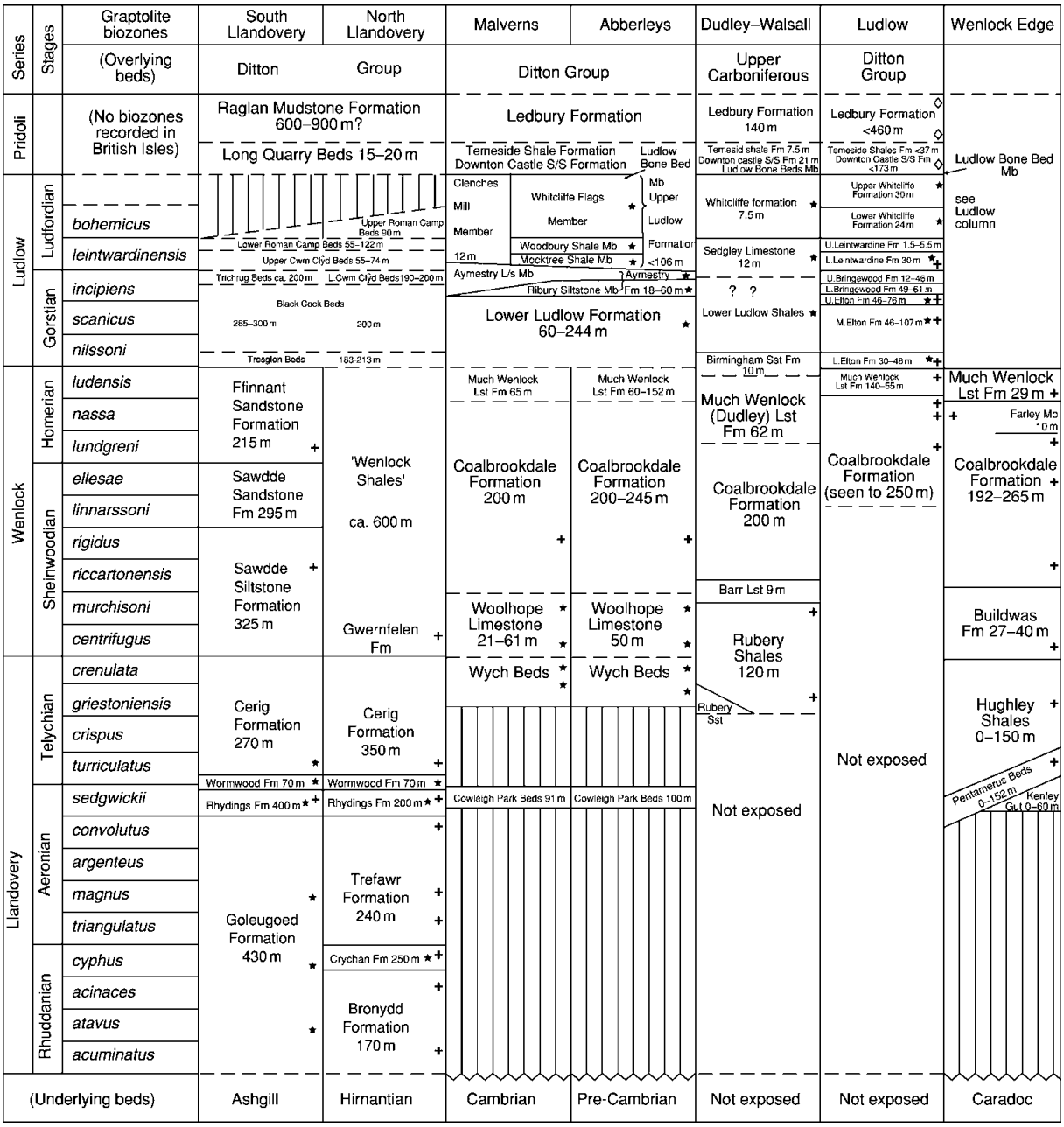


Figure 1 Correlation of key Silurian sections in the British Isles. (Reproduced from Cocks LRM, Holland CH, and Rickards RB (1992) *A Revised Correlation of Silurian Rocks in the British Isles*. Special Report 21. London: Geological Society.)

clastic rocks varying from sandstone through siltstone to shale. These sediments contain a wide variety of fossils. Macrofossils are dominated by brachiopods, including the early members of the *Stricklandia* lineage, whose species are so important for correlating Llandovery and Early Wenlock rocks internationally. Over 35 brachiopod genera are recorded from the Rhuddanian Bronydd, Crychan, and Lower Trefawr formations and their lateral equivalents at Llandovery, together with numerous trilobites, molluscs, and echinoderms and rare corals, bryozoans, and dasyclad algae. Microfossils are represented by conodonts of the *Icriodella discreta–I. deflecta* Assemblage Biozone, as well as by various acritarchs, chitinozoa, and scolecodonts.

The Aeronian Stage has its 'golden spike' base in the Forestry Commission Trefawr track near Cwm Coed Aeron Farm, Llandovery, and its base lies within the blocky mudstones of the Trefawr Formation at a horizon corresponding to the base of the *triangulatus*

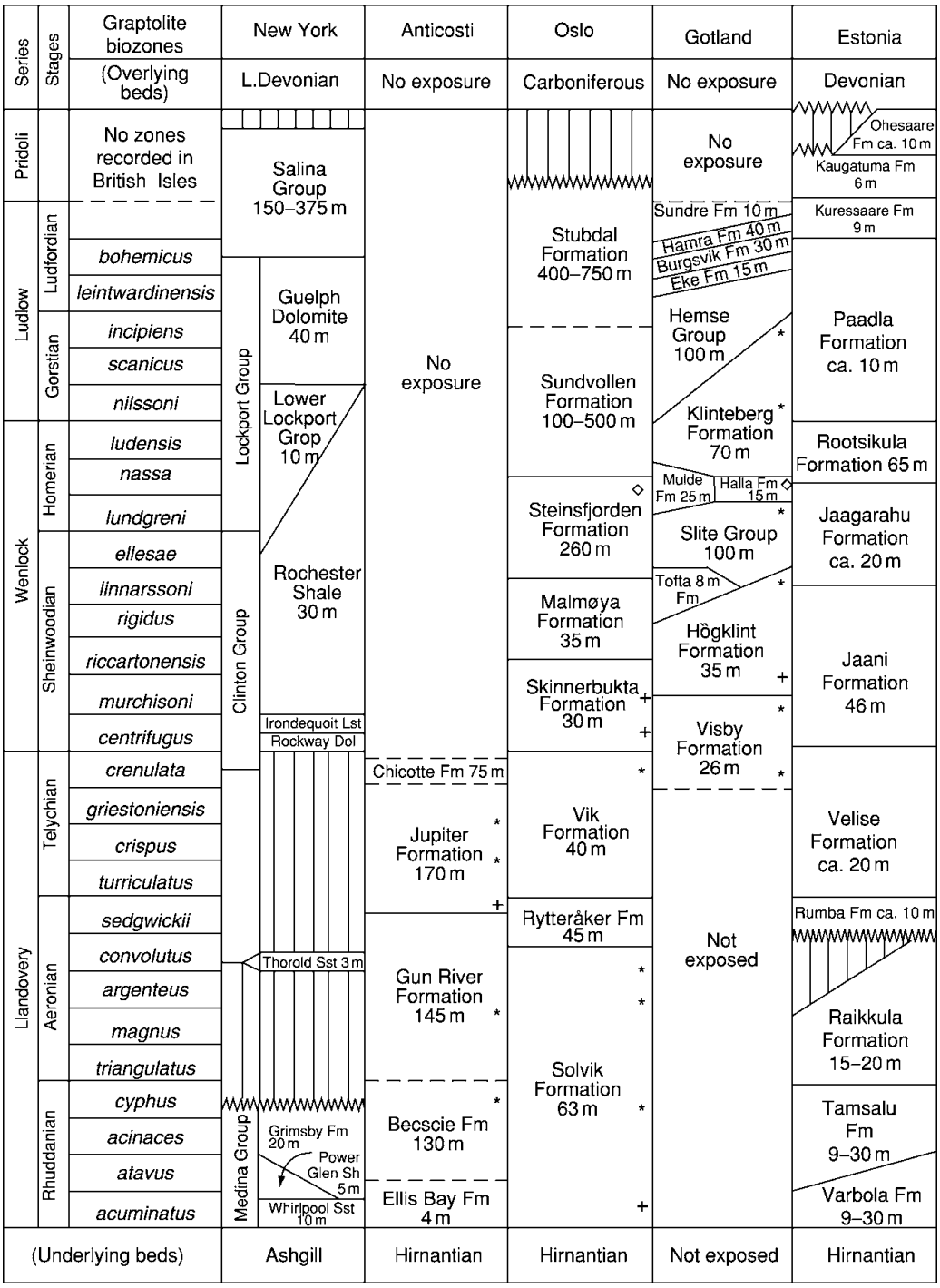


Figure 2 Correlation of key Silurian sections in New York State, Anticosti Island (Canada), the Oslo Region of Norway, the island of Gotland (Sweden), and the northern area of Estonia. (Reproduced from Cocks LRM, Holland CH, and Rickards RB (1992) *A Revised Correlation of Silurian Rocks in the British Isles*. Special Report 21. London: Geological Society.)

graptolite Biozone. The Aeronian is represented at Llandovery by over 450 m of rocks carrying, once again, an abundant macrofauna, microfauna, and microflora, including two more *Stricklandia* brachiopod species. Also present for the first time are two species of the rhynchonellide *Eocoelia*, a brachiopod genus that, in the upper half of the Llandovery, evolved even more quickly than *Stricklandi* and whose species are therefore invaluable for detailed correlation. The Aeronian formations at

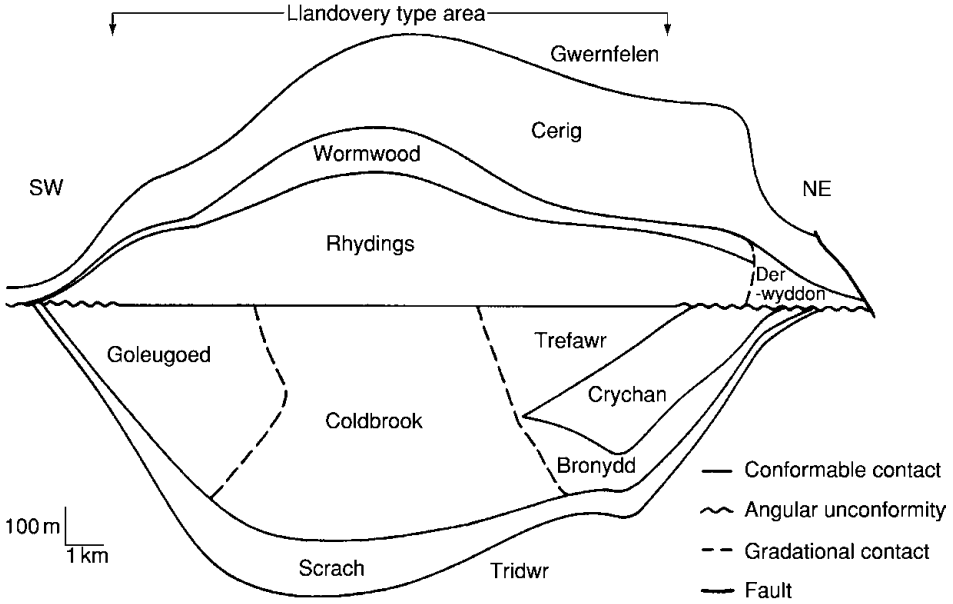


Figure 3 Schematic cross-section of the formations and their thicknesses in the type Llandovery area, Wales. (Reproduced from Cocks LRM, Woodcock NH, Rickards RB, Temple JT, and Lane PD (1984) The Llandovery series of the type area. *Bulletin of the British Museum* (*Natural History*) *Geology* 38: 131–182.)

Llandovery are the Upper Trefawr, Rhydings, and Lower Wormwood Formations and their lateral equivalents.

The Telychian Stage has its formal base within a 30 cm sandstone in a small quarry near Pen-lan-Telych Farm, east of Llandovery. This sandstone occurs just below the top of the Wormwood Formation at a horizon that correlates with the base of the *turriculatus* graptolite Biozone. It is also just above the last occurrence of the brachiopod *Eocoelia intermedia* and just below the earliest record of *Eocoelia curtisi*. The higher levels of the Telychian in the type Llandovery area, the Cerig Formation, consist of over 350 m of monotonous and rather poorly fossiliferous purple mudstones. The thickness of this sequence, along with the fact that the overlying Gwernfelen Formation carries the basal Wenlock *centrifugus* Biozone graptolites, suggests that the great majority of Llandovery time is represented by rocks in the type Llandovery area.

The Wenlock Series

The Wenlock has its type area near the town of Much Wenlock in Shropshire, England, after which the series is named. The Wenlock is divided into two stages, a lower Sheinwoodian Stage and an upper Homerian Stage. The base of the Wenlock Series and the Sheinwoodian Stage is defined in Hughley Brook, Shropshire, where the Llandovery Purple (or Hughley) Shale Formation grades upwards into the overlying Buildwas Formation. This horizon, which is in an area of apparently continuous sedimentation, correlates with the base of the *centrifugus* graptolite Biozone and is accompanied by abundant and superbly preserved microfossils as well as many macrofossils, chiefly brachiopods and molluscs. The Buildwas Formation, which contains sporadic large calcareous nodules, was deposited fairly deep on the Silurian continental shelf and is about 35 m thick. This is followed conformably by the 192–265 m thick Coalbrookdale Mudstone Formation. The base of the type Homerian Stage lies within the Coalbrookdale Formation in a stream section at Whitwell Coppice, Shropshire, at a horizon corresponding to the base of the *lundgreni* graptolite Biozone. The Coalbrookdale Formation becomes more limey near its top, formally termed the Farley Member, and is overlain by the famous Much Wenlock Limestone Formation, which caps the escarpment of Wenlock Edge for more than 30 km. The Wenlock Limestone (as it is more commonly called) is only 29 m thick, but contains massive bioherms exposed in the many quarries along Wenlock Edge. The Wenlock Limestone is very extensive laterally both to the south and east of Shropshire and crops out substantially near Birmingham, where it is known locally as the Dudley Limestone, in particular at a locality named Wren's Nest. The Coalbrookdale Formation carries an abundant macrofauna, especially cephalopods and brachiopods, and varied microfossils, but the Wenlock Limestone fossils are even more varied and are exquisitely

preserved: the trilobites and corals are the best known. The bioherms consist mostly of stromatoporoids and calcareous algae, with subsidiary tabulate corals. Above the Wenlock Limestone on Wenlock Edge the strata pass up without substantial break into the overlying Elton Formation, which is of Ludlow age and extends laterally into the type Ludlow area, some 25 km away.

The Ludlow Series

The mediaeval town of Ludlow lies south-west of Much Wenlock in Shropshire, England, and the low cliffs to the south and west of the town form a shallowly plunging anticline, which includes the type rocks of the Ludlow Series and its two stages, the lower Gorstian and the upper Ludfordian. The formal base of the Ludlow Series and Gorstian Stage is located in Pitch Coppice quarry, 4.5 km south-west of Ludlow, where the Much Wenlock Limestone Formation is succeeded conformably by the Elton Formation. This boundary correlates with the base of the *Neodiversograptus nilssoni* graptolite Biozone. The 110–225 m thick Elton Formation consists largely of blocky fine siltstones, which were deposited in the middle and deeper parts of the shelf and contain many macrofossils, particularly brachiopods, trilobites, and cephalopods, and a large variety of microfossils, particularly conodonts, acritarchs, chitinozoa, and ostracods. Above the Elton Formation is the 62–105 m thick Bringewood Formation, which includes the Aymestrey Limestone of Murchison, famous for its banks of the large pentameride brachiopod *Kirkidium knightii*. The formal base of the Ludfordian Stage is at Sunnyhill Quarry, 2.5 km south-west of Ludlow, where the Bringewood Formation is overlain by the Leintwardine Formation. The base of this unit is defined in beds correlating with the base of the *Saetograptus leintwardinensis* graptolite Biozone. The 35 m thick Leintwardine Formation and the overlying 55 m thick Whitcliffe Formation both consist of siltstones and mudstones with, again, common brachiopods and cephalopods and rich microfaunas and microfloras. At the top of the Whitcliffe Formation is the Ludlow Bone Bed (see above), which is overlain by the Downton Castle Sandstone. The latter can be correlated indirectly with beds of the Pridoli Series in its type section in the Czech Republic.

The Pridoli Series

The Pridoli (technically spelt Přídolí, but the Czech accents are often omitted) Series was of much shorter duration than the other three series of the Silurian and as a result is not subdivided into stages. Its type section is in the disused Požáry Quarry at Řeporyje, in the western suburbs of Prague, Czech Republic. There the Požáry Formation is a uniform facies of dark platy limestones with intercalations of calcareous shale. Brachiopods occur commonly throughout, with abundant cephalopods at certain horizons and rich microfossil floras and faunas. The base of the Pridoli Series is defined at the first appearance of the graptolite *Monograptus parultimus*, a key constituent of the *Monograptus ultimus* graptolite Biozone. This horizon is close to the gradational boundary between the underlying Kopanina Formation and the overlying Požáry Formation. The Pridoli Series lies entirely within one conodont biozone, that of *Ozarkodina eosteinhornensis*, whose first and last appearances fall below and above the Pridoli rocks. The upper limit of the series is about 8 m above the base of the Lochkov Formation, which overlies the Požáry Formation. The top of the Pridoli Series is defined by the internationally agreed base of the Devonian System and its basal Lochkovian Stage, which is taken at the base of the *Monograptus uniformis* graptolite Biozone, whose 'golden spike' is at Klonk, also in the Czech Republic and about 23 km south-west of Požáry.

Other Key Silurian Areas

New York

Niagara Falls, in New York State, USA, and bordering Ontario, Canada, runs over a Silurian escarpment of relatively flat-lying rocks. There the Medina Group, consisting of the Whirlpool Sandstone, the Power Glen Shale, the Grimsby Formation, and the Thorold Sandstone, is of Llandovery age and is overlain paraconformably by the Clinton Group, which is of Wenlock age. The latter unit includes the very fossiliferous Rochester Shale, famous for its trilobites and other invertebrate fossils, and other formations (Figure 4). Above this lie the Lockport Group, of late Wenlock and Ludlow age, and the Salina Group, which includes many anhydrites of latest Ludlow and Pridoli age. Much of the sequence is relatively unfossiliferous, since the limestones have been largely replaced by dolomites.

Norway and Sweden

The graben running north–south through the Oslo region of Norway includes a rich succession of Silurian rocks. The Solvik, Rytteråker, and Vik formations form a Llandovery sequence that conformably overlies the latest Ordovician rocks and underlies the Wenlock rocks. The latter include the Skinnerbukta, Malmøya, and Steinsfjorden formations. The Solvik

Figure 4 Section through most of the Silurian in southern Ontario, Canada, near Niagara Falls. The lower limestones are Late Llandovery and Early Wenlock; they are overlain by the Late Wenlock Rochester Shale Formation, which is in turn overlain by Ludlow, Pridoli and earliest Devonian limestones.

to the Steinsfjorden formations were all deposited in shallow-marine environments on the continental shelf and include abundant brachiopods and other macrofossils and microfossils. Above these, the Late Wenlock and Ludlow Sundvollen and Stubdal formations were deposited under freshwater or continental non-marine conditions. The Pridoli is not known in this region. On the Baltic island of Gotland, Sweden, there are flat-lying outcrops of Late Llandovery to Ludlow rocks, in many cases forming imposing sea cliffs and containing notable bioherms. The visible sequence starts with the Visby Marl, which spans the Llandovery–Wenlock boundary, and this is succeeded by the bioherms of the Högklint Formation. Ten further Wenlock and Ludlow formations follow, including the Klinteberg Formation, which spans the Wenlock–Ludlow boundary and is famous for its reefs of the pentameride brachiopod *Conchidium*, originally described by Linnaeus in the early eighteenth century.

Anticosti

Anticosti Island lies near the mouth of the St Lawrence River, Canada, and is nearly 100 km long, with a relatively flat-lying Late Ordovician and Early Silurian succession. The Silurian follows the latest Ordovician with only a minor break within the Ellis Bay Formation, which is followed by the Becscie, Gun River, Jupiter, and Chicotte formations, all of which are of Llandovery age. These rocks consist of alternating limestones and shales, all deposited on a shallow-marine continental shelf and yielding excellently preserved macrofossils (particularly brachiopods, corals, and bryozoans) and microfossils (particularly condonts, ostracods, acritarchs, and chitinozoans). Since graptolites are poorly represented, the principal correlations have been achieved by using conodont biozones.

Podolia

The area bordering the Dnestr River has been part of both Poland and the USSR in the past and is now in the west of Ukraine. A very thin band of Middle Llandovery rocks lies unconformably on Middle Ordovician rocks and is followed paraconformably by the Wenlock Kitaigorod Formation, the Bagovitsa Formation, which spans the Wenlock–Ludlow boundary, the Ludlow Malmovtsy Formation, and the Pridoli Rashkov and Dzwinogorod formations, which are themselves followed conformably by earliest Devonian strata. All are marine, with many limestones but much clastic fine-grained input, and have yielded very abundant macrofossils and microfossils. The latest Silurian rocks, locally termed the Skala Stage, were a strong candidate for the stratotype of the latest formal series of the Silurian, but the vote went to the Pridoli area of the Czech Republic.

Methods of Dating and Correlation

The system is well-constrained by radioisotopic ages down to the level of the series (see above for the dates in Ma), but within the series the dating and correlation are achieved by using quickly evolving fossils. The best of these are the graptolites, upon which more than 30 biozones have been based, covering the entire Silurian. Because of their planktonic mode of life, these quickly evolving lineages dispersed very rapidly, enabling correlation across all continents and terranes. However, graptolites are often uncommon or even absent in rocks deposited under very shallow-water conditions, and, in these rocks, microfossils, particularly conodonts, are used. However, there are often four or five successive graptolite biozones during the period of a single Silurian conodont biozone. Conodonts, acritarchs, chitinozoans, and spores can often be recovered from small samples, for example the rock chips produced through boring by

the hydrocarbon industry. In the Early Silurian (Llandovery and Early Wenlock), quickly evolving brachiopods, such as *Stricklandia* and *Eocoelia* (mentioned above), are also used for local and international correlation.

Life in the Silurian

Within the marine realm, invertebrates were abundant and diverse throughout the Silurian. The benthos was dominated by brachiopods, bivalves, gastropods, corals, stromatoporoids, bryozoans, various echinoderms (particularly crinoids), calcareous algae, and trilobites. The plankton included cephalopods (most of which swam), graptolites, and a variety of microplankton, particularly acritarchs, chitinozoans, and spores. Ostracods and conodonts were certainly mobile and in some cases dispersed quickly, but their detailed ecologies and modes of life are less certain. Fishes became progressively more abundant during the period. There is debate as to whether all Silurian fishes were marine or whether some were entirely freshwater: some fish fossils are to be found in apparently non-marine rocks. Fossils of land-living creatures are very rare, and the land was clearly not yet much colonized. Various groups of arthropods probably lived onshore at the time, as did some algae whose remains are equivocally preserved. Judging by the occurrence of fossil spores in marine rocks, some land plants probably lived as early as the Ordovician, but the oldest land plants yet represented by convincing body fossils are from rocks of Middle Silurian (Wenlock) age.

Tectonic Activity

The Silurian was considerably shorter than either the Ordovician or the Devonian, and thus there was a smaller variety of tectonic activity during the duration of the System. The docking between Avalonia and Baltica at the start of the period was relatively mild tectonically, since the collision was oblique and therefore relatively 'soft' in tectonic terms. Subsequently, during most if not all of Silurian time, there

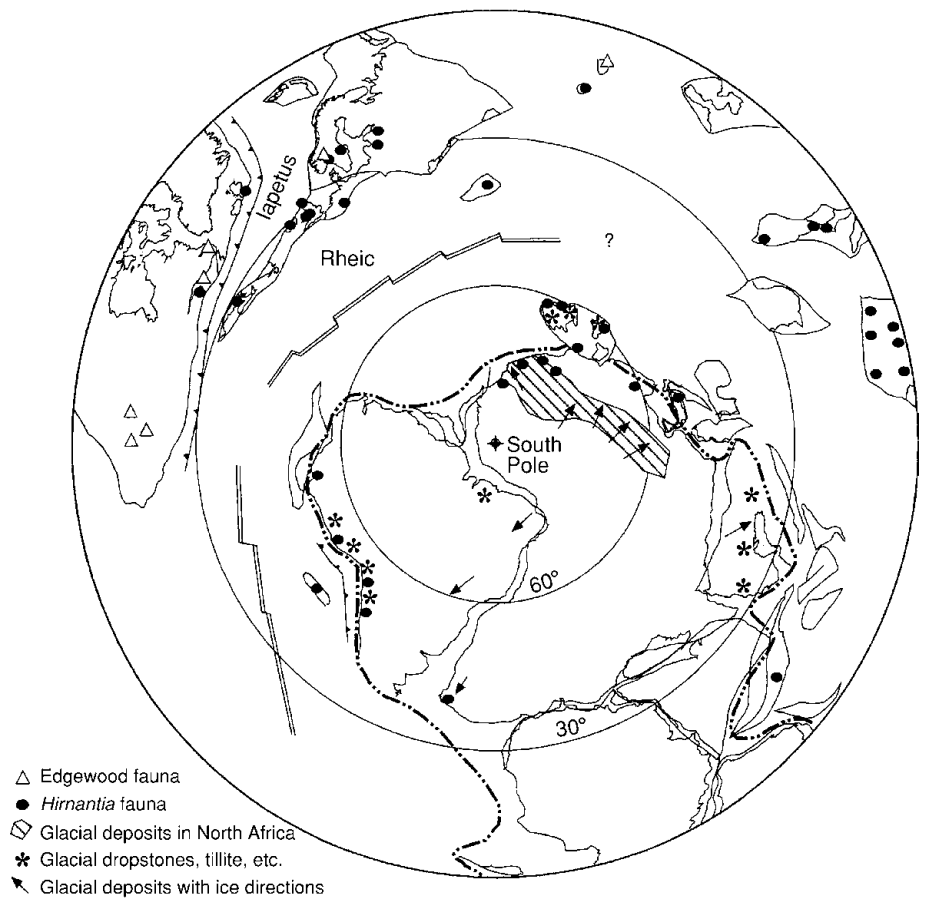


Figure 5 Palaeogeography at the Ordovician–Silurian boundary (approximately 440 Ma), showing the distribution of the latest Ordovician glacial deposits and the *Hirnantia* brachiopod fauna. (Reproduced from Cocks LRM and Torsvik TH (2002) Earth geography from 500 to 400 million years ago: a faunal and palaeomagnetic review. *Journal of the Geological Society, London* 159: 631–644.)

was a major collision between Avalonia–Baltica and Laurentia, resulting in a major tectonic event termed the Caledonian Orogeny (*see* **Europe:** Caledonides of Britain and Ireland; Scandinavian Caledonides (with Greenland)), which led to mountain building and metamorphism along a belt that today embraces all of the North Atlantic rim from the eastern USA (Carolina) to Norway, including eastern North America, Ireland and Britain, and the Scandinavian Caledonides of Norway and Sweden. Away from this area, there appears to have been less tectonism, apart from in today's central Asia, where a number of small terranes and their marginal island arcs were breaking up and regrouping, particularly in Kazakhstan and adjacent countries. Around the vast Gondwanan margin there were also island arcs and tectonic activity in south-east Australia (Victoria and New South Wales) and New Zealand.

Palaeogeography and Climate

Silurian geography was dominated by the vast supercontinent of Gondwana (*see* **Gondwanaland and Gondwana**), which included today's South America, Africa, peninsular India, Antarctica, and Australasia, as well as adjacent terranes in southern Europe and the Middle and Far East, which may or may not have been physically attached to the supercontinent (Figure 5). Other substantial terranes were Laurentia (which included most of North America, Greenland, and the easternmost part of Siberia), Siberia (which was only part of today's political entity), and Baltica–Avalonia. Baltica, which today comprises most of northern Europe eastwards to the Ural Mountains, had been an independent terrane until about the Ordovician–Silurian boundary (443 Ma), when it collided with Avalonia, whose long narrow

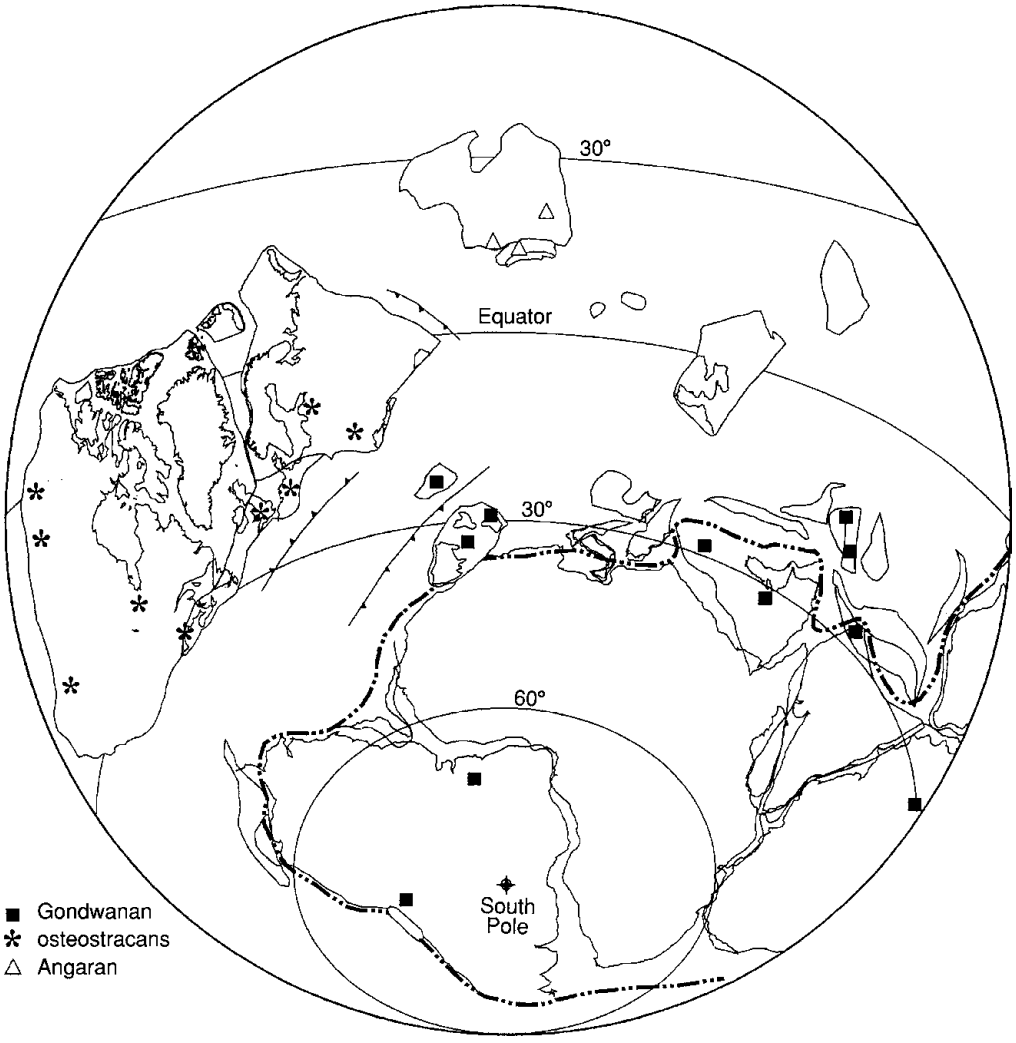


Figure 6 Palaeogeography at the Silurian–Devonian boundary (approximately 420 Ma), showing the distribution of the earliest Devonian fishes. (Reproduced from Cocks LRM and Torsvik TH (2002) Earth geography from 500 to 400 million years ago: a faunal and palaeomagnetic review. *Journal of the Geological Society, London* 159: 631–644.)

area today includes the Maritime States of Canada and the USA as far south as Cape Cod, Massachusetts, eastern Newfoundland, southern Ireland, England, Wales, Belgium, Holland, and parts of northwestern Germany. North and South China were also independent terranes, as was Sibumasu, which stretched from Burma through Thailand and western Malaysia to Sumatra in Indonesia. There were also many smaller terranes, including Perunica (Bohemia, Czech Republic).

The oceans were dominated by the vast Panthalassic Ocean, which was comparable in size to today's Pacific. The Iapetus Ocean between Laurentia and Avalonia–Baltica, which had been very substantial in Cambrian and Ordovician times, was steadily closing. During the Late Ordovician some of the island arcs within Iapetus collided with and accreted to the eastern margin of Laurentia or the western margin of Avalonia–Baltica. The Silurian saw the major continent–continent collision between Laurentia and Avalonia–Baltica, which resulted in the closure of Iapetus and the Caledonian Orogeny (see above). This orogeny caused considerable uplift and formed the land supporting the Old Red Sandstone Continent in the Late Silurian and Devonian (Figure 6). The combined terrane that resulted from the Laurentia–Avalonia–Baltica collision is termed Laurussia. To the south of the Avalonian part of Laurussia the widening ocean between it and Gondwana is termed the Rheic Ocean.

Over a vast area of the continental shelf of Gondwana, today including France, Bohemia, the Iberian Peninsula, Italy, and much of northern Africa, anoxic conditions prevailed, precluding the colonization by benthos. The fossils preserved in the black shales deposited there consist largely of plankton, including graptolites and cephalopods. It was not until Late Silurian times in some of these areas, for example Bohemia, that the seafloor became oxygenated, enabling its colonization by brachiopods, crinoids, and other benthos.

The climate of the Silurian started cold. The latest Ordovician (Hirnantian) ice-cap was very widespread, but gradually receded in Llandovery times; the latest known Silurian glacial sediments are in the early Wenlock of Brazil, which was near the then south pole. It is not until the Late Llandovery (Telychian) that warmer-water carbonate deposits become common, but from then on into the Wenlock, Ludlow, and Pridoli the global temperature was obviously warm to hot. This can be deduced from the carbonate build-ups and bioherms present in the Wenlock type area of England and in Gotland and elsewhere and also from the Late Silurian evaporite deposits, which are widespread but best known from New York State, USA.

See Also

Europe: Caledonides of Britain and Ireland; Scandinavian Caledonides (with Greenland). **Famous Geologists:** Murchison; Sedgwick. **Fossil Invertebrates:** Brachiopods; Graptolites. **Gondwanaland and Gondwana**. **Palaeozoic:** Ordovician; Devonian. **Time Scale**.

Further Reading

Bassett MG, Cocks LRM, Holland CH, Rickards RB, and Warren PT (1975) The type Wenlock Series. *Institute of Geological Sciences Report* 77(13): 1–33.

Cocks LRM and Rickards RB (eds.) (1988) *A global analysis of the Ordovician–Silurian boundary. Bulletin of the British Museum (Natural History) Geology* 43: 1–394.

Cocks LRM and Torsvik TH (2002) Earth geography from 500 to 400 million years ago: a faunal and palaeomagnetic review. *Journal of the Geological Society, London* 159: 631–644.

Cocks LRM, Woodcock NH, Rickards RB, Temple JT, and Lane PD (1984) The Llandovery series of the type area. *Bulletin of the British Museum (Natural History) Geology* 38: 131–182.

Cocks LRM, Holland CH, and Rickards RB (1992) *A Revised Correlation of Silurian Rocks in the British Isles*. Special Report 21. London: Geological Society.

Holland CH and Bassett MG (eds.) (1989) *A global standard for the Silurian System National Museum of Wales Geological Series* 9: 1–325.

Holland CH, Lawson JD, and Walmsley VG (1963) The Silurian rocks of the Ludlow District, Shropshire. *Bulletin of the British Museum (Natural History) Geology* 8: 93–171.

Landing E and Johnson ME (eds.) (2003) Silurian lands and seas: paleogeography outside of Laurentia. *New York State Museum Bulletin* 493: 1–400.

Martinsson A (ed.) (1977) *The Silurian–Devonian Boundary*. Stuttgart: Schweizerbart'sche.

Sengor AMC and Natalin BA (1996) Paleotectonics of Asia: Fragments of a synthesis. In: Yin A and Harrison M (eds.) *The Tectonic Evolution of Asia*, pp. 486–640. Cambridge: Cambridge University Press.

Devonian

G R McGhee, Rutgers University, New Brunswick, NJ, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Devonian Period of the Palaeozoic Era was only slightly shorter than the entire Cenozoic Era (63 million years as opposed to 65) and was a time of many 'firsts' in Earth history. The Devonian saw the explosive diversification of vertebrate life (e.g., new fish species that proliferated in both marine and freshwater environments) including the evolution of the first terrestrial vertebrates, the amphibians. In the oceans, the Devonian saw the evolution of the largest reef ecosystems in Earth history. Both plant and animal groups were rapidly evolving and invading the terrestrial realm, and in the Devonian the Earth's first forests evolved with trees that towered some 30 metres high. Yet it is also in the Devonian that the Earth suffered one of its greatest biodiversity crises, one of the 'Big Five' mass extinctions of life, and it is in the Devonian that the entire climate of the planet switched from a hot greenhouse state to a cold icehouse phase.

The Devonian System is named after fossiliferous marine strata exposed in Devonshire, southwest England. The English palaeontologists Adam Sedgwick (*see* **Famous Geologists:** Sedgwick) and Roderick Murchinson (*see* **Famous Geologists:** Murchison) first proposed that the fossils in Devon were younger than Silurian fossils, yet older than Carboniferous, in 1839. Subsequent stratigraphic work demonstrated that the Old Red Sandstone of Scotland and Wales, long thought to lie in the basal Carboniferous, was the non-marine facies equivalent of Devonian marine strata. The Devonian Period is divided into seven Ages, from oldest to youngest: the Lochkovian, Pragian, and Emsian (constituting the Early Devonian), the Eifelian and Givetian (Middle Devonian), and the Frasnian and Famennian (Late Devonian). Lochkovian and Pragian are in the Czech Republic, Emsian and Eifelian in Germany, and Givetian, Frasnian and Famennian in Belgium.

Life in the Sea

The Devonian was a time of major evolutionary innovation in the marine realm. The Devonian has been called the 'Age of Fishes', as it is in this period that the evolution of the major fish groups of today's oceans occurred (*see* **Fossil Vertebrates:** Fish). The explosive diversification of bony, ray-finned fishes (the Actinopterygii) took place in the Devonian, as did the origination of the first true sharks. Of particular note is the origination of the first lobe-finned fishes (the Sarcopterygii) as this group later gave rise to the amphibians, also in the Devonian. The ammonoids – externally shelled, swimming cephalopods distantly related to the modern-day chambered nautilus – first evolved in Devonian seas (*see* **Fossil Invertebrates:** Cephalopods (Other Than Ammonites)). These animals were active swimmers with widespread geographic ranges and possessed a characteristic high rate of speciation, two traits that made them valuable for biostratigraphic dating of sedimentary rocks for over 300 million years before their lineage perished in the Cretaceous–Tertiary mass extinction.

The largest development of reefal ecosystems in Earth history occurred in the Devonian. The great Devonian reefs are estimated to have covered over 5 000 000 square kilometres of seafloor at their maximum development, almost ten times the areal extent of reefal ecosystems seen in modern oceans. The Devonian was also the 'Golden Age of Biconvex Brachiopods' (*see* **Fossil Invertebrates:** Brachiopods), the period of maximum diversification of these 'shellfish of the Palaeozoic', an ecological role that is dominated by the bivalve molluscs today.

We are terrestrial animals, and we tend to think of life in terms of life on land, rather than in the seas. It was during the Devonian that life staged its dramatic invasion of the land, and that the origination of the first complex terrestrial ecosystems occurred.

The Invasion of the Land

The Plant Invasion

Life existed in the Earth's oceans for over three billion years before finally venturing out onto the harsh environments of dry land. The biological invasion of the land has deep roots, beginning with terrestrial autotrophic microbes in the Cambrian and progressing to simple mosses and liverworts in the Ordovician and Silurian. It was only in the Late Silurian to Early Devonian, however, that life's invasion of the land dramatically accelerated with the evolution of vascular land plants.

In the Lochkovian vascular land plants were confined mostly to lowland floodplains and stream margins. These plants were simple in structure, many with short vertical stems sometimes covered

with scale-like leaves (resembling asparagus spears), others with short vertical stems that branched into progressively smaller forks towards the top of the plant. The tallest of these plants was less than half a metre high, and they would remain the tallest terrestrial organisms on Earth for the first 17 million years of the Devonian (Figure 1).

By the time of the Emsian, competition for sunlight and space resulted in the development of woody tissues by several plant groups. Such support structures enabled plants that possessed them to grow taller than their neighbours. These included very primitive ferns and small tree-like lycopods (relatives of modern club mosses) that could reach two to three metres in height, and they were the tallest plants on Earth for the next 20 million years (Figure 1).

Vegetative height increased dramatically again in the Late Eifelian to Early Givetian with the evolution of arborescent plants that are characterized by secondary wood. The Earth's first forests evolved in the Givetian, and rapidly spread across the continents in the hot climates of the Frasnian, from tropical to almost subpolar regions. In the Early Frasnian the first true woody trees appeared, the progymnosperms of the genus *Archaeopteris* (ancient relatives of modern nonflowering evergreens). *Archaeopteris* trees towered from 10 to 30 metres high in Late Devonian forest landscapes (*see* **Fossil Plants:** Gymnosperms). These were accompanied by strange sphenopsid trees (which looked like gigantic reeds) and lycopsid trees 10 to 20 metres tall.

The last major Devonian advance in plant evolution was the evolution of seed plants in the Late Famennian. Reproduction by seeds (rather than spores) freed land plants from their confinement to moist lowland areas and allowed them to spread into much drier inland areas and into the harsher climates of highlands (Figure 1).

The evolution of increasing vertical height in land plants during the Devonian is mirrored in the evolution of deeper and deeper root penetration into the ground below, and the progressive formation and spread of new soil types. In the Late Pragian some of the herbaceous lycopods penetrated just a single centimetre into the ground. Root penetration increased to 20 centimetres with the evolution of arborescence in the Givetian, and in the Frasnian *Archaeopteris* trees had roots that penetrated to depths of over 100 centimetres. Both the evolution of land plants themselves, and their effects on soil formation (*see* **Soils:** Modern) and rock weathering on the land, were to have a profound impact on both the Earth's atmosphere (*see* **Atmosphere Evolution**) and hydrosphere, and precipitated the global climatic shift from the Early Devonian 'greenhouse' climate to the Late Devonian 'icehouse' climate.

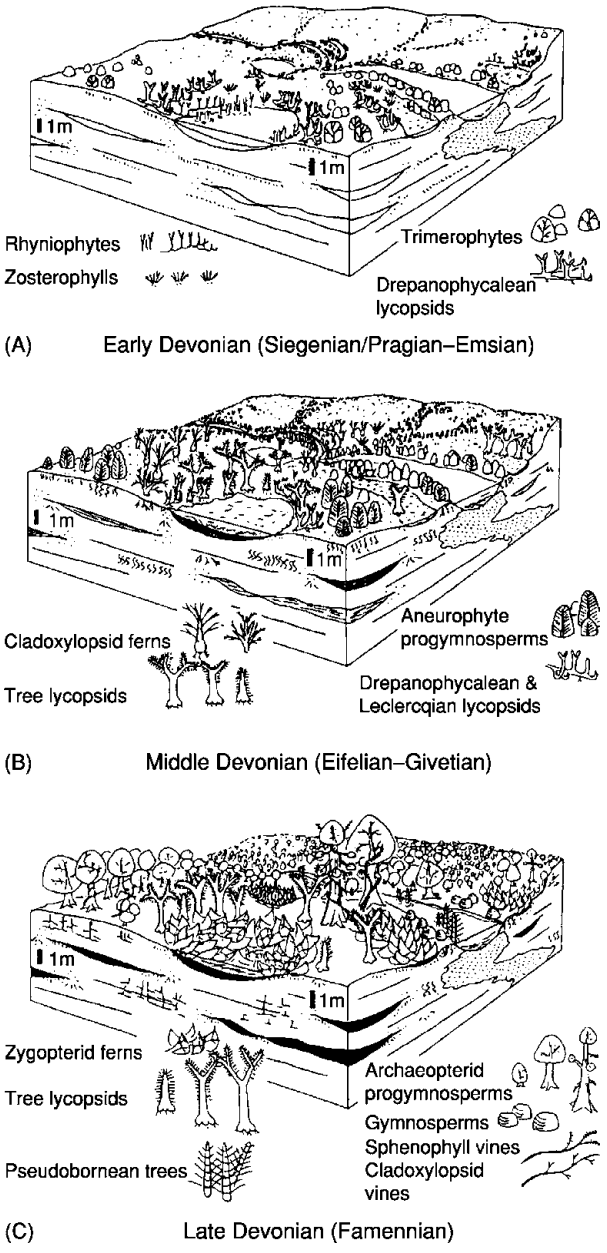


Figure 1 Palaeobotanical and palaeoecological reconstructions of Early Devonian (A), Middle Devonian (B), and Late Devonian (C) upland floodplains. (Adapted from Algeo, Scheckler and Maynard (2001), used with permission.)

The Animal Invasion

Plants are autotrophic organisms, producing their own food by photosynthesis. Plants are used as a food source by animals, and it is no coincidence that the evolution of land plants was quickly followed by the appearance of land animals.

Although millipedes may have made the transition to land as early as the Ordovician, it is in the Late

Silurian and Early Devonian that the animal invasion of the land began in earnest. Millipedes, centipedes, and extinct spider-like arachnids were present on land in the Late Silurian, and became much more numerous in the Early Devonian. Scorpions successfully invaded freshwater streams in the Late Silurian, and emerged onto dry land in the Early Devonian. The earliest known insect-like hexapods appeared in the Pragian, ancient ancestors of the insects that today comprise over 70% of all animal species.

Larger vertebrates were soon to follow, as many of these used the smaller arthropods (*see* **Fossil Invertebrates:** Arthropods) as a food source. The Devonian is justly called the 'Age of Fishes' due to the rapid diversification of numerous fish species that occurred during this period. Many of these species invaded freshwater habitats on land. The most important of these is *Eusthenopteron*, a lobe-finned fish that is anatomically very similar to – and has been cited as the direct ancestor of – the well-known Famennian amphibian *Ichthyostega*. Fossil trackways and fragmentary skeletal material suggests that the evolution of amphibians, the ancestors of all land dwelling vertebrates today, probably took place in the Frasnian.

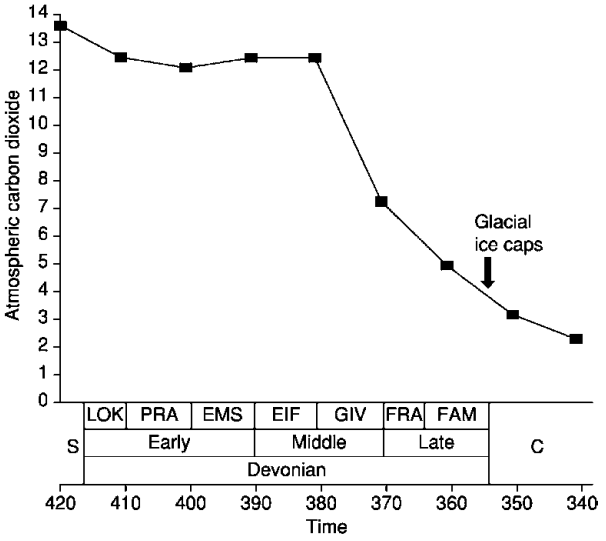


Figure 2 Atmospheric carbon dioxide concentrations during the Devonian, where the amount of carbon dioxide in the atmosphere is measured in factors of the carbon dioxide concentration in the atmosphere at the present (taken as the preindustrial level of 300 ppm). Timescale is in million of years before the present. Abbreviations: S = Silurian, LOK = Lochkovian, PRA = Pragian, EMS = Emsian, GIV = Givetian, FRA = Frasnian, FAM = Famennian. Data from Berner (2001), timescale after Gradstein and Ogg (1996).

The Devonian Global Climatic Shift

The Devonian is a critical period in the Earth's climate history. During this period the global climate switched from the hot greenhouse phase of the earlier Palaeozoic to the cold icehouse phase of the later Palaeozoic, a phase that would lead to continental glaciations in the succeeding Carboniferous and Permian.

The Devonian cooling trend was driven by a sharp decline in the partial pressure of carbon dioxide in the atmosphere from the Givetian to the Early Carboniferous (Figure 2), a decline driven by the establishment of extensive terrestrial land plant cover (discussed in the previous section) and associated soil weathering. Land plants themselves fixed atmospheric carbon in their tissues (*see* **Carbon Cycle**), and nutrient run-off produced by their activity on land triggered blooms of phytoplankton growth in the Devonian shallow seas. This, in turn, led to further atmospheric carbon fixing and organic carbon burial in extensive Middle and Late Devonian black-shale deposits. Chemical weathering of rocks by land plants also produced massive run-off of calcium and magnesium-rich minerals which precipitated as carbonates in the oceans, further drawing down the carbon dioxide partial pressure of the atmosphere.

From the Lochkovian through to the Eifelian the carbon dioxide concentration in the Earth's atmosphere was some 12 to 14 times greater than that of the modern atmosphere (Figure 2). By the end of the Givetian the carbon dioxide concentration had dropped to a little over seven times as great as the present, and in the Early Carboniferous the carbon dioxide content of the atmosphere was only about twice that of the modern atmosphere (Figure 2). Carbon dioxide is the most well-known of the greenhouse gasses, gasses that promote heat retention in the Earth's atmosphere. It is clear that the dramatic decline of carbon dioxide partial pressures in the atmosphere during the 40 million years from the Givetian to the Early Carboniferous must have made the atmosphere more transparent to heat loss, and it is no coincidence that glacial ice-caps formed and spread on the Earth's southern continent of Gondwana in the latest Devonian (Figure 2).

Late Devonian Biodiversity Crises

Life diversified and proliferated in the world's oceans during the first 30 million years or so of the Devonian (Figure 3). Then the trend reversed. Marine biodiversity began to decline in the Emsian and continued to decline, first slowly and then very abruptly, from the Eifelian through the Frasnian. Life rediversified briefly during the Famennian, but then marine biodiversity declined to a very substantial low at the close of the Devonian (Figure 3).

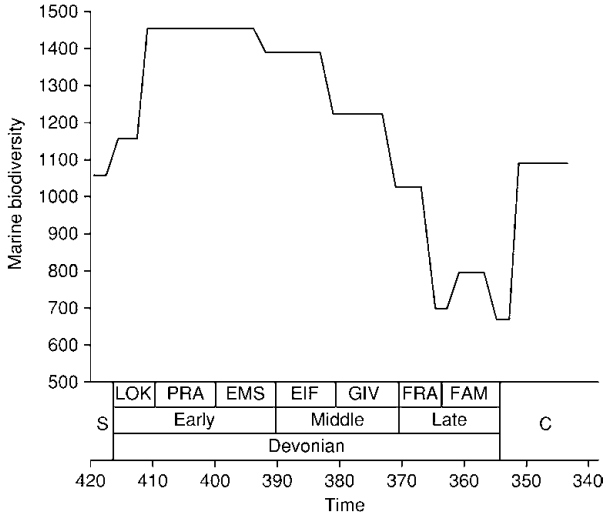


Figure 3 Marine biodiversity during the Devonian, where biodiversity is measured as the number of genera of marine organisms that are well preserved in the fossil record. (Data from Sepkoski (1996).)

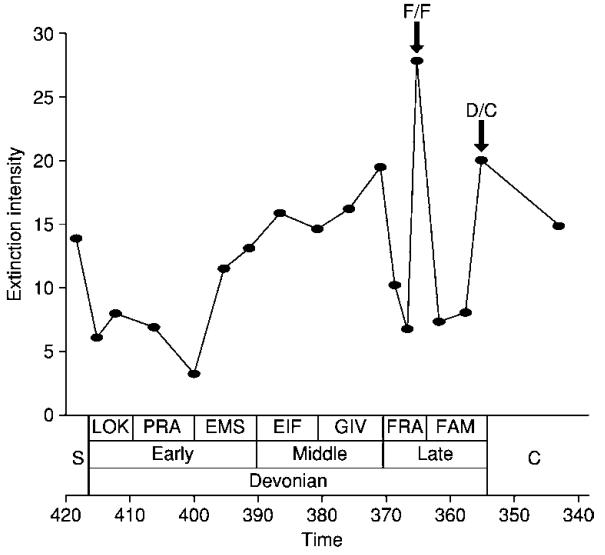


Figure 4 Extinction intensities during the Devonian, where intensity is measured as the percent of marine genera that go extinct per time interval. Positions of the Late Devonian (F/F) and end-Devonian (D/C) extinction peaks are marked with arrows. (Data from Sepkoski (1996).)

The factors controlling the Devonian pattern of biodiversity gain and loss are complex, and not all factors are agreed upon to the present day. In general, the biodiversity gain seen in the Lochkovian was a recovery of biodiversity lost in extinctions that occurred in the Late Silurian. The Pragian saw the evolution of many new taxa, pushing Devonian marine biodiversity to levels higher than any that had existed previously in the Palaeozoic, but many of these new taxa were endemic, however, and having restricted geographic distributions and confined to local regions. Global sea-level began to rise in the Emsian and continued to rise through the Frasnian (perhaps due to the initiation of continental rifting in the Ukraine region of Europe), gradually breaking down barriers between marine regions and facilitating species migration. With the progress of rising sea-levels and increased species migration in the oceans, endemic local faunas successively became extinct and were replaced by more cosmopolitan migrants. By the Early Frasnian, biodiversity in the Devonian seas had dropped to levels comparable to those seen at the beginning of the Devonian (Figure 3), but these seas were now populated by cosmopolitan species with wide geographic distributions.

The gradual decline in marine biodiversity from the Emsian through to the Givetian was mirrored in the gradual increase in extinction intensity within marine faunas during this same interval (Figure 4). Extinction intensity reached a peak in the Late Givetian, with the elimination of most of the previously existing endemic faunas, and then fell to levels comparable with those seen in the Lochkovian and Pragian and within the cosmopolitan faunas of the Frasnian (Figure 4). Very abruptly, extinction intensity within these same cosmopolitan faunas jumped to a level seen only five times in Earth history, and precipitated an equally abrupt loss of marine biodiversity (Figure 3). This is the Late Devonian mass extinction, also termed the Frasnian/Famennian (F/F) mass extinction or the Kellwasser bioevent, and was the first of the Late Devonian biodiversity crises.

The second Late Devonian biodiversity crisis occurred at the very end of the Devonian, and is variously termed the end-Devonian extinction, the Devonian/Carboniferous (D/C) extinction, or the Hangenberg bioevent. The extinction intensity of the end-Devonian extinction was only slightly more severe than that seen in the Late Givetian (Figure 4), but was markedly different in that it was an abrupt jump in extinction, rather than the culmination of a long trend of increasing extinction magnitudes (Figure 4).

Causes of the Late Devonian Biodiversity Crises?

The End-Devonian Biodiversity Crisis

It is generally agreed that the end-Devonian biodiversity crisis, or Hangenberg bioevent, was triggered by global cooling and associated continental glaciation. In the very latest Famennian, just before the end of the Devonian, extensive glacial ice-caps formed in regions

of Gondwana (present day South America), ice-caps that extended to sea-level. Global sea level had begun to fall earlier in the Famennian, and continued to fall throughout the Famennian, indicating the probable formation of high-altitude mountain glaciers considerably before the end of the Famennian.

The end-Devonian extinctions began slightly earlier in the marine realm than the terrestrial, but the entire period of extinctions appears to have spanned only 100 000 years. Although the observed extinctions of marine animals were not as great as those in the Late Devonian mass extinction (Figure 4), the effect of the end-Devonian extinction on land plants and on phytoplankton in the sea was much more severe. In particular, the towering *Archaeopteris* trees that had been so widespread and characteristic of Late Devonian forests rapidly declined in abundance during the Late Famennian climatic cooling, barely surviving into the Early Carboniferous before dying out.

The end-Devonian ice caps were clearly the result of global cooling, which itself was most likely triggered by the dramatic decrease in atmospheric carbon dioxide throughout the latter half of the Devonian (Figure 2). By the end-Devonian, carbon dioxide concentrations in the atmosphere were only three to four times higher than those at present (Figure 2).

The Late Devonian Biodiversity Crisis

The principal cause of the Late Devonian extinction also appears to have been global cooling, but the pattern of cooling has a very different signature than that seen at the end of the Devonian, and the cause of the cooling remains controversial. Palaeobotanical evidence shows that the Frasnian world was quite warm, with extensive shallow seas covering much of the continental landmasses. In these warm seas the great Devonian reefs (*see* **Fossil Invertebrates:** Corals and Other Cnidaria), and other high-diversity marine ecosystems, proliferated. In contrast to Frasnian climates, the Early Famennian world appears to have been cold and arid. The great forests shrank in their areal extent, and were confined to low-latitude equatorial regions. The uplands and high-latitude regions of the Earth supported only very sparse vegetation in conditions that were not only very cold but were also very dry, without perennial snow. In the oceans, huge reef tracts died out all over the world during the latest Frasnian. The few reef organisms that managed to survive into the Famennian were confined to what few warm waters still existed on the Earth, in restricted regions along the equator.

It is still not clear what triggered the collapse of the Frasnian hot climates, and the rapid fall of global temperature into the cold Famennian world. The rapidity of the Late Frasnian cooling is particularly noteworthy, and quite different from the gradual pattern of cooling seen at the end of the Devonian that culminated in continental glaciation. Geochemical evidence indicates that tropical sea-surface temperatures fell by as much as 5° to 7°C in two abrupt pulses, separated by about 300 000 years, in the Late Frasnian to Early Famennian interval.

Three competing hypotheses are at present being debated concerning the ultimate cause of the Late Devonian biodiversity crisis. The 'Devonian plant hypothesis' proposes that both the Late Devonian and end-Devonian biodiversity crises were triggered by the evolution and spread of extensive land plant cover (Figure 1), and the effects of that new terrestrial plant biomass on the atmosphere (Figure 2) and on the oceans (particularly the massive increase in nutrient input to marine ecosystems). However, the Late Devonian biodiversity crisis occurred some 10 million years before the end of the Devonian (Figure 4). All attempts to find geologic evidence of glaciation in the Late Frasnian to Early Famennian interval to date have failed. Global sea-level was at an all-time high for the Devonian, indicating there was no major build-up of glacial ice in this time interval (although apparently rapid oscillations in sea-level during the Late Frasnian and Early Famennian may indicate the formation of small, high-altitude alpine-type glaciers). Atmospheric carbon dioxide concentrations were six times higher than at the present day (Figure 2), and approximately twice as high as those present during the formation of the end-Devonian glacial ice-caps. Many argue that these facts, particularly the rapidity of the temperature drops during the Late Devonian biodiversity crisis, indicate a substantially different triggering mechanism than that seen for the end-Devonian extinctions.

The 'volcanic winter hypothesis' proposes that an interval of catastrophic flood-basalt volcanism occurred during the Late Frasnian. One of the climatic effects of such extensive volcanism is hypothesized to have been rapid global cooling, due to the vast amount of debris, ash, and gasses injected into the Earth's atmosphere. A volcanic-produced global dust and gas cloud would have blocked light from the sun, from reaching the Earth's surface, triggering planet-wide lethally cold temperatures even at the equator.

The catastrophic volcanism scenario is based upon the recognition that there have been other flood-basalt episodes in Earth history, during which enormous amounts of lava, gasses, and volcanic dust have been produced in volcanic eruptions of almost unimaginable magnitude. These flood-basalt fissure eruptions are produced by gigantic plumes of molten rock that originate deep in the Earth's mantle, which then

slowly rise to produce paroxysms of volcanic eruptions over huge geographic areas when they intersect the Earth's surface. Several flood-basalt episodes are known to have occurred at other times of biotic crisis, such as the eruption of the Siberian flood basalts during the Permo-Triassic mass extinction and the Deccan flood basalts during the Cretaceous–Tertiary mass extinction.

Vast regions of the bottoms of the Earth's oceans were depleted in oxygen during the Late Devonian. It has been argued that the great geographic extent of these anoxic water masses were produced in part by extensive submarine volcanism, and hence might be evidence for catastrophic volcanic episodes during the Late Devonian. In addition, a major continental rift-system is now known to have been active in the Ukraine region of Europe during the Late Devonian (the Pripyat–Dnieper–Donet rift), and many of the Earth's flood-basalt fissure eruptions are associated with rifting and spreading of the Earth's tectonic plates.

However, the volume of volcanic material erupted in the Pripyat–Dnieper–Donet rift appears to be relatively small (less than 10 000 cubic kilometres), and thus the intensity of the volcanism associated with the rift was not near the magnitude associated with the Siberian or Deccan flood basalt fissure eruptions (both of which produced well over a million cubic kilometres of volcanic material). Likewise, the extensive spread of anoxic bottom waters in the Late Devonian seas, and the characteristic black-shale deposits within these seas, has been argued to have been due more likely to greatly enhanced organic productivity in the phytoplankton. The greatly increased influx of nutrients into the oceans, produced by the effects of extensive land plant cover on the land (Figure 1), is believed to have produced widespread eutrophication in Devonian shallow seas. Bacterial degradation of phytoplankton rain and burial of the massive amounts of organic carbon fixed by eutrophication is a more probable cause of the Late Devonian black-shale deposits than submarine volcanism.

Another hypothetical way to produce a rapid drop in global temperature would be to impact the Earth with a large asteroid or comet, in essence to produce an 'impact winter' instead of a 'volcanic winter'. As in the case of the hypothesised Late Frasnian glaciations, extensive searches for geologic evidence of major impact events in the Late Frasnian have produced very few hard data. In contrast, considerable evidence exists for several impact events in the Early Frasnian, but about three million years before the biodiversity crisis in the Late Frasnian. In particular, three impact events are currently dated to have occurred between 369 to 367 million years ago: the Alamo impactor, which produced ejecta debris that covers over 19 000 square kilometres of southern Nevada (USA) and which is estimated to have produced a crater 70 kilometres in diameter, the Siljan impactor, which produced a 52-kilometre-diameter crater in Sweden, and the Flynn Creek impactor, which produced a 3.5-kilometre-diameter crater in Tennessee (USA). These impacts all occurred on the Earth's continents; however, an unknown number of additional impacts may also have occurred in the Earth's oceans, for which we have no geologic record.

The 'lag-time multiple impacts hypothesis' proposes that these Early Frasnian impacts produced an abrupt increase in global temperature rather than a decrease. The Alamo impactor in particular struck carbonate target rock, and would have produced a massive injection of carbon dioxide into the atmosphere. The climatic effect of these known Early Frasnian impactors, and of other impacts as yet poorly dated but known to have occurred during the span of the Late Devonian, is argued to have produced an anomalous greenhouse interval that interrupted the Givetian to Carboniferous gradual decline in global temperatures (Figure 5). The collapse of this anomalous greenhouse interval, and the rapid drop in global temperatures produced by the resumption of the pre-existing global cooling trend, is proposed to have produced the observed rapidity of temperature fall seen during the Late Devonian biodiversity crisis (Figure 5). This hypothesis would also account for

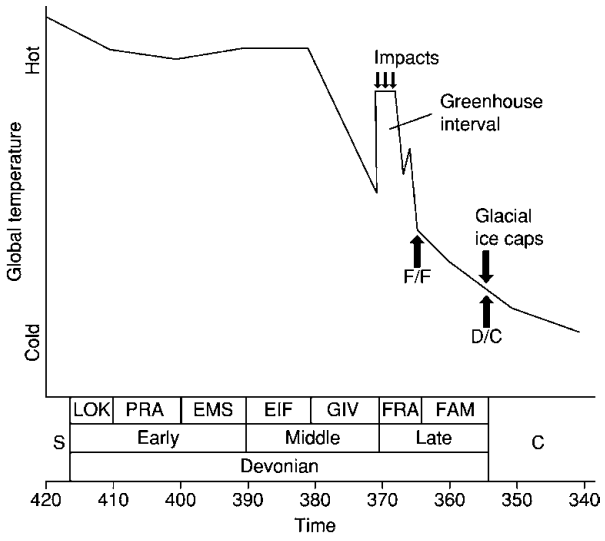


Figure 5 A proposed temperature curve for the Late Devonian, based upon the lag-time multiple impacts hypothesis. Several known impacts that occurred in the Early Frasnian are proposed to have produced an anomalous greenhouse interval, interrupting the gradual cooling trend from the Givetian to the Carboniferous. Positions of the Late Devonian (F/F) and end-Devonian (D/C) biodiversity crises are marked with arrows.

the apparent anomaly of hot Frasnian climates immediately preceding cold Famennian ones, indicated by the palaeobotanical evidence, whereas a gradual cooling trend produced by gradually decreasing atmospheric carbon dioxide levels (Figure 2) would predict Late Frasnian climates only slightly warmer, and different, from the Early Famennian.

The ultimate test of any of the proposed causal scenarios for the Late Devonian biodiversity crisis rests upon the construction of an accurate and detailed temperature curve for the entire Late Devonian period of time. Such a task will be laborious, but it is now technically feasible with oxygen-isotope analyses and laser-based sampling of conodont elements.

See Also

Atmosphere Evolution. **Carbon Cycle**. **Earth Structure and Origins**. **Famous Geologists:** Murchison; Sedgwick. **Fossil Invertebrates:** Arthropods; Brachiopods; Corals and Other Cnidaria; Cephalopods (Other Than Ammonites); Porifera. **Fossil Plants:** Gymnosperms. **Fossil Vertebrates:** Jawless Fish-Like Vertebrates; Fish. **Microfossils:** Conodonts. **Soils:** Modern. **Unidirectional Aqueous Flow**.

Further Reading

Algeo TJ, Scheckler SE, and Maynard JB (2001) Effects of the Middle to Late Devonian spread of vascular land plants on weathering regimes, marine biotas, and global climate. In: Gensel PG and Edwards D (eds.) *Plants Invade the Land*, pp. 213–236. New York: Columbia University Press.

Berner RA (2001) The effect of the rise of land plants on atmosphere CO_2 during the Paleozoic. In: Gensel PG and Edwards D (eds.) *Plants Invade the Land*, pp. 173–178. New York: Columbia University Press.

Dineley DL (1984) *Aspects of a Stratigraphic System: the Devonian*. New York: John Wiley & Sons.

Droser ML, Bottjer DJ, Sheehan PM, and McGhee GR (2000) Decoupling of taxonomic and ecologic severity of Phanerozoic marine mass extinctions. *Geology* 28: 675–678.

Filer JK (2002) Late Frasnian sedimentation cycles in the Appalachian basin – possible evidence for high frequency eustatic sea-level changes. *Sedimentary Geology* 154: 31–52.

Gradstein FM and Ogg J (1996) A Phanerozoic time scale. *Episodes* 19: 3–5.

Joachimski MM and Buggisch W (2002) Conodont apatite $\delta^{13}C$ signatures indicate climatic cooling as a trigger of the Late Devonian mass extinction. *Geology* 30: 711–714.

Long JA (1995) *The Rise of Fishes*. Baltimore: The Johns Hopkins University Press.

McGhee GR (1996) *The Late Devonian Mass Extinction*. New York: Columbia University Press.

McGhee GR (2001) The 'multiple impacts hypothesis' for mass extinction: a comparison of the Late Devonian and the late Eocene. *Palaeogeography, Palaeoclimatology, Palaeoecology* 176: 47–58.

Murphy AE, Sageman BB, and Hollander DJ (2000) Eutrophication by decoupling of the marine biogeochemical cycles of C, N, and P: a mechanism for the Late Devonian mass extinction. *Geology* 28: 427–430.

Racki G (1998) Frasnian-Famennian biotic crisis: undervalued tectonic control? *Palaeogeography, Palaeoclimatology, Palaeoecology* 141: 177–198.

Sepkoski JJ (1996) Patterns of Phanerozoic extinction: a perspective from global data bases. In: Walliser OH (ed.) *Global Events and Event Stratigraphy*, pp. 35–51. Berlin: Springer.

Streel M, Caputo MV, Loboziak S, and Melo JHG (2000) Late Frasnian-Famennian climates based on palynomorph analyses and the question of the Late Devonian glaciations. *Earth-Science Reviews* 52: 121–173.

Walliser OH (1996) Global events in the Devonian and Carboniferous. In: Walliser OH (ed.) *Global Events and Event Stratigraphy*, pp. 226–250. Berlin: Springer.

Carboniferous

A C Scott, Royal Holloway, University of London, Egham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Carboniferous, spanning 60 Ma, was an important period in Earth history, with major changes in atmospheric composition (in terms of both oxygen and carbon dioxide), the development of an icehouse world, the formation of vast tropical peat (coal) deposits, and the radiation of plants and animals on land. In particular, vertebrates got a major foothold on land and the air was conquered by the insects. The vast supercontinent of Pangaea was also formed during this period. The significance of the Carboniferous can be demonstrated by the fact that it was one of the first geological periods to be recognized. Despite this and the fact that Carboniferous rocks and fossils

have been amongst the most widely studied, very few formal definitions of the units of the Carboniferous have been internationally ratified.

Historical Setting

The Carboniferous was named by William Conybeare and William Phillips in 1822, based upon their work in northern England. From their studies, particularly in Yorkshire, they recognized a distinctive coal-bearing sequence, which had an important limestone sequence below. The frequent coals, which were economically important, are reflected in the name. The lower limestone unit became known as the Mountain Limestone, and the coal series became known as the Coal Measures.

This sequence of rocks was found across much of Western Europe, and the Carboniferous was divided into the Lower Carboniferous, comprising predominantly marine limestones, and the Upper Carboniferous, comprising clastic and coal-bearing strata. In continental Europe these two major divisions were termed the Dinantian and Silesian. In North America, however, whilst a similar sequence of limestones and coals was found, the divisions were named the Mississippian by Winchell and the Pennsylvanian by Williams. Unfortunately, the boundary between them did not coincide with the boundary defined in western Europe. To complicate matters even further, a three-fold division of the Carboniferous was recognized in Russia.

Chronostratigraphy and Biostratigraphy

Through the Heerlen congresses, and later the International Carboniferous Congress, held in Europe, North America, South America, and China through the twentieth century, a number of attempts to divide the Carboniferous into stages and to correlate local stages worldwide were made. Local stages were erected with formal boundary stratotypes, but there is often confusion as to usage in the literature. Currently, whilst the Carboniferous stage nomenclature has been agreed, few boundaries have yet been defined by the International Commission on Stratigraphy (ICS) (Figure 1).

Devonian–Carboniferous Boundary

The base of the Carboniferous System (also the base of the Tournaisian Stage) has now been formally defined as the first appearance of the conodont *Siphonodella sulcata* within the evolutionary lineage from *Siphonodella praesulcata* to *Siphonodella sulcata*.

Figure 1 Current statigraphical nomenclature for the Carboniferous. Reproduced from the International Commission on Stratigraphy. Taken from ICS website.

The Global Standard Section and Point is at the base of Bed 89 in Trench E′ at La Serre in southern France. The rocks in the La Serre sequence are marine, belonging to the 'klippen of cabrières', comprising biodetrital oolitic limestones within a pelagic matrix of shale and cephalopod-bearing calcilutites. The fauna is varied and includes trilobites.

Carboniferous–Permian Boundary

The base of the Permian System (and hence the top of the Carboniferous System) is defined as the first occurrence of the 'isolated-nodular' morphotype of *Streptognathodus* '*wabaunsensis*' conodont chronocline, 27 m above the base of Bed 19, Aidarabsh Creek, northern Kazakhstan, which is in the southern Ural Mountains. The strata comprise hemipelagic silt and clay, with occasional lenses of sand and very course sand. The marine fauna, in addition to conodonts, includes ammonoids and fusilinid benthic foraminifera.

Mississippian–Pennsylvanian Boundary

Following considerable debate, the division of the Carboniferous System into two subsystems has been agreed. These two subsystems have been named the Mississippian and Pennsylvanian, although their usage may not always correspond with previous usage. The mid-Carboniferous boundary does not coincide with the boundary between either the Lower and Upper Carboniferous or the Dinantian and Silesian of previous schemes. In Europe the boundary is within the Namurian of older usage (Figure 2).

The base of the Pennsylvanian is also the base of the Bashkirian stage. The Global Standard Section and Point has been located at the lowest occurrence of the conodont *Declinognathodus nodiliferus* s.l., 82.9 m above the top of Battleship Wash Formation, Arrow Canyon, southern Nevada, USA.

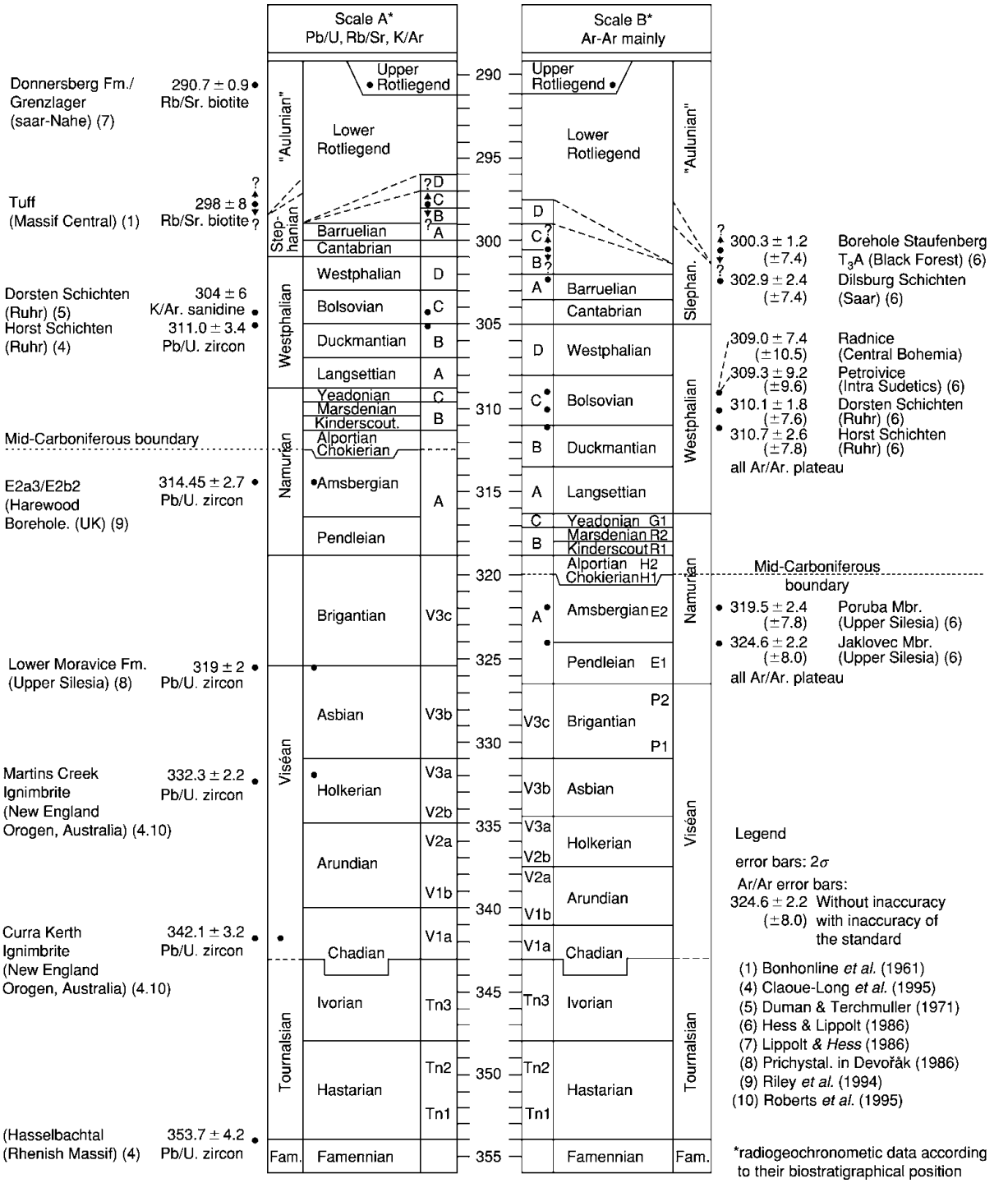


Figure 2 Absolute ages of Carboniferous strata obtained using different dating techniques and using a traditional Carboniferous chronostratigraphical nomenclature. Reproduced from Menning M, Weyer G, Drozdzewski G, Van Amerom HWJ, and Wendt I (2000) A Carboniferous time scale 2000: discussion and use of geological parameters as time indicators from central and western Europe. *Geologische Jahrbuch*, Hannover, A 156: 3–44.

Carboniferous Subdivisions

The Mississippian is now formally divided into three Stages (although these may become Series): Tournaisian, Viséan, and Serpukhovian. The Tournaisian–Viséan and Serpukhovian–Viséan boundaries have still to be formally placed. The Pennsylvanian is now formally divided into four: Bashkirian, Moscovian, Kasimovian, and Gzhelian. The Bashkirian–Moscovian,

Moscovian–Kasimovian, and Kasimovian–Gzhelian boundaries have yet to be decided. The division of the Carboniferous into formally internationally agreed series and stages may take a considerable time.

Absolute Ages

There is currently considerable debate over the absolute ages of Carboniferous sequences. Traditionally, most dates have been obtained from tonsteins (volcanic ash deposits) within coal seams. The ICS currently provide a number of dates. The base of the Carboniferous is stated by the ICS as 359 Ma and the top of the Carboniferous as 299 Ma, making the Carboniferous 60 Ma long. The mid-Carboniferous boundary is given as 318 Ma, making the Mississippian 41 Ma and the Pennsylvanian 19 Ma long. Menning, in his review of Carboniferous absolute time-scales, showed that there is a significant discrepancy between different radiometric-dating techniques (Figure 2). Traditional dates using uranium–lead dating techniques are in conflict with those obtained using argon–argon techniques. In scale A of Figure 2 the dates are from a series of techniques using uranium–lead, rubidium–strontium, and potassium–argon. A number of these involve the use of uranium–lead dates of zircons. Here the mid-Carboniferous boundary is placed at approximately 313.5 Ma. In contrast, scale B uses mainly argon–argon dates and the mid-Carboniferous boundary is placed at 320 Ma. The current trend is to place more reliance on the argon–argon dates, but it must be made clear that there are relatively few data points.

Biostratigraphy

Many groups of fossils have been used for biostratigraphical zonation and correlation. In the predominantly marine strata of the Mississippian, corals and brachiopods were widely used in Europe. Goniatites (cephalopods) have also been extensively used, particularly in North America. On land both vascular plants and non-marine bivalves have been widely used since the end of the nineteenth century. Advances in micropalaeontology in the mid-twentieth century brought the use of microfossils to the fore. Both foraminifera and conodonts became widely used in the later part of the twentieth century. Conodonts have, in marine strata, become the zonal fossils of choice, being used to define major chronostratigraphical boundaries. On land macrofossil plants and, more recently, pollen and spores (palynomorphs) have been widely used. Palynomorphs were first used to help correlate Pennsylvanian coal seams, and subsequent work on their distribution in clastic sediments has allowed them to be used to establish zones that can be correlated across continents. The use of pollen and spores for zonation and correlation was extended into the Mississippian, and they have the advantage not only of being found in non-marine strata but also of being washed or blown into marine rocks. This has allowed correlation between marine and non-marine sequences.

Lithologies and Environments

Within the Carboniferous equatorial belt, warm shallow-water carbonates dominate Mississippian sequences. This is what Derek Ager termed the 'persistence of facies.' Mississippian facies differ markedly in temperate and higher latitudes. Carboniferous sequences in South America, Australia, and South Africa are characterized by cold-water and often glacial deposits. In Europe, later Mississippian and Early Pennsylvanian deposits are commonly clastic shales and sandstones, which were lain down in thick deltaic sequences. By the later part of the Mississippian, commercial coals are found, such as in the Viséan of the Moscow Basin.

Widespread regression, possibly related to the onset of a major glaciation, allowed extensive mires to form throughout the tropics (Europe and North America) (Figure 3). Later Carboniferous deposits in Euramerica are characterized in many places by coal-bearing sequences. It was during the Pennsylvanian that many of the world's economic bituminous coals were formed. Some of these seams are several metres thick, taking tens of thousands of years to form. Many of the larger coal basins show evidence of regular marine influxes. The resulting marine bands, often containing goniatites, have been widely used for basinal and even intercontinental correlation. The regular sequences of non-marine rocks, in some cases with associated marine bands, have, on various scales, been called cyclothems and mesothems. The regular transgressive–regressive cycles, so well displayed in the mid-west of the USA, are thought to be the result of eustatic changes caused by the waxing and waning of the southern ice-sheet.

The later part of the Pennsylvanian is characterized by increasing aridity. Coal-forming mires shrank and peat-forming plants became extinct. In China, however, some of these environments and vegetation continued into the Permian. The drier environments supported very different vegetation from that which characterized the wet lowland habitats. Interbasin correlation, therefore, becomes difficult in some of these areas.

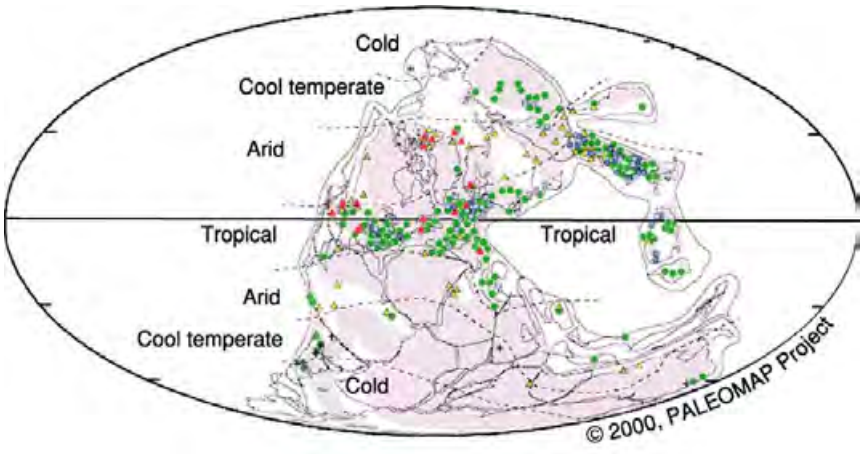


Figure 3 Lithological indicators of palaeoclimate in the Pennsylvanian. Reproduced from Scotese 2000 Paleomap project.

The Carboniferous World

Plate Reconstructions

The Carboniferous saw the gradual accretion of many continental landmasses. At the beginning of the Mississippian, two or three major continents had formed (Figure 4A). Europe and North America had collided to form Euramerica, which has also been termed Laurussia or Laurasia. Accretion was still active, with collision and uplift of the Appalachian mountain belt, which may have been a response to the northward migration of Africa. The Rheic Ocean, south of Britain, continued to close. The far-east Asian plate assemblies were more complex. Siberia was in the temperate northern hemisphere and is also known as Angara. China, comprising a northern block and a southern block, was also north of the equator, but there is dispute as to whether these two blocks had collided by the Carboniferous. This area is known as Cathaysia.

Many of the southern continents formed the supercontinent of Gondwana (*see* **Gondwanaland and Gondwana**). In the Carboniferous there continued to be continental accretion, for example in South America. Through the Carboniferous the northern and southern continents continued to amalgamate (Figure 4B) so that by the Permian the giant supercontinent of Pangaea (*see* **Pangaea**) had formed.

Palaeobiogeography

There have been several recent attempts to assess the palaeobiogeography of early Carboniferous vegetation using data from either macrofossils or pollen and spores. Taking the Mississippian as a whole, it is generally considered that the world's floras were predominantly cosmopolitan, characterized by assemblages including the lycopod genus *Lepidodendropsis* (Figure 5A), which was a widespread arborescent form that did not shed its leaves. The only regionally separate flora that has been recognized is the Angaran flora, which comprises mainly endemic genera. However, an additional temperate floral belt has recently been suggested. A larger number of regional floras have been recognized using mathematical techniques, but these analyses suffer from many problems, not least concerning the identity and age of the plants in question, some of which were first studied in the nineteenth century. These attempts of course have taken into consideration only compression plant genera, and there has been much recent work on anatomically preserved taxa, which can rarely be correlated with compression taxa. In addition, as might be expected, within any given region floras may vary because of ecological or preservational controls.

By the end of the Carboniferous, clear regional floras can be recognized (Figure 5B). The best-studied flora is that of Euramerica. This region comprised Europe (including eastern Europe) and North America and was, during this period, equatorial with a warm humid climate. The floras comprise diverse taxa typified by the species preserved in the Pennsylvanian Coal Measures, in particular the lycopods *Lepidodendron*, *Sigillaria*, and *Lepidophloios*, the sphenopsid *Calamites*, the ferns *Corynepteris*, *Renaultia*, and *Pecopteris*, the seed-ferns *Neuropteris*, *Alethopteris*, and *Mariopteris*, the cordiate *Cordaites*, and the conifer *Lebachia* (*Walchia*).

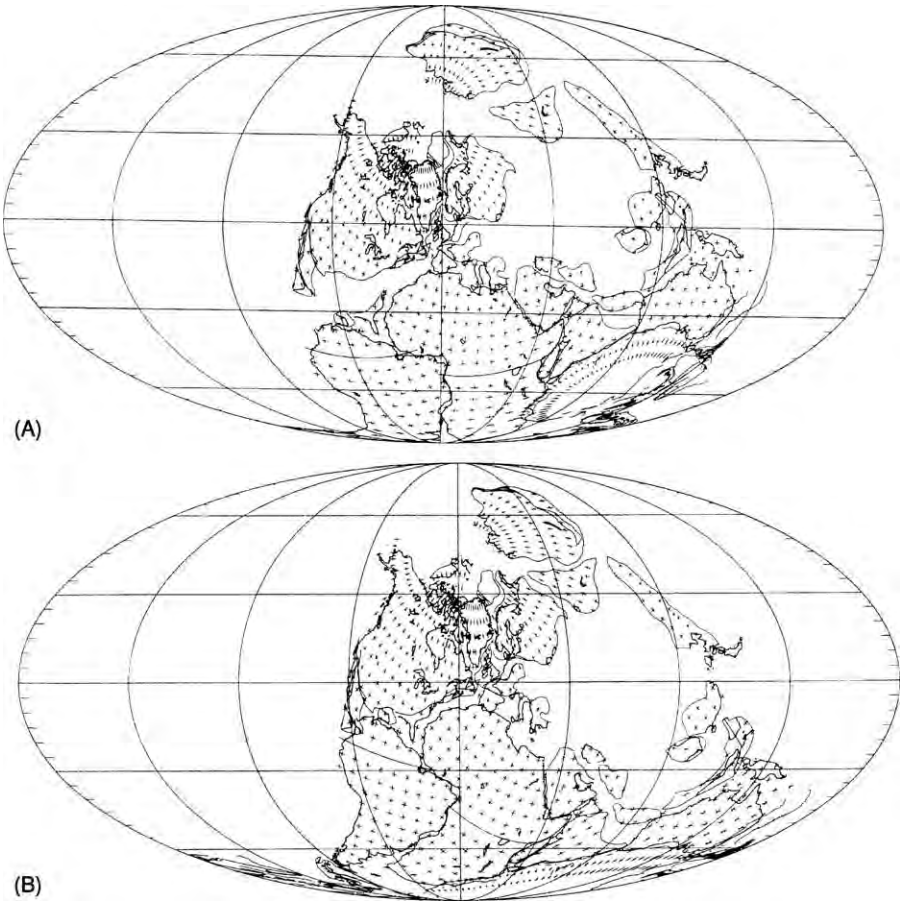


Figure 4 Plate reconstructions for (A) the Mississippian (Viséan) and (B) the Pennsylvanian (Moscovian). Reproduced from Scotese CR and McKerrow WS (1990) Revised world maps and introduction. In: McKerrow WS and Scotese CR (eds.) *Palaeozoic Palaeogeography and Biogeography*, pp. 1–21. Memoir 12. London: Geological Society.

The Gondwanan flora, characterized by the seed-fern *Glossopteris* and its relatives, is found during the Early Permian in Australia, Brazil, Africa, Antarctica, and India.

The Late Carboniferous flora from Australia is problematic. It comprises several genera with fern-like foliage, which have been linked to Early Carboniferous northern genera. There are two possible explanations for these occurrences: either they represent a relict population, which migrated southwards during the Carboniferous and survived in Australia after becoming extinct elsewhere, or, more likely, they belong to different plants with a similar leaf morphology. These plants include *Rhacopteris* and *Triphyllopteris*.

This regionalization of floras may partly reflect latitude – the *Glossopteris* flora is found in the high-latitude southern hemisphere, the Euramerican flora is found in equatorial areas, and the Angaran flora is found at high latitudes in the northern hemisphere – and may also reflect different climatic regimes (Figure 5C). This is supported by the fact that strata bearing the *Glossopteris* flora of Gondwana often overlie tillite horizons. The Cathaysian flora is also problematic, with the main regionalization in China occurring in the Permian.

The Carboniferous Atmosphere

The atmospheric composition changed markedly through the Carboniferous, and this period was one of the most unusual in Phanerozoic history (Figure 6).

Models of atmospheric oxygen for the Carboniferous by Berner and others show that the present atmospheric level of oxygen, 21%, was probably reached by the end of the Devonian. Oxygen levels are thought to have risen considerably during the Carboniferous, reaching a peak well above 30% by the Pennsylvanian. This rise in atmospheric oxygen may have been a result of the occurrence of

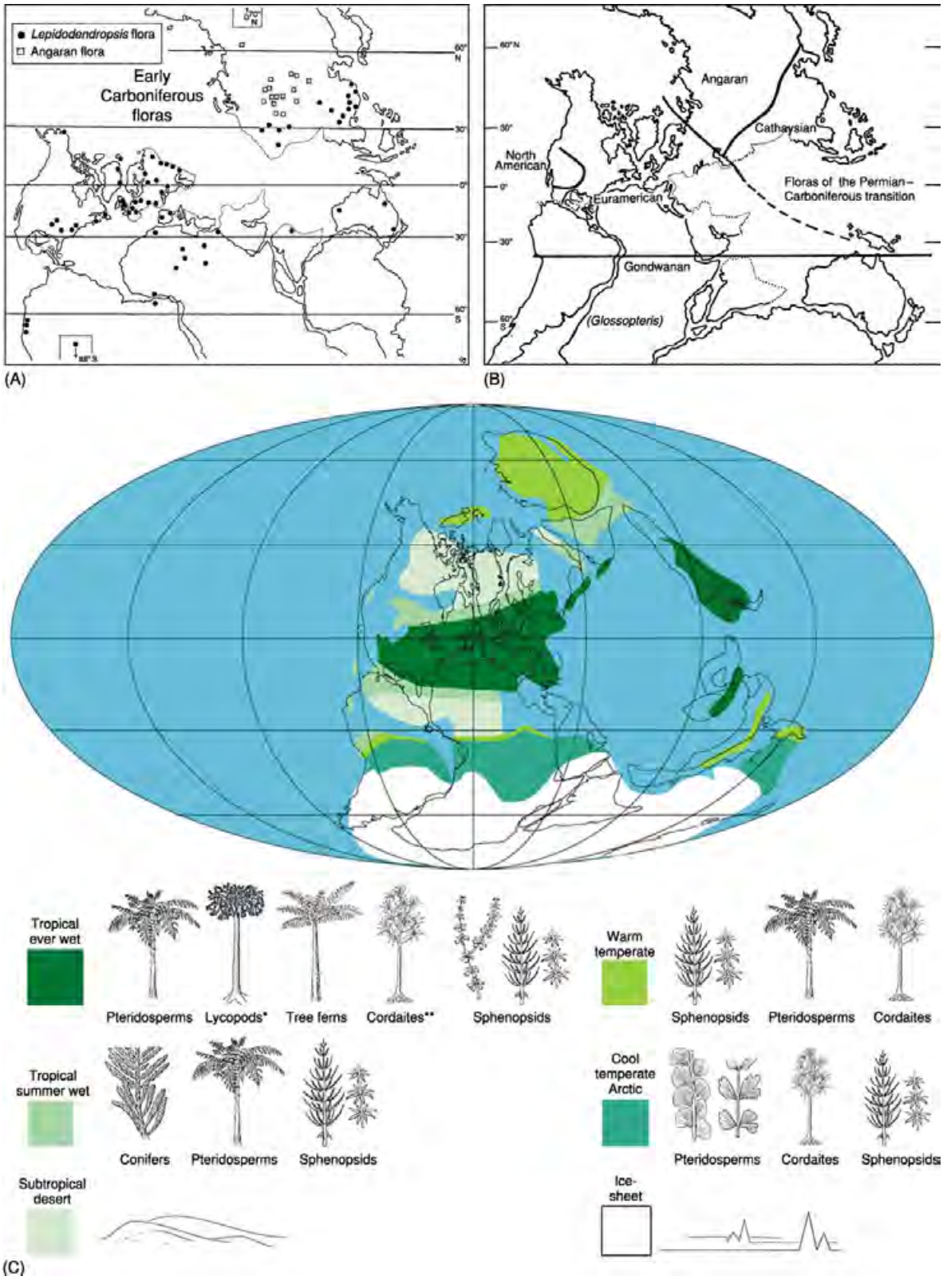


Figure 5 Palaeobiogeographical reconstructions of the Late Carboniferous showing (A) the floral provinces of the Mississippian (reproduced from Chaloner WG and Lacey WS (1973) *The Distribution of Late Palaeozoic Floras*. Special papers in Palaeontology 12. pp. 271–289), (B) the floral provinces of the Carboniferous–Permian transition (adapted from Chaloner WG and Lacey WS (1973) *The Distribution of Late Palaeozoic Floras*. Special papers in Palaeontology 12. pp. 271–289.), and (C) Pennsylvanian biomes (reproduced from Willis KJ and McElwain JC (2002) *The Evolution of Plants*. Oxford: Oxford University Press).

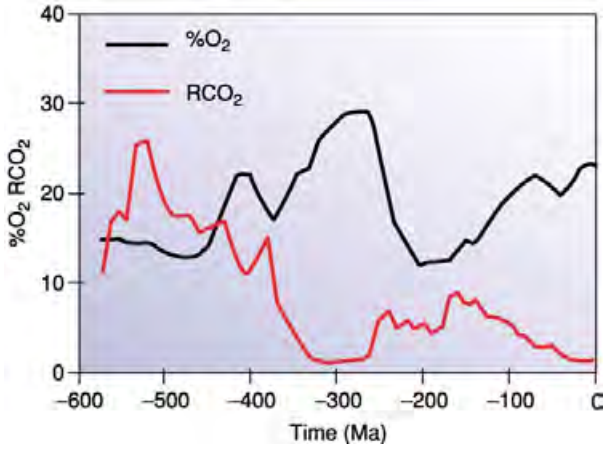


Figure 6 Changes in atmospheric composition over the Late Proterozoic and Phanerozoic, showing falling carbon dioxide and rising oxygen levels during the Carboniferous (360–300 Ma). Reproduced from Berner RA (2003) The long-term carbon cycle, fossil fuels and atmospheric composition. *Nature* 426: 323–326.

widespread lowland tropical rainforest. The later postulated rapid fall in oxygen may have been a result of the demise of these extensive tropical coal-forming forests. The oxygen-rich atmosphere may have had a significant effect on animal evolution, allowing the growth of giant arthropods including insects. In addition, fires may have been more widespread and intense.

In contrast, levels of carbon dioxide fell rapidly at the end of the Devonian and during the Early Carboniferous. This coincided both with the evolution and initial spread of trees and with the onset of extensive peat formation and, hence, burial of carbon, in addition to the burial of extensive shallow-water carbonates. All of these factors combined to cause a major drawdown of carbon dioxide. The major climatic changes in the Late Carboniferous–Early Permian may be linked to increasing levels of carbon dioxide. The impact of these low levels of carbon dioxide on the Carboniferous world was severe, plunging it into a major icehouse phase, as evidenced by the spread of the polar ice-cap in the southern continents.

The Carboniferous Climate: Icehouse Conditions

One of the major glaciations in Phanerozoic history occurred in the Late Palaeozoic, beginning in the Carboniferous and ending in the Permian (Figure 7). The world, therefore, went from greenhouse conditions to icehouse conditions and back to greenhouse conditions. Striated pavements and glacial tillites both indicate the presence of a southern polar ice-cap across the Gondwanan continents (Figure 8). All other aspects of this glacial episode, however, are open to at least some debate.

- Onset. The onset of glacial conditions has been difficult to date. Palynological dating of South African sequences suggests that glaciation may have begun as early as the Devonian–Carboniferous boundary. Unequivocal evidence from tillites and inferences from carbon isotope excursions indicate that a significant ice-cap existed by Late Mississippian times. It is believed that the earliest glacial deposits are those of South America.
- Spread and extent. It has been shown that the onset of glaciation spread to South Africa and Antarctica and thence to Australia and India. Typically glacial tillites (e.g. Dwyka of South Africa) overlie striated pavements and are overlain by temperate coal-bearing strata. Whilst on palaeogeographical maps a large polar ice-cap is often shown across much of southern Gondwana, it is not clear whether this was the case or whether a smaller ice-sheet moved as the continents moved.
- Demise. Just as the bases of the glacial deposits across Gondwana are diachronous, so too are the overlying non-glacial deposits. Remnant ice-caps may have been present in upland areas, but Permian coal-bearing strata are found in South America, South Africa, Antartica, India, and Australia. The cessation of glaciation also appears to be diachronous, being older in South America and younger in Australia.
- Causes. Whilst a number of key facts about the Gondwanan glaciation and the Carboniferous world have been established, there is little agreement about cause and effect.

Climate change can be seen over several different time-scales, all of which have been well documented for the Carboniferous. The longest cycle, causing a change from greenhouse to icehouse cycles, may have been caused by the combination of plate movements, atmospheric change, orbital cycles, and the evolution of land plants. These factors continued to act throughout the Carboniferous. There are changes in both local and global climate through the Carboniferous. The occurrence of evaporates in some Mississippian deposits indicates a drier environment than that seen in the Pennsylvanian. Widespread wetland systems occurred in tropical regions during the Pennsylvanian (Figure 9).

Orbital-eccentricity cycles and axial-tilt and precession cycles have intermediate- and short-term climatic effects. However, these effects may be modified by local tectonic or orographic conditions. Studies of coals and vegetation show major wet–dry cycles

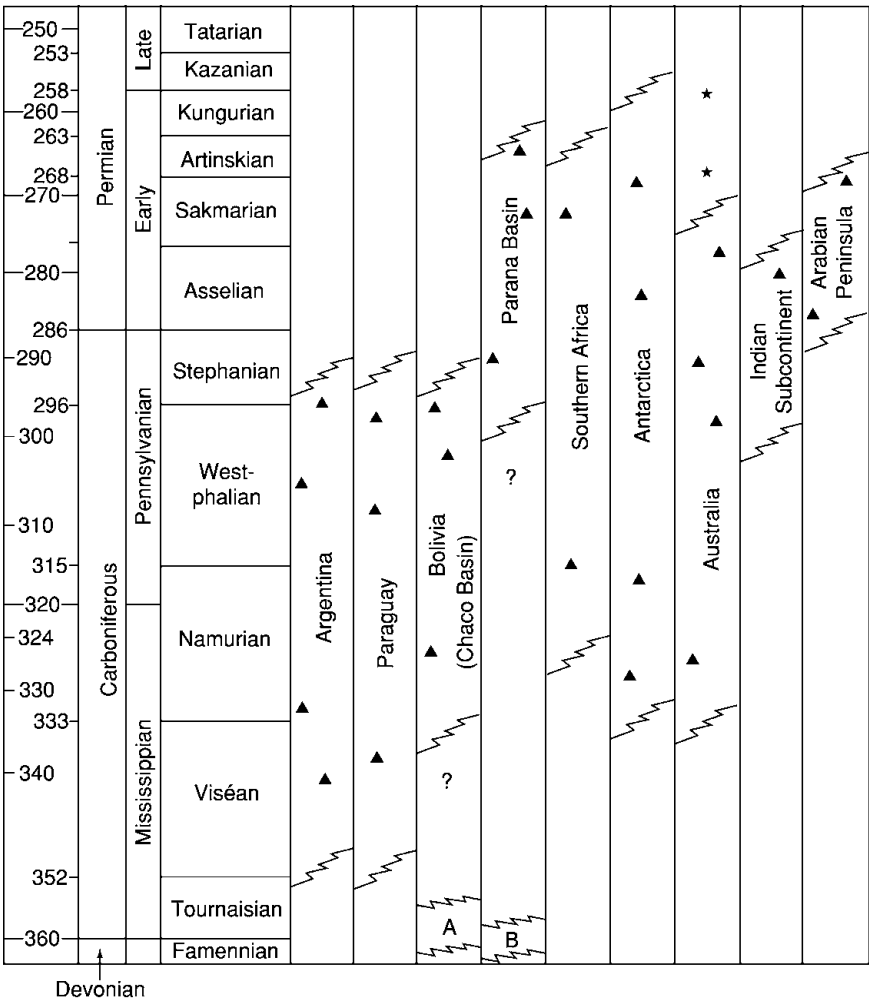


Figure 7 Timing of glaciation in the Late Palaeozoic across Gondwana. Reproduced from Eyles N (1993) Earth's glacial record and its tectonic setting. *Earth Science Reviews* 35: 1–248.

during the Pennsylvanian of Illinois. These changes in rainfall had a dramatic effect on vegetation and on peat (coal) formation. In particular, a major drying phase at the Westphalian–Stephanian boundary caused the extinction of a number of peat-forming lycophytes (Figure 9). Increasing aridity towards the end of the Carboniferous is also widely documented. The broad wet–dry cycles seen in Illinois are not always seen in other Euramerican basins, although the broad picture is confirmed (Figure 10).

Milankovitch cycles have been proposed to have caused waxing and waning of the ice-sheet over the southern continents, leading to rises and falls in sea-level (*see* **Earth:** Orbital Variation (Including Milankovitch Cycles)). It has been suggested that these sea-level changes may have been responsible for the cyclothemic and mesothemic units that are seen particularly well in the Pennsylvanian.

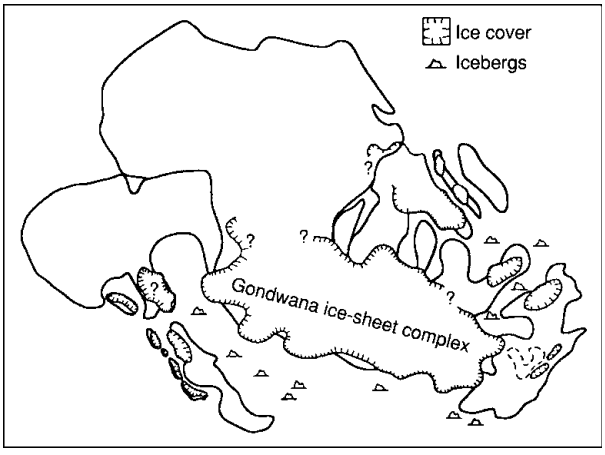


Figure 8 Maximum glaciation of Gondwana in the Carboniferous and Permian. Reproduced from Eyles N (1993) Earth's glacial record and its tectonic setting. *Earth Science Reviews* 35: 1–248.

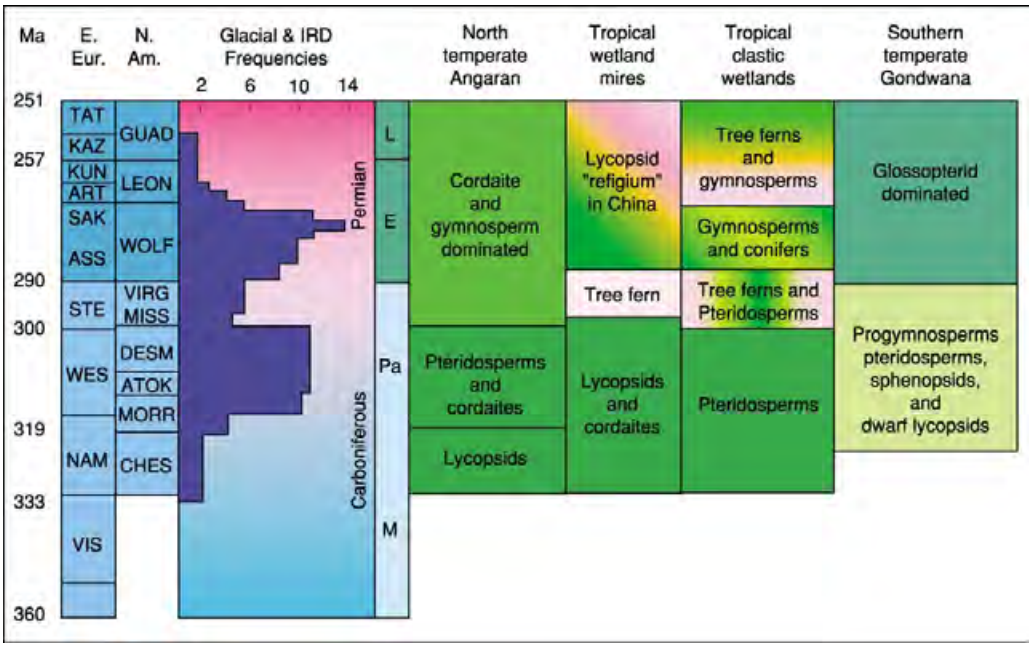


Figure 9 Vegetational change through the Late Palaeozoic in relation to climate change. IRD = Ice rafted deposits; M = Mississippian; P = Pennsylvanian; E = Early; L = Late; VIS = Visean; NAM = Namurian; WES = Westphalian; STE = Stephanian; ASS = Asselian; SAK = Sakmarian; ART = Artinskian; KUN = Kungurian; KAZ = Kazanian; TAT = Tatarian; CHES = Chesterian; MORR = Morrowan; ATOK = Atokan; DESM = Desmoinian; MISS = Missourian; VIRG = Virgillian; WOLF = Wolfcampian; LEON = Leonardian; GUAD = Gudalupian. Reproduced from Gastaldo RA, DiMichele WA, and Pefferkorn HW (1996) Out of the icehouse into the greenhouse: a Late Paleozoic analogue for modern global vegetational change. *GSA Today* 10: 1–7.

The Nature and Influence of Fire in Carboniferous Ecosystems

Fires may range from commonly seasonal to rare and catastrophic. Climate may play a significant role in controlling the frequency of fires. It is possible that, in addition to large-scale climatic changes, smaller-scale changes in climate may play a significant role in the occurrence of fires. It is possible that in some ever-wet environments fires are unusual. Disturbance of the normal climatic regime (e.g. El Niño events) may cause temporary changes in the rainfall pattern, for example, leading to a dry period when a catastrophic fire may occur. This may be the reason for the Great Fire of Borneo in 1982–1983. Small fluctuations in rainfall pattern may allow frequent small fires in a mire-forming environment. Storms may follow a major fire, and the increased erosion may give rise to the transportation and deposition of a sediment–charcoal mix at several sites. The occurrence of fusain (fossil charcoal) layers in the fossil record may be used to distinguish between events of different timescales.

The effects of large-scale fires are illustrated by the Lower Carboniferous marginal-marine fusain deposits at Shalwy Point, Donegal, Ireland. Here, there is no doubt that the fire was catastrophic. The widespread nature of the fire not only yielded significant amounts of charcoal but also would have led directly to increased soil erosion, which in turn would have led to a major increase in bed load in the alluvial system.

Life on Land

There were many major changes in the terrestrial biota during the Carboniferous.

Plants

The Early Carboniferous saw a major radiation in many groups of vascular land plants. Amongst spore-bearing plants, lycopsids underwent a major radiation into several new ecological niches. Extensive peat mires were developed for the first time. At this time sphenopsids also became abundant, with forms such as *Archaeocalamites* being particularly widespread. There was a dramatic diversification of ferns during the Mississippian, with zygopterid ferns being particularly abundant. It has been proposed that these plants were the primary colonizers of volcanic terrains throughout Europe.

The most important plant radiation at this time was in the seed plants. In particular, pteridosperms

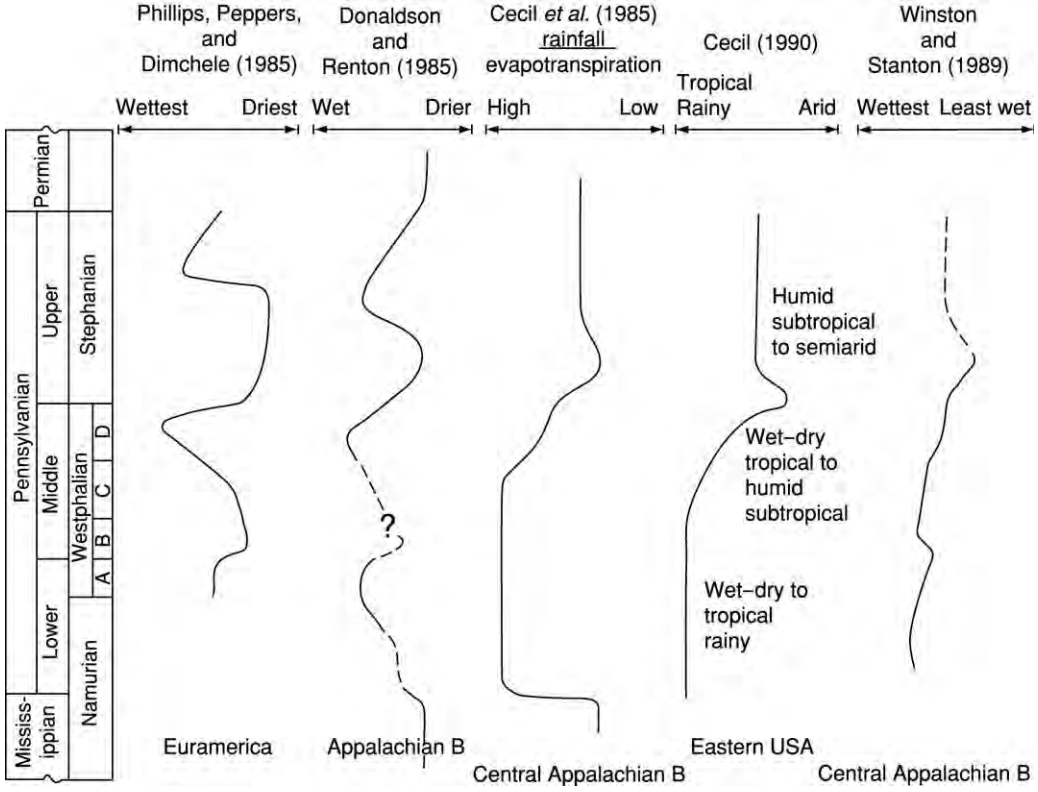


Figure 10 Interpreted palaeoclimatic trends for the Euramerican coal basins during the Pennsylvanian. Reproduced from Calder JH and Gibling MR (1994) The Euramerican Coal Province: controls on Late Paleozoic peat accumulation. *Palaeogeography, Palaeoclimatology, Palaeoecology* 106: 1–21.

underwent a rapid radiation into a variety of lowland habitats. The abundance of anatomically preserved specimens in the Mississippian of Scotland has allowed detailed phylogenetic studies of these plants.

Several plant groups diversified during the Pennsylvanian. The lycopods are the main coal-forming plants of the Carboniferous. They are also known as club mosses but are, of course, true vascular plants. Today only a few genera of herbaceous forms survive, but in the Carboniferous they showed a vast array of life habits from small herbaceous forms to large arboresent (tree) forms. Much of our recent data on these important plants has come from the study of coal balls – anatomically preserved fossils where the internal anatomy can be elucidated. As with the lycopods, Carboniferous sphenopsids are represented by both arborescent and herbaceous forms.

Ferns are plants with well-developed megaphyllous leaves. All the Carboniferous forms are homosporous. Fern-like foliage is abundant in the Pennsylvanian Coal Measures but most belongs to the pteridosperms (seed-ferns). The identification of true ferns in the compression fossil record requires fertile specimens. The Carboniferous was once known as the age of ferns because of the abundance of fern-like foliage amongst the compression fossils. It was shown, however, that most of this foliage belongs to an extinct group of plants that bore seeds and pollen organs, known as the pteridosperms or seed-ferns. Three major families are known from the Upper Carboniferous: Lyginopteridales, Medullosales, and Callistophytales. These were either shrubs or climbers in the vegetation.

The most advanced plants of the late Carboniferous were the closely related cordaites and conifers. It is still unclear whether the cordaites gave rise to the conifers or whether they had a common ancestor. Both, however, are gymnospermous and are represented by both trees and shrubs. The conifers were probably major elements in drier upland floras.

Terrestrial Invertebrates

The East Kirkton terrestrial Mississippian Lagerstätten has opened a major window on a terrestrial ecosystem. Several invertebrate groups are first found here or there is evidence of their terrestrial nature. The fauna includes eurypterids, scorpions, myriapods, and a spider, in addition to aquatic arthropods such as ostracods. The eurypterids were large and

include *Hibbertopterus*. The scorpions show unequivocal evidence of terrestrial adaptations and specimens over 30 cm long have been found; they may have reached up to 1 m in length and been a top predator. Other arthropods include millipedes and an opilionid or Harvestman spider.

No fossil insects are known from East Kirkton; the first unequivocal evidence of true insects (Hexapoda: Pterygota) does not occur until the Early Pennsylvanian, although there are recent claims of early Devonian insects. Insects of the Pennsylvanian had wings.

The terrestrial fauna of the Pennsylvanian Coal Measure forests included many carnivorous animals. Centipedes (Myriapoda: Chilopoda) were common, especially in the Mazon Creek fauna of Illinois, as were many spiders (Arachnida: Aranae), some of which had inward-striking (dianial) fangs, indicating an ability to kill insects and other groups of spiders that possessed downward-striking (paraxial) jaws for attacking prey on a firm substrate, either on the ground or on tree trunks.

The first insects, which make an appearance in the Namurian, were members of the extinct order Palaeodictyoptera. Together with five other orders they comprise the Palaeoptera, and all have their origins in the Pennsylvanian. Two orders, the Ephemeroptera (mayflies) and Odonata (dragonflies and damselflies) are extant.

The other Carboniferous orders are, like the Palaeodictyoptera, extinct, surviving only until the end of the Permian. The Protodonata were very similar to the first dragonflies except for a variation in wing venation and their great size. The Megasecoptera and Diaphanopterodea were very similar to the Palaeodictyoptera.

Five orders of exopterygote neopteran insects also appear in the Pennsylvanian. Three of these orders became extinct at the end of the Palaeozoic. These were the Protorthoptera, which was the most diverse of all the extinct orders, the Caloneurodea, which some workers place in the Endopterygota, and the Miomoptera, which were very small insects, perhaps related to the Protorthoptera. The other two orders, the Blattodea (cockroaches) and Orthoptera (grasshoppers, katydids, and crickets), are extant. The Blattodea were apparently the commonest group of insects during the Pennsylvanian. There is increasing evidence of plant–insect interactions in the Carboniferous.

Vertebrates

The Carboniferous saw the diversification and spread of vertebrates onto land. Earliest Mississippian deposits contain relatively few tetrapod fossils. Some late Tournasian deposits in Scotland and Canada have yielded a few isolated fragments. The later Viséan rocks, mainly of Scotland, have yielded many important new skeletons. A single small specimen of tetrapod, named *Casineria kiddi*, has been described from mid-Viséan deposits in eastern Scotland. The anatomical features of the specimen indicate that it was fully terrestrial. The most spectacular finds in recent years have come from the Late Viséan deposits of East Kirkton. Amongst a wide range of taxa, all less than 1 m long, are the temnospondyls. Other forms include early anthrocosaurs. This group is better known from Pennsylvanian strata. *Westlothiana lizziae* (Lizzie the lizard) is a species of particular note as it shows many of the characteristics of reptiles.

Several tetrapod groups show evidence of major diversification in the late Mississippian and early Pennsylvanian (*see* **Fossil Vertebrates:** Palaeozoic Non-Amniote Tetrapods). Most of the important tetrapod sites are found in tropical Euramerica and include the North American sites of Linton (Ohio), Mazon Creek (Illinois), and Joggins (Nova Scotia), several British and Irish sites, including Jarrow and Newsham, and continental European sites, such as Nyrany (Czech Republic).

The temnospondyls underwent a major radiation in the early Pennsylvanian and were the largest group of 'amphibians.' The most important of these was the small *Dendrerpeton*, which was first found in the tree stumps at Joggins. Amongst the temnospondyls were several aquatic forms. The anthracosaurs were another important group of tetrapods, also with a Mississippian origin. They included the crocodile-like embolomeres such as *Eoherpeton*. The most fearsome predator in this group was *Anthracosaurus*, which is known only from its skull.

Microsaurs also have a Mississippian origin; they include several important genera and have been confused with true amniotes. Other related groups include the lysorophids and aiistopods. Amniotes (including the reptiles for some authors) are first found in the Mississippian but radiated in the Pennsylvanian. Whilst several reptiles are known from the Pennsylvanian, it is clear that terrestrial tetrapod faunas were dominated by amphibians. It is significant to note that, with the possible exception of *Scincosaurus*, all the major Upper Carboniferous tetrapods were probably carnivores or insectivores.

The first true amniote (as opposed to ancestral group) was also found at Joggins. *Hylonomus* was amongst the animals found within the Joggins tree stumps. Other forms include *Captorhinus*, but it was not until the latest Carboniferous that herbivory evolved in larger forms, such as *Edaphosaurus*. In the Early Permian of Gondwana there is evidence of the

first aquatic amniote, *Mesosaurus*, which is found on several continents. It is in the Permian that the first major radiation of amniotes occurred.

Life in the Sea

Much of our knowledge about Carboniferous marine life comes from the extensive tropical shallow-water carbonates of the Mississippian. These warm shallow seas supported a rich benthos dominated by filter-feeding brachiopods and crinoids. Plankton diversity appears, however, to have been low, with the fauna dominated by only a few groups of acritarchs (marine aquatic algae). In some areas reefs dominated by rugose and tabulate corals together with sponges, bryozoans, and calcareous algae were important. Fusiline foraminifera became important for the first time. Nektonic animals included goniatites and such vertebrates as conodonts, which are important for biostratigraphical correlation. Undisputed conodont animals were discovered in the Mississippian rocks of Eastern Scotland associated with various aquatic arthropods. Fishes are also found in many marine and non-marine environments. A variety of sharks are known from Carboniferous deposits, and some of the best examples (such as *Stethacanthus* from Bearsden in Scotland) are from brackish marine deposits. Ray-finned and lobe-finned fishes also diversified in the oceans at this time.

The structure of Pennsylvanian marine communities was similar to that in the Mississippian. However, Pennsylvanian marine strata are not as widespread. In Europe the spread of marine facies across low-lying mires is well documented. The marine rocks are predominantly muddy and yield goniatites, bivalves, and conodonts, in contrast to the limestones that yield corals and brachiopods in addition to the goniatites and conodonts known from North America and Ukraine. Fusiline foraminifera became increasingly abundant through the Pennsylvanian in these carbonate environments.

The End of the Carboniferous and Mass Extinctions

Major changes took place at the end of the Carboniferous. Plate movements resulted in the continued collision of Laurasia and Gondwana to form the supercontinent Pangaea. The formation of Pangaea led to major climatic change, with tropical drying and the spread of continental red beds. The decrease in rainfall prevented peat formation in more tropical areas and was accompanied by the apparent extinctions of many terrestrial plant and animal groups. The floras of dryer upland habitats began to spread towards the lowlands. These floras were dominated by the newly evolved conifers. In the southern latitudes, the floras that had been typical of the Carboniferous, composed of progymnosperms, pteridosperms, spenopsids, and dwarf lycopsids, gave way to the glossopterid-dominated floras of the Permian. Recent research in China has shown that many of the extinctions in Euramerica at the end of the Carboniferous were local and not global. Several Carboniferous plant and animal groups are found in the Permian of China and become extinct only at the end of the Permian.

In the oceans there was continued faunal turnover, with some groups becoming extinct and others, such as the fusiline foraminifera, diversifying and becoming more important and widespread – they are particularly useful for defining the base of the Permian and for correlating subsequent Permian stages. There is no evidence, however, of a mass extinction at the end of the Carboniferous.

See Also

Analytical Methods: Geochronological Techniques. **Atmosphere Evolution**. **Earth:** Orbital Variation (Including Milankovitch Cycles). **Fossil Vertebrates:** Palaeozoic Non-Amniote Tetrapods. **Gondwanaland and Gondwana**. **Palaeoclimates**. **Pangaea**. **Sedimentary Processes:** Glaciers.

Further Reading

Ager DV (1984) *The Nature of the Stratigraphical Record*, 3rd edn. Basingstoke: MacMillan Publishers Ltd.

Calder JH and Gibling MR (1994) The Euramerican Coal Province: controls on Late Paleozoic peat accumulation. *Palaeogeography, Palaeoclimatology, Palaeoecology* 106: 1–21.

Clack JA (2002) *Gaining Ground: The Origin and Evolution of Tetrapods*. Bloomington: Indiana University Press.

Cleal CJ (1991) Carboniferous and Permian biostratigraphy. In: Cleal CJ (ed.) *Plant Fossils in Geological Investigation: The Palaeozoic*, pp. 182–315. Chichester: Ellis Horwood.

Cleal CJ and Thomas BA (1991) Carboniferous and Permian palaeogeography. In: Cleal CJ (ed.) *Plant Fossils in Geological Investigation: The Palaeozoic*, pp. 154–183. Chichester: Ellis Horwood.

Davydov VI, Glenister BF, Spinosa C, *et al.* (1998) Proposal of Aidaralash as Global Stratotype Section and Point (GSSP) for base of the Permian System. *Episodes* 21: 11–18.

DiMichele WA, Pfefferkorn HW, and Phillips TL (1996) Persistence of Late Carboniferous tropical vegetation during glacially driven climatic and sea-level fluctuations. *Palaeogeography, Palaeoclimatology, Palaeoecology* 125: 105–128.

Gastaldo RA and DiMichele WA (eds.) (2000) *Phanerozoic Terrestrial Ecosystems*. Special Papers 6. Paleontological Society.

Gastaldo RA, DiMichele WA, and Pefferkorn HW (1996) Out of the icehouse into the greenhouse: a late Paleozoic analogue for modern global vegetational change. *GSA Today* 10: 1–7.

Hower J, Scott AC, Hutton AC, and Parekh BK (1995) Coal availability, mining and preparation. In: Bisio A and Boots SG (eds.) *Encyclopaedia of Energy Technology and the Environment,* pp. 603–684. New York: J Wiley & Sons.

Iannuzzi R and Pfefferkorn HW (2002) A pre-glacial, warm-temperate floral belt in Gondwana (Late Viséan, Early Carboniferous). *Palaios* 17: 571–590.

Labandeira CC (1998) Early history of arthropod and vascular plant associations. *Annual Review of Earth and Planetary Sciences* 26: 329–377.

Lane HR, Brenckle PL, Baesemann JF, and Richards B (1999) The IUGS boundary in the middle of the Carboniferous: Arrow Canyon, Nevada, USA. *Episodes* 22: 272–283.

Menning M, Weyer G, Drozdzewski G, Van Amerom HWJ, and Wendt I (2000) A Carboniferous time scale 2000: discussion and use of geological parameters as time indicators from central and Western Europe. *Geologische Jahrbuch,* Hannover, A 156: 3–44.

Paproth E, Feist R, and Flajs G (1991) Decision of the Devonian–Carboniferous boundary stratotype. *Episodes* 14: 331–336.

Phillips TL and Peppers RA (1984) Changing patterns of Pennsylvanian coal-swamp vegetation and implications of climatic control on coal occurrence. *International Journal of Coal Geology* 3: 205–255.

Rolfe WDI, Durant GP, Fallick AE, *et al.* (1990) An early terrestrial biota preserved by Viséan vulcanicity in Scotland. In: Lockley MG and Rice A (eds.) *Volcanism and Fossil Biotas,* pp. 13–24. Special Publication 244. Boulder: Geological Society of America.

Scheffler K, Hoernes S, and Schwark L (2003) Global changes during Carboniferous–Permian glaciation of Gondwana: linking polar and equatorial climate evolution by geochemical proxies. *Geology* 31: 605–608.

Scott AC (1980) The ecology of some Upper Palaeozoic floras. In: Panchen A (ed.) *The Terrestrial Environment and Origin of Land Vertebrates,* pp. 87–115. Systematic Association Symposium Volume 15. London and New York: Academic Press.

Scott AC (2000) The Pre-Quaternary history of fire. *Palaeogeography, Palaeoclimatology, Palaeoecology* 164: 281–329.

Scott AC and Galtier J (1996) A review of the problems in the stratigraphical, palaeoecological and palaeobiogeographical interpretation of Lower Carboniferous (Dinantian) floras from Western Europe. *Review of Palaeobotany and Palynology* 90: 141–153.

Scott AC, Stephenson J, and Chaloner WG (1992) Interaction and co-evolution of plants and arthropods in the Palaeozoic and Mesozoic. *Philosophical Transactions of the Royal Society of London B* 335: 129–165.

Scott AC, Mattey D, and Howard R (1996) New data on the formation of Carboniferous coal balls. *Review of Palaeobotany and Palynology* 93: 317–331.

Wagner RH and Winkler PCF (1991) Major subdivisions of the Carboniferous system. In: *Compte Rendu XI^{e} Congrès International de Stratigraphie et de Géologie du Carbonifère, Beijing 1987,* pp. 213–245.

Wagner RH and Winkler PCF (1997) Carboniferous chronostratigraphy: quo vadis? *Proceedings of the XIII International Congress on the Carboniferous and Permian*. In: *Prace Panstwowego Instytutu Geologicznego CLVII,* pp. 187–196.

Wnuk C (1996) The development of floristic provinciality during the Middle and Late Paleozoic. *Review of Palaeobotany and Palynology* 90: 5–40.

Permian

P B Wignall, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The Permian Period, spanning the 43 Ma interval between 294 Ma and 251 Ma, witnessed several major turning points in Earth's history. Pangaea, the greatest supercontinent of all time, finally amalgamated in the Permian, and began to break up again almost immediately in a prolonged disintegration that would finally finish in the Cretaceous. Climate also changed markedly: the extensive southern-hemisphere glaciation that had begun in the Carboniferous concluded in the Middle Permian, to be replaced by a warming trend that was to peak in the earliest Triassic with some of the warmest conditions the world has ever experienced. The Permian was also a time of extremes in the fortunes of plants and animals. The Early to Middle Permian was generally a successful time for life, with low extinction rates and the evolution of the first large land animals and of plant communities that invaded the continental interiors. In contrast, the end of the Middle Permian was marked by an extinction crisis that primarily affected the inhabitants of shallow tropical seas. This was the first such crisis since the end of the Devonian, 100 Ma previously, and it brought to an end one of the most benign intervals of Earth history. Much worse was to follow with the end-Permian mass extinction, by far the biggest disaster in the history of life.

Subdivisions of Permian Time

Dividing the Permian into smaller time units has always proved surprisingly controversial. Whereas the subdivision of other intervals, such as the Jurassic and Cretaceous, has been stable and established for many decades, the current chronostratigraphical scheme for the Permian (Table 1) is only a few years old, and numerous other schemes can be found in the older literature. The problem stems principally from the lack of useful fossils available to correlate Permian strata in different regions of the world. In contrast, Ordovician stratigraphers can easily subdivide time using a succession of extremely widespread graptolite fossils, and Jurassic stratigraphers similarly use ammonites for subdivision and global correlation. For a long time Permian stratigraphers used ammonoids to define biozones, but these tend to be restricted to specific regions and in many areas are absent altogether. For example, there are extensive exposures of Permian rocks in western and Arctic Canada, but ammonoids are extremely rare in these deposits. In particular, since there are no ammonoids diagnostic of Late Permian time in this region, it was thought that Late Permian rocks were not present there. It now appears more likely that such rocks are present but the diagnostic fossils are not – a crucial distinction. Much Permian correlation is now achieved using conodonts (*see* **Microfossils:** Conodonts), a group of microfossils that are widespread in marine sediments of this age and therefore more useful than the ammonoids in establishing global time units.

Despite recent advances, there are still problems with correlating Permian rocks, primarily because different intervals of time are defined in different regions. Thus, the Early Permian (Cisuralian) is based on sections from Russia, the country where the period was originally defined; the Middle Permian (Guadalupian) is based on US sections, mostly in Texas; and the Upper Permian (Lopingian) is based on excellent marine sections in South China. Two Russian names, the Dzhulfian and the Dorashamian, were commonly used for the Late Permian interval, but these proved very difficult to identify outside the region, once again because of problems with the localized occurrences of ammonoids. As a result, the Dorashamian Stage – identified using ammonoids – was thought to be absent in most regions, but the essentially equivalent Changxingian Stage – identified using conodonts – is known to be widespread.

Tectonics

For nearly 200 Ma, from the Ordovician to the Permian, the Earth's continents gradually coalesced until eventually most continental crust had been amalgamated to form the supercontinent of Pangaea (Figure 1) (*see* **Pangaea**). The final stage of closure occurred during the Permian with collision of the Siberian continent with the Laurussian portion of northern Pangaea, the Eurasian Orogeny. This produced the Urals, a narrow and very long mountain belt that extends from the Volga region of southern Russia to Novaya Zemlya in the Arctic Ocean. During the Permian this newly forming mountain range shed vast quantities of sediment to the west of the advancing deformation front, into a deep-water basin in the eastern Barents Sea area, to the north of Norway, and into a series of terrestrial molasse

basins: the Volga-Urals and Timan-Pechora basins. To the east of the Urals the vast West Siberian Basin was also receiving large quantities of sediment and for much of the Late Permian sediment influx and subsidence kept pace so that the region remained around sea-level. In the latest Permian this area was the site of eruption of one of the largest volcanic provinces ever to form: the Siberian Traps. Subsidence rates continued to be high and the region remained around sea-level despite the eruption of several kilometres thickness of lava.

By the Middle Permian the areal extent of Pangaea was at its zenith, and it stretched in a giant arcuate shape from the north pole, which lay in eastern Siberia, to the south pole, which then, as now, lay in Antarctica (Figure 1). Pangaea was narrow at its waist owing to the presence of Tethys, an equatorially centred ocean that separated the northern and southern arms of the supercontinent. The eastern opening of Tethys was partially closed by the presence of some of the few continental masses that were not incorporated in Pangaea, notably South China and several smaller continents that are now part of south-east Asia. On the other side of the world from Tethys a vast ocean, called Panthalassa, stretched across 220° of longitude, making it 20% larger than the modern Pacific. Viewed from space in the Permian one side of the world would have looked entirely blue!

The break up of Pangaea began even as the Siberia–Laurussia collision was nearing completion, with the rifting of several small continents from the southern margin of Tethys, known as the Perigondwanan margin, in the Middle Permian (Figure 1). As these drifted northwards they opened up a narrow ocean behind them, known as Neotethys; the older Tethys to their north is often known as Palaeotethys. These slivers of continents included the present-day regions of Antalaya (Turkey), Armenia, parts of Iran and Afghanistan, and central Tibet; they are often depicted as forming an elongate narrow continent called Cimmeria. However, it is uncertain whether these ever formed a continuous continent. The Tibetan Block in particular seems to have begun its drift somewhat later than the other continents and to have lagged behind them during their northern passage. The Middle Permian rifting activity was accompanied by some voluminous volcanic episodes, notably that which formed the Panjal volcanics of Kashmir, India. This formerly extensive province of basaltic lavas was subsequently much deformed and metamorphosed during the formation of the Himalayas. However, the biggest Permian volcanic province is in the Emeishan region of south-western China and seems to have occurred during a poorly understood rifting episode around the Middle Permian–Late Permian boundary. Extensive sheets of basalt dominate the Emeishan

Table 1 Subdivisions of Permian time (names in brackets are those formerly used to describe Late Permian intervals)

Series	*Stages*
Lopingian base 262 Ma, top, 251 Ma (Tatarian)	Changxingian (Dorashamian) Wuchiapingian (Dzhulfian)
Guadalupian base 272 Ma	Capitanian Wordian Roadian
Cisuralian base 294 Ma	Kungurian Artinskian Sakmarian Asselian

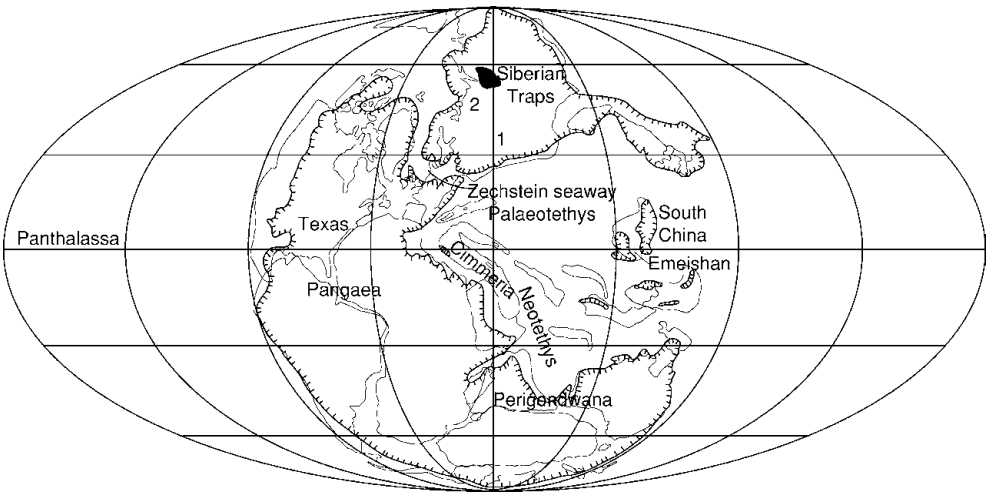


Figure 1 Global palaeogeography of the Late Permian showing the supercontinent of Pangaea. 1, Volga-Urals Basin; 2, Timan-Pechora Basin.

volcanic pile, and they may exceed a million cubic kilometres in volume. This basalt was extruded quietly from long fissures, but, during the later stages of the volcanism, these eruptions were replaced by explosive eruptions that produced more acidic lavas.

One consequence of supercontinent formation is that the area of fringing shallow shelf seas becomes minimal. This was indeed the case in the Permian, when the area of shallow-marine deposition was probably at its lowest level of the past 600 Ma. Truly extensive shallow seas were restricted to the continents that were not part of Pangaea, notably South China. However, several major marine embayments pierced the coast of Pangaea and brought shallow seas into the heart of the supercontinent. Two of these embayments were associated with the development of rift systems that would eventually contribute to the break up of Pangaea. Thus, down the east coast of Greenland and along the east coast of Africa (west coast of Madagascar) two long and narrow seaways developed.

A major invasion of the sea also occurred in the western and southern USA. In this region one of the most spectacular reef systems of all time developed. This is now superbly exposed in the Guadalupe Mountains of Texas. Much of the original topography of this reef system has been re-exposed, and it is possible to walk from the reef flat down the steep reef front to the deep basinal setting in front of the reefs – now a desert floor in front of a mountain escarpment. The other major area of Permian marine deposition in the USA occurs in the Idaho–Wyoming–Utah region. Here the rocks are of a very different character from the Texan rocks: dark shales and phosphates dominate the succession. Phosphates form today where cold deep waters upwell on western-facing continental margins, and a similar origin appears likely for the Permian phosphates.

At low to intermediate latitudes the deep interior of Pangaea was very arid (see below) and so it is not surprising that many of the seaways that penetrated into the supercontinent were subject to intense evaporation. The biggest of these evaporitic basins occurred in the north-west European area, where the Zechstein Sea was gradually infilled with a vast thickness of sea salts – one of the thickest evaporite deposits in the geological record.

Climate

Vast areas in the centre of Pangaea were a long way from any seaway and so it is not surprising that arid conditions were common over many areas of the continent. This was particularly true in the northern hemisphere, where much of Europe and North America lay in the rain shadow of the Appalachian–Variscan mountain chain (**Figure 1**). Many geologists live in these regions today and so a general impression of the Permian as a time of aridity has tended to enter the literature. However, the view from the Permian polar regions is rather different, and extensive coal deposits – testimony to humid climates – are known from the Permian of Siberia and Antarctica. Many of these coals formed in forested areas that were at latitudes of over 65° (e.g. eastern Australia, Antarctica, and eastern Siberia). These deposits indicate substantial rainfall and a lack of permanent snow cover at the Permian poles, conditions that were very different from those encountered in polar areas today.

The 43 Ma duration of the Permian also saw substantial changes in climate, the most important of which was the waning of the southern-hemisphere glaciation that had begun in the Early Carboniferous. At the beginning of the Permian extensive ice-sheets covered large areas of Antarctica and South Africa and deposited a jumble of boulders and clays known as the Dwyka Tillite. Ice also advanced as far north as southern Australia and parts of South America, at palaeolatitudes of approximately 45° S, before gradually shrinking back. By the end of the Middle Permian there is no evidence of permanent ice anywhere. The reason for the termination of this major glacial interval, known as the Permo-Carboniferous glaciation, is not understood but it may relate to the drift of southern Pangaea away from the south pole. At the start of the Permian the pole lay within Antarctica but by the Late Permian it was within the southernmost Panthalassa Ocean.

Marine Fossils

Life in the Permian seas and oceans was remarkably similar to that in the preceding Carboniferous. In the tropical carbonate seas of equatorial latitudes the bottom-dwelling fauna was diverse and consisted of brachiopods, stalked echinoderms (especially crinoids), bryozoans, corals, lesser numbers of molluscs, and the occasional rare trilobite. Foraminifera (single-celled organisms that secrete a tiny calcareous skeleton) were also common and they produced a group of 'giants' in the Permian – the fusulinids – that reached up to a centimetre in diameter. These are abundant and highly diverse in some Permian limestones. Reefs were common in equatorial locations and included the magnificent examples from the Guadalupe Mountains (mentioned above) and many others from Tethyan settings. Calcareous sponges and calcareous algae (especially a form known as *Tubiphytes*) were the principal constructors of these

reefs, although the associated fauna was diverse. This included a bizarre group of brachiopods known as the richthofenids, which evolved a conical valve with a lid-like upper valve. In gross morphology they resembled the contemporaneous rugose corals, and they were undoubtedly pursuing a similar life strategy.

Land Animals

Prior to the Late Carboniferous–Early Permian it is probable that very few land animals ever strayed far from water, but this changed dramatically during the Permian with the colonization of the continental interiors, a major ecological breakthrough that was almost certainly triggered by the evolution of eggs that did not have to be laid in water. The majority of Permian tetrapods (four-legged land dwellers) belong to two major groups, the synapsids, a group that ultimately gave rise to mammals (but not in the Permian), and the diapsids, a group that gave rise to reptiles and later (in the Triassic) to the dinosaurs.

Unlike the relatively stable Permian marine invertebrate faunas, the Permian tetrapod faunas underwent several phases of turnover. The initial radiation was dominated by the pelycosaurs, a synapsid group that included the large sail-backed predator *Dimetrodon*, one of the few 'famous' Permian tetrapods. The pelycosaurs included both herbivores and carnivores of a broad range of sizes although all had a characteristic sprawling pose and dragged long tails. They would have undoubtedly looked clumsy compared with many modern mammals. In the later Permian the pelycosaurs gave way to an even more diverse group of synapsids known as the therapsids. These included the dicynodonts, which were large (up to 3 m long) powerfully built herbivores, the cynodonts, which were smaller herbivores with distinctive tusks in their upper jaws, and the rather fearsome-looking gorgonopsids, which were predators of around a metre in length. The diapsids contributed the pareiasaurs, a group of large powerful herbivores, to the impressively diverse Late Permian tetrapod communities.

Land Plants

Just as the conquest of the land by animals was probably facilitated by the evolution of eggs, so the parallel conquest by plants was probably driven by the evolution of seed-bearing plants. Thus, the Permian saw the decline of tree ferns, which dominated Carboniferous floras, and the rise of several groups of seed plants (particularly cycads and ginkgos) and seed ferns (glossopterids). Four floral realms, with a clear climate-controlled distribution, are readily distinguished in the Permian. The flora of the Angaran Realm, centred on Siberia, was adapted to the cool-to-cold climates of this region. Cordaites, a primitive group of seed ferns, and the fern-like sphenopsids dominated the flora as they had done in this region in the preceding Late Carboniferous. These two groups were also common, together with ginkgos, in the Euramerican flora, which occupied the semiarid and possibly monsoonal locations of some low northerly palaeolatitudes. The Cathaysian flora, centred on northern China, was adapted to the warm ever-wet climate of this equatorial part of the world. It was similar in composition to the tree fern-dominated flora of the preceding Carboniferous, although with the addition of the large-leaved gigantopterids, another group of seed ferns. The most distinctive Permian flora occurred in the Gondwanan realm, where deciduous glossopterid tree fern forests were extensive (Figure 2). These thrived in the cool temperate conditions that developed in high-latitude locations such as Australia and South Africa after the melting of the ice-caps in the Early Permian. Most of the coal reserves of these countries, together with the less exploitable coals of Antarctica, formed at this time.

Extinction

For a long time the end-Permian mass extinction (*see* **Palaeozoic:** End Permian Extinctions) was thought to be a protracted affair that had begun well before the end of the Permian. However, recent work has demonstrated that there were in fact two separate extinction events in the latter part of the Permian, separated by a phase of radiation and recovery in the Late Permian. The end-Middle Permian extinction, sometimes known as the end-Guadalupian event, is poorly understood, primarily because it has only recently been discovered and has been little studied, but also because there are few complete successions recording the transition from Middle Permian to Upper Permian strata. The effects of this crisis are best known from equatorial sections, particularly in southern China where many species of fusulinid and brachiopod went extinct. Reef development also ceased in many areas of the world, including the Guadalupe reefs of Texas, at around this time. What is far from clear is whether this extinction event can be detected at higher latitudes and whether there were any corresponding extinctions on land. The plant record does not appear to show any major change around this time.

Despite its recent discovery, several theories have already been proposed to explain the end-Guadalupian mass extinction. As noted above, complete sections spanning the Middle Permian–Upper Permian boundary are rare. This is because of a major fall in sea-level

Figure 2 Slab of rock from the Permian of Antarctica, showing abundant fronds of the tree fern *Glossopteris*.

at this time, possibly to one of the lowest levels of the past 600 Ma. The result was that shallow-marine habitats became very limited, with drastic consequences for the inhabitants of such environments. Climate change triggered by volcanic activity is also a possible factor in this extinction event (and in many other extinction events). The Emeishan volcanic traps of southern China are roughly contemporaneous with the extinction event, raising the possibility that the emissions of volcanic gas associated with the lava eruptions may have changed the climate. However, precise synchronicity of the eruptions and the extinction has yet to be demonstrated. Recently, a Middle Permian–Late Permian succession was discovered in a limestone succession that formed on a small island that was formerly within the Panthalassa Ocean (the rocks have now been accreted to Japan). Remarkably, an acidic volcanic ash is developed at the boundary. This is a product of the kind of explosive volcanism normally characteristic of continental rather than oceanic locations. A huge volcanic eruption in a continental location (perhaps Emeishan) may have blown ash out into the far reaches of the Panthalassa Ocean. If ash could reach this distant location, then it is possible that it could also have triggered drastic climate change. Only further study will demonstrate the veracity of the currently proposed mechanisms of the end-Guadalupian extinction.

See Also

Fossil Invertebrates: Porifera. **Fossil Vertebrates:** Palaeozoic Non-Amniote Tetrapods. **Mesozoic:** Triassic. **Microfossils:** Conodonts. **Palaeoclimates**. **Palaeozoic:** Carboniferous; End Permian Extinctions. **Pangaea**. **Sedimentary Environments:** Reefs ('Build-Ups'). **Sedimentary Rocks:** Evaporites.

Further Reading

Benton MJ (1997) *Vertebrate Palaeontology*, 2nd edn. London: Chapman & Hall.

Hallam A and Wignall PB (1997) *Mass Extinctions and Their Aftermath*. Oxford: Oxford University Press.

Parrish JT (1993) Climate of the supercontinent Pangaea. *Journal of Geology* 101: 215–233.

Willis KJ and McElwain JC (2002) *The Evolution of Plants*. Oxford: Oxford University Press.

End Permian Extinctions

R J Twitchett, University of Plymouth, Plymouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The extinction event at the close of the Permian period was the largest of the Phanerozoic. Understanding this event is crucial to understanding the history of life on Earth, yet it is only since the late 1980s that scientists have begun to study this event in detail. This is partly because of the discovery of many new geological sections spanning the time of the extinction crisis, partly because of a renewed public and scientific interest in extinction, and partly because of increasing dialogue, travel, and exchanges between geologists and palaeontologists from the West (Europe and America) and those from Russia and China. Over the years, a bewildering number of hypotheses have been suggested to explain the end-Permian extinction crisis, and new data are constantly appearing. There is currently no consensus as to the precise cause, and the current leading contenders are considered below.

Definition and Dating

The Permian–Triassic boundary is now defined as the point at which a specific species of fossil conodont (an extinct type of primitive vertebrate) called *Hindeodus parvus* first appears in a geological section at Meishan in South China (Figure 1). This definition was accepted only after many years of debate: its significance is immense. Palaeontologists from around the world are now able to correlate their local sections with this single reference section, thus unifying a confusing plethora of names that geologists working independently, in different countries, had erected to describe rocks of the same age. It also confirmed that the main extinction event occurred in the latest Permian, before the Permian–Triassic boundary (i.e. before the appearance of *H. parvus*), in the Changhsingian Stage (Figure 2).

The presence of layers of volcanic ash interbedded with the fossiliferous sediments at Meishan have also allowed absolute dating of the boundary interval through the precise measurement of radiogenic elements, such as ^{206}Pb and ^{238}U, in minerals such as zircons. In 1991, Claoue-Long and colleagues dated the Permian–Triassic boundary at 251.2 ± 3.4 Ma. Seven years later, Sam Bowring and his group reanalysed the same beds and arrived at a date of 251.4 ± 0.3 Ma, confirming the earlier result but with far higher resolution and cementing 251 Ma in the subsequent literature.

However, accurate dating is notoriously difficult, and the further back in time one goes the worse the resolution becomes and the greater is the potential error. In 2001, Roland Mundil and colleagues confirmed that there were problems with the Bowring dates. Mundil's group reassigned the ash bed above the Permian–Triassic boundary an (older) age of 252.7 ± 0.4 Ma. However, the ash bed below the Permian–Triassic boundary, approximating to the level of the mass extinction, proved far more difficult to date: that it is probably older than 254 Ma was the best that the geochronologists could achieve.

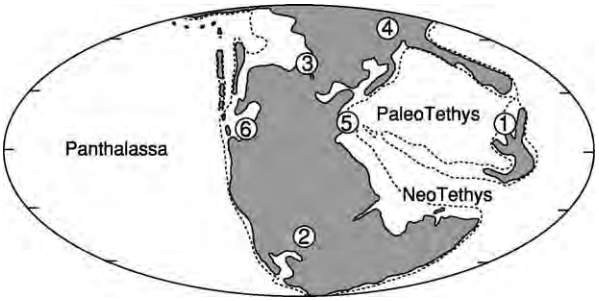


Figure 1 Sketch of Permian–Triassic palaeogeography, showing the major regions discussed in the text: 1, South China; 2, Karoo Basin, South Africa; 3, eastern Greenland; 4, Siberia; 5, northern Italy; and 6, western USA.

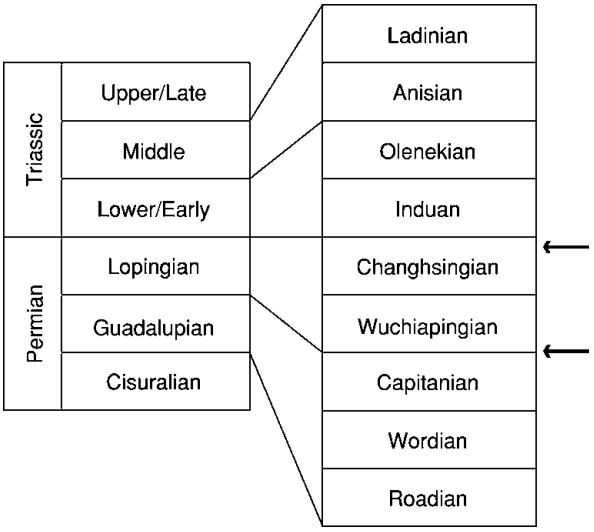


Figure 2 Permian and Triassic stratigraphy. Arrows indicate the stratigraphical positions of the extinction events discussed in the text.

Is this change in date and loss of resolution important? For many Permian–Triassic studies it is largely irrelevant. Correlation between marine sections relies on well-established methods of biostratigraphy and is unaffected by such results: the relative sequence of events can be deduced regardless. However, it is a significant blow for correlation between fossiliferous marine sediments and distant terrestrial sections, and for the hypothesis that the Siberian flood basalts were responsible for the extinction crisis (see below). It also means that absolute rates of change, both faunal and environmental, cannot be determined.

What Went Extinct?

Estimating the severity of past extinction events is not easy. Owing to the vagaries of preservation and fossil recovery and the difficulty of recognizing biological species from fossils, palaeontologists investigating extinctions tend to work at higher taxonomic levels. Thus, estimates of extinction magnitude are derived from large databases of family-level diversity through time, which show that 49% of marine invertebrate families disappeared during the Permian–Triassic interval.

Extinction at the species level is then estimated, rather than directly measured, using a statistical technique called reverse rarefaction. From this method, the often-quoted figure of a 96% loss of marine species is derived. However, such a calculation makes a number of assumptions, including, for example, the assumption that species extinction was random, with no selectivity against certain groups. The balance of evidence indicates that this assumption is incorrect, and, thus, the true magnitude of species loss may be closer to 80%. However, this still makes the end-Permian extinction event the most severe crisis of the Phanerozoic!

Marine Extinctions

Despite these shortcomings, it is clear that several groups were severely diminished in number in the Late Permian, some groups went completely extinct, and not all groups were affected equally. For example, the diverse and successful fusulinid foraminifera disappear suddenly during the Changhsingian stage with the loss of some 18 families, whereas other benthic foraminifera suffered much lower levels of extinction. The diversity of the stenolaemate bryozoans also decreased markedly through the latter half of the Permian, but only the Order Fenestrata became extinct. By contrast, other bryozoan groups, such as the gymnolaemate order Ctenostomata, maintained their diversity at the family level.

A pattern of gradual decline throughout the Permian, particularly during the latter half, followed by final extinction of the last few remaining taxa in the Changhsingian, is typical of many groups. Examples include the Palaeozoic corals (Rugosa and Tabulata), trilobites, and goniatites, all of which became extinct, and the articulate brachiopods, which were reduced to a handful of surviving genera. This pattern implies that longer-term (possibly climate driven) changes in the marine realm may have been largely to blame for the loss of diversity.

A few groups, such as the echinoderms, suffered dramatic losses prior to the end-Permian. All of the extant echinoderm groups experienced severe bottlenecks during the Permian–Triassic interval: only a couple of lineages of crinoids and echinoids survived into the Mesozoic. However, the major crisis interval appears to have been in the Late Guadalupian, when crinoids experienced over 90% loss and other groups of echinoderms, such as the Blastoidea, became extinct.

Terrestrial Extinctions

The extinction on land was just as severe as in the oceans, with some 77% of vertebrate families becoming extinct. In the Karoo Basin of South Africa (Figure 1), only seven out of 44 tetrapod genera (16%) cross the Permian–Triassic boundary. When all fossil terrestrial organisms are considered together, the Permian–Triassic event is the single largest extinction episode in the otherwise exponential rise of terrestrial diversity through the Phanerozoic.

A long-term (more than 20 Ma) change in terrestrial vegetation occurred during the Permian–Triassic interval. Termed the Palaeophytic–Mesophytic transition, it probably reflects gradual shifts in climate and palaeogeography. However, superimposed on this longer-term trend is a catastrophic collapse of the dominant gymnosperm forests in the latest Permian, as evidenced by the disappearance of pollen taxa such as *Vittatina*, *Weylandites*, and *Lueckisporites*. Studies in eastern Greenland (Figure 1) have shown that this ecological crisis occurred simultaneously with the marine extinction event.

Some workers have suggested that a peak in the abundance of *Reduviasporonites* (also called *Tympanicysta*), observed in some sections, marks the sudden destruction of these forests. They have interpreted this taxon as a saprophytic fungus, which thrived on the piles of dead and dying vegetation. Unfortunately, this attractive scenario must now be rejected. Recent geochemical, structural, and biomarker studies have shown that *Reduviasporonites* is definitely not a fungus and is probably a photosynthetic alga.

One Event or Two?

Early databases of global marine biodiversity recorded a broad peak of elevated extinction rates spanning the entire Late Permian. As the data improved, two distinct peaks were resolved: one in the latest Permian (the major event) and one at the end of the Guadalupian. This latter event marks the point at which many diverse groups (such as sponges and stenolaemate bryozoans) begin to decline in diversity. For the brachiopods and echinoderms, especially the crinoids, it marks the main episode of diversity loss.

While there is some evidence of climate change at this time, there are also good reasons why the end-Guadalupian event might not be a 'real' biotic extinction event. The Middle Permian was a time of incredible biodiversity, but the majority of the fossil taxa are endemic to the comprehensively monographed and exquisitely preserved silicified faunas of the Guadalupian reef belts of the southern USA, particularly western Texas. At the close of the Guadalupian, sea-level fell across that region, and the overlying sediments are unfossiliferous evaporites. The disappearance of the fossils is therefore a result of facies change and not real extinction.

Globally, Upper Permian fossiliferous marine rocks are relatively poorly known, and the quality (i.e. completeness) of the Late Permian fossil record is correspondingly poor. Many taxa that were common in the Guadalupian have not been found in Upper Permian rocks but must have been living at that time because they reappear in the Triassic. These missing taxa are called Lazarus taxa. The presence of many Lazarus taxa indicates that the completeness of the fossil record is low (Figure 3).

The reality of the end-Guadalupian event is still debated. If it is simply the result of bias in the rock record, then newly described sections should contain many of the taxa previously supposed to have become extinct. Certainly this is true of the brachiopods: newly described Upper Permian sections from Tibet contain some taxa that were thought to have vanished earlier. Further detailed analysis of this interval is clearly needed.

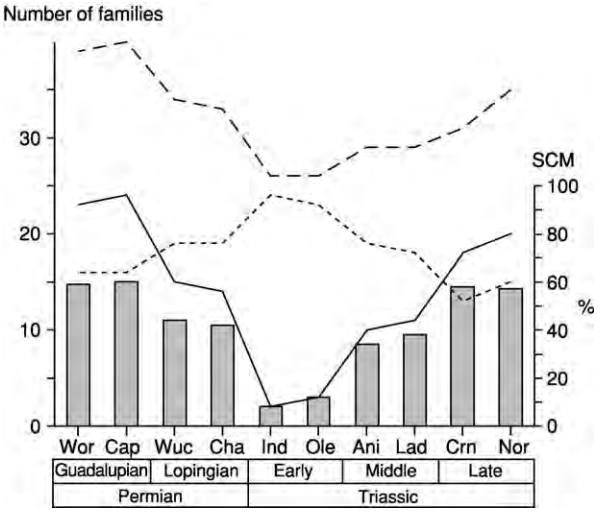


Figure 3 Fluctuations in fossil diversity and the quality of the fossil record of marine sponges through the Permian–Triassic extinction and recovery interval: solid line, families of sponges represented by actual fossils; dotted line, number of Lazarus taxa; dashed line, total diversity (sum of fossil taxa and Lazarus taxa). Grey bars show the simple completeness metric (SCM), which is the number of actual fossil taxa as a percentage of the total diversity. The SCM gives an estimate of the completeness of the fossil record. Note that, as diversity declines after the Guadalupian, the number of Lazarus taxa increases and so the completeness of the fossil record decreases. Stratigraphy from Figure 2. Data from *The Fossil Record 2* (Benton, 1993).

Possible Causes

Extraterrestrial Impact

Initial attempts to find evidence of a Permian–Triassic impact were made in the early 1980s, following the proposal that a massive meteorite impact had caused the end-Cretaceous extinction event (*see* **Mesozoic:** End Cretaceous Extinctions). Three pieces of evidence are considered crucial in identifying a meteorite impact event: a crater, an enrichment ('spike') of the rare metallic element iridium, and a layer of impact debris containing shocked quartz and tektites. However, the few reports of these pieces of evidence in latest Permian rocks are considered to be, at best, equivocal. Certainly, there is no Permian–Triassic iridium spike comparable to that found at the Cretaceous–Tertiary boundary, no widespread impact layer, and no large crater that can be accurately correlated with the extinction horizon. For an impact event to be a possible cause of the extinction crisis, it must also be demonstrated that the extinctions were unequivocally instantaneous and coincident with the impact itself.

In 2001, Luann Becker and colleagues reported the presence of helium and argon trapped in the cage-like molecular structure of fullerenes from the Permian–Triassic boundary in China and Japan. Isotopic profiles indicated that the helium and argon, and hence the fullerenes, had to have come from an extraterrestrial source. However, there are persistent problems with the acceptance of Becker's data, as other scientists have consistently failed to replicate her results, despite using samples from exactly the same sites and exactly the same laboratory procedures. A Japanese expert, Yukio Isozaki, has also argued that the Permian–Triassic boundary is missing from the Japanese section studied by Becker and colleagues

and that their samples came from below the extinction level.

Later in 2001, Kaiho and colleagues described Permian–Triassic sediment grains that were supposedly formed by impact, as well as geochemical shifts that they interpreted as indicative of a huge impact. However, their data are far from conclusive and were rapidly and severely criticized by other geochemists. In 2003, Asish Basu and colleagues revived interest by claiming to have found 40 tiny (50–400 μm) unaltered fragments of meteorite in a sediment sample from a terrestrial Permian–Triassic boundary section in Antarctica. Although Basu and colleagues were quick to dismiss contamination as a source of these grains, meteorite experts were immediately sceptical of the findings, as meteoritic metals are highly reactive and, in terrestrial settings, oxidize extremely quickly.

Doubtless scientists will continue to produce evidence of extraterrestrial impact at or near the Permian–Triassic boundary, if only because such hypotheses readily appeal to the wider public and journal editors alike. However, current data are highly controversial and lack the key criteria of major impact that have been recorded time and again in Cretaceous–Tertiary sections the world over. Independent replication of results is crucial for scientific acceptance, especially when the data are unusual and controversial. So far, all the evidence proposed for an impact at the Permian–Triassic boundary has failed this necessary test.

Eruption

The largest outpouring of continental flood basalts in the Phanerozoic occurred in Siberia during the Permian–Triassic interval. Including both the Siberian Platform basalts and the newly discovered coeval deposits buried in the West Siberia Basin, described by Marc Reichow and colleagues, the flood basalts covered an area of 1.6×10^6 km^2 to maximum depths of 3.5 km (Figure 1). If all other igneous rocks, such as pyroclastic flows, are included, then this coverage increases to 3.9×10^6 km^2. Dating the top and bottom of the lava pile shows that the eruptions occurred over a relatively short period of time, maybe just 600 000 years. Was this huge volcanic event a cause of the Permian–Triassic extinction crisis?

Radiometric dating is the only way to answer this question because no fossils have yet been collected from sediments interbedded with the basalts that provide correlation with other regions. Early efforts at dating the Siberian Traps produced an array of ages, from 160 Ma to 280 Ma. By contrast, more recent results, by different scientists using a variety of methods, cluster around 250 ± 1 Ma. In the late 1990s, this was considered to be exactly the date required, and the flood basalts were thought to be the primary trigger for the catastrophic extinction. However, the recent redating of the Meishan beds now implies that the Permian–Triassic event occurred before 253 Ma, and the Siberian Traps are therefore several million years too young.

On a Phanerozoic time-scale, the excellent correlation between extinction episodes and flood-basalt provinces means that it is difficult to accept the possibility that the Siberian Traps played no role in the end-Permian extinction event. There is still a chance. The oldest date that Reichow's team have for the onset of volcanic activity (dating the emplacement of intrusive gabbros) is 253.4 ± 0.8 Ma.

Another intriguing consequence of the 253 Ma age for the Permian–Triassic boundary is that the Emeishan flood basalts of western South China now need to be considered. Dating by Ching-Hua Lo and colleagues (published in 2002) shows that the main eruptions, which were largely marginal marine events, occurred around 251–253 Ma, with an initial phase of activity at around 255 Ma; thus, they predate the Siberian Traps and are dated close to the new dates for the Permian–Triassic crisis. Interestingly, Ching-Hua Lo's new dates preclude the possibility that the Emeishan flood basalts played a role in the end-Guadalupian crisis (at about 256–259 Ma; see above). Mei-Fu Zhou and colleagues had championed this particular correlation following their dating of the Emeishan basalts to 259 ± 3 Ma, which was published just a few months before Ching-Hua Lo's results in the same journal! Absolute dating is a difficult business (*see* **Analytical Methods:** Geochronological Techniques).

Global Warming

Animals and plants do not curl up and die just because a volcano is erupting on the other side of the world. Species become extinct when their population falls below viable levels. Population decline in natural systems is a response to local changes, such as loss of habitat, drought, and temperature change, many of which are ultimately driven by climate. What climate changes are recorded around the extinction interval and how might they be related to the flood-basalt volcanism?

Global cooling was proposed early on as a cause of the Permian–Triassic event; this largely hinged on the identification of latest-Permian glacial deposits in Siberia and eastern Australia. However, reanalysis of the biostratigraphy proved that these deposits are,

in fact, Middle Permian in age. With evidence of extensive ice sheets confined to the Early Permian and the last vestiges of glaciation now dated as Middle Permian, it is evident that the Permian as a whole records a warming trend.

At the culmination of this long-term warming trend, there is also evidence for an additional rapid greenhouse episode at the time of the extinction crisis. A large drop in the proportion of heavy oxygen isotopes (^{18}O) in carbonates from the Gartnerkofel-1 core of southern Austria has been interpreted as representing an increase in temperature of 6°C. However, interpretation of the oxygen isotope record is problematic as it is very sensitive to alteration during burial and diagenesis. The limestones of Gartnerkofel-1 have been largely recrystallized, especially around the boundary interval, and thus the oxygen isotope data are surely suspect.

However, there is growing independent support for global warming at the boundary. Gregory Retallack has demonstrated, from studies of the stomatal index of fossil leaves, that peaks in atmospheric CO_2 (a proven greenhouse gas) occur at both the end-Permian and end-Guadalupian boundaries. His studies of fossil soils also indicate a change to humid greenhouse conditions in the Early Triassic. Finally, there is a large (usually 3–6‰) increase in the proportion of light ^{12}C, relative to heavier ^{13}C, in carbonates deposited at this time. This shift is interpreted by most geochemists to be the result of methane (CH_4) release, as this is the only mechanism currently known that could rapidly deliver the required amount of light ^{12}C to the atmosphere and oceans. This methane, which is also a potent greenhouse gas, is assumed to have been produced by the melting of methane hydrate deposits in shallow marine shelves and polar tundra. Thus, a runaway greenhouse scenario is envisaged, whereby warming led to melting of gas hydrates and release of methane, which fuelled further warming and further methane release, in a positive-feedback loop that presumably continued until the reserves of gas hydrate were depleted (Figure 4).

Flood basalts are assigned a triggering role in this runaway greenhouse model, by venting enough initial CO_2 to drive temperatures above the threshold for gas-hydrate breakdown. However, volcanic eruptions also reduce global temperatures by releasing particulates and SO_2. Some authors have incorporated a cooling episode into their extinction models, but regard it as a brief event, which was subsequently eclipsed by the later warming. There is currently no geological evidence to support an episode of global cooling at the onset of the Permian–Triassic extinction crisis.

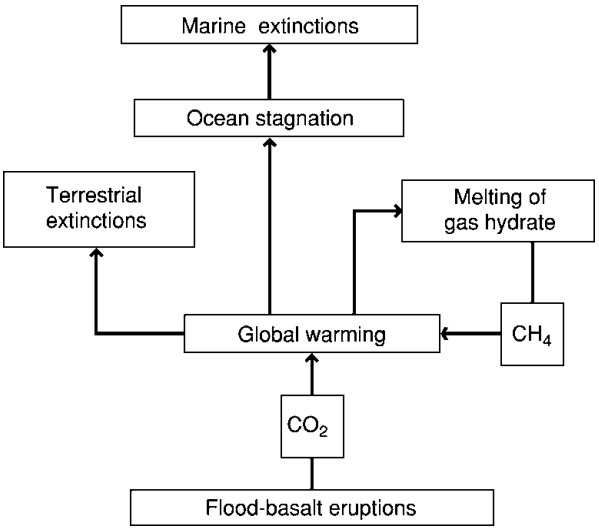


Figure 4 The runaway greenhouse model for the end-Permian extinction event. See text for details.

Post-Extinction Recovery

Douglas Erwin has remarked that the duration of the post-Permian recovery interval is the greatest of any Phanerozoic mass extinction and is proportionally longer than would be expected given the magnitude of the species loss. Certainly, a literal reading of the fossil record shows that it took some 100 Ma (until the mid-Jurassic) for marine biodiversity at the family level to return to pre-extinction levels. However, ecological recovery was somewhat quicker, with complex communities such as reefs being re-established some 10 Ma after the Permian–Triassic boundary. Thus, how one views the duration of the recovery interval, measured on a local or global scale, depends on how one defines 'recovery'.

Details of the marine recovery are best known from the low palaeolatitudes of Panthalassa (the present-day western USA) and Palaeotethys (e.g. northern Italy) (Figure 1). Epifaunal benthic communities in the immediate aftermath of the event (the Early Induan) comprise low-diversity assemblages of small suspension feeders (typified by the bivalves *Promyalina* and *Claraia* and often the inarticulate brachiopod *Lingula*). Bedding planes are dominated by a single taxon, which may occur in prodigious numbers. These taxa are considered by many, with varying degrees of evidence, to be pioneering r-selected opportunists (i.e. animals that mature early and produce many offspring). Stromatolites and other microbial mats are commonly encountered and may have provided grazing opportunities for the depauperate microgastropod community. Occasional horizons of mm-diameter *Planolites* burrows attest to temporary colonization events by a scarce infauna

of small deposit feeders living just 1–2 cm below the sediment surface. In the higher palaeolatitude regions of the Boreal Ocean (Greenland, Spitsbergen) and NeoTethys (Madagascar, western Australia) a fairly diverse but small-sized nekton of fishes and ammonoids is recorded.

Sedimentary and geochemical data indicate that environmental conditions were particularly harsh. Most recent palaeoenvironmental studies have concluded that the Early Triassic oceans contained very low concentrations of dissolved oxygen and may have been anoxic or even euxinic for considerable periods of time. Low oxygen levels were probably the result of a combination of factors. First, levels of atmospheric oxygen had fallen steadily from 30% in the Early Permian to approximately 15% (i.e. 70% of present-day values) by the Permian–Triassic boundary. Second, global warming in the latest Permian (see above) would have increased sea-surface temperatures, and warm water holds less dissolved oxygen than cool water. Finally, climate models indicate that global warming would also lead to a reduction in the temperature gradient between the poles and the equator, thus slowing down thermohaline circulation. In present-day oceans, the temperature difference between the warm equator and the cold poles drives a relatively vigorous thermohaline circulation, which maintains oxygenation of the ocean floors. In the sluggish earliest Triassic oceans, oxygen would have been used up faster than it could be replaced, resulting in stagnation and anoxia.

The ecology of the post-extinction survivors and the long interval of low diversity are interpreted as being the result of this anoxic event. In the few places where the oceans were well-oxygenated in the immediate aftermath of the extinction (e.g. Oman), a far more diverse benthic community, comprising upwards of 20 genera, was soon established. Unfortunately, these localized experiments in rapid post-extinction recovery were snuffed out by the arrival of anoxia in the later Induan.

With the disappearance of benthic oxygen restriction later in the Early Triassic, larger organisms and more diverse communities reappeared. The ecological complexity of infaunal communities began to increase with the reappearance of deeper-burrowing suspension feeders (producing *Arenicolites*, *Skolithos*, and *Diplocraterion* burrows). Burrowing crustaceans (producing traces such as *Thalassinoides*) are the last significant component of the infauna to reappear. In the palaeotropics, *Thalassinoides* first reappears in the Middle Triassic. In higher palaeolatitudes (East Greenland, Spitsbergen), small *Thalassinoides* reappear in the later Induan, suggesting faster recovery in the Boreal Realm.

In most parts of the world, epifaunal complexity did not significantly recover until the Late Olenekian, with an increase in tiering (vertical stratification of organisms above the sediment surface) as crinoids and bryozoans returned. However, the most complex ecological structures – metazoan reefs – do not return until the Middle Triassic, accompanied by a significant increase in faunal diversity and the reappearance of many Lazarus taxa.

Our understanding of the recovery of terrestrial ecosystems is less refined, with most data deriving from the Karoo Basin of South Africa. Like the marine survivors, terrestrial vertebrate survivors tended to be small: in the Karoo Basin five small carnivorous therocephalians and one small anapsid (*Procolophon*) survived the Permian–Triassic crisis. The other survivor, and the dominant terrestrial vertebrate for several million years, was the herbivore *Lystrosaurus*, which, like *Claraia*, was globally widespread.

More is known of the recovery of the terrestrial vegetation. The most detailed studies are those of Cindy Looy and colleagues, who have analysed fossil spores and pollen from Europe and eastern Greenland. Following the collapse of the Late Permian gymnosperm forests, open herbaceous vegetation rapidly took over, with short-lived blooms of pioneering opportunistic lycopsids, ferns, and bryophytes – stress-tolerant forms that were subordinate members of the pre-crisis vegetation. Pollen from woody gymnosperms seems to indicate that a few surviving elements of the Permian forests lingered on for a while, but equally these records could just represent reworking of the pollen. Certainly, by the earliest Triassic no tree-like gymnosperms remained, and a stable low-diversity open-shrubland vegetation of cycads and lycopsids was established. Complex and diverse forest communities did not reappear, at least in Europe, until the latest Olenekian and early Middle Triassic. Thus, the return of ecological complexity on land closely mirrors that in epifaunal marine communities.

Conclusions

The greatest mass-extinction event of the Phanerozoic is receiving an unprecedented level of scientific interest. New sections are being discovered and described, allowing new hypotheses of causes and consequences to be tested and old ones to be modified or rejected. The 1990s witnessed a number of major advances, culminating in the acceptance of Meishan, China, as the appropriate locality to define the Permian–Triassic boundary and the subsequent formalization of Permian–Triassic stratigraphy. The

picture presented here is a current (2004) view of this crisis: a snapshot that doubtless will be refined in the decade to come.

See Also

Analytical Methods: Geochronological Techniques. **Carbon Cycle**. **Impact Structures**. **Large Igneous Provinces**. **Mesozoic:** Triassic; End Cretaceous Extinctions. **Palaeoclimates**. **Palaeozoic:** Permian. **Sedimentary Environments:** Anoxic Environments.

Further Reading

Benton MJ (ed.) (1993) *The Fossil Record 2*. London: Chapman & Hall.

Benton MJ (2003) *When Life Nearly Died: The Greatest Mass Extinction of all Time*. London: Thames and Hudson.

Benton MJ and Twitchett RJ (2003) How to kill all life: the end-Permian extinction event. *Trends in Ecology and Evolution* 18: 358–365.

Erwin DH (1993) *The Great Paleozoic Crisis: Life and Death in the Permian*. New York: Columbia University Press.

Erwin DH (1994) The Permo-Triassic extinction. *Nature* 367: 231–236.

Erwin DH (1996) Understanding biotic recoveries: extinction, survival and preservation during the end-Permian mass extinction. In: Jablonski D, Erwin DH, and Lipps JH (eds.) *Evolutionary Paleobiology*, pp. 398–418. Chicago: University of Chicago Press.

Hallam A and Wignall PB (1997) *Mass Extinctions and Their Aftermath*. Oxford: Oxford University Press.

PANGAEA

S G Lucas, New Mexico Museum of Natural History, Albuquerque, NM, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Geologists apply the term Pangaea (from the Greek words meaning 'all Earth', and usually pronounced pan-JEE-uh) to the supercontinent that existed during the Late Palaeozoic and Early Mesozoic, about 300–200 Ma ago. Pangaea was an amalgamation of continental blocks that were the precursors of the existing continents, though much of eastern Asia was an archipelago of islands loosely connected to eastern Pangaea.

Components of Pangaea

Pangaea (Figure 1) was accreted from continental blocks that differed from today's continents. The southern supercontinent of Gondwana (also called Gondwanaland) (*see* **Gondwanaland and Gondwana**) was the Palaeozoic amalgamation of South America, Africa, Antarctica, Australia, and the Indo-Pakistani subcontinent, as well as several smaller terranes. The northern supercontinent of Laurussia (or Laurasia) consisted of North America and most of Europe.

Asia, however, did not exist as a single block. The continental nucleus of present-day Asia was the Siberian block, but other blocks included Kazakhstan, Tarim, and a loose archipelago of blocks that were to become much of China and South-east Asia. Many of these eastern Asian blocks originated as slivers of northern Gondwana and drifted northwards during the Late Palaeozoic and earliest Mesozoic to accrete to the larger Asian blocks during the Late Triassic–Jurassic.

Pangaea was surrounded by a universal ocean called Panthalassa (from the Greek words meaning 'all sea'), which was the ancestor of today's Pacific Ocean. The Atlantic Ocean did not exist even in ancestral form, because of the fusions of North America to Europe and of South America to Africa.

An arm of Panthalassa formed a deep east–west embayment in the eastern edge of Pangaea (Figure 1). This was the Tethys Sea, which is named after the sister and consort of Oceanus in Greek mythology. Tethys was the ancestor of the present Mediterranean Sea. Successive ocean basins within it are termed Palaeotethys (Devonian–Permian) and Neotethys; the latter opened during the Late Permian as a result of rifting between Gondwana and the smaller central and southern Asian terranes.

Late Carboniferous Accretion of Pangaea

The Pangaean supercontinent came together (accreted) at the end of the Carboniferous (*see* **Palaeozoic:** Carboniferous), when Laurussia and Gondwana were sutured along what has been termed the Hercynian megasuture. Very old mountain ranges mark the

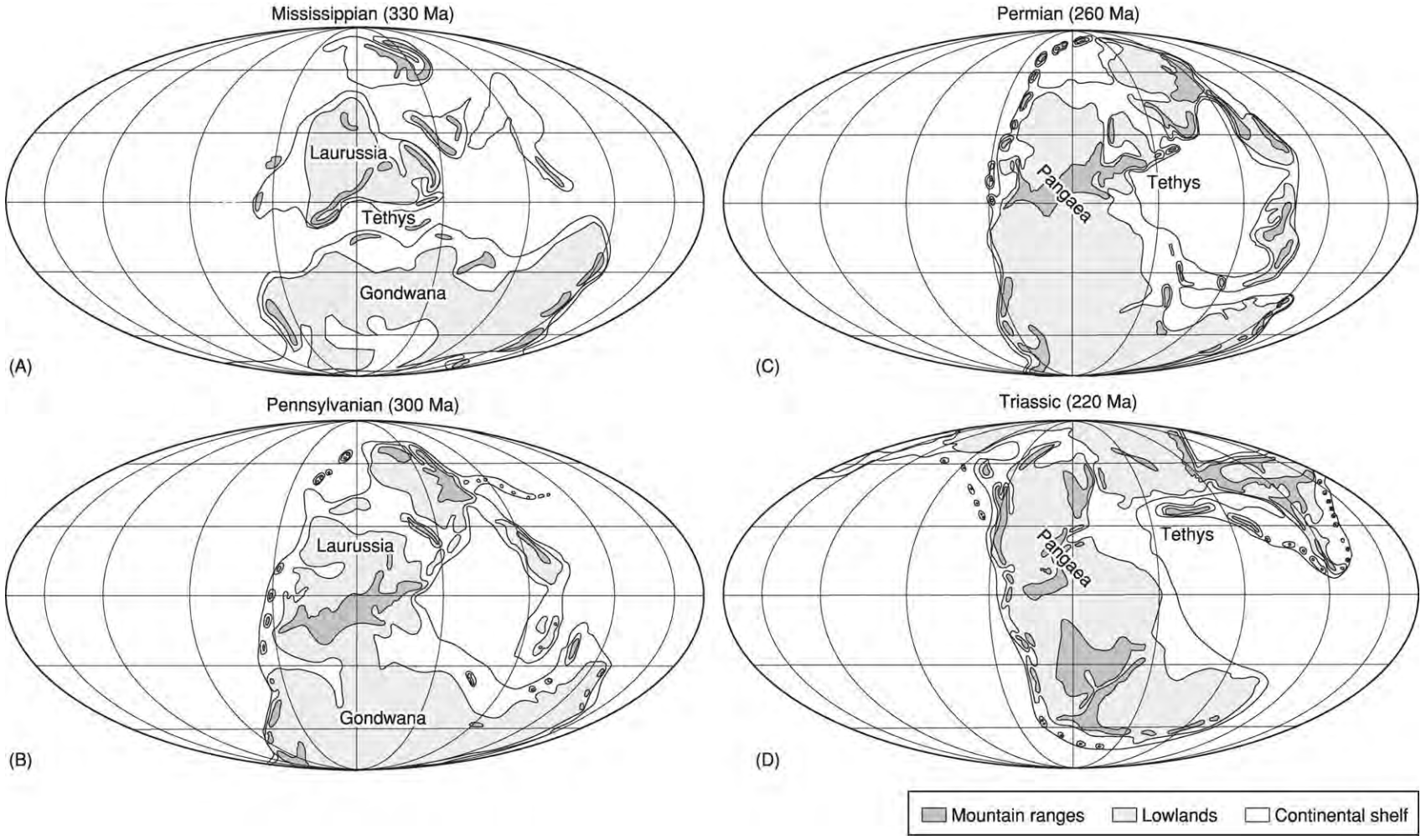


Figure 1 Palaeogeographical maps. (A,B) In the Carboniferous, Pangaea was accreted by the suturing of Gondwana and Laurussia near the palaeoequator – (A) Early Carboniferous (Mississippian), and (B) Late Carboniferous (Pennsylvanian). Pangaea was fully united during (C) the Permian and (D) throughout much of the Triassic.

collisional boundaries of the Pangaean blocks. These are the Appalachian Mountains of eastern North America (*see* **North America:** Northern Appalachians; Southern and Central Appalachians), the Ural Mountains of European Russia (*see* **Europe:** The Urals), and the Hercynian Mountain ranges of southern Europe and parts of North Africa.

The Late Carboniferous was a time of vast coal swamps in the tropical latitudes. There was a steep temperature gradient during the Late Carboniferous, from icy poles to hot tropics, which was more similar to today's world than perhaps at any other time in Earth history. In Gondwana, ice ages pushed glacial ice to within 30° of the palaeoequator.

Permian Pangaea

Once assembled, at the beginning of the Permian, Pangaea stretched from pole to pole in a single hemisphere (Figure 1C). The ocean of Panthalassa covered the other hemisphere.

Permian Pangaea was a relatively diverse place in terms of climate and topography. Permian glacial deposits found in South America, Africa, India, and Australia provide evidence of continuing glacial ages in southern Gondwana during the Early Permian. Along the sutures of Pangaea, huge mountain ranges towered over vast tropical lowlands. Interior areas were dry deserts where dune sands accumulated. Evaporites (particularly gypsum and halite), deposited in the south-western USA and northern Europe, resulted from the evaporation of hot shallow seas and form the most extensive salt deposits in the geological record. Perhaps the best testimony to the diversity of Permian Pangaea are its fossil plants, which identify several floral provinces across the vast supercontinent.

The coastlines of the Pangaean tropics were warm, and the waters were teeming with life. The cold offshore waters of the poleward portions of Pangaea had a very different and less diverse biota. The biotic extinctions (turnover) at the end of the Permian were the largest of the entire Phanerozoic (*see* **Palaeozoic:** End Permian Extinctions); they were undoubtedly influenced by the vast volcanic outpourings in Russia (the Siberian traps) and South China (the Emeishan traps).

Triassic Pangaea

At the beginning of the Triassic, the Pangaean supercontinent still sat in a single hemisphere surrounded by the enormous Panthalassan Ocean. Subduction zones that dipped beneath the continents nearly encircled Pangaea, and elevations were relatively high owing to regional uplift across Pangaea during the Late Permian and Triassic. Intense mountain building and foreland-basin deformation took place around the Pangaean margins during the Late Permian and earliest Triassic. The maximum accretion (integration) of Pangaea occurred at about the end of the Middle Triassic, approximately 230 Ma ago.

Early–Middle Triassic Pangaea was a much more cosmopolitan world than Permian Pangaea. Triassic fossil plants indicate only two vast provinces, north and south of the palaeoequator. However, some land animals, such as the mammal-like reptile *Lystrosaurus* (Figure 2), roamed across much of Pangaea. Indeed, recognition during the 1970s of the widespread nature of Early Triassic fossils of *Lystrosaurus* (which are found in Russia, China, India, Africa, and Antarctica), a land animal, provided compelling evidence of the unity of Pangaea.

During the Triassic, the vast Pangaean supercontinent drifted northwards and rotated counterclockwise. The landmass was nearly symmetrical about the palaeoequator. However, no sooner had the maximum integration of Pangaea occurred, than the supercontinent began to fragment.

Thus, by the Late Triassic, separation of Laurussia and Gondwana had begun, with the onset of rifting in the Gulf of Mexico basin. Indeed, the breakup of Pangaea began in the western Tethys and severed the supercontinent into two pieces, north and south, by the Jurassic. The east–west rifting that opened up the Atlantic Ocean basin was not really achieved until the Cretaceous. However, around the nascent North Atlantic basin, a huge basaltic volcanic province, called the

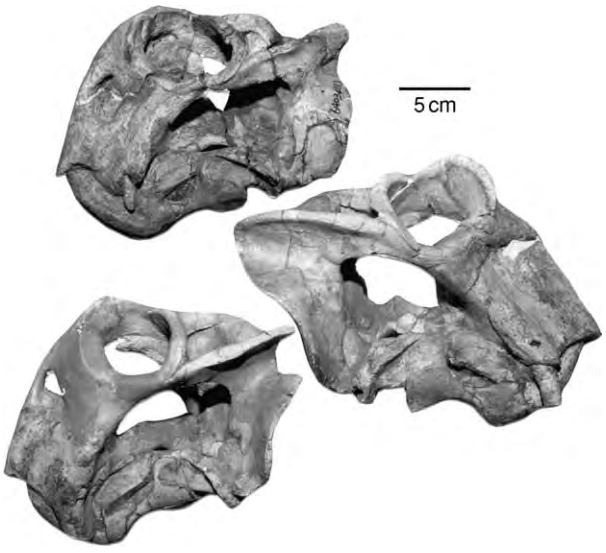


Figure 2 Skulls of the Early Triassic mammal-like reptile *Lystrosaurus* from northern China. Fossils of *Lystrosaurus* are also known from Russia, India, Africa, and Antarctica, providing strong evidence of the unity of Pangaea.

circum-Atlantic magmatic province, appeared at about the beginning of the Jurassic, as lavas poured forth along fractures that formed as Pangaea rifted apart.

An important aspect of Triassic Pangaea was the large number of microplates (also called displaced or suspect terranes) that were to accrete to the Pangaean nucleus during the Late Triassic and Jurassic. These microplates were of variable size and can be identified by the sharp contrasts between their rocks and fossils and those of adjacent rocks. They are fragments of other continents, island arcs, or pieces of oceanic crust that moved long distances to be accreted to continents. More than 100 such terranes have been identified in the Cordilleran belt of western North America. Numerous terranes in the Panthalassa Ocean moved eastwards to join the western ends of the Americas. In eastern Tethys, as noted above, rifted slivers of northern Gondwana moved northwards to accrete to and form much of eastern and South-east Asia during the Late Triassic–Jurassic.

See Also

Asia: Central; South-East. **Europe:** The Urals. **Gondwanaland and Gondwana**. **Mesozoic:** Triassic. **North America:** Northern Appalachians; Southern and Central Appalachians. **Palaeozoic:** Carboniferous; Permian; End Permian Extinctions. **Terranes, Overview**.

Further Reading

Behrensmeyer AK, Damuth JD, DiMichele WA, *et al.* (eds.) (1992) *Terrestrial Ecosystems Through Time*. Chicago: University of Chicago Press.

Embry AF, Beauchamp B, and Glass DJ (eds.) (1994) *Pangea: Global Environments and Resources*. Memoir 17. Calgary: Canadian Society of Petroleum Geologists.

Lucas SG (2000) The epicontinental Triassic, an overview. *Zentralblatt für Geologie und Paläontologie Teil I* 7–8: 475–496.

Marvin UB (1973) *Continental Drift*. Washington DC: Smithsonian Institution Press.

Sherlock RL (1948) *The Permo-Triassic Formations*. London: Hutchinson's Scientific and Technical Publications.

Smith AG and Briden JC (1977) *Mesozoic and Cenozoic Palaeocontinental Maps*. Cambridge: Cambridge University Press.

Tarling D and Tarling M (1971) *Continental Drift*. Garden City: Doubleday.

Ziegler PA (1989) *Evolution of Laurussia*. Dordrecht: Kluwer Academic Publishers.

PETROLEUM GEOLOGY

Contents

Overview
Chemical and Physical Properties
Gas Hydrates
The Petroleum System
Exploration
Production
Reserves

Overview

J Gluyas, Acorn Oil and Gas Ltd., Staines, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Petroleum geoscience is defined as the disciplines of geology and geophysics applied to the understanding of the origin, distribution, and properties of petroleum and petroleum-bearing rocks.

Petroleum geoscience can be described as the study and understanding of five key components: source, seal, trap, reservoir, and timing (of petroleum migration). These are sometimes known as the 'magic five ingredients' without which a basin cannot become a petroleum province (Figure 1). This article examines each of these components.

Source Rock

Petroleum (oil and gas) forms from organic matter: dead plants and animals. Burial, and thus heating of such organic matter induces reactions leading to the generation of gas, then oil and gas, and, finally, gas alone as the temperature and residence time at high temperature increase. Continued burial and heating turn the residual organic matter into pyrobitumen and eventually into graphite (*see* **Petroleum Geology:** The Petroleum System).

Seal

Oil and gas are less dense than water and, following expulsion from the source rock, they rise towards the Earth's surface unless movement is arrested by a seal. Seals tend to be fine-grained or crystalline, low-permeability rocks, such as mudstone/shale, cemented limestones, cherts, anhydrite, and salt (halite). Seals can also develop along fault planes, faulted zones, and fractures.

The presence of seals is critical for the development of petroleum pools. In the absence of seals, petroleum will rise to the Earth's surface and be destroyed. Although seals are critical for the development of petroleum pools, none are perfect. All leak. Combinations of regionally extensive seals and underlying reservoir complexes are commonly referred to as 'plays', and the areas within which the quality of seals and reservoirs is such that petroleum accumulations could occur (given an appropriate trapping configuration) are commonly referred to as 'play fairways'.

The most common subdivision of seals distinguishes between seals in which petroleum is unable to force its way through the largest pores (membrane seals) and seals in which petroleum can escape only by creating fractures (hydraulic seals). Attributes which favour a rock as a seal include a small pore size, high ductility, large thickness, and wide lateral extent. The physical properties of the water and petroleum are also important. Water salinity, petroleum density, and interfacial tension between petroleum and water are the most important, and these properties will change according to changing pressure and temperature conditions.

The most common lithology that forms a petroleum seal is mudrock. Mudrocks are composed of either carbonate or siliciclastic minerals (or both), and mudrock sequences are often thick (>50 m) and laterally continuous (>1.0 km^2). Examples of mudrock seals are found in all deltaic settings (including the Gulf of Mexico, Niger, and Nile Delta petroleum

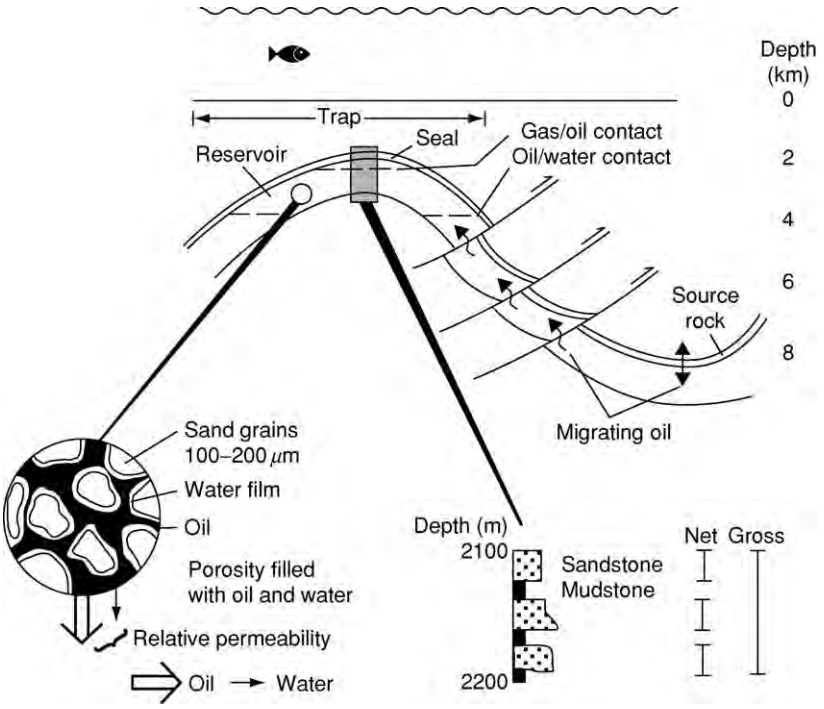


Figure 1 Diagram of a cross-section of a petroleum-bearing basin, illustrating the five key components: source rock, seal, reservoir, trap, and petroleum migration. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

provinces) and many interior, rift, and passive continental margin basins (including north-west Australian shelf, north-west Europe, North Sea, and south-east Asia). Halite can form a more effective seal, but is a relatively rare lithology found only where conditions of high evaporation of seawater have taken place. The Upper Permian rocks of north-west Europe contain Zechstein halite that is known to have trapped gas for long periods of geological time. Halite forms part of the sealing lithology in many of the large Middle East petroleum accumulations.

Membrane Seal

When petroleum is trapped beneath a seal, there is a buoyancy pressure (P_b). The magnitude of the buoyancy pressure is a function of the contrast in density between the water (ρ_w) and petroleum (ρ_p), and its height (h) above the free water level where both are in equilibrium (normally close to the petroleum–water contact). The relationship can be written as

$$P_b = (\rho_w - \rho_p)h$$

where P_b is expressed in units of pressure (e.g., bar, psi, MPa) and the fluid densities are expressed as pressure gradients (e.g., bar m^{-1}, psi ft^{-1}, Pa m^{-1}) (Figure 2). The maximum petroleum column is controlled by the capillary entry pressure of the petroleum into the largest pores in the seal. The capillary entry pressure (P_d) of a water-wet rock is given by the equation

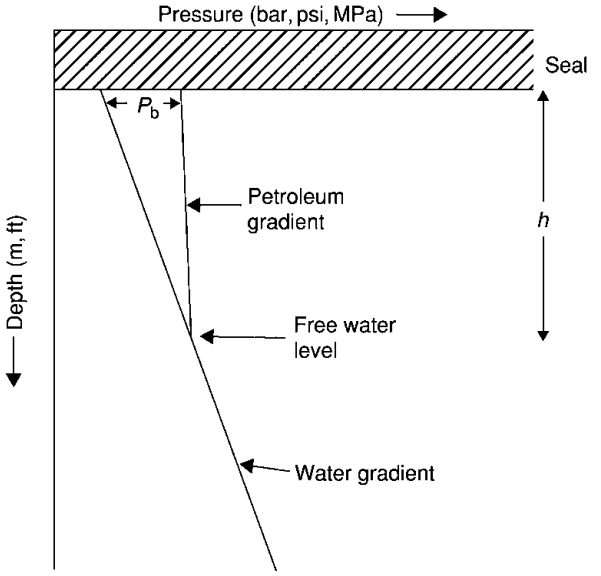


Figure 2 A pressure vs. depth plot, illustrating a typical water gradient (aquifer) supporting a petroleum column, whose steeper gradients lead to a pressure difference (P_b) at its maximum beneath the seal. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

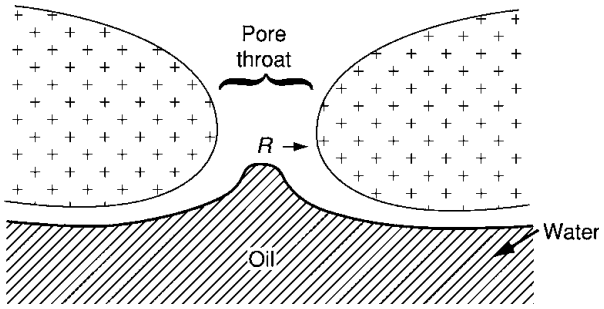


Figure 3 A schematic illustration of a pore throat between two grains. The radius of the pore throat and the buoyancy pressure, plus the interfacial angle and surface tension between oil and water, determine whether oil can migrate through the pore throat or remain trapped beneath. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

$$P_d = 2\gamma \cos\theta / R$$

where γ is the interfacial tension between the water and the petroleum, θ is the contact angle, and R is the radius of the largest pore (Figure 3). The interfacial tension and contact angle change with increasing temperature and pressure and are related to fluid type and density. These properties are routinely established from laboratory experiments on rocks. The seal capacity determines the height of a petroleum column that can be trapped beneath it, and the seal will be breached when the buoyancy pressure (P_b) exceeds the seal capillary entry pressure (P_d).

Hydraulic Seal

When the capillary entry pressure of the rock is extremely high, for example in evaporites, the failure of the seal is controlled by the strength of the rock that, if exceeded, creates natural tension fractures (*see* **Tectonics:** Faults; Fractures (Including Joints)). The rock will fracture when the pore pressure is greater than both the minimum stress and the tensile strength of the rock. Structural geologists describe the stresses in rock in terms of three orthogonal components of stress (Figure 4), one oriented vertically (S_v) and the other two oriented horizontally (S_{hmin} and S_{hmax}). In relaxed sedimentary rocks found in an extensional basin or a young delta, S_v is usually greater than both S_{hmin} and S_{hmax}, and so the minimum stress (σ_3) is horizontal. Under these conditions, the rock will fracture by creating vertically oriented fractures which will propagate horizontally. In rocks under horizontal push, or compression, the vertical stress (S_v) is the minimum stress (σ_3), and the rock will fail by the opening of horizontal fractures.

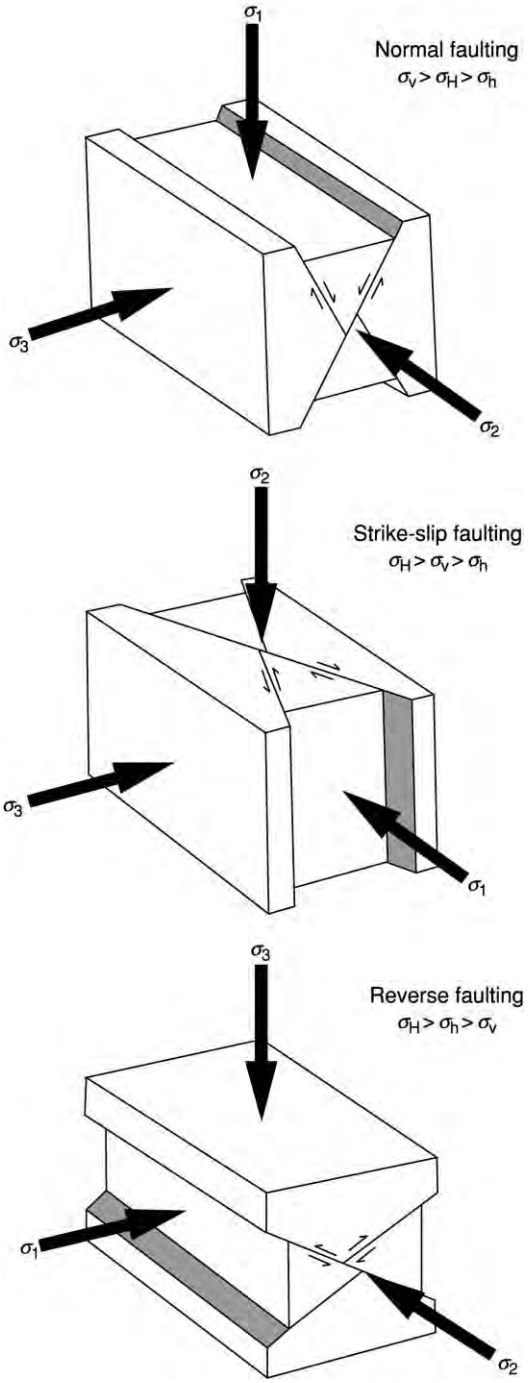


Figure 4 The relative magnitudes of the three principal stresses, one vertical (v) and two horizontal (H, h), acting in a rock mass, and the associated direction of shear, for normal strike-slip and reverse faulting regimes. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

Fault

Faults can act as both conduits (migration pathways) and seals, depending on the hydraulic conditions, the rock properties of the faults, and the properties of the

rocks juxtaposed across the faults. The consideration of faults as seals follows the same reasoning as for cap-rock seals above, i.e. the sealing capacity of a fault relates to its membrane strength and hydraulic strength. Membrane fault seals fail when the pressure of the petroleum can exceed the entry pressure of the largest pores along the fault plane. Hydraulic fault seals fail when the fault is opened mechanically by high pore pressure which exceeds the minimum stress.

Reservoir

For a rock to be a petroleum reservoir, it need only be porous, i.e., capable of holding petroleum, and permeable, i.e., able to flow petroleum.

Intrinsic Properties

The following properties must be known or estimated in order for the petroleum volume to be calculated.

1. Net to gross.
2. Porosity.
3. Permeability.
4. Petroleum saturation.

The question regarding whether any discovered petroleum will flow from its reservoir into the well bore is only partially addressed in exploration. This is commonly because permeability estimations are rarely accurate.

Net to gross Net to gross is a measure of the potentially productive part of a reservoir, commonly expressed as the percentage of producible (net) reservoir within the overall (gross) reservoir package (Figure 5). The percentage net reservoir can vary from just a few per cent to 100%. Net pay is used to describe the petroleum-bearing net reservoir.

It is common to define net sand (or limestone) using a permeability cut-off (typically 1 mD for gas and 10 mD for light oil). Such information on permeability is only available when the reservoir has been cored or a petroleum flow test completed. For uncored intervals and uncored wells, a combination of data on lithology and porosity from wireline logs is used. These data are calibrated to permeability data in a cored interval (*see* **Petroleum Geology:** Exploration).

Porosity Porosity is the void space in a rock (Figure 6). It is commonly measured as a volume percentage (*see* **Sedimentary Rocks:** Sandstones, Diagenesis and Porosity Evolution). In the subsurface, this volume may be filled with petroleum, water, a range of non-hydrocarbon gases (CO_2, H_2S, N_2), or some combination of these. Most reservoirs have porosities in the range 20–30%.

Not all pores are alike; there are big pores and little pores, pores with simple shapes, and others with highly complex three-dimensional morphologies. A knowledge of the size and shape of the pores and the way in which they are interconnected is

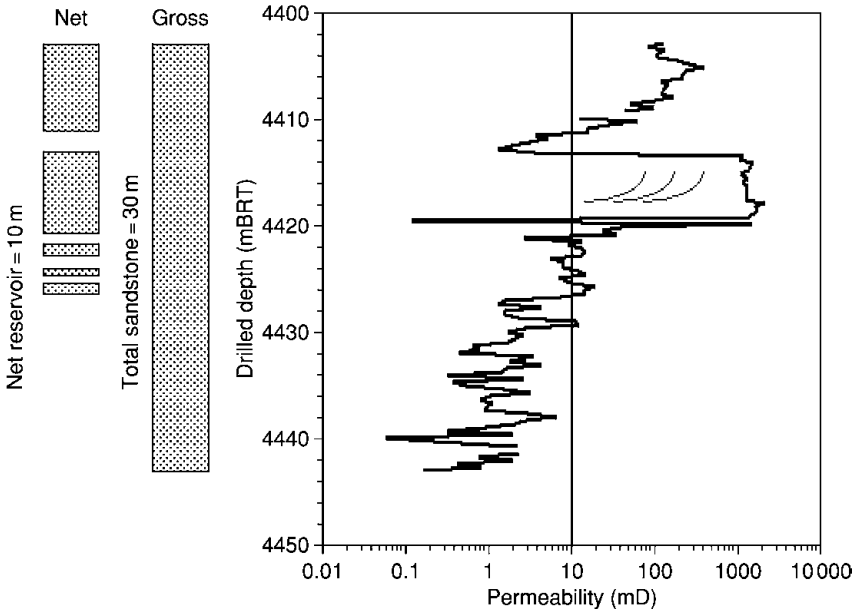


Figure 5 Net to gross is the ratio between reservoir rock capable of flowing petroleum and the gross reservoir interval. It is commonly defined using a single permeability cut-off. The example here is a Jurassic oil-bearing sandstone from the North Sea. Thirty metres of sandstone was cut in one core, but only 19 m had a permeability greater than the 10 mD cut-off chosen, a net to gross of 63%. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

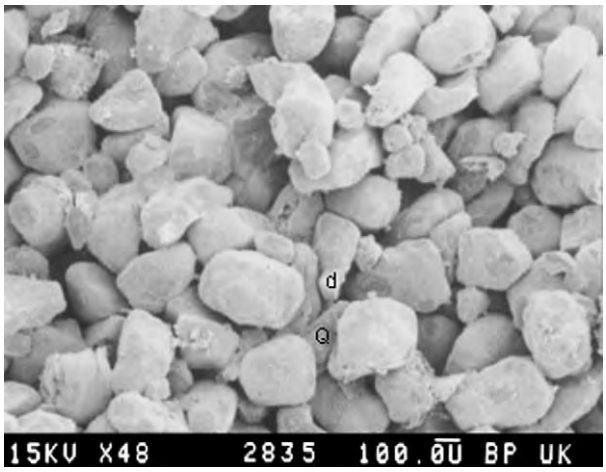


Figure 6 Scanning electron photomicrograph of a porous (28%) and permeable (2200 mD) Permian, Rotliegend reservoir sandstone, southern North Sea. Field of view, 2.7 mm × 1.8 mm. Photograph by A. J. Leonard, reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

important, because it is these factors that determine the permeability of the rock.

For sands and sandstones, a threefold description of porosity is used: intergranular, intragranular, and microporosity (Figure 7). Intergranular porosity occurs between grains. Individual pores in clean sand will occupy approximately 40% of the total volume. For coarse sands, the pores are larger than in fine sands. In most sandstones, the intergranular porosity is primary, a residuum of that imparted at deposition. Some intergranular porosity may be created in sandstones by the dissolution of mineral cements. Most intragranular porosity is secondary in origin, created on partial dissolution of grains. Microporosity simply means small pores, those associated with depositional or diagenetic clay or other microcrystalline cements.

Porosity development in limestones and dolomites is much more variable than that for sandstones. Both rock types are much more prone to mineral dissolution and precipitation than sandstones. This, coupled with the often varied suite of shell and other bioclastic material in the carbonates, makes for a wealth of pore types (Figure 8): intergranular, intragranular, intercrystalline, intracrystalline, biomouldic, vuggy, fracture, cavernous. The size range for pores is also much greater for limestones than for sandstones: from micropores (a few micrometres) in individual oolite grains to giant cave systems.

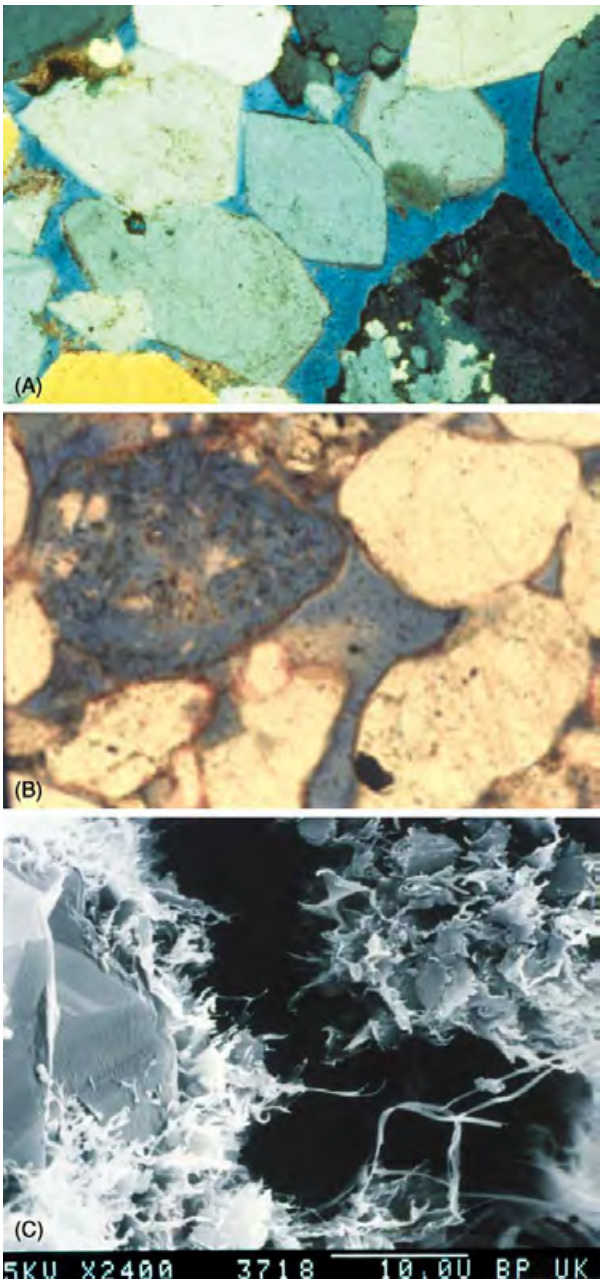


Figure 7 Porosity in sandstones. (A) Intergranular porosity (arrowed) between quartz grains with quartz overgrowths, Jurassic Brent sandstone, northern North Sea. Field of view, 1.3 mm × 0.9 mm. (B) Intragranular porosity within partially dissolved feldspar, Permian Rotliegend sandstone, southern North Sea. Field of view, 650 μm × 450 μm. (C) Microporosity (arrowed) between illitized kaolinite plates, Jurassic Brent sandstone, northern North Sea. Photographs by J. G. Gluyas.

Permeability Permeability is an intrinsic property of a material that determines how easily a fluid can pass through it. In the petroleum industry, the darcy (D) is the standard unit of permeability, but millidarcies (mD) ($1\,\text{mD} = 10^{-3}\,\text{D}$) are more commonly used. A darcy is defined as a flow rate of $10^{-2}\,\text{m}\,\text{s}^{-1}$ for a fluid of 1 cP (centipoise) under a pressure of $10^{-4}\,\text{atm}\,\text{m}^{-2}$. Permeability in reservoir rocks may range from 0.1 mD to more than 10 D. Permeability

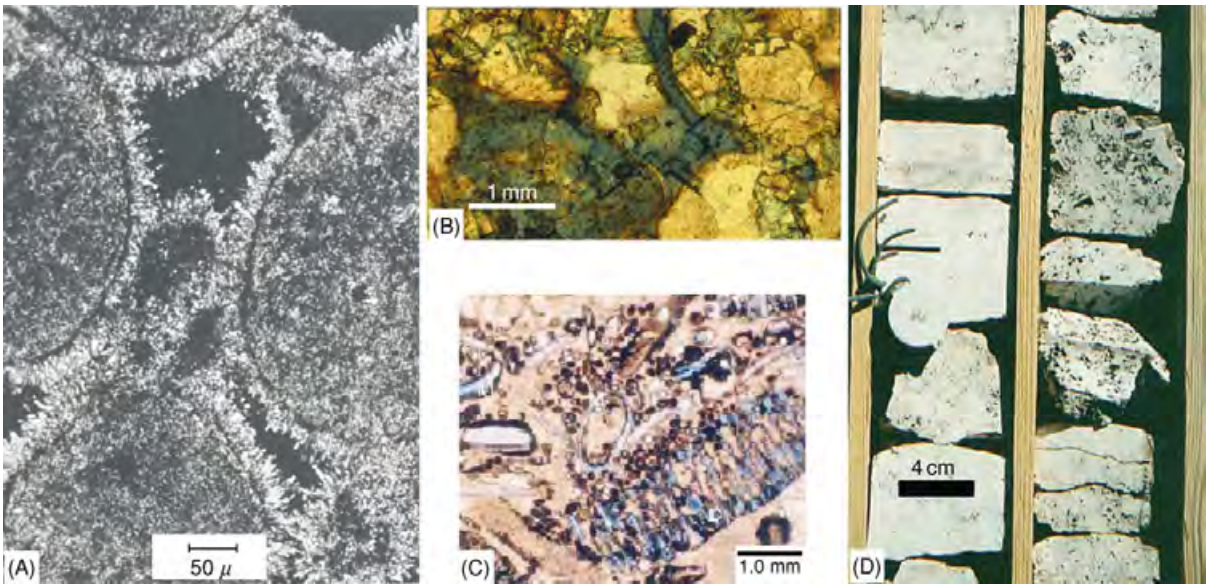


Figure 8 Porosity systems in carbonate reservoirs. (A) Intergranular porosity in limestone, beach rock, Bahamas. Reproduced from Bathurst RGC (1976) *Carbonate Sediments and Their Diagenesis, Developments in Sedimentology 12*. Oxford: Elsevier. (B) Intercrystalline porosity within dolomitized limestone, Permian Zechstein reservoir, southern North Sea, Dutch sector. Field of view, 3.25 mm × 2.50 mm. Photograph by J. G. Gluyas. (C) Biomoldic porosity within algal and mollusc moulds, Pennsylvanian limestone, Texas. Field of view, 5 mm × 4 mm. Reproduced with permission from Dickson JAD and Saller AH (1995) Identification of subaerial exposure surfaces and porosity preservation in Pennsylvanian and Lower Permian shelf limestones, eastern central Basin Platform, Texas. In: Budd DA, Saller AH, and Harris PM (eds.) *Unconformities and Porosity in Carbonate Strata, American Association of Petroleum Geologists Memoir 63*, pp. 239–258. Tulsa, OK: American Association of Petroleum Geologists. (D) Vuggy, oil-stained porosity within Cretaceous Bangestan limestones, Zagros Mountains, Iran. Field of view, 10 cm × 8 cm. Photograph by J. G. Gluyas.

measurements made at the Earth's surface are commonly greater than those in the subsurface, and a pressure correction must be made to restore the value of permeability to reservoir conditions. This intrinsic rock property is called the absolute permeability when the rock is 100% saturated with one fluid phase.

Water, oil, and gas saturation It is rare in nature to find a reservoir entirely oil (or gas) saturated. More commonly, the pore system contains both oil and water. The proportions of each phase are commonly expressed as percentages linked to the abbreviations: S_w for water, S_o for oil, and S_g for gas. Water and petroleum saturations are not constant across a reservoir. They vary in response to the position in the oil column, the permeability of the rock, and the mineralogy of the rock. Oil and water saturations will also change as petroleum is produced.

Reservoir Lithologies

Sandstone and limestone (including dolomite) are the most common reservoir lithologies. Sandstones dominate as important reservoirs in the USA (including Alaska), South America, Europe, Russian Asia, north Africa, and Australia. Limestones form the dominant reservoirs in the Middle East. They are also important in the Far East, western Canada, and some of the former Soviet states.

Sandstones, limestones, and dolomites of any age can make fine reservoirs. However, most of the best reservoirs in the world are relatively young. Petroleum fields are more common in Cenozoic and Mesozoic sediments than in Palaeozoic reservoirs. Precambrian age reservoirs are rare. There is no intrinsic reason why old rocks are more or less likely to be reservoirs than younger ones; it is simply that older reservoirs have had greater chance to be involved in tectonism or cementation, so destroying their porosity or permeability.

In addition to sandstone and limestone, fractured rock of any type can form a reservoir. The fractures alone may form the total pore volume of the reservoir. Alternatively, the fractures may help drain petroleum from the intervening lower permeability rock.

Reservoir; sandstone depositional systems Sediments, including those which may one day form a petroleum reservoir; can accumulate in many environments on the Earth's surface (*see* **Sedimentary Environments:** Depositional Systems and Facies). This includes sands deposited both on land and beneath the sea (Table 1). The overall architecture and internal

Table 1 Clastic reservoirs

Depositional system	*Architectural elements*	*Size range*	*Reservoir properties*	*Example oil/gas field(s) or province*
Alluvial fan	Low-angle half cones, linear and sheet sand bodies	1–10 km diameter	Heterogeneous, poorly sorted	Quiriquire Field, Pliocene E. Venezuela, 750 mmbbl
Aeolian deposit	Dune, sand sheet	100s km^2	Dune well-sorted porous sands high quality sandsheet moderate quality	Permian Rotliegend, Europe; Jurassic Norphlet, US Gulf Coast
Lake deposit	Half cone (fan)	Few km diameter	Poor to good, function of sediment input	Thailand & China
Fluvial system	Channel fill, crevasse splay	Channel belts 10s km × few km, crevasse splay few km diameter	Channel fills in braided and meandering systems commonly good; braided net to gross > meandering net to gross	Prudhoe Bay Field, Triassic Alaska, >10 bn bbl; Wytch Farm Field, Triassic UK onshore, 300 mmbbl
Delta and coastline	Channel mouthbar, shoreface, beach	**Figure 9**	Commonly good in reworked sandstones, variable net to gross	Niger Delta, W. Africa; Brent system, North Sea; Mahakan Delta, Indonesia; multiple billion barrels fields
Shallow marine	Shoreface to offshore sandstone bars	**Figure 9**	Good to excellent, high net to gross	Shannon Sandstone, Cretaceous USA; Fulmar/Ula Sandstones, Jurassic UK and Norwegian North Sea; Toro Sandstone, Jurassic/Cretaceous Papua New Guinea
Deep marine	Fan lobe, fan channel	10–100s km long, 10s km across	Depends upon sediment source area	Tertiary formations, North Sea; Plio-Pleistocene US Gulf Coast; Tertiary Congo Fan

bn bbl, billion barrels; mmbbl, million barrels.

geometry of the sand bodies (**Figure 9**) control the performance of a reservoir during petroleum production (*see* **Petroleum Geology:** Production).

Reservoir; carbonate depositional systems Limestones and dolomites form some of the largest petroleum reservoirs in the world. Many of the largest occur in the Middle East. Other areas in which carbonate reservoirs deliver large quantities of oil and gas are western Canada, Mexico, Texas (USA), Norway (central North Sea), Poland, Kazakhstan, western and south-eastern China, Iran, and Libya.

The sediment that forms most carbonate reservoirs accumulated in shallow marine environments (*see* **Sedimentary Environments:** Carbonate Shorelines and Shelves). The important exceptions are pelagic chalks (Ekofisk Complex of the North Sea) and deep-water resedimented carbonates of the Poza Rica Trend in Mexico.

Like their clastic counterparts, there is a clear link between the reservoir potential of a carbonate body and the environment in which the host sediment accumulated. High-energy ooid and shell shoals can make excellent reservoirs. Framework reef complexes are also prime reservoir targets. However, unlike siliciclastics, carbonates can undergo almost complete transformation on weathering and/or diagenesis to produce reservoirs from former seals and seals from former potential reservoirs.

Dolomite (*see* **Sedimentary Rocks:** Dolomites) Producing dolomites range in age from Precambrian to Tertiary. It is estimated that about 80% of the recoverable petroleum in carbonate-hosted reservoirs of the USA occurs in dolomite and only about 20% in limestone. The same ratio probably applies to the producible reservoirs in the Permian Zechstein of Europe, whilst older carbonate plays in Europe and Russian Asia are almost wholly dolomite. The dolomitization of limestones commonly leads to an increase in both porosity and permeability (**Figure 10**).

Karst (*see* **Sedimentary Processes:** Karst and Palaeokarst) Karstified limestones and dolomites represent the second major group of carbonate reservoirs not directly linked to depositional environments. Karst is a product of mineral dissolution (**Figure 11**). It develops where carbonates are exposed to meteoric water, often linked with episodes of sea-level fall. Karst features are well known to geologists and geographers alike: caves, collapse breccias, dissolution-enhanced joints and fractures, and vugs.

Fields producing from karstified limestones and dolomites include the Liuhua Field in the South China

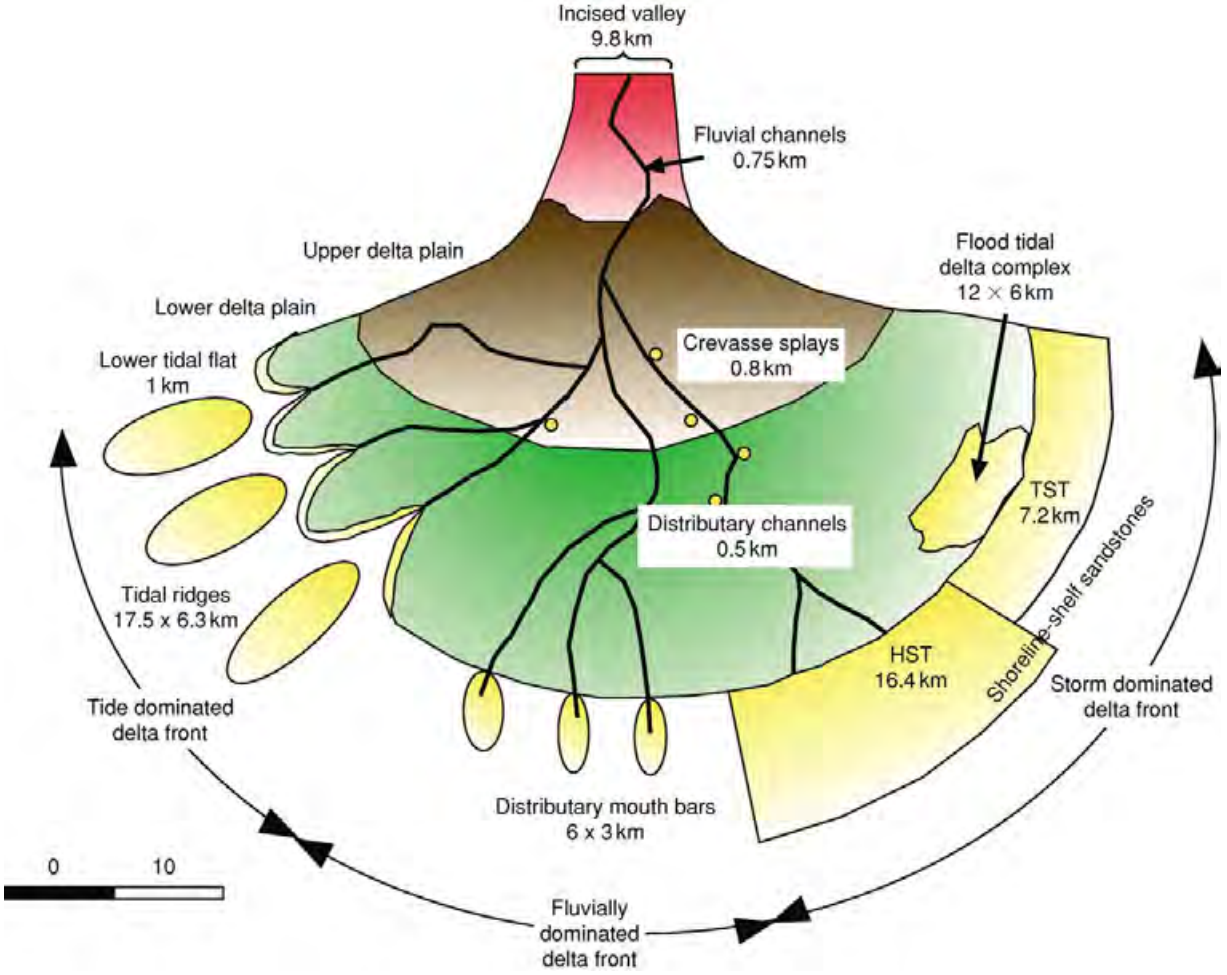


Figure 9 The average size, shape, and location of sand bodies in wave, tidal, and fluvially influenced reservoirs. Reproduced from Reynolds AD (1994) Sequence stratigraphy and the dimensions of paralic sandstone bodies. In: Johnson SD (ed.) *High Resolution Sequence Stratigraphy: Innovations and Applications*, pp. 69–72. Liverpool: Liverpool University.

Sea, the Permian reservoirs of Texas and New Mexico, and parts of the Upper Permian in the Zechstein Basin in Europe. Thermal karst, produced when hot fluids dissolve limestones at depth, may also become reservoirs. The Albion Scipio Field of Michigan is of this type.

Trap

Trap is the term to describe the body, bounded by seals and containing reservoir, in which petroleum can accumulate as it migrates from the source rock to the Earth's surface. There are many different trap geometries. These can be grouped into three categories: structural, stratigraphical, and hydrodynamic (Table 2). Structural traps are created by tectonic, diapiric, compactional, and gravitational processes (Figure 12). Almost the entire world's discovered petroleum is in traps that are largely structural. Stratigraphical traps are

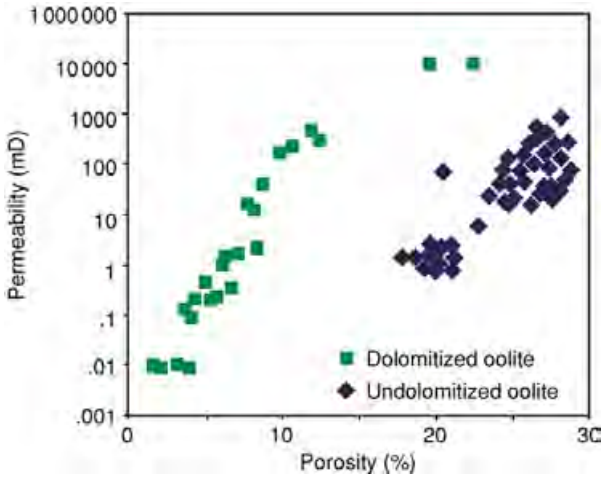


Figure 10 A comparison of porosity and permeability for dolomitized and undolomitized oolite, Cretaceous Middle East. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

formed by lithological variations or property variations generated by alteration of the sediment or fluid through diagenesis (Figure 13). Much of the world's remaining undiscovered petroleum will be found in stratigraphical traps. Purely hydrodynamic traps are rare. Such traps rely on the flow of water through the reservoir horizon to 'drag' the petroleum into a favourable trapping configuration, such as the plunging nose of a fold. The trapping mechanism for many fields is commonly a combination of structural and stratigraphical elements or, more rarely, structural elements and hydrodynamic conditions.

Figure 11 Tower karst containing fracture and cavernous porosity, Palaeozoic limestones, Zhaoquing, Guangdong Province, China. Photograph by J. G. Gluyas.

Structural Trap

Compressive tectonic regimes commonly lead to the development of large-scale contractional folds and thrusts. This is common at convergent plate boundaries and transpressional strike-slip plate boundaries (wrench systems). The El-Furrial Trend of eastern Venezuela is an example of such a system. The anticlinal traps of the trend were developed during convergence of the Caribbean and South American plates during the Neogene. Many of the traps are large ramp anticlines (Figure 14). They have oil columns of, on average, 400 m and reservoirs formed from high net to gross shallow marine sandstones.

In North America, thrust-linked rollover anticlines form the major trap type in the Wyoming–Utah thrust belt fields and the southern foothills of the Alberta Basin, Canada. Compressional anticlines also form giant traps within the Zagros fold belt of Iran.

Traps formed by extensional tectonics are common in rift basins. The East Shetland Basin of both the UK and Norwegian sectors of the North Sea contained about 15 billion barrels of recoverable oil. Much of this oil was trapped in tilted fault blocks formed during Late Jurassic rifting. In the pre-rift section, oil is reservoired in the sandstones of the Middle Jurassic Brent Group, together with other sandstones of both Jurassic and Triassic age (Figure 15). Traps formed through tectonic extension are also important in the Gulf of Suez, the Haltenbanken area, offshore mid-Norway, and in the pre-rift sections of the Gippsland Basin (Australia).

Traps can also be formed by diapiric processes. The specific gravity of salt (halite) is about 2.2 g cm^{-3} and that of fully consolidated rock is about 2.5–2.7 g cm^{-3}. Thus salt is buoyant relative to most other sediments

Table 2 Structural and stratigraphical traps

Structural	Tectonic	Extensional
		Compressional
	Diapiric	Salt movement
		Mud movement
	Compactional	Drape structures
	Gravitational	Listric fault movement
Stratigraphical	Depositional	Pinchouts (dunes, bars, reefs, channels, etc.)
		Unconformities (erosional, subcrop, karst, etc.)
	Diagenetic	Mineral precipitation
		Mineral dissolution (thermal karst, dolomitization)
		Tar mats
		Permafrost
		Gas hydrate crystallization

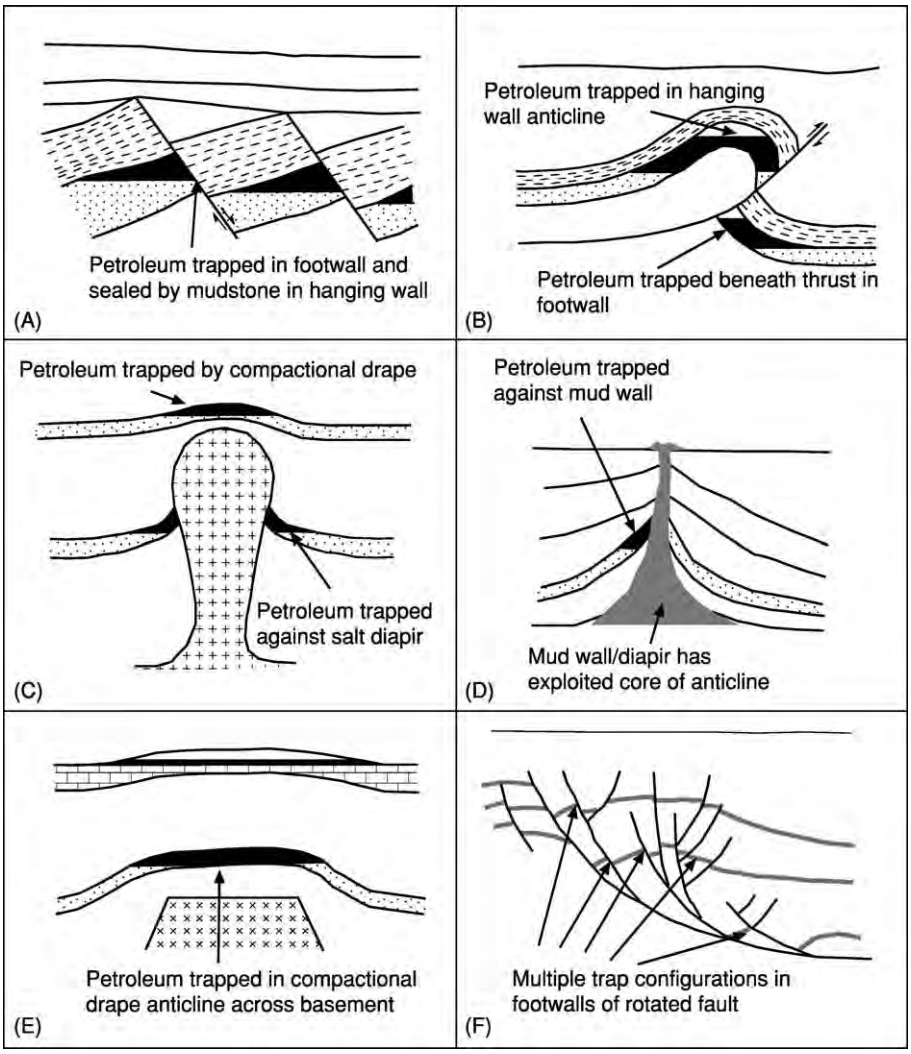


Figure 12 Structural traps. (A) Tilted fault blocks in an extensional regime. The seals are overlying mudstones and cross-fault juxtaposition against mudstones. (B) Rollover anticline on thrust. Petroleum accumulations may occur on both the hanging wall and the footwall. The hanging wall accumulation is dependent on a subthrust fault seal, whereas at least part of the hanging wall trap is likely to be a simple, four-way, dip-closed structure. (C) Lateral seal of a trap against a salt diapir and compactional drape trap over the diapir crest. (D) Diapiric mudstone associated trap with lateral seal against mud wall. Diapiric mud associated traps share many common features with that of salt. In this diagram, the diapiric mud wall developed at the core of a compressional fold. (E) Compactional drape over a basement block commonly creates enormous low-relief traps. (F) Gravity-generated trapping commonly occurs in deltaic sequences. Sediment loading causes gravity-driven failure and produces convex-down (listric) faults. The hanging wall of the fault rotates, creating space for sediment accumulation adjacent to the fault planes. The marker beds (grey) illustrate the form of the structure that has many favourable sites for petroleum accumulation. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

and sedimentary rocks. Over geological time, salt deforms plastically. With loading caused by continued sedimentation, layers of salt may aggregate into swells and eventually pillows. Subsequently, a salt diapir may rise through the overburden. Very similar processes to those associated with salt diapirism can occur in association with muds. Rapidly deposited muds are commonly water rich, overpressured, and, in consequence, highly mobile. Mud lumps (Niger Delta), shale walls, diapirs, and mud volcanoes (Trinidad, Azerbaijan) are all products of mass mud movement.

Diapiric movement of both salt and mud can create anticlinal structures that could form petroleum traps. Trap configurations can also develop in the areas of salt withdrawal. The 'turtle' structure anticline develops via increased sedimentation in areas of salt withdrawal. Later, as salt continues to feed the diapir, the structure flounders and flips into an anticline.

Greater Burgan (Kuwait), the second largest oilfield in the world (>75 billion barrels of reserves), developed over a large, low-amplitude salt swell.

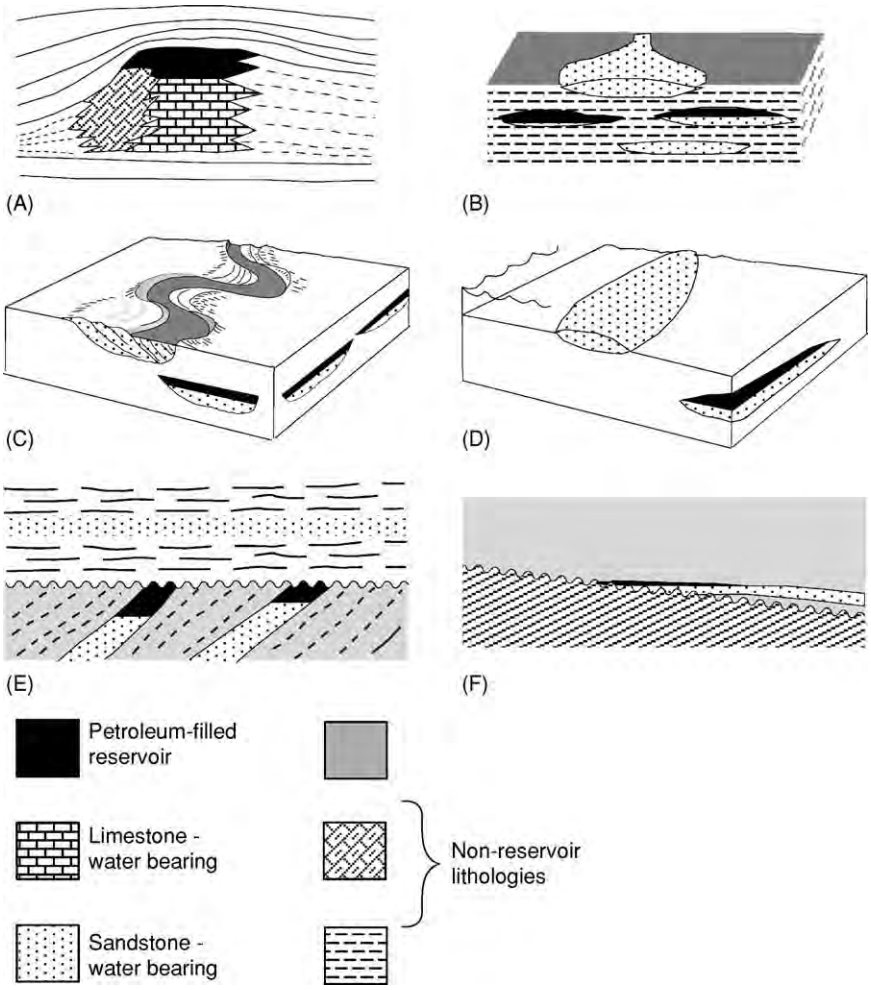


Figure 13 Stratigraphical traps. (A) 'Reef' oil is trapped in the core of the reef, with fore-reef talus and back-reef lagoonal muds acting as lateral seals and basinal mudstones as top seals. (B) Pinchout (sandstone) trap within stacked submarine fan sandstones. The upper surface of the diagram shows the plan geometry of a simple fan lobe. Lateral, bottom, and top seals are the surrounding basinal mudstones. (C) Channel-fill sandstone trap. The oil occurs in ribbon-shaped sandstone bodies. The top surface of the diagram shows the depositional geometry of the sandstone. Total seal may be provided by interdistributary mudstones or a combination of these and marine flooding surfaces. (D) Shallow marine sandstone bar completely encased in shallow marine mudstone. The upper surface of the diagram shows the prolate bar. (E) Subunconformity trap. The reservoir horizon is truncated at its up-dip end by an unconformity and the sediments overlying the unconformity provide the top seal. Lateral and bottom seals, like the reservoir interval, pre-date the unconformity. (F) Onlap trap. A basal or near-basal sandstone onlaps a tilted unconformity. The sandstone pinches out on the unconformity and is overstepped by a top seal mudstone. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

Similar, simple anticlinal dome traps typify the Cretaceous Chalk fields of the Norwegian North Sea. As with the Middle East examples, the key controlling structures are the underlying salt pillows.

Traps associated with diapirs rather than swells tend to be much smaller in aerial extent than the giants described above. They also tend to be much more structurally complex, commonly containing both radial and concentric fault patterns. The Machar Field (STOOIP (standard barrels of oil originally in place) about 228 mmstb) of the North Sea is roughly circular in outline with a diameter of about 4 km (Figure 16).

Anticlinal traps created through compaction occur across basement highs, tilted fault blocks, carbonate shelf rims, reefs, or isolated sand bodies. Some of the simplest are also the largest. The world's biggest field, Ghawar in Saudia Arabia, is such a trap. Oil occurs in Jurassic carbonates draped over and compacted around a north–south-trending basement high.

Traps formed by gravity-driven processes are particularly important in large recent deltas. The best-described examples are from the US Gulf Coast and West African deltas (Niger, Congo). The gravity structures form independently of basement tectonics and owe their existence to shallow detachment along

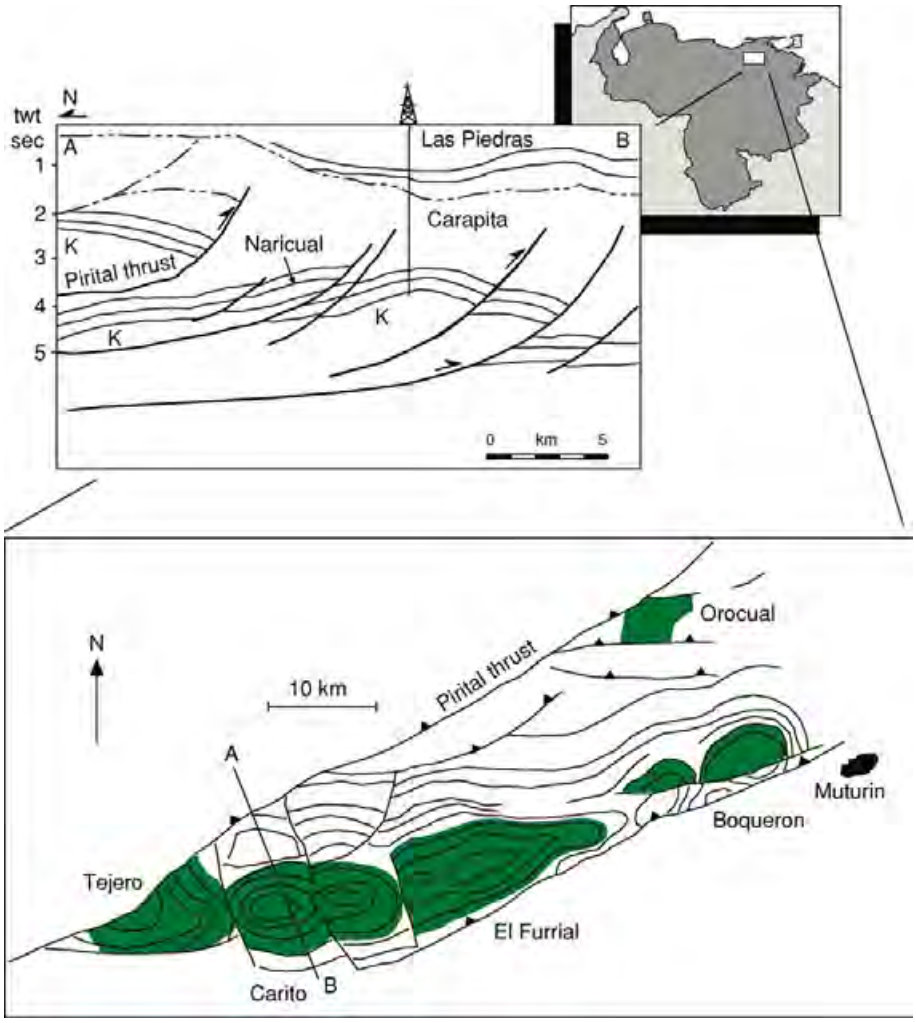


Figure 14 El Furrial Trend, eastern Venezuela. The petroleum traps are large rollover anticlines on the hanging walls of thrusts. Most are four-way, dip-closed structures, whilst some have a dependence on a fault component along their south-eastern thrust margin. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

low-angle, basinward-dipping planes. The drive mechanism is provided by the weight of sediment deposited by the delta at the shelf–slope break or on the slope itself. In the Niger Delta, the detachment planes are highly mobile muds, whereas, in the Gulf Coast (Mississippi Delta), detachment occurs on both muds and the Louanne Salt (Jurassic). The key detachment surfaces are commonly listric, concave-up, and concave-basinward in plan view. The main faults are commonly large, being tens of kilometres tip to tip.

Stratigraphical Trap

From top to bottom of a systems tract, each depositional environment is capable of producing a juxtaposition of permeable and impermeable sediments which might one day form a stratigraphical trap for petroleum. In practice, the reservoir geometry becomes the trap geometry. Examples include aeolian dunes encased in lacustrine mudstone, sand-filled fluvial channels cut into mud-rich overbank deposits, shallow marine bar sandstones surrounded by marine shales, carbonate reefs isolated by enclosing marls, and submarine fan sands trapped within the domain of pelagic mud.

The Paradox Basin (Colorado and Utah, USA) contains a large array of small oil and gas fields in stratigraphical pinchout traps. Devonian reservoirs occur within shallow marine bar sandstones and Carboniferous reservoirs within carbonate mounds. The Paradox Basin traps are difficult to find, but have relatively simple shapes. Their geometries are either prolate bar forms or more equidimensional carbonate mounds. Pinchout traps formed in deltaic settings are often much more complex in outline and, because

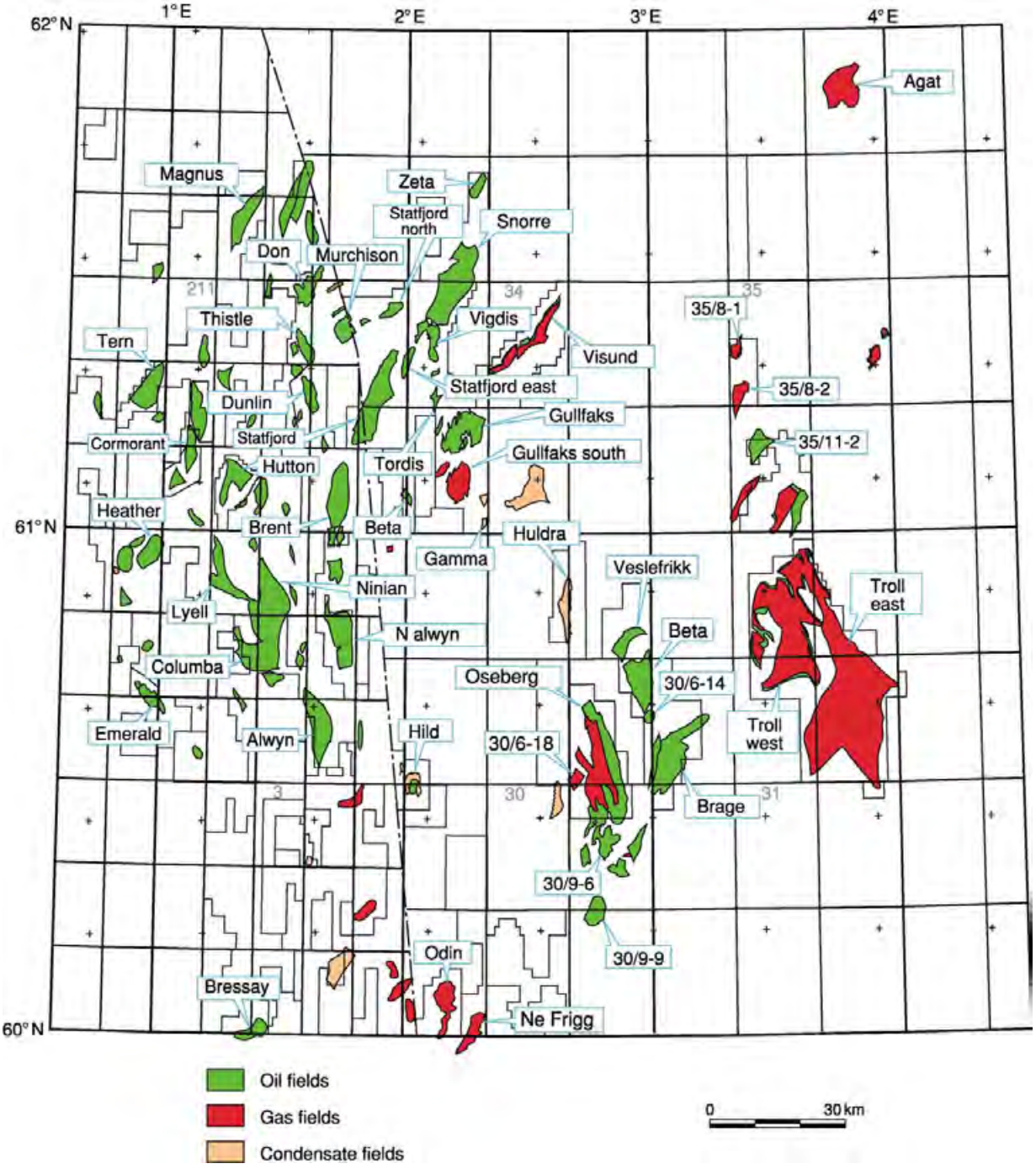


Figure 15 UK and Norwegian Brent Province. The elongate shapes reflect the geometry of the tilted fault blocks that form the traps. The reservoir is largely Middle Jurassic Brent sandstones together with Triassic and Upper Jurassic sandstones in some fields. The traps for neither Troll (Upper Jurassic reservoir in low-relief anticline) nor Agat (Lower Cretaceous reservoir, stratigraphically trapped) are tilted fault blocks. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

potential reservoir sandstones are commonly discontinuous, multiple pools (clustered fields) are common (Figure 17).

Attenuation of the up-dip portions of a potential reservoir interval by an unconformity can create massive traps with enormous petroleum catchment (drainage) areas. The largest oilfield in North America, Alaska's Prudhoe Bay, is an unconformity trap. It has about 25 billion barrels of liquid and more than 20 trillion cubic feet of gas in place. East Texas, the

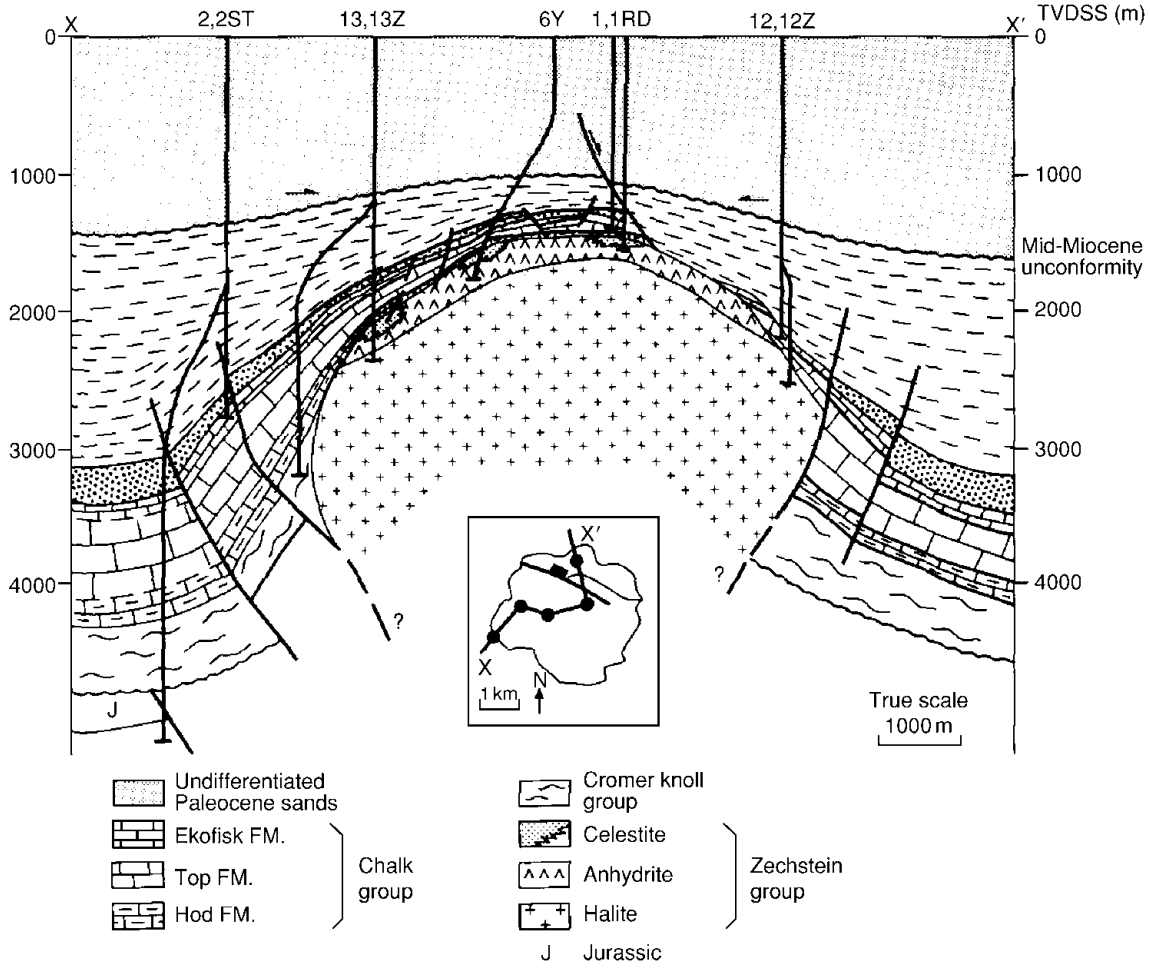


Figure 16 A structural cross-section of the Machar Field, central North Sea, showing the circular outline of the field and the distribution of reservoirs around the head of the salt diapir. Reproduced from Foster PT and Rattey PR (1993) The evolution of a fractured chalk reservoir: Machar Oilfield, UK North Sea. In: Parker JR (ed.) *Petroleum Geology of Northwest Europe: Proceedings of the 4th Conference*, pp. 1445–1452. London: Geological Society.

largest oilfield in the USA Lower 48, is also a stratigraphical trap. The productive Woodbine Sandstone reservoir, with its initial reserves of about 6.8 billion barrels, is sandwiched between two unconformities. The sand rests upon the Washita Group mudstones and is itself truncated beneath the Austin Chalk. The field, some 40 miles long and 5 miles wide, is a simple homoclinal dip to the west.

Each of the unconformity traps described above relies on a combination of trapping mechanisms, which rely in large part on a planar or gently folded unconformity. Unconformities come in a variety of shapes. The most spectacular of the unconformity-bounded traps are those commonly referred to as 'buried hills'. Such hills are residual topography of a one-time land surface. Thus it is the unconformity surface that has the trapping geometry (Figure 18). Buried hill traps are most common in karstified areas, such as northern China.

Mineral cements are known to form top, lateral, and even bottom seals to reservoirs. Examples in carbonate systems are more numerous than those in clastic systems. In the Albion-Scipio Field of Michigan (USA), all surrounding rock to the trap is thoroughly cemented limestone and dolomite. A comparable situation exists for many of the carbonate-hosted oilfields of Abu Dhabi; porosity only exists where there is oil. Areas that at one time must have been the aquifers to the oilfields have been thoroughly cemented. For a few fields, such cementation has allowed trap integrity to be maintained despite tilting of the field after petroleum accumulated.

Tar mat seals are common in the shallow subsurface. They also act as cap rock for the largest single accumulation of heavy (viscous) oil in the world; the Faja of south-eastern Venezuela, which has about 1.2 trillion barrels of oil in place. Tar seals and tar sands

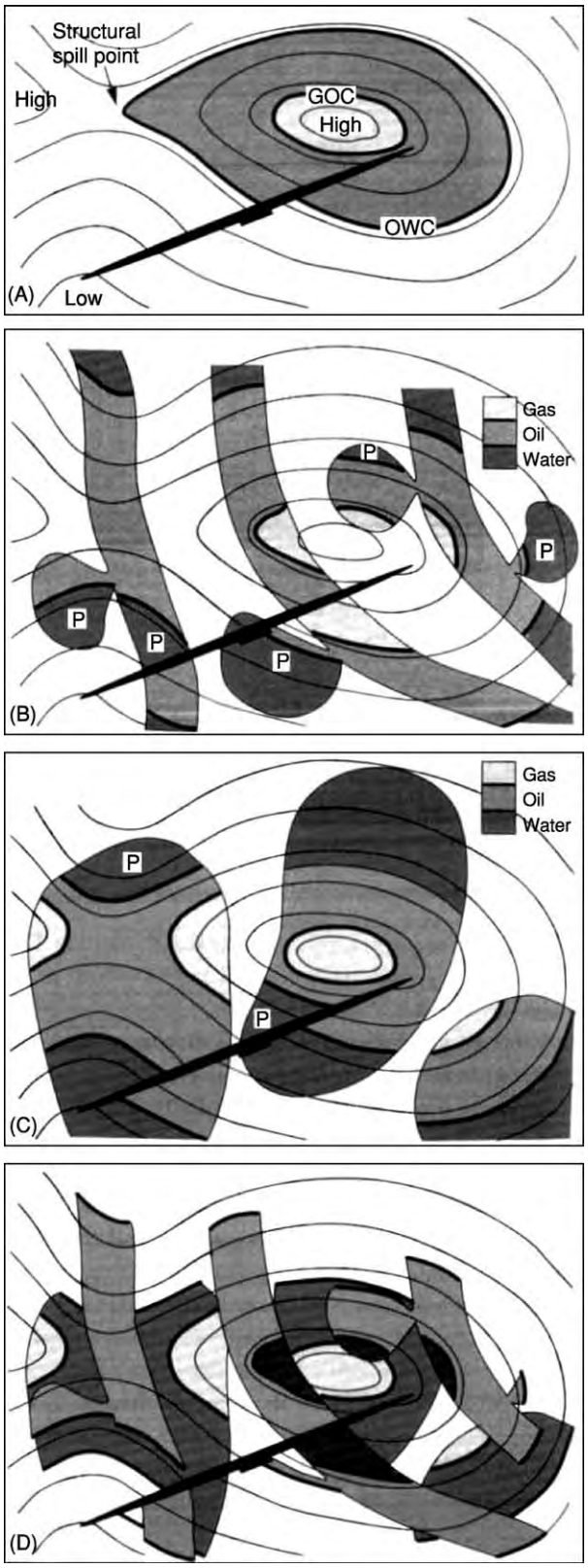


Figure 17 Paralic field outlines commonly have complex shapes because of the interaction between structure and sediment bodies. This complexity is multiplied because individual paralic sandstones tend to be stacked. The four examples show: (A) field shape on a simple faulted anticline for which the reservoir interval is much larger than the anticline; (B) the same structure as in (A), but with the reservoirs developed in channel and crevasse splay sandstones that are smaller in area than the structure; (C) the same structure as in (A), but with mouthbar sandstones which are also smaller than the structure; (D) a combination of channel and mouthbar sandstones at different levels. A. Reynolds, personal communication, 1994. Reproduced courtesy of BP.

are also common within the Western Canada Basin and Californian basins.

Gas trapped beneath permafrost forms large fields in the northern part of the West Siberia Basin, adjacent to the Kara Sea. In cold regions, gas (methane) is also trapped as gas hydrate.

Hydrodynamic Trap

The idea that moving water could and would control the distribution of both oil and gas traps was first advocated in 1909. The hypothesis had a number of supporters until the 1930s, when the number of publications on the topic dwindled and the anticlinal theory of petroleum accumulation reassumed its position as the only favoured theory. Twenty years later, the idea was resurrected, although it remains controversial.

Those traps with undoubted hydrodynamic credentials tend to be in foreland basins where subsurface reservoir units commonly crop out in adjacent mountain belts. The outcropping reservoir units are recharged with meteoric water and the hydraulic head drives the flow through the basin. Two of the best-documented examples are the Frannie Field of the Big Horn Basin, Wyoming, and the East Colinga Extension Field of San Joaquin Valley, California. In both instances, there is sufficient information to map the tilted oil–water contacts, rule out the possibility of significant permeability barriers in the systems, and explain the water flow in terms of the adjacent topography and subsurface structure.

Migration

Migration is the process (or processes) whereby petroleum moves from its place of origin, the source rock, to its destruction at the Earth's surface. Along the route, the petroleum's progress may be temporarily arrested and the petroleum may accumulate within a trap. The timing of trap formation relative to that of petroleum generation and migration is critical. The trap has to form at the same time or earlier than petroleum migration if it is to capture petroleum.

Migration may be divided into three stages (Figure 19).

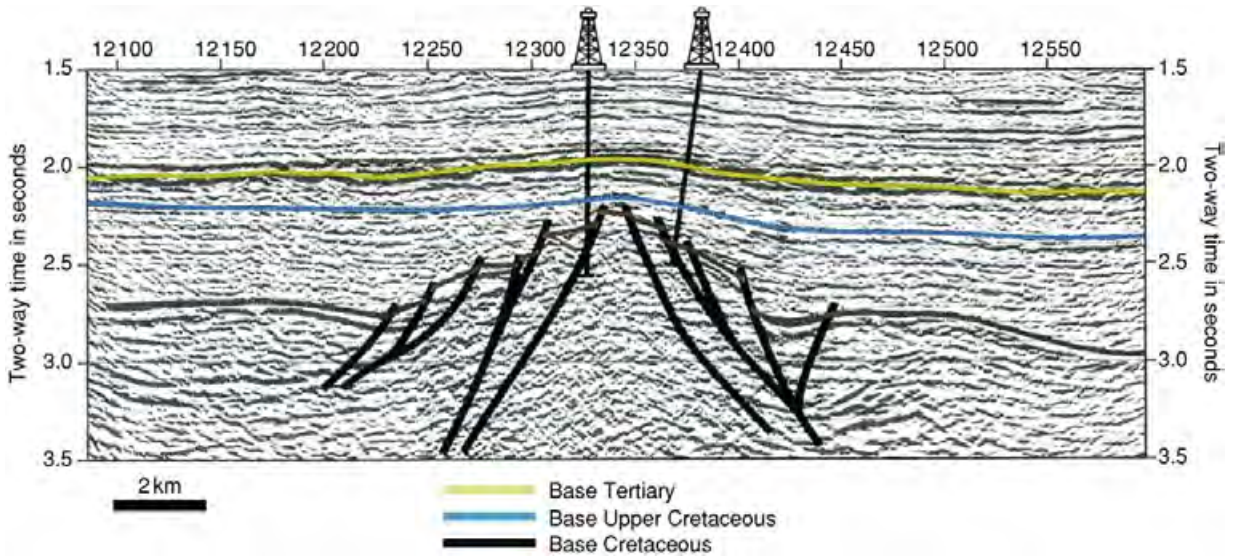


Figure 18 Subunconformity trap beneath the base Cretaceous unconformity, Buchan Field, UK North Sea (fractured Devonian sandstone reservoir). Reproduced from Abbots IL (1991) *United Kingdom Oil and Gas Fields, 25 Years Commemorative Volume, Geological Society Memoir No. 14*. London: Geological Society.

- Primary migration: expulsion of petroleum from the source rock.
- Secondary migration: the journey from source rock to trap.
- Tertiary migration: leakage and dissipation of the petroleum at the Earth's surface.

Primary Migration

There have been many hypotheses created to explain the migration of petroleum out of the source rock. Most researchers now favour processes whereby petroleum is expelled from the source rock as a separate phase within a water-wet rock matrix.

Analyses have been performed on a source rock (Kimmeridge Clay, North Sea) which is actively expelling petroleum. The aim was to elucidate the precise primary migration mechanisms. The analytical results could best be explained by invoking pressure-driven flow of a petroleum-rich phase as the main expulsion mechanism for source rocks. Specifically, it was demonstrated that petroleum was first expelled when the volume of generated petroleum approximately matched the volume of pore space within the mudstone. That is, the mudstone was almost fully saturated with petroleum before expulsion occurred. This supported earlier observations on lean source rocks. Those which yield less than 5 kg petroleum per tonne tend not to achieve sufficient saturation for expulsion to occur.

Gas expulsion may occur in a similar fashion to that of oil, albeit at higher temperatures. Clearly, the volume increase associated with gas generation is massive, be it directly from kerogen or from the thermal decomposition of previously formed oil. Pressure-driven expulsion will occur either through

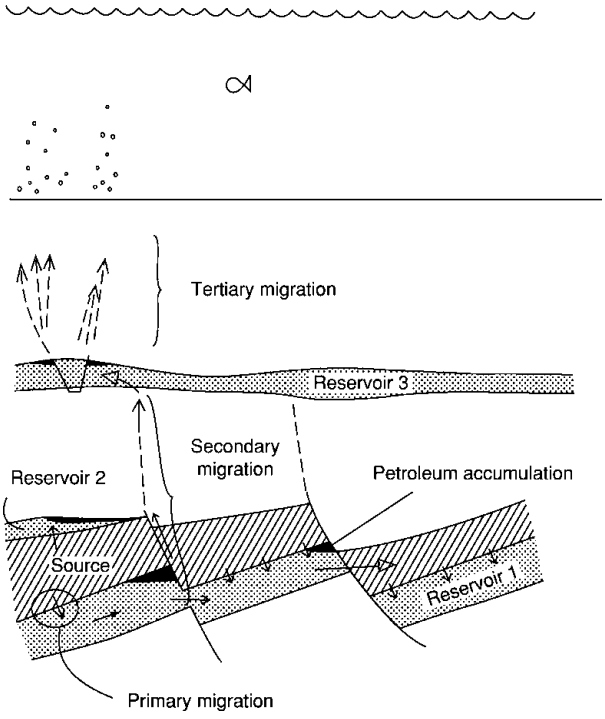


Figure 19 Diagram of the three stages of migration. Primary migration out of the source rock and into a trapped reservoir (reservoir 2) or a carrier bed (reservoir 1). Secondary migration in carrier reservoir 2 and up faults into reservoir 3. Tertiary migration (dissipation) from reservoir 3 to the surface. Reproduced from Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

the existing pore network or through induced fractures. During gas generation, previously generated, short-chain liquid hydrocarbons may become dissolved in the gas and expelled with it. This mechanism has been used to explain the production of condensate from Type III kerogen in overpressured mudstone.

Secondary Migration

Secondary migration takes petroleum from the source location to trap or traps via carrier beds. The defining aspect of secondary migration is that it concentrates or focuses the petroleum. On escape from the source rock, petroleum is dispersed over a large area. By the time petroleum reaches the relatively restricted area of a trap, it can occupy more than 90% of the pore volume in the reservoir. Secondary migration is temporarily arrested once the migrating petroleum enters a trap. Disruption of the trap or overfilling of the trap can lead to remigration of the petroleum to a higher structural level under the same secondary migration process. Such secondary migration ends when petroleum approaches the Earth's surface.

The medium through which the petroleum travels during secondary migration is also quite different from that of the source rock. The pore size and thus permeability in a carrier bed, be it a sandstone, carbonate, or fractured lithology, is much larger than that in a source rock. The driving mechanism for secondary migration is the density difference between the petroleum (less dense) and water (more dense). The density difference is expressed through the buoyancy force generated by the pressure difference between a point in a continuous petroleum column and the adjacent pore water.

The restricting force to petroleum migration is the capillary injection pressure. A slug of petroleum migrates from pore to pore in a carrier bed, squeezing through the intervening pore throats. The force required to move petroleum through a pore throat is a function of the radius of the pore throat, the interfacial tension between the petroleum and the water, and the wettability of the rock–petroleum–water system.

The buoyancy effect means that petroleum will tend to rise within the sediment column. The capillary effect dictates that, in the absence of other forces, petroleum will migrate from small pores to large pores. Furthermore, petroleum (and water) will attempt to equilibrate with respect to pressure. That is, flow can be induced by pressure differential (either overpressure or hydrodynamics).

It is possible to estimate the likely migration directions from source bed to reservoirs by mapping the orthocontours of the likely carrier systems (Figure 20). Orthocontours are simply lines constructed on a map at right angles to the contours. Instead of displaying areas of equal height (or depth), they depict lines of maximum dip. The buoyancy effect dictates that the rising petroleum will follow such orthocontours. Clearly, such an exercise must be attempted on the geometry of the carrier bed(s) as it was during the phase of petroleum migration. This clearly leads to attempts to reconstruct the basin history in terms of deposition, structuring, and source rock maturation.

The capillary effect controls how much of a carrier bed becomes petroleum saturated. Rarely are carrier beds of a uniform grain size distribution. Thus, petroleum will tend to migrate along the coarsest, high-permeability pathways (Figure 21). These may occupy 10% or less of any particular formation. Open fractures have the same effect as coarse beds. Petroleum will exploit them. Temporarily open fracture systems are commonly invoked as the mechanism whereby migrating petroleum 'jumps' upward in the stratigraphy of a particular basin.

The rate at which petroleum migrates can be calculated using Darcy's law

$$q = -(k/\mu)(\mathrm{d}\theta/\mathrm{d}z)$$

where q is the volume of flow rate ($\mathrm{m^3\,m^{-2}\,s^{-1}}$), k is the permeability, μ is the viscosity ($\mathrm{Pa\,s^{-1}}$), and $\mathrm{d}\theta/\mathrm{d}z$ is the fluid potential gradient.

Typical permeability values are: sandstones, 10^{-12}–$10^{-15}\,\mathrm{m^2}$ (1 D to 1 mD); limestones, 10^{-14}–$10^{-17}\,\mathrm{m^2}$ (10 mD to 10 μD).

From these data, it is possible to calculate that the migration rate for petroleum in sandstone will be 1–1000 km per million years and, in limestone, 0.01–10 km per million years.

Phase changes will occur in petroleum as a result of its migration upwards to regions of lower pressure and temperature. This is most important for high-temperature, high-pressure condensates, but any oil will exsolve some gas if the pressure in the formation drops below the bubble point. The residual petroleum and generated gas are then likely to behave differently with respect to subsequent migration.

At low temperatures (<70°C), and in regions in which there is significant water flow, petroleum may be degraded by bacterial action or by water washing. The bacterial process follows a systematic loss of the *n*-alkanes, branched alkanes, isoprenoids, alkylcyclohexanes, and polycyclic alkanes. This progressive destruction of the petroleum leads to increases in the pour point and viscosity of the oil and a lowering of the API (American Petroleum Institute) gravity.

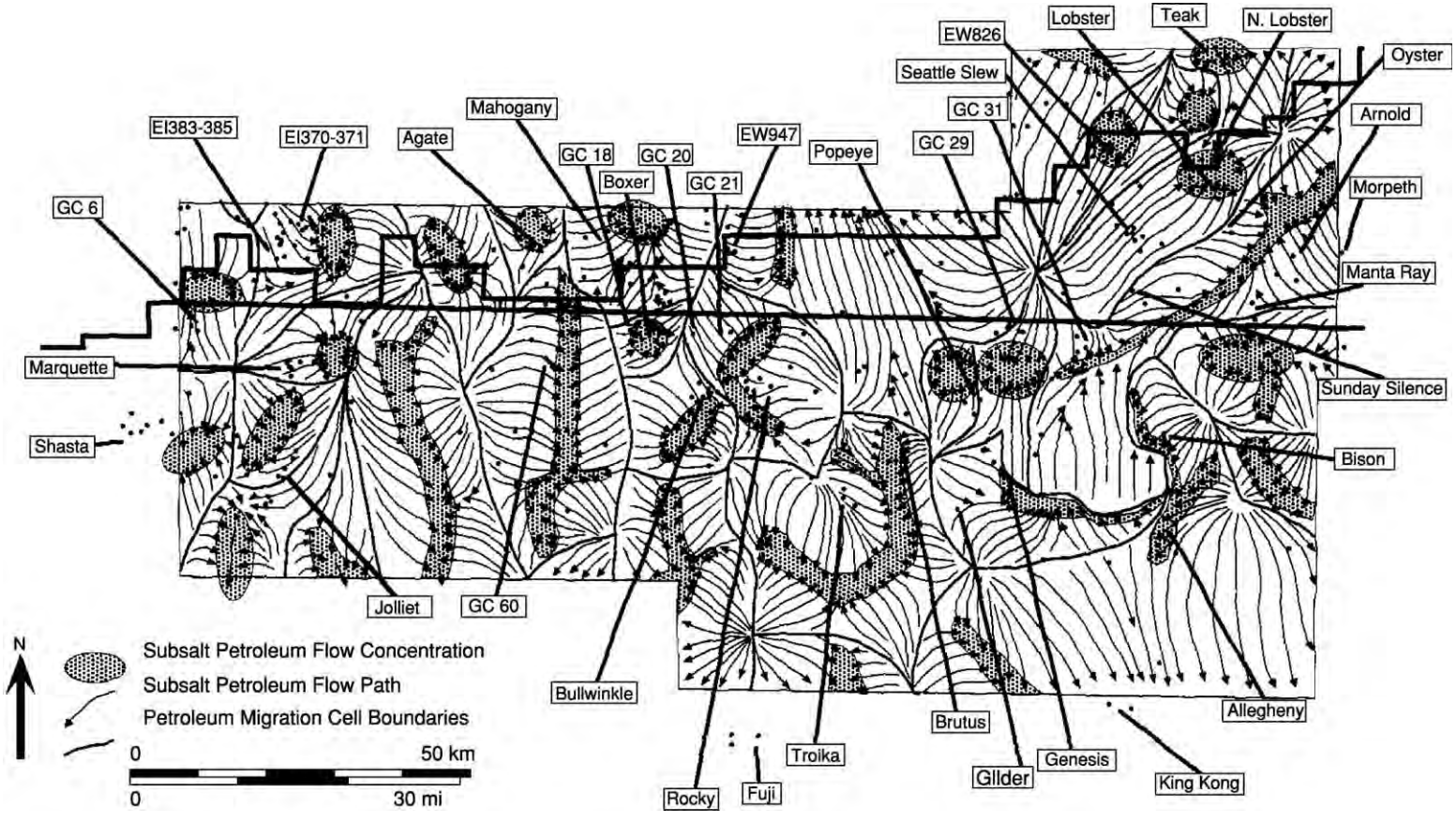


Figure 20 Orthocontours – reconstructed subsalt petroleum migration pathways, Ewing Bank to Green Canyon areas, Gulf of Mexico, USA. Reproduced with permission from McBride BC, Weimer P, and Rowan MG (1998) The effect of allochthonous salt on the petroleum systems of the Northern Green Canyon and Ewing Bank (Offshore Louisiana), Northern Gulf of Mexico. *American Association of Petroleum Geologists Bulletin* 82: 1083–1112.

Figure 21 Petroleum migration along high-permeability sandstone beds within a stacked sequence of turbidite sandstones and siltstones. The migration route was exposed during the excavation of a road cutting in Ecuador. Photograph by M. Heffernan. Reproduced from England WA, Mackenzie A, Mann D, and Quigley T (1987) The movement and entrapment of petroleum fluids in the subsurface. *J. Geol. Soc.*, vol. 144, p. 327. London.

Tertiary Migration

Tertiary migration includes leakage, seepage, dissipation, and alteration of petroleum as it reaches the Earth's surface. The products of seepage may be gas chimneys in the shallow sediment, gas hydrate layers and mounds, cemented pock marks and mud volcanoes, effects on vegetation, and live oil and gas seepage at the surface.

The physical processes that drive tertiary migration are the same as those that operate during secondary migration. Buoyancy drives the petroleum to the surface. This may be helped or hindered by overpressure gradients or hydrodynamics. The only major difference that can be used to separate tertiary migration from secondary migration is the rate of petroleum supply. Trap failure, through capillary leakage, hydraulic fracture, or tectonism, supplies petroleum into a new carrier system much more rapidly than does a maturing source rock.

See Also

Petroleum Geology: The Petroleum System; Exploration; Production. **Sedimentary Environments:** Depositional Systems and Facies; Carbonate Shorelines and Shelves. **Sedimentary Processes:** Karst and Palaeokarst. **Sedimentary Rocks:** Chalk; Dolomites; Sandstones, Diagenesis and Porosity Evolution; Limestones. **Tectonics:** Faults; Fractures (Including Joints).

Further Reading

Abbots IL (1991) *United Kingdom Oil and Gas Fields, 25 Years Commemorative Volume, Geological Society Memoir No. 14*. London: Geological Society.

Allen PA and Allen JR (1990) *Basin Analysis, Principles and Applications*. Oxford: Blackwell Science.

Archer JS and Wall PG (1986) *Petroleum Engineering, Principles and Practice*. London: Graham & Trotman.

Bathurst RGC (1976) *Carbonate Sediments and Their Diagenesis, Developments in Sedimentology 12*. Oxford: Elsevier.

Dickson JAD and Saller AH (1995) Identification of subaerial exposure surfaces and porosity preservation in Pennsylvanian and Lower Permian shelf limestones, eastern central Basin Platform, Texas. In: Budd DA, Saller AH, and Harris PM (eds.) *Unconformities and Porosity in Carbonate Strata, American Association of Petroleum Geologists Memoir 63*, pp. 239–258. Tulsa, OK: American Association of Petroleum Geologists.

England WA and Fleet AJ (1991) *Petroleum Migration, Special Publication 59*. London: Geological Society.

Foster PT and Rattey PR (1993) The evolution of a fractured chalk reservoir: Machar Oilfield, UK North Sea. In: Parker JR (ed.) *Petroleum Geology of Northwest Europe: Proceedings of the 4th Conference*, pp. 1445–1452. London: Geological Society.

Glennie KW (1998) *Petroleum Geology of the North Sea, Basic Concepts and Recent Advances*, 4th edn. Oxford: Blackwell Science.

Gluyas JG and Hichens HM (2003) *United Kingdom Oil and Gas Fields Commemorative Millennium Volume, Memoir 20*. London: Geological Society.

Gluyas JG and Swarbrick RE (2003) *Petroleum Geoscience*. Oxford: Blackwell Science.

McBride BC, Weimer P, and Rowan MG (1998) The effect of allochthonous salt on the petroleum systems of the Northern Green Canyon and Ewing Bank (Offshore Louisiana), Northern Gulf of Mexico. *American Association of Petroleum Geologists Bulletin* 82: 1083–1112.

Reynolds AD (1994) Sequence stratigraphy and the dimensions of paralic sandstone bodies. In: Johnson SD (ed.) *High Resolution Sequence Stratigraphy: Innovations and Applications*, pp. 69–72. Liverpool: Liverpool University.

Selley RC (1996) Elements of Petroleum Geology, 2nd edn. San Diego: Academic Press.

Chemical and Physical Properties

C Clayton, Eardiston, Tenbury Wells, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Definitions

Geochemists define petroleum as any subsurface material which, when produced, yields crude oil and/or gas. Most petroleum forms from the thermal degradation of kerogen during burial, although, strictly speaking, biogenic or bacterially formed gases should also be included. Hydrocarbons are molecules consisting only of carbon and hydrogen atoms. Although they are one of the principal constituents of petroleum, they are not the only components present, and so the two terms are not, strictly speaking, equivalent. Oil is the fraction of petroleum which is liquid at standard temperature and pressure, whilst the remainder is gas. Gas typically consists of compounds with up to five carbon atoms (C_1–C_5), whereas oil has six or more (C_{6+}), although some smaller molecules are sometimes present in oil and some C_{6+} compounds may be present in gas. Condensate is petroleum that is liquid at surface conditions, but exists in the reservoir in the gas phase as a gas condensate. There is no genetic connotation to this definition, although, in general, condensates tend to be of high thermal maturity.

Oils

Chemical Composition

Oil is a complex mixture, dominated by hydrocarbons, but with varying amounts of compounds containing sulphur, nitrogen, oxygen, and metals, principally nickel and vanadium. The chemistry of an oil is determined both by the nature and quality of the source rock and the depositional environment of the kerogen from which the oil formed, and by the thermal maturity of the source rock (i.e., the maximum temperature encountered in the source rock). For example, oil sourced from a marine carbonate is typically very heavy and sulphur rich, but becomes lighter and progressively sulphur depleted with increasing maturity. Clastic marine source rocks tend to produce lower sulphur, lighter oils. In addition, the composition of the petroleum may subsequently become altered in the reservoir by biological processes (biodegradation), thermal cracking, or physical separation (e.g., precipitation of high molecular weight components to form a tar mat).

At the most basic level, oil components can be divided into four main compound groups: saturated compounds, aromatic compounds, an NSO, Nitrogen-, Sulphur-, and Oxygen-bearing compounds, (sometimes called a residue or resin) fraction, and asphaltenes (Figure 1). In practice, the last two groups form a continuum, as their separation is based on solubility in the laboratory, rather than on a unique chemical composition. They are often referred to together as the polar fraction, as their molecules often have a significant surface charge, unlike saturated and aromatic compounds.

Saturated compounds These are generally the most abundant components of oil, comprising up to 90% of some light crudes. A large number of compounds are present within the saturated fraction, but these can be broadly divided into three groups: normal alkanes, branched alkanes, and cyclic compounds (Figure 2). All of these are said to be saturated as they contain no double chemical bonds.

Normal, or straight-chain, alkanes, denoted by the prefix '*n*', range in carbon number from five up to several hundred, but are dominantly in the range *n*-C_5 to *n*-C_{40}. Alkanes greater than about *n*-C_{20} are often referred to as the wax fraction, although here again the definition is based on laboratory solubility rather than on strict chemistry. In practice, industrial 'paraffin wax' also contains branched and cyclic alkanes. Within these higher molecular weight *n*-alkanes, oils derived from higher land plants usually show a dominance of odd carbon numbers over even, reflecting their origin from cuticular waxes of continental higher plants. With increasing maturity, the distribution of *n*-alkanes becomes progressively skewed towards shorter chain lengths and any generic information is lost. Normal alkanes are also particularly vulnerable to bacterial attack and are usually absent in biodegraded oils.

Branched alkanes contain one or more side-chains attached to the straight-chain backbone. These are quantitatively the most abundant fraction within the saturates. A special subgroup of the branched alkanes, called the isoprenoids, is worthy of note. These range from C_9 to C_{25} and are believed to be derived from the phytol side-chain of chlorophyll. The C_{19} and C_{20} compounds, pristane (Pr) and phytane (Ph), are of particular diagnostic value in determining the source of the oil. Coal-sourced oils

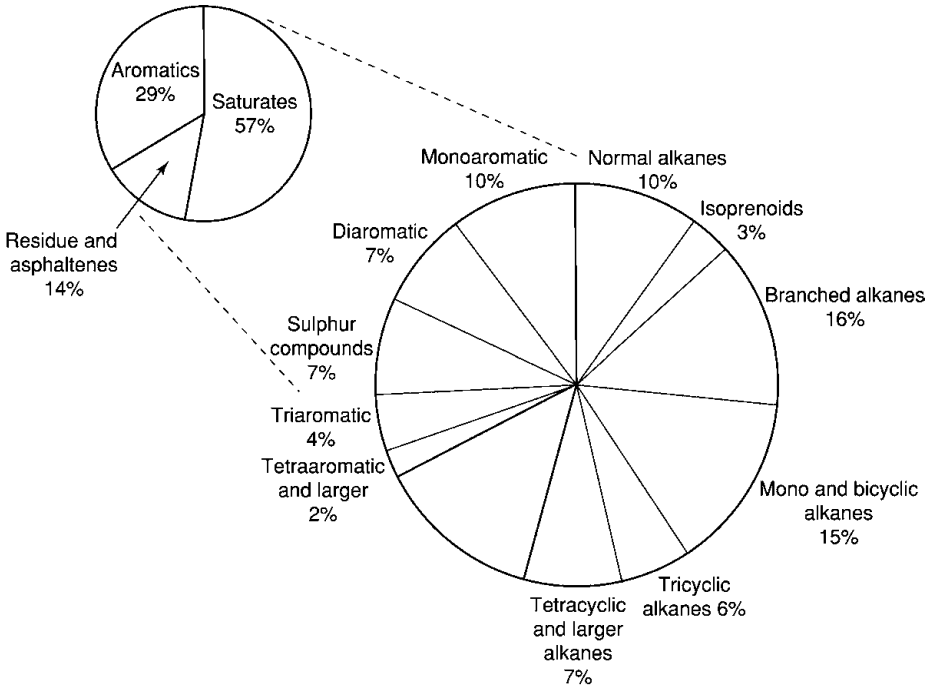


Figure 1 Average composition of oils based on between 100 and 500 samples depending on fraction.

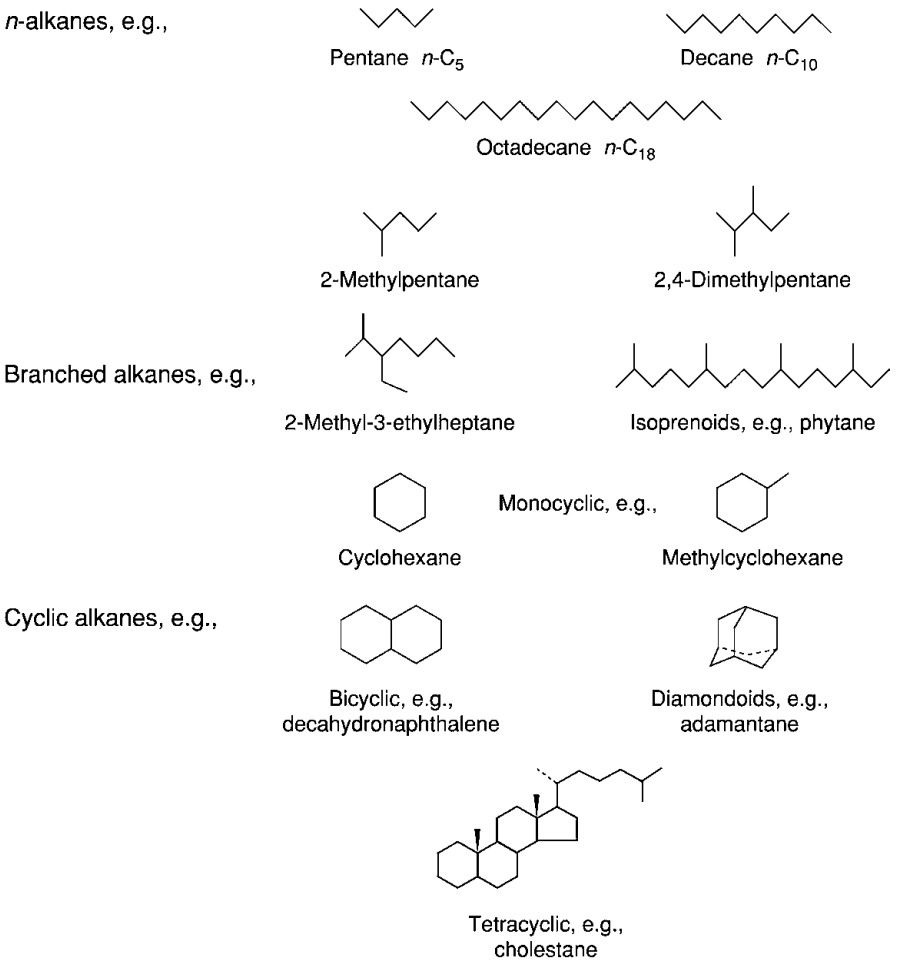


Figure 2 Typical saturated compounds in oil.

generally have a Pr/Ph ratio above three, whereas almost all aquatic kerogens produce oils with a Pr/Ph ratio less than this.

Cyclic alkanes contain carbon rings, usually with five or six carbon atoms in each ring. These are sometimes referred to as the naphthenes, in contrast with the straight and branched alkanes which are paraffins. These compounds may be monocyclic (one ring), or contain up to five or, in rare cases, more rings.

Amongst the saturated compounds (and, to a lesser extent, the aromatic and polar compounds too), some of the more complex molecules can have their origin traced back to specific biomolecules deposited in the source kerogen. They are therefore useful to determine the source of the oil. These so-called biomarkers (or geochemical fossils) are derived ultimately from bacterial or algal cell membranes, or from specific metabolites in algae, bacteria, or higher land plants, and a very large number are known. Some typical examples are shown in Figure 3. Particularly important amongst the polycyclic biomarkers are the pentacyclic steranes and tetracyclic triterpanes (hopanes), as, although present in only very low abundance, many of them are highly characteristic of the source input to the kerogen from which the oil was derived. In addition, subtle variations in the chemical structure of many of the biomarkers (such as their stereochemistry and degree of aromatization) are temperature sensitive, and hence can be used as molecular maturity indicators (see Figure 4).

Aromatic fraction Aromatic compounds contain one or more benzene rings within their structure (Figure 5). These generally consist of roughly equal amounts of one-ring, two-ring, and organosulphur compounds (mainly benzothiothenes and dibenzothiothenes; Figure 6), although lesser amounts of larger aromatic compounds are also present. Some of the more complex compounds, such as the so-called aromatic steranes, are formed by thermal alteration of the saturated equivalent, and hence their relative abundance is indicative of maturity (Figure 4).

Polar fractions The NSO and asphaltene fractions are made up of more complex compounds, usually containing nitrogen, oxygen, or sulphur (Figure 6). The larger molecules can be thought of as small fragments of kerogen. Asphaltenes are the highest molecular weight components present and are defined as the fraction that is not soluble in hydrocarbon solvents (usually hexane). In some heavy oils (such as those formed at low maturity or by bacterial alteration of an originally lighter oil), asphaltenes may comprise up to 50% of the total oil, but they are generally present as only a few per cent. It is the abundance of these polar fractions that determines in most part the viscosity of the oil.

The detailed chemistry of the polar fractions is less well understood than that of the other compound groups because of the complexity of many of the species present. However, recent research suggests that a number of the nitrogen- and oxygen-bearing molecules may be of value as indicators of migration distance, as their ratios potentially reflect the extent to which the oil has interacted with water since expulsion from the source rock.

Analytical Methods

Chemical analyses of oils are most frequently presented in graphical form rather than as tables of results, and so it is relevant to consider the most important analytical methods used to produce these plots. Although a very wide range of analytical tools are now used, by far the most important methods are based on chromatography.

Asphaltene separation and liquid chromatography The largest and most polar components are separated chemically as, by definition, asphaltenes are insoluble in hydrocarbon solvents (e.g., hexane or heptane). No further measurements are carried out on the asphaltene fraction other than perhaps carbon isotope measurements (see below) or occasionally pyrolysis to release smaller, more easily identifiable, components. After asphaltene separation, the remaining mixture is usually separated into compound groups using high-performance liquid chromatography (HPLC). This separates the saturated compounds, aromatic compounds, and residue fraction (the NSO fraction or resins).

Gas chromatography The saturated fraction (and sometimes the aromatic fraction) is then further analysed by gas chromatography (GC) (*see* **Analytical Methods:** Geochemical Analysis (Including X-Ray)). In this method, gas is used to force a sample through a long narrow tube. Small molecules elute first, whilst larger molecules are held up. It is therefore possible to resolve and quantify individual compounds of interest in the oil. The resulting gas chromatograms (Figure 7) show the concentration of each (resolved) compound plotted against the retention time in the chromatographic column. These show a number of features of interest, such as the distribution of the *n*-alkanes, the Pr/Ph ratio, the amount of wax (high molecular weight alkanes), and sometimes characteristic individual compounds of value in correlation studies. Gas chromatography is the method also used to determine natural gas composition.

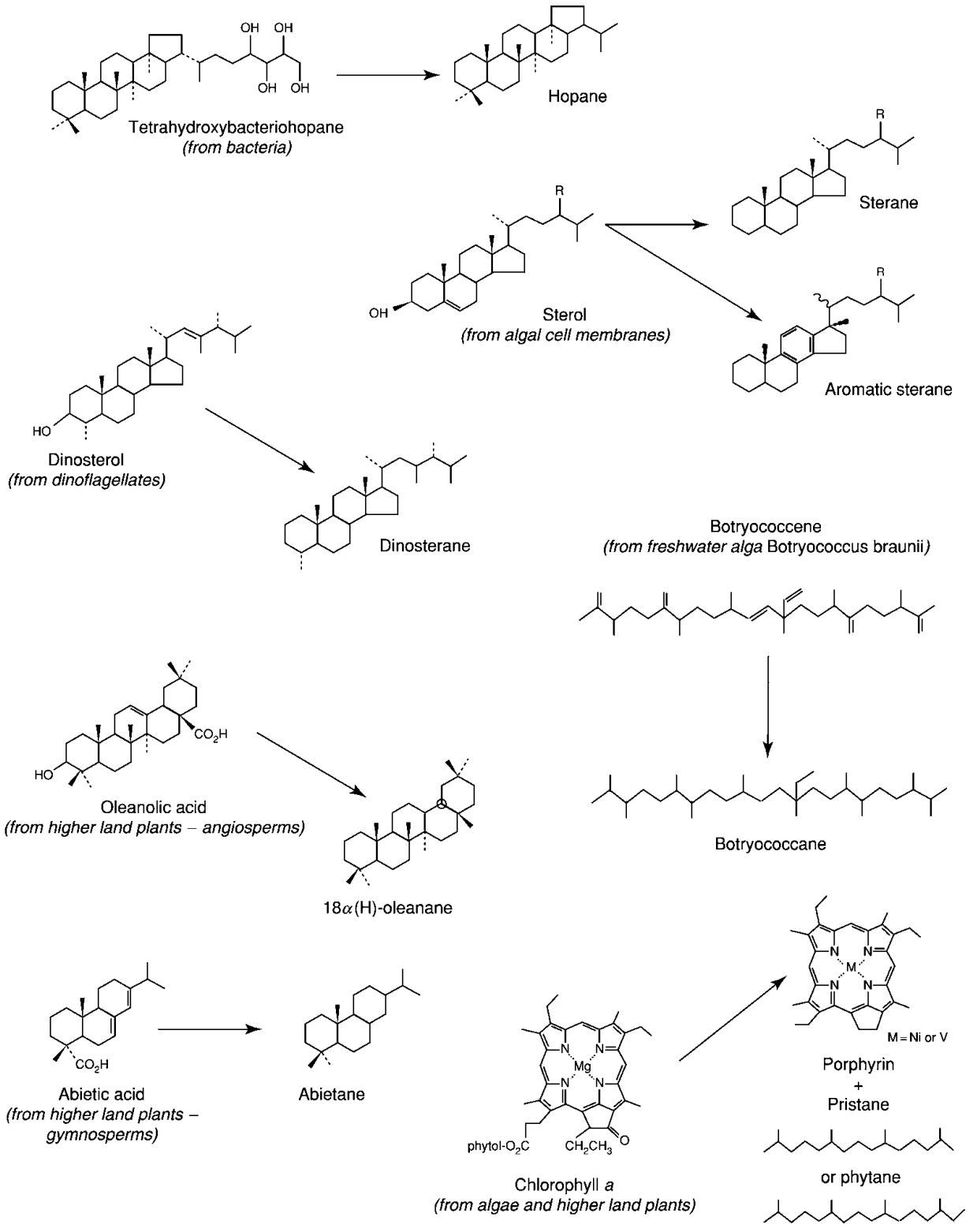


Figure 3 Origin of some common biomarkers.

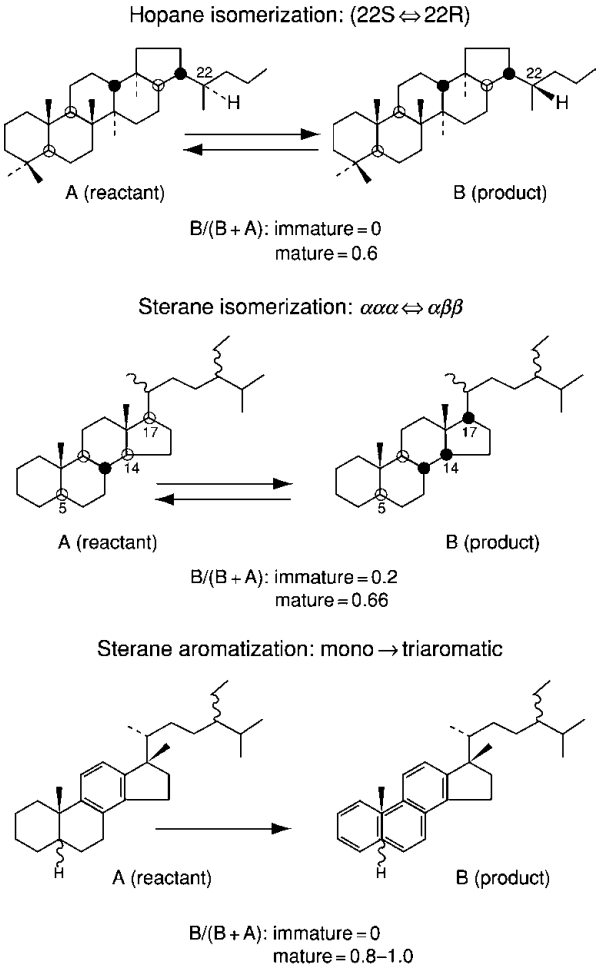


Figure 4 Examples of molecular maturity parameters.

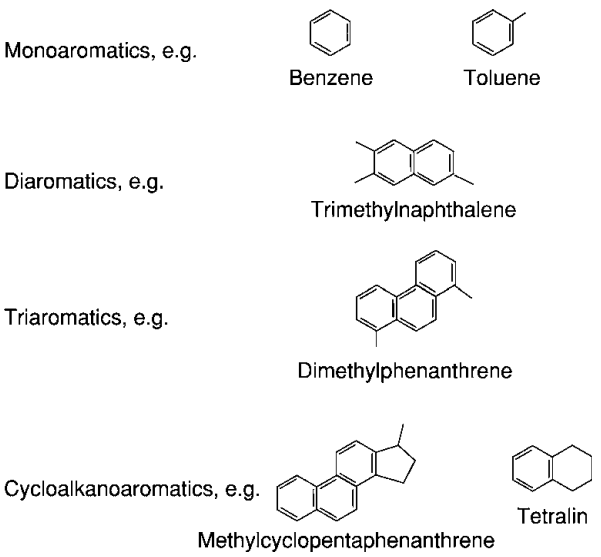


Figure 5 Typical aromatic compounds in oil.

Gas chromatography-mass spectrometry Most of the biomarkers of interest are present in only trace quantities, and so special methods are needed to measure their concentration. Central amongst these is gas chromatography-mass spectrometry (GCMS). This is similar to conventional GC, but a mass spectrometer is used as a detector. When a sterane, for example, enters the mass spectrometer, it fragments into many parts. However, all steranes produce a fragment with a characteristic mass (mass/charge ratio, *m/z*; *m/z* 217 in this case). Thus, by monitoring all fragments of *m/z* 217 as they come off the GC into the mass spectrometer, we are essentially resolving all of the steranes into individual compounds. The result is a plot similar to a GC trace but specific to the steranes. Similarly, the triterpanes (dominantly the hopanes) produce a characteristic fragment at *m/z* 191, and so on for all the other compound groups of interest. Figure 8 shows an example of mass 191 and 217 mass fragmentograms for a typical North Sea oil with some of the important peaks labelled.

Classification of Crude Oils

There are numerous schemes for classifying crude oils, but two main approaches dominate. The first is based on the relative concentrations of the main chemical classes and the most widely used of these was developed by the Institut Francais du Pétrole (Figure 9). Alternatively, and more useful for exploration purposes, oils can be classified according to the depositional environment in which their source kerogen was formed. A typical example of this is shown in Table 1, which includes some of the most relevant chemical characteristics for each environment. It should be noted, however, that, with increasing maturity, the chemical signatures are progressively lost as the characteristic compounds become progressively broken down to smaller, less diagnostic compounds.

Bulk Properties

The economic value of an oil is usually dictated by its bulk properties rather than its detailed chemistry, although the two are obviously inextricably linked. A number of standard measurements are used to define the bulk properties.

API gravity API gravity (from American Petroleum Institute) is a measure of the specific gravity of the oil at 60°F or 16°C. It is somewhat arbitrarily defined by the equation

$$\text{API}^\circ = (141.5/\text{specific gravity}) - 131.5$$

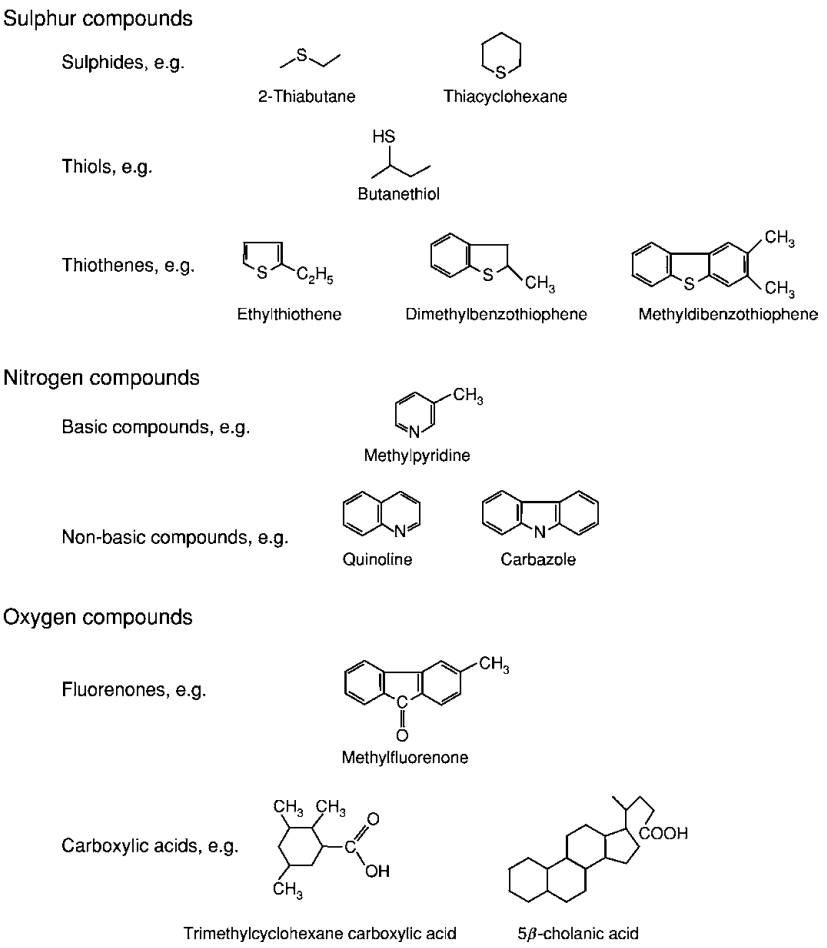


Figure 6 Typical sulphur-, nitrogen-, and oxygen-bearing compounds found in oil.

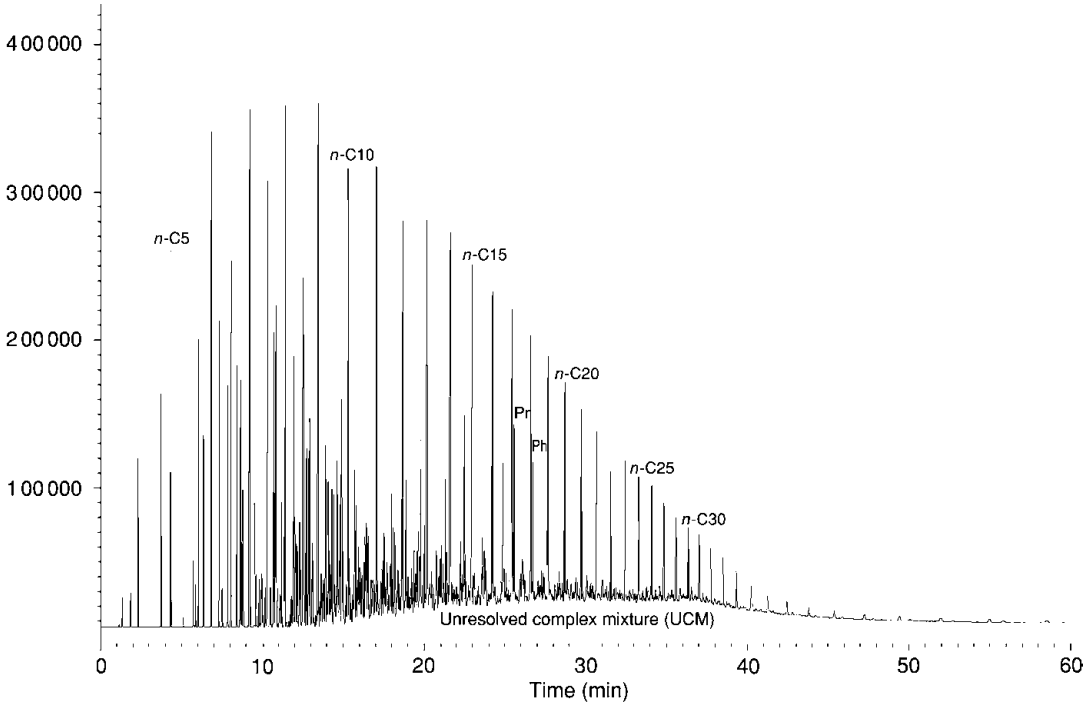


Figure 7 Typical gas chromatogram of saturates fraction. Ph, phytane; Pr, pristane.

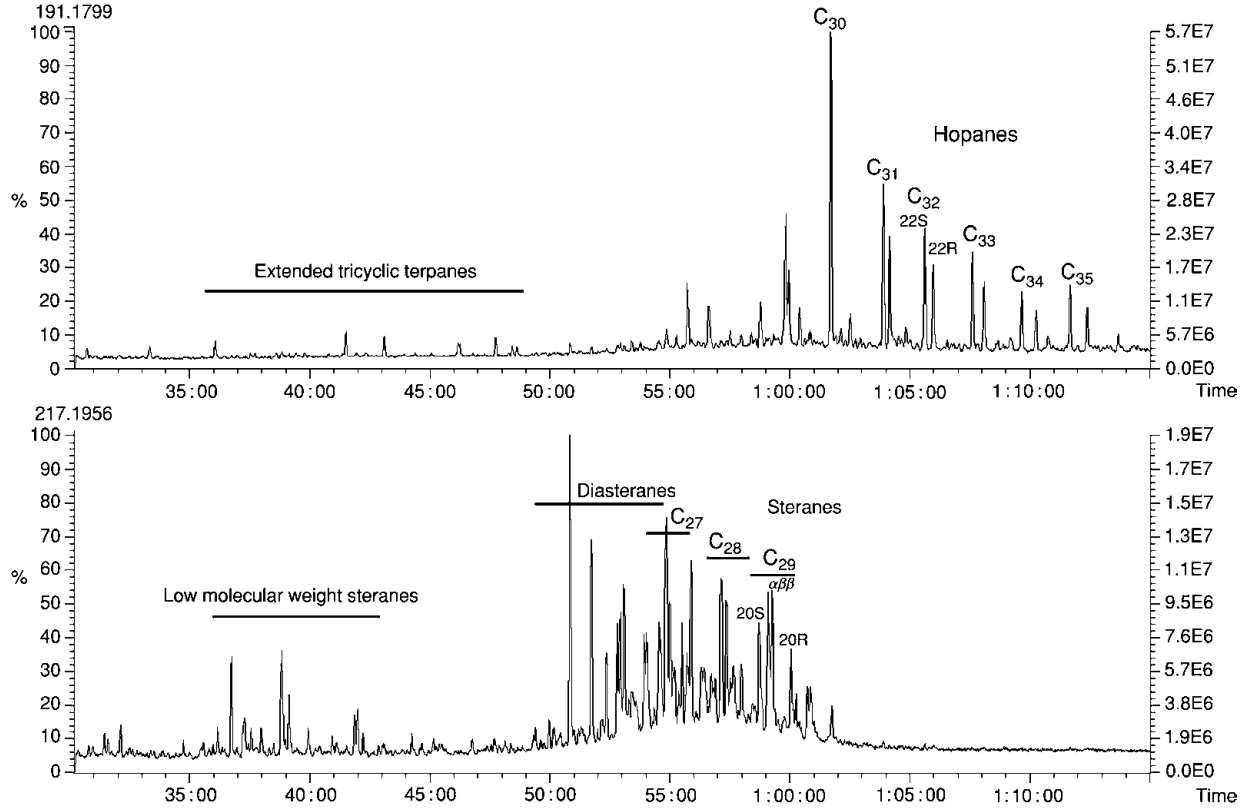


Figure 8 Typical gas chromatography-mass spectrometry (GCMS) fragmentograms. Top part shows the *m/z* 191 trace; bottom part shows the *m/z* 217 trace.

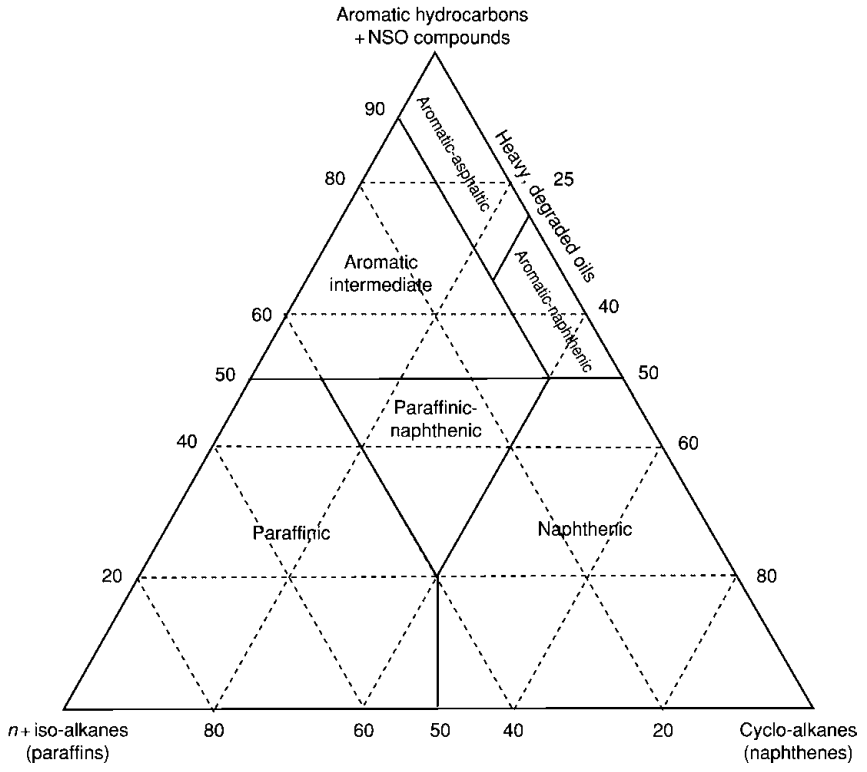


Figure 9 Classification of crude oils based on bulk composition. NSO, Nitrogen-, Sulphur-, and Oxygen-bearing compounds.

Table 1 Oil classification according to source type

	Typical characteristics at moderate maturity[a]					
	API°	*Wax*	*Sulphur*	*Nitrogen*	*Ni + V*	*Ni/V*
Marine carbonate or siliceous	25–30	Low	High	High	50–300	<2
Marine siliciclastic	35–40	Low	Moderate–low	Moderate	0–50	<2
Lacustrine	30–35 and 45–50[b]	High	Low	Moderate	0–50	>2
Coal	30–45	High	Low	Low	~0	–

API, American Petroleum Institute.
[a]With increasing maturity, API increases, and wax, sulphur, nitrogen, and metal contents decrease.
[b]Bimodal distribution observed.

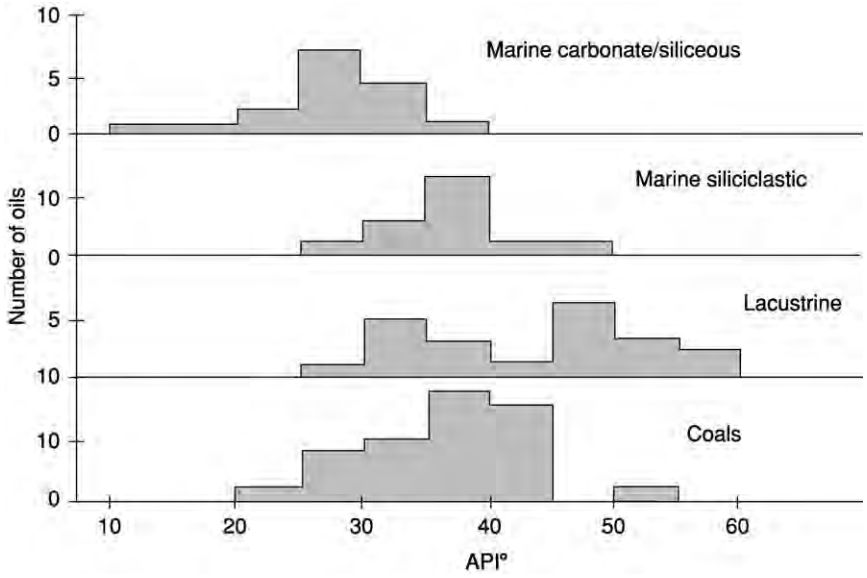


Figure 10 API (American Petroleum Institute) gravity variation of oils according to the depositional environment of the source kerogen.

This system is used as it gives an integer value that is easier to display than a four decimal place number (e.g., pure hexane with a specific gravity of 0.6594 is equivalent to an API of 83°). A heavy crude will generally have an API <25°, whilst a light crude will generally have an API between 35° and 45°. Most condensates have an API >45°. Pure water, by definition, has an API of 10°, and thus crudes with API <10° will not float on water.

The API gravity varies somewhat depending on the source of the oil (Figure 10), and also increases significantly with maturity regardless of source. Oils derived from carbonate or siliceous source facies are usually of lower API than those from other lithologies. In addition, bacterial alteration of oil in the reservoir can result in oils with especially low API values, and gravity segregation in thick homogeneous reservoir formations or highly compartmentalized reservoirs can lead to significant vertical and lateral heterogeneity, respectively.

Gas/oil ratio The gas/oil ratio (GOR) is the ratio of components that exist in the gas phase at surface conditions (dominantly compounds containing less than five carbon atoms) to the rest of the petroleum. It is traditionally expressed as standard cubic feet of gas per barrel of oil (sft^3/bbl), and 1 sft^3/bbl is equal to 0.178 $m^3\,m^{-3}$. When dealing with gas condensates, it is often more convenient to express this ratio as the condensate/gas ratio (CGR) in terms of barrels of condensate per million standard cubic feet of gas (bbl/MMscf).

GOR varies depending on the source of the oil, but is influenced mainly by maturity. GOR increases enormously with increasing maturity.

Wax content The wax content is a reflection of the amount of high molecular weight saturated compounds (*n*-alkanes $>C_{20}$). Most oils range from 0% up to about 40% wax. There are two distinct sources of wax in oil. It may be derived ultimately

from cuticular waxes on the leaves of higher land plants or from long-chain unsaturated molecules synthesized by some freshwater algae. Consequently, many lacustrine or coal-sourced oils have a high wax content (Figure 11), although this decreases with increasing maturity. Wax deposits in pipelines and other equipment can cause mechanical problems.

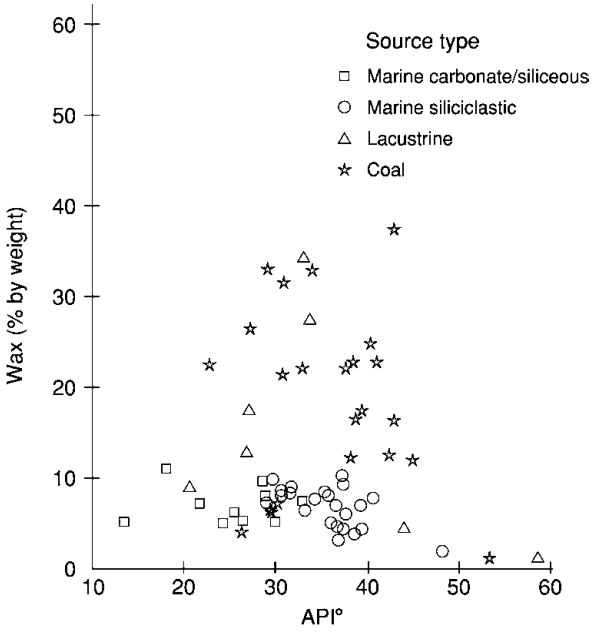


Figure 11 Wax content of oil types vs. API (American Petroleum Institute) gravity as a proxy for maturity.

Sulphur content The sulphur content of an oil has an important impact on its economic value as it causes problems during refining. The main control on the sulphur content is the kerogen type. Oils derived from carbonate or siliceous source rocks are higher in sulphur than those from other lithologies (Figure 12). This itself is a consequence of bacterial sulphate reduction: there is little free iron available in carbonate or siliceous sediments, and so the hydrogen sulphide formed reacts with the organic matter to give a sulphur-rich kerogen. All oils show a marked decrease in sulphur content with maturity.

Nitrogen content The nitrogen content of an oil is a reflection of the abundance of the NSO fraction and hence usually correlates with API gravity and viscosity. The nitrogen content is generally very low in oils sourced from coal, and highest in those derived from marine carbonate and siliceous sediments, but decreases markedly with increasing maturity (Figure 13).

Viscosity The viscosity of oils ranges from around 0.2 cP (2×10^{-4} Pa s) for a very light oil up to around 100 P (10 Pa s) for some very heavy oils. The viscosity is controlled dominantly by the abundance of the high molecular weight polar components (NSOs and asphaltenes), and hence shows a correlation with API gravity and sulphur content.

Distillation fractions Of interest to the refining process are the distillation fractions of an oil (i.e., the distribution of different boiling point fractions). The

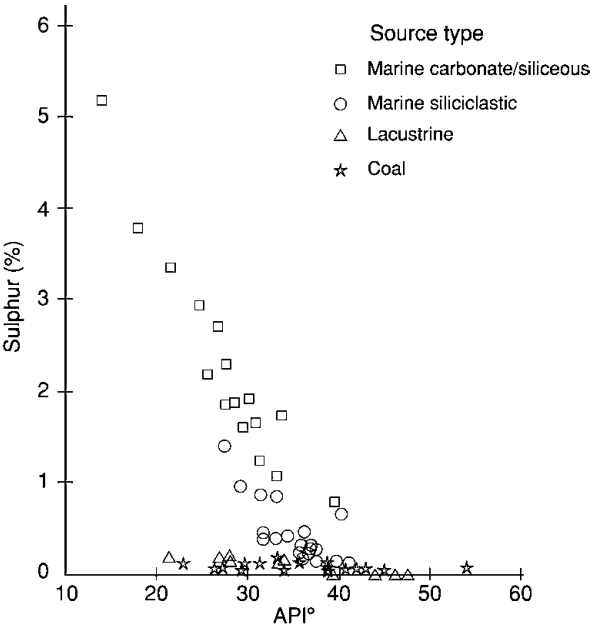


Figure 12 Sulphur content of oil types vs. API (American Petroleum Institute) gravity as a proxy for maturity.

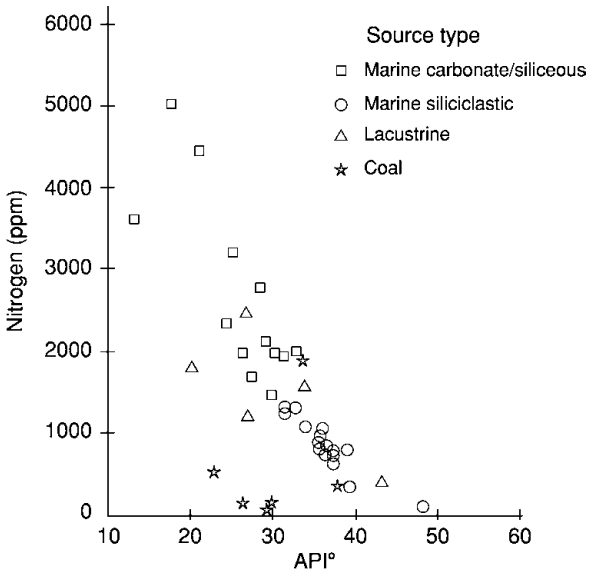


Figure 13 Nitrogen content of oil types vs. API (American Petroleum Institute) gravity as a proxy for maturity.

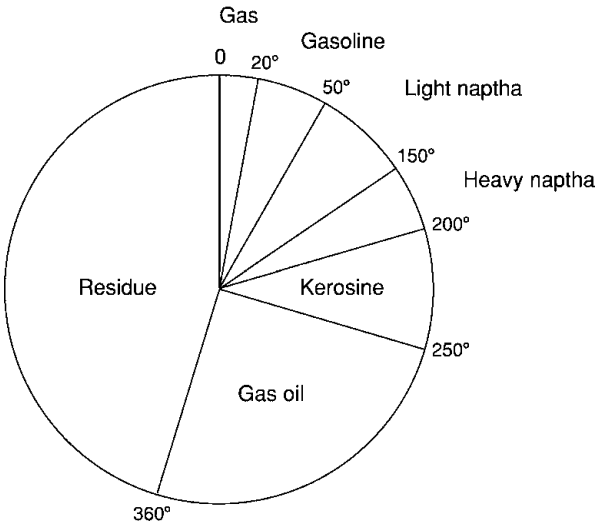


Figure 14 Distribution of boiling point cuts for a typical crude oil.

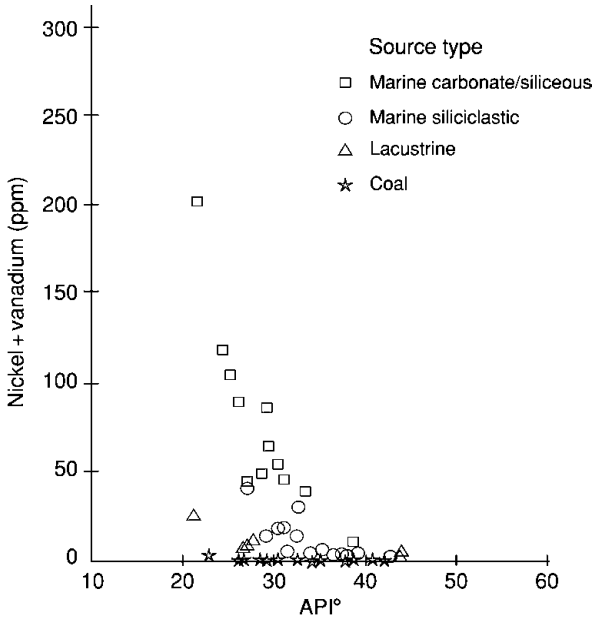


Figure 15 Metal content of oil types vs. API (American Petroleum Institute) gravity as a proxy for maturity.

higher the API gravity of an oil, the larger the yields of the lower boiling point cuts. Various names are applied to the different fractions, an example of which is summarized in Figure 14.

Metal content The metal content of oils is dominated by nickel and vanadium, although other metals, such as iron, zinc, copper, lead, molybdenum, chromium, manganese, cobalt, and arsenic, are also sometimes present in trace quantities. These metals are present dominantly in a specific group of cage-like compounds, called porphyrins (Figure 3), which are ultimately derived from chlorophyll. High nickel and vanadium contents significantly reduce the value of an oil, as they adversely affect the catalysts used in the cracking of oil at the refinery. Metal contents vary enormously, from less than 1 ppm in some Palaeozoic crudes up to 1200 ppm vanadium and 150 ppm nickel in the Boscan crude from Venezuela, and are again controlled by both the source type and maturity (Figure 15). In addition, the ratio of nickel to vanadium is of some diagnostic value in determining the source of the oil (Figure 16).

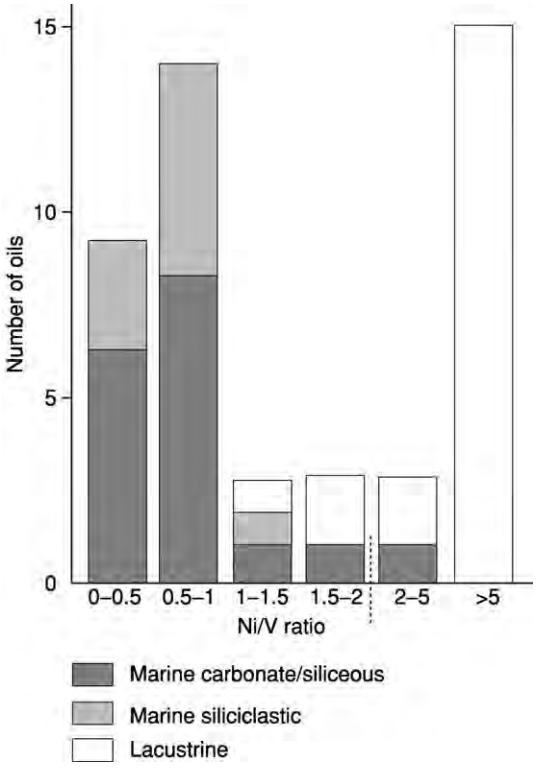


Figure 16 Nickel/vanadium ratios of different oil types.

Gases

Natural gases consist dominantly of hydrocarbons up to C_5 (pentane), together with varying amounts of inorganic components, principally nitrogen (N_2), carbon dioxide (CO_2), and hydrogen sulphide (H_2S).

Hydrocarbon Gases

Natural hydrocarbon gases can be either 'biogenic' (arising from bacterial processes, usually during early diagenesis) or 'thermogenic' (resulting from the effects of heat on oil or kerogen) in origin.

Biogenic (or bacterial) gases are formed in organic-rich sediments during diagenesis by a specialized

group of bacteria, the 'methanogens'. Methanogenic bacteria require anoxic, sulphate-free environments at temperatures below approximately 75°C. These are met, for example, in paddy fields, the intestinal tracts of higher organisms, and in marine sediments below the zone of sulphate reduction. Chemically, biogenic gases are almost pure methane.

Under the right conditions, biogenic gases can form substantial economic accumulations. They account for 15–20% of the world's gas resources, including the Cenomanian pay zones of the giant Urengoy field in western Siberia with ~180 trillion cubic feet recoverable reserves. They occur at depths down to 4600 m (Antonella, Po Valley) and in sediments from Holocene to Cretaceous age.

The thermogenic gases can be subdivided depending on their origin. 'Oil-associated gases' form from kerogen at the same time as oil is generated. They can be thought of as the short-chain component of the oil and are believed to be derived from the lipids (essential oils, fats, and waxes) of organisms. 'Non-associated gases' are derived from lignin, a component of higher land plants. They are the dominant component of gases derived from coals, although they are not strictly synonymous with coal-gas which may also contain cracked oil-derived gas, particularly where oil-prone coals are involved. 'Cracked oil gases' are formed from the thermal destruction of oil, either in the reservoir or by cracking of residual unexpelled oil in the source rock.

The chemistry of natural gases is basically simple. Hydrocarbon components are based on methane (CH_4) and the higher homologues ethane (C_2H_6), propane (C_3H_8), iso- and normal butane (C_4H_{10}), and iso- and normal pentane (C_5H_{12}). Hydrocarbon components containing more than five carbon atoms are generally referred to as the gasoline fraction and considered to be part of oil rather than gas. Unsaturated compounds, containing a double bond, such as ethene (ethylene), are rarely found in natural gases, although can form as a product of laboratory pyrolysis.

Because of the simplicity of gas composition, very little can be deduced from the composition alone, although two ratios are in common usage: the gas 'wetness' (strictly speaking the gas dryness!), $C_1/\Sigma C_{1-5}$, and the iso- to *n*-butane ratio, iso-C_4/*n*-C_4. Some authors also use simpler ratios, such as the methane/ethane ratio (C_1/C_2), or methane relative to ethane and propane ($C_1/C_2 + C_3$). In general, these are not as useful as the gas wetness index as they are open-ended scales (i.e., they extend to infinity). All of these ratios are affected by the gas origin, maturity, and sampling methodology, and generally also by secondary alteration processes, such as biodegradation. Dry gases are generally considered to contain less than 10% heavy hydrocarbons (i.e., $C_1/\Sigma C_n > 0.9$), although different authors use different cut-off points. It is therefore necessary to exercise caution when describing gases as 'wet' or 'dry'.

The best evidence of the source of a natural gas is from the stable carbon isotope ratios of the individual hydrocarbons (methane to butanes and sometimes pentanes). The bulk fractions of oils, and increasingly a number of individual compounds, are often also analysed for their stable carbon isotopic composition. The carbon isotope ratio is based on the ratio of the stable carbon isotopes ^{13}C and ^{12}C and is universally expressed in a standard 'del' (or delta) notation

$$\delta^{13}C = [(R_{sample} - R_{standard})/R_{standard}] \times 1000‰$$

in which R is the ratio $^{13}C/^{12}C$ and the units are parts per thousand or 'per mil'. The standard used for carbon is PDB, calcite from a fossil belemnite from the Cretaceous Peedee Formation of South Carolina, although secondary standards are now used in the laboratory. Positive values contain more ^{13}C than the standard, and are said to be isotopically heavier or enriched. Negative values are isotopically lighter or depleted.

When combined with basic chemical data for the gases, these allow the differentiation of the different sources of natural gas, estimates of their thermal maturity, and, under favourable conditions, the identification of the specific source rock involved (see examples in Figures 17 and 18).

Non-Hydrocarbon Gases

In many reservoirs, non-hydrocarbon gases (notably CO_2, N_2, and H_2S) are encountered in varying proportions. A number of sources for these may be proposed, although it is often difficult to be confident about their origin, even though isotope ratios of carbon, nitrogen, and sulphur can provide useful information. Table 2 shows composition data for a number of natural gases. The non-associated gases in particular contain large amounts of non-hydrocarbons. These reduce the economic value of an accumulation, partly by diluting the hydrocarbons and partly by causing problems in production (e.g., H_2S is both corrosive and highly toxic).

Nitrogen In some cases, nitrogen can be derived from basement rock, particularly from igneous sources. It can also be concentrated in biogenic

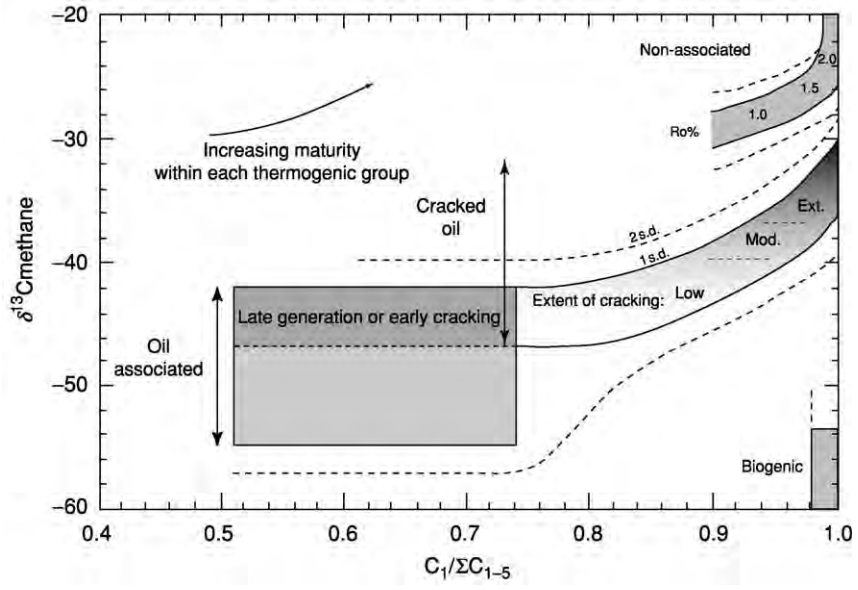


Figure 17 Differentiation of gas origins using gas wetness and carbon isotope ratios of methane.

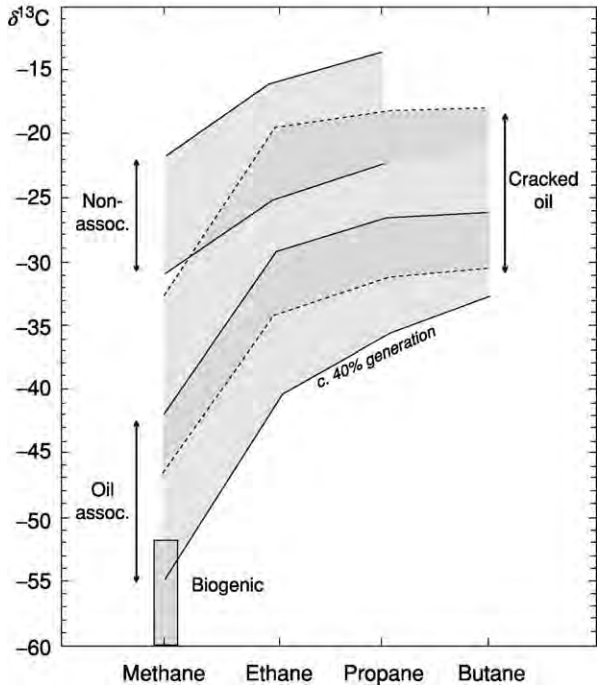


Figure 18 Isotope distributions in gases of different types.

gases by exsolution from flowing groundwater. However, by far the most important source is believed to be from kerogen breakdown (kerogen typically contains about 2% nitrogen by weight).

Table 2 Typical inorganic compositions of oil and gas fields (mol.%)

Area/field	CH_4	C_2–C_5	CO_2	N_2	H_2S
Oil-associated					
Abu Dhabi/Zakum	76.0	20.3	2.3	1.1	0.3
Iran/Agha Jari	66.0	31.5	1.5	1.0	–
N. Sea/Forties	44.5	53.6	0.6	1.3	–
N. Sea/Brent	82.0	16.4	0.7	0.9	–
Non-associated					
Algeria/Hassi-R'Mel	83.5	10.2	0.2	6.1	–
France/Lacq	69.3	5.5	9.6	0.4	15.2
Holland/Groningen	81.3	3.5	0.9	14.3	Trace
New Zealand/Kapuni	46.2	7.9	44.9	1.0	–
N. Sea/West Sole	94.4	4.0	0.5	1.1	–

Nitrogen is generated from organic matter, particularly during the advanced stages of maturation. It has also been proposed that different types of kerogen yield different amounts of nitrogen. Experimental cracking of coals under laboratory conditions indicates that coals can generate up to 6% of nitrogen in total. Kinetic studies, based on laboratory experiments, suggest that gases containing more than 40–50% nitrogen are generated in excess of 300°C, after the main phase of hydrocarbon generation. These experiments imply that, in nature, gas with 50% nitrogen must have been formed by preferential trapping of late-stage generation products, whereby

the gas from the initial stages of generation has not been trapped.

Carbon dioxide Carbon dioxide (CO_2) has a number of possible sources, although most of these usually only contribute small quantities of gas. Although bacterial action produces some CO_2 under specific conditions, the most important sources are kerogen breakdown, inorganic reactions in the reservoir, and deep-seated CO_2.

CO_2 sourced from kerogen occurs as a result of thermal decarboxylation of kerogen. This occurs typically between about 50 and 100°C, before the main phase of oil and hydrocarbon gas generation. This reaction generally contributes only 1 or 2% CO_2 to the total gas. Inorganic reactions are limited to relatively hot reservoirs and are the result of chemical buffering between carbonates and silicates. These can either increase or decrease the CO_2 concentration, depending on the specific mineralogy involved and on the reservoir temperature and pressure. Nevertheless, many tens of per cent of CO_2 may result. Deep-seated CO_2 is often associated with igneous activity, but is also commonly the result of carbonate decomposition in the basement. This is probably the most important source of CO_2 in very high CO_2 (>30%) regions.

Hydrogen sulphide Hydrogen sulphide (H_2S) is an unpleasant, toxic gas occasionally found in large quantities in reservoirs. There are three main sources: bacterial sulphate reduction (of only limited occurrence), thermal cracking of kerogen or oil, and gas souring (thermochemical sulphate reduction). In addition, in an oil reservoir, H_2S preferentially resides in the water phase. This means that, even if there is little H_2S produced when the field is initially brought on stream, the field may subsequently sour as it goes into decline and the water cut increases.

A major source of H_2S is via the thermal cracking of kerogen or oil. Many compounds in oil contain sulphur, and the thermal degradation of these at very high temperatures leads to the formation of H_2S. Such a process becomes more important with increasing temperature. It is frequently associated with carbonate-sourced oils or kerogens as these are richer in sulphur.

The production of H_2S in deep, hot carbonate/evaporite reservoirs is probably the most important source of sour gas. Above approximately 140°C, the ratio of H_2S to methane often increases dramatically in proportion to temperature due to reaction between methane and gypsum or anhydrite, represented in its simplest form by

$$CaSO_4 + CH_4 \rightarrow CaCO_3 + H_2S + H_2O$$

However, in practice, numerous other parallel and intermediate reactions also occur.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray). **Economic Geology**. **Petroleum Geology:** Overview; Gas Hydrates; The Petroleum System; Production; Reserves.

Further Reading

Hunt JM (1996) *Petroleum Geochemistry and Geology*, 2nd edn. New York: Freeman and Company.

Miles JA (1989) *Illustrated Glossary of Petroleum Geochemistry*. Oxford: Clarendon Press.

Peters KE and Moldowan JM (1991) *The Biomarker Guide*. Englewood Cliffs, NJ: Prentice-Hall.

Tissot BP and Welte DH (1984) *Petroleum Formation and Occurrence*, 2nd edn. Berlin: Springer Verlag.

Vially R (ed.) (1992) *Bacterial Gas*. Paris: Edition Technip.

Gas Hydrates

M Hovland, Statoil, Stavanger, Norway

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Oceanic gas hydrates have attracted massive scientific and, not least, political attention over the past years. This is because of their dual role as a potential energy resource and as a geohazard/potential climate mediator. Interest in these compounds is reflected not only in the number of recent scientific projects aimed at studying and sampling oceanic hydrates, but also in the virtual flood of books, articles, and websites dedicated to these 'mysterious' substances.

Gas hydrates are crystalline, ice-like compounds composed of water and gas; the gas molecules are trapped within a cage-like framework of hydrogen-bonded water molecules. These structures are formed under very specific temperature and pressure conditions in an environment with adequate water and gas. The right conditions can be found on land (in polar regions, where surface temperatures are very cold) and within water or seabed sediments at depths exceeding 300–500 m and where the temperature is adequately low (generally $<10^{\circ}$C). Mathews, in *The Log Analyst*, described hydrates as follows:

> Hydrates form when one species (known as the host) forms a crystal lattice containing voids or cavities that physically entrap other guest molecules. In natural gas hydrates, the host is water and the guest can be any common constituent of natural gas such as CH_4, C_2H_6, C_3H_8, C_4H_{10}, CO_2 or H_2S. Natural gas hydrates can have two possible structures, depending on the size of the guest molecules. Both structures consist of a network of water molecules hydrogen-bonded to each other in a manner similar to water ice. (Mathews (1986))

Figure 1 shows a diagram of gas hydrates; the water molecules form a lattice containing 'guest' (gas) molecules – this is also called a clathrate substance in chemistry. The conditions under which gas hydrates are stable (Figure 2) are tightly controlled by temperature and pressure, but are also affected by the chemistry (i.e., including the salinity) of the pore water and the composition of the 'guest' gases. Chemicals (such as salts and methanol) that reduce the pressure/temperature envelope of hydrate stability are called hydrate inhibitors.

Because they tend to block pipes and flowlines in oil and gas fields, gas hydrates have been the focus of well-controlled laboratory studies for many years. Laboratory observations illustrate clearly how dynamic gas hydrates can be: at times these compounds are permeable and at times they present a complete barrier to flow. In a laboratory hydrate reactor at a Statoil facility in Norway, an initial attempt at forming hydrate plugs was carried out with a gas mixture of methane and isobutane, and tap water in the absence of hydrocarbon liquids. Hydrates were initially formed by injecting water droplets into the gas atmosphere at a pressure of 60 bar. A total of 13.5 litres of water was injected during 1.5 h at a subcooling temperature of 13°C. Gas was periodically circulated through the hydrate plug from the bottom of the reactor and the plug was periodically allowed to mature undisturbed for several days. After a static maturation period, an additional 4 litres of water was injected at the top of the plug. Most of the water was drawn into the plug, which acted as a sponge, and remained in the plug for the rest of the experiment. The plug remained porous throughout the entire 20-day-long experiment and grew to a length of 40 cm. When a pressure difference was introduced over the plug, the video recordings demonstrated that gas bubbles easily moved through the internal channels of the plug. Hence, the nature of the channels determined the permeability of the plug. Gas bubbled through the channels until more hydrates were formed and blocked further migration.

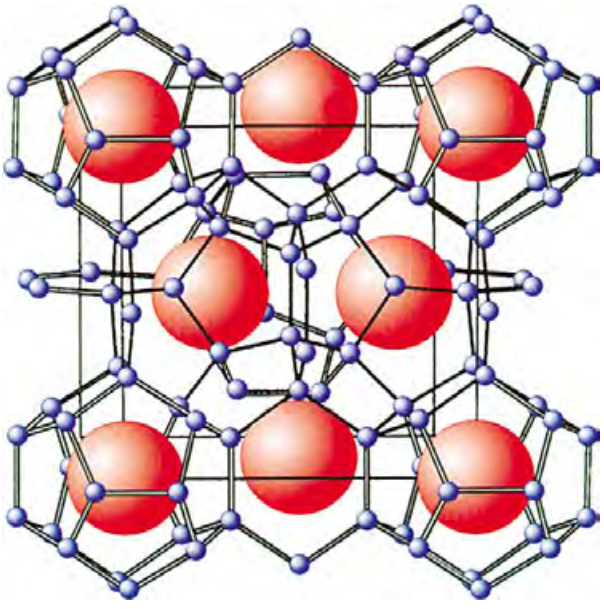


Figure 1 Diagram of a gas hydrate, showing how water molecules form a lattice around 'guest' gas molecules. The chemical name for this cage-like solid structure is 'clathrate'. Diagram courtesy of Bjørn Kvamme.

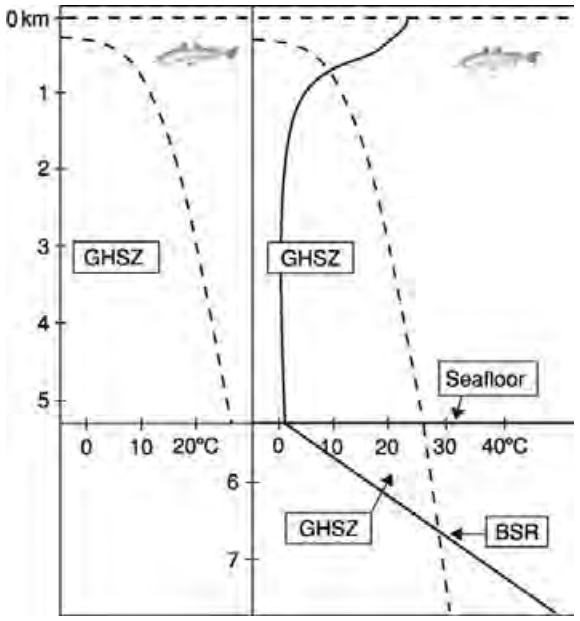


Figure 2 Methane hydrate stability field. The gas hydrate stability zone (GHSZ) in ocean water is determined by pressure and temperature. The stability envelope of gas hydrates in seawater is shown schematically on the left. Gas hydrates usually will not form at water depths of less than about 300 m and will not form in open seawater systems because of seawater circulation and the relatively high solubility of methane in seawater. However, in the confines of sedimentary pores in ocean-bottom sediments (right-hand diagram), gas hydrates will be stable and may form more or less permanently. Because of the increase in temperature with water depth (the geothermal gradient), hydrates will not be able to form beyond certain depths, indicated by the bottom-simulating reflection (BSR), where the temperature gradient curve intersects the hydrate stability curve.

Figure 3 Core sample number 1249F-8h-2, from the Ocean Drilling Program Leg 204, drilling on Hydrate Ridge, the Cascadia Accretionary Prism, off the coast of Oregon, USA. The sample shows how gas hydrates (white) may occur in shallow, porous sediments. Photograph courtesy of Anne M Tréhu and the Shipboard Scientific Party, Ocean Drilling Program Leg 204, 2003.

The gas then found new channels until these were blocked as well. Thus, the plug gradually became less permeable. By measuring the gas released during melting and the water volume after complete dissociation, the plug content was calculated. The volume percentages of hydrates, free water, and free gas in the plug were 12, 38, and 50%, respectively. Hence, the actual hydrate content in the plug was low and the porosity of the plug was concluded to be as high as 88%.

In early work on gas hydrates, it was suspected that they were impervious to fluid flow, and that the entire sediment column within the gas hydrate stability zone (GHSZ) was cemented by gas hydrates. The active migration of free gas through a GHSZ and the porous nature of gas hydrates have since been amply demonstrated experimentally and are known to occur in nature. Gas hydrates may form rapidly when small-molecule gases leak into the sediments, and they may dissociate ('melt') rapidly when disturbed by heating or depressurization. When sediment-hosted hydrates dissociate, they return free gas and water to the environment, which may result in a total change of the physical sediment conditions and even of the seabed topography. Consequently, at water depths and sediment depths within the GHSZ, any planned interaction with the seabed has to take into consideration the possible existence and potential formation of gas hydrates.

When high-porosity sediments such as buried sands and gravels are present in regions with hydrocarbon flux, within the GHSZ, these sediment layers are likely to have developed into massive gas hydrate reservoirs. This has been demonstrated by the Mallik gas hydrate research well in the Mackenzie Delta, Canada. Also, with increasing hydrocarbon flux and increasing tectonic activity, the amount of gas hydrates within the near-surface sediments will increase, as has been demonstrated by, among others, the Geochemical Exploration and Research Group (GERG), in the Gulf of Mexico. In addition, the seafloor topography will be increasingly affected, and the presence of gas hydrates may even lead to very large slope failures and seafloor collapse events.

In regions with a very high fluid (hydrocarbon) flux through the seabed, masses of gas hydrates may form within the sediments, as found at Hydrate Ridge, off the coast of Oregon, USA (Figure 3). Here, seabed sampling showed that pure gas hydrates occurred in layers or in joints several millimetres to decimetres thick, generally oriented parallel to bedding, but sometimes cutting the bedding planes obliquely.

There was not one instance observed where the hydrates occupied the original sediment pore spaces and cemented the sediment framework on a microscale level. Rather, they filled large secondary pore spaces, either as fractures or joints, or seemed to create their own spaces during growth, by fracturing the sediment framework, most commonly along bedding planes. The internal fabric of the pure hydrates also showed a peculiar structure, with large, filled pores, resembling bubble-pack material.

Another important aspect of gas hydrates is that they chemically desalinate the ambient pore water. This characteristic is used by some geochemists to determine how much gas hydrate has been formed in low-permeability sediments, by studying the chlorinity of the pore waters. During Ocean Drilling Program (ODP) Leg 204 on Hydrate Ridge, this phenomenon was studied in detail. Among the most surprising findings of this leg was the rate at which hydrates form near the ridge summit. Salts are excluded when hydrates form, and the excess salt in the pore water diffuses away from the gas hydrates with time. However, if the hydrates form very rapidly, the salts do not have time to diffuse and the water in the sediment pore spaces will become saltier than seawater. Because of the hydrate inhibitor effect, gas hydrates will not be able to form within such brine-filled sediment pockets in the future, another factor that causes sedimentary inhomogeneities.

Detection and Distribution of Gas Hydrates

It is relatively easy to map the 'potential gas hydrate' distribution (i.e., the distribution of locations in which suitable physical conditions occur in the ocean sediments). However, the expected distribution frequency may be overestimated because, even within the GHSZ, gas hydrates will form only if there is adequate water within the sediments to form the 'host' molecules, and an adequate supply of gas 'guest molecules' (for example, methane). The origin of the gas is immaterial, so methane, being the most common gas, may be of either thermogenic or bacterial origin.

Indirect evidence for gas hydrates in ocean-floor sediments first came from anomalous reflections observed on seismic profiles recorded during the Deep Sea Drilling Program (DSDP), Leg 11, on the Blake Outer Ridge, off the eastern coast of the USA (*see* **North America:** Atlantic Margin). The special reflection event (Figure 2), called a bottom-simulating reflection (BSR), is an acoustic interface generated by a (diagenetic) phase boundary. It is anomalous owing to its highly reflective appearance, representing an abrupt change in acoustic impedance. It is also anomalous in that it cross-cuts bedding planes and is of reverse acoustic polarity. Sedimentary strata below the BSR are usually more strongly defined (i.e., enhanced) compared to those above it, where lithological layers tend to generate only weak reflections. This upper zone is therefore referred to as 'seismically transparent' or as 'blanking' and is suspected to be caused by hydrate 'cementation' and relatively high acoustic velocity.

Although the mapping of BSRs can aid in determining the presence of gas hydrates below the seafloor, some areas with gas hydrates do not have BSRs. It was long thought that a BSR was a necessary and adequate indicator of *in situ* gas hydrates, but one important result of the dedicated ODP BSR investigations is that this assumption was found to be partly false. In some places, drilling through BSRs has failed to find gas hydrates; in other cases, gas hydrates have been found where there is no BSR.

A prominent BSR occurring adjacent to the Storegga Slide scarp, off the coast of Norway, was investigated by drilling in 1997, as part of the geotechnical investigations for field development of the Ormen Lange field. However, no gas hydrates or temperature anomalies were found in the sampled sediments. Furthermore, only very little methane occurred in the pore waters below the BSR. These 'negative' results indicate that the existence of a BSR is no guarantee that there are detectable amounts of gas hydrates or free gases in the sediments. The question about what actually produces the seismically detectable BSR is therefore valid. Off the coast of Norway, it was concluded that the feature perhaps represents a palaeo-BSR, and that although the gases and gas hydrates may once have been present, they are no longer there in quantities that are of any practical consequence.

The true nature of BSRs and how they develop are therefore not yet fully understood. There is some consensus that they develop mainly as a consequence of the acoustic contrast between relatively high-velocity, gas-hydrate-charged sediments above low-velocity sediments. The low-velocity characteristic is attributed to free gas beneath a hydrate layer. A BSR may therefore be indicative of gas hydrates above the BSR, free gas beneath the BSR, and the diffusive migration of gas through the whole sediment package, i.e., from below the GHSZ to the seabed.

One of the most surprising results from ODP Leg 164, at Blake Ridge, was the finding of bacterial activity near the BSR. This finding is thought to demonstrate that the sediments near and below the BSR form a biogeochemically dynamic zone, with carbon cycling occurring through methane, acetate, and carbon dioxide. At Site 995, pore water acetate was present in very high concentrations, reaching

$\sim$15 mmol l^{-1} at 691 mbsf (metres below seafloor), $\sim$1000 times higher than 'typical' near-surface concentrations (2–20 μmol l^{-1}). Potential rates of acetate metabolism were extremely high and could not be sustained without influx of organic carbon into the sediment; hence, *in situ* rates are likely to be lower than these potential rate measurements indicate. However, there is evidence for upward migration of high concentrations of dissolved organic carbon into the sediments at these sites. Based on these observations, it can be concluded that the BSR represents a physicochemical and perhaps even a biogeochemical 'reaction zone' brought about by the advection of light hydrocarbons; the hydrocarbons accumulate as free gas below the local GHSZ and form hydrates above – thereby rendering an acoustic reflector that can be imaged by seismic methods.

In 2001, Kvenvolden and Lorenson provided an updated overview of worldwide gas hydrate occurrences:

> An inventory of global occurrences shows 77 places in which the presence of gas hydrate is inferred by geophysical, geochemical, and geological methods. This inventory includes 19 places where samples of the substance have actually been recovered. Most natural gas hydrate contains methane that was generated by microbial processes. Methane with a thermal origin occurs in gas hydrate only where deeper deposits of thermogenic natural gas provide a source of the methane. The potential amount of methane in global gas-hydrate occurrences is very large, with current estimates converging at about 10 teratonnes of methane carbon.

Gas Hydrates as an Energy Source

It has long been speculated that methane hydrates in ocean sediments may represent the greatest reservoir of methane on the planet, offering a clean and abundant energy source for the future. Historically, there has been a great disparity between various estimates of methane locked up as hydrates in ocean sediments. This can be explained by the fact that previous estimates had assumed that the gas hydrate-bearing sediments contained massive hydrates, whereas sampling has shown that hydrates normally occur as small, finely dispersed crystals and aggregates within the sediments; average concentrations may be as little as 3% by volume – more like the ore of a metallic mineral than a traditional hydrocarbon reservoir. But integrated over the world's oceans and Arctic tundras, the total methane stored in gas hydrates is about 10^{19} g, i.e., about 10 Mt, which represents an order of magnitude more carbon than exists in the biosphere, according to Nisbet.

In the Outer Niger Delta, potential BSR-associated gas hydrates have been considered for production

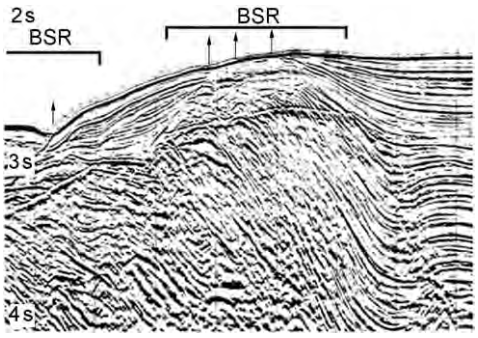


Figure 4 The bottom simulating reflection (BSR): Four of the main BSR characteristics are evident on this 2D-seismic record from off Nigeria: (1) depth below the seafloor is more or less constant, (2) the BSR is anomalously strong, (3) the BSR cross-cuts bedding planes, and (4) the BSR is of reverse polarity with respect to the seafloor reflection. These four aspects, respectively, mean that the BSR (1) is temperature dependent rather than pressure dependent, (2) is caused by a relatively large impedance contrast (i.e., jump in density and/or sound velocity with depth), and (3) is fluid/gas related (i.e., diagenetic rather than lithologically related), and that (4) the P waves reflect from a boundary with an inverse impedance contrast (i.e., where the density is reduced and/or the material's sound velocity is reduced with depth).

(Figures 4 and 5). However, their suspected dispersed nature suggests that they could be extracted only by open-pit mining, a method not currently viable, because the water depth is between 1000 and 2000 m! But attempts to find commercial oceanic gas hydrates have not yet been abandoned. There is hope of finding thick deposits of sandy turbidites containing massive methane hydrates at a 900-m water depth in the Nankai Trough off the coast of Japan. Japanese scientists and the Ministry of Economy, Trade, and Industry (METI) will be going ahead with a three-dimensional seismic acquisition and a full-scale exploration drilling campaign in the Nankai Trough over the next few years.

Oceanic methane hydrate deposits may be of strategic interest. Even if the growth of fossil-fuel consumption continues to slow in Europe, North America, and Japan as conservation and alternative energy sources come to play larger roles, the rest of the world inevitably will require much more energy than it now uses. Creating the means to exploit gas hydrates could be a matter of national security. This is why Japan has long devoted a substantial effort to studying natural-gas hydrates, including the drilling and analysis of a number of test wells. Strategic planning also accounts for the rising interest in hydrates within the US Department of Energy.

Several years ago, Japanese researchers teamed up with Canadian and US scientists to explore commercial exploitation of gas hydrates by drilling through

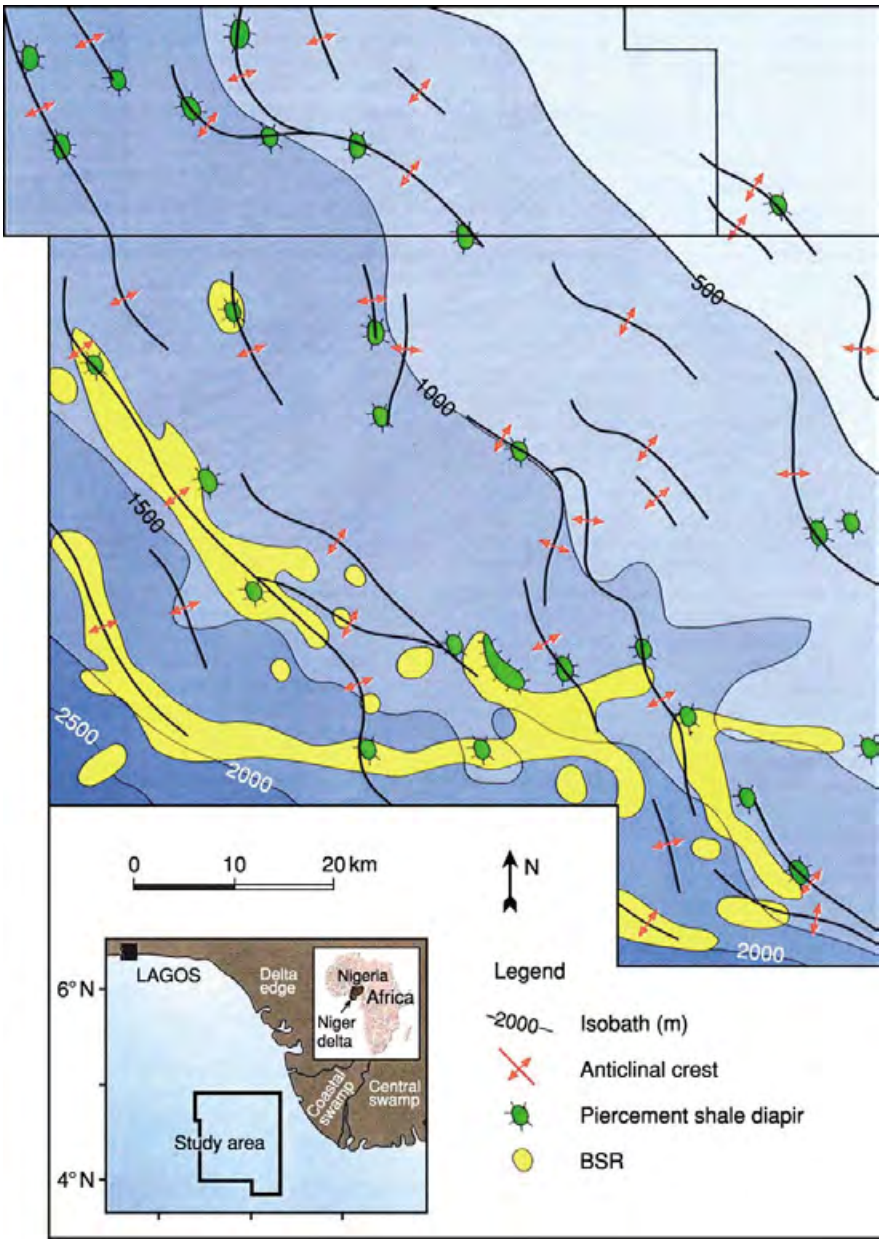


Figure 5 Map showing an area in the Outer Niger Delta Front where resource estimates of gas hydrates have been made. The yellow regions define areas with bottom-simulating reflections (BSRs), i.e., strong gas hydrate indicators. The green circular features represent mud volcanoes, where fluids and mud erupt on the seafloor and build low-lying sediments cones. Depth contours are in metres. Note that the BSRs occur only in water depths greater than about 900 m.

known, thick hydrate occurrences in Arctic Canada, the so-called Mallik well. In 2001, this project was expanded into a multinational campaign, including personnel from India and Germany. They have investigated both warm water circulation and pressure release as production means of freeing up the subsurface hydrate-locked gas. In 2002, Adam wrote that

> Those involved said that they were encouraged by the amount of methane produced – enough to ignite a flare similar to those seen burning over oil rigs. But whether the yellow flame is merely symbolic or a genuine step forwards remains to be seen. The researchers are confident that they will get more energy out than they put in, but not [enough] to pipe it to a power station.

The most promising areas to look for available gas hydrate-associated energy are probably at active deep-water mud volcanoes. Large volumes ($\sim$25 Mt year^{-1}) of gas escape from deep-water mud volcanoes, suggesting that global gas flux from the

seafloor may be underestimated. The European Community is currently investing money into research on gas hydrates in order to assess the potential benefits of or hazards posed by these resources. As part of a European research programme (HYDRATECH), a specially designed laboratory apparatus (a resonant column) is to be constructed to simulate the conditions that allow gas hydrates to be grown artificially. The resonant column will also be used in experiments conducted to characterize the seismic signature of sediments containing gas hydrates. The results of the experiments will allow scientists to better identify and estimate the uses of or threats posed by gas hydrates. However, because even one of the world's best mapped and documented seabed areas still keeps its geological secrets well hidden, the race for exploitable hydrates may take a long time. As expressed by Lorenson and co-workers in 2002:

> Despite years of geophysical prospecting for hydrocarbons, the spatial and vertical distribution of deep gas hydrate in the Gulf of Mexico is not well known. This uncertainty hinders both the determination of the economic potential of gas hydrate resources in the region . . . and their potential as a geohazard.

Gas Hydrates as a Geohazard and Climate Mediator

Extensive data collection efforts on the Louisiana continental slope in the Gulf of Mexico have demonstrated that ambient gas hydrates in sediments are sensitive to small temperature and pressure changes. A temperature increase in the water column as small as 1.0–1.5°C caused distinct dissociation of the gas hydrate, leading to outgassing. Smaller fluctuations in fluid flow were also recorded at the tidal frequency, implying that very small perturbations in the ambient pressure conditions may lead to dissociation. The assessment of a potential drilling site or deep-water construction site, including potential pipeline routes, therefore, has to include an evaluation of the possible presence and formation of gas hydrates. It is therefore necessary to use and interpret all of the available signs of gas hydrate presence, including indirect means, such as interpretation of seismic, sonar, and topographic features, and direct ground-truthing means, such as visual observation and seabed sampling. Ground truth is required to calibrate wide-area surveys and derived results. Within a large dataset, small areas must be examined at the highest resolution and geotechnical standards. The interpretation of the wide-area dataset can then be constrained and the data made relevant within the whole of the wide-area survey. In addition, small areas that have been selected for exploration drilling, or because they constitute a representative seafloor, require optical inspection using video cameras; inspection should be made over relatively wide areas and for every sampling point.

An assessment of potential gas hydrates should also go hand in hand with a theoretical consideration and analysis. Because the regional and local diffusive and focused flux of light hydrocarbons through the sediments is of such importance, it may also be important to quantify the hydrocarbon flux from the seafloor into the ocean, and because of the dynamic nature of the hydrates, long-term monitoring of subcropping or outcropping hydrates could be useful at certain deep-water construction sites. Experiments for performing such studies in the Gulf of Mexico are being planned.

Recent interpretation of high-resolution seismic records of seafloor sediments on the Cascadia margin, off the coast of Vancouver Island, Canada, suggests a periodic disintegration of oceanic gas hydrates. Thus, vertically oriented acoustically anomalous 'chimneys', also called 'wipeouts', which are perturbations penetrating through the GHSZ, indicate fine-scale fluid flux through the upper seafloor. It has been suggested that these vertical intrusions are associated with upward flow of warmer fluids. Therefore, where seafloor fluid expulsion and methane hydrate deposits coincide, the base of the hydrate stability zone might exhibit significant roughness and increased surface area. Increased area implies that significantly more methane hydrate is close to being unstable and hence is closer to dissociation in the event of a lowering of pressure due to sea-level fall. On the Norwegian continental slope, north-east of the Storegga slide scarp, similar features are common on multichannel and high-resolution seismic records. In this region, many of the columnar 'chimneys' penetrate up to the seabed surface and are clearly associated with large pockmarks containing heaps of methane-derived limestone blocks.

'The clathrate gun' hypothesis, developed by Kennett and co-workers, is based on sediment sampling in the Santa Barbara Basin during ODP Leg 146B. Abrupt changes in the $\delta^{13}C$ values for marine organisms were observed in the samples, and it was speculated that the reason was abruptly changing methane contents in the atmosphere and oceanic water due to the dissociation of gas hydrates in ocean sediments. In association with the publication of the hypothesis, James Kennett made the following conclusions on this matter:

> Although no single, unequivocal proof exists for the hypothesis we present, a broad range of evidence suggests it. We expect reluctance by scientists to support such an

idea at this stage without further testing, but if the hypothesis holds, it will require a major shift in thinking about what drives Quaternary climate change.

However, it seems Kennett does not need to worry about reluctant scientists, because recent geological and palaeoclimatic results support the notion that methane released from oceanic gas hydrates is the main culprit in rapid warming events in the past. Smith and co-workers, for example, in 2003, reported suspected massive methane release during deglaciation, only about 14 000 years ago. Their conclusions are based on high-resolution stable isotope records from marine cores on the East Greenland shelf; these records exhibit three rapid light $\delta^{13}C$ events (-3 to -7%) in benthic and planktic foraminifera that are spatially and temporally transgressive. Smith and colleagues invoke that a combination of pressure release (due to ice melting) and temperature increase (associated with ice shelf retreat) has caused large masses of oceanic methane hydrates to dissociate suddenly. A similar but much more severe episode occurred about 55 million years ago, during the Paleocene/Eocene thermal maximum (PETM) global warming event. Based on seabed coring near Antarctica, researchers concluded that there must have been a massive release of carbon-12 into the ocean and atmosphere – an injection that must have occurred because of methane hydrate destabilization.

The main challenge now is to find out what triggers such destabilization. There are several possibilities, including ocean warming ('melting' of oceanic hydrates), lowering of mean sea-level and submarine landslides (depressurizing oceanic hydrates), and large impact events, all of which are equally difficult to document. Perhaps the triggering mechanism is quite feeble, such as a very slight warming or a very slight lowering of sea-level, but at some point, the balance is tipped, leading to a cascade of events. It is known, for example, that there are thousands of submarine mud volcanoes located on the ocean floor. These are known to represent 'pressure release valves' that may suddenly erupt violently, if disturbed by earthquakes, etc. The most dramatic and freak observation of an underwater mud volcano blowout (violent eruption) occurred in the Caspian Sea, on 18 November 1958. By chance, the event was photographed from the city of Baku, 20 km away from the eruption site, the Bachar Bank mud volcano, located at a water depth of 150 m. The flaming jet of gases at first shot several km into the sky. Later, it was estimated to stabilize at a height of 500 m and was topped by a large, 120-m-long horizontal flame. Judging from these dimensions, the jet must have been issuing at supersonic velocity from the sea surface, and the pressure must have been released directly from a depth several thousand metres below the seabed. It is known that the sea-level of the Caspian Sea is about 28 m lower than mean sea-level (MSL). Furthermore, the Caspian Sea drains the Volga River system and is known to adjust its level over long time-spans of 20 to 40 years, in response to suspected long-period, gradual, and cyclic climate changes. The Caspian Sea experienced a high-stand in the late 1930s, followed by a low-stand in 1974, which was followed by a new high-stand in 1996. Therefore, the triggering of the Bachar Bank natural blowout event could possibly have been the gradual (feeble) lowering of the Caspian sea-level. Although mud volcanoes are associated with gas hydrates, their formation mechanism is generally poorly understood.

All-in-all, it may be concluded that sediment-hosted gas hydrates and free gas located below capping layers of hydrates are dynamic, interesting, natural substances that may promise future wealth, but also wrath – if not treated with respect.

See Also

Atmosphere Evolution. **Economic Geology**. **Engineering Geology:** Natural and Anthropogenic Geohazards. **North America:** Atlantic Margin. **Palaeoclimates**. **Petroleum Geology:** Exploration.

Further Reading

Adam D (2002) Fire from ice. *Nature* 415: 913–914.

Austvik T, Li X, and Gjertsen LH (1999) *Hydrate Plug Properties – Formation and Removal of Plugs*. Third International Conference on Natural Gas Hydrates, Salt Lake City, July, 1999.

Dallimore SR, Uchida T, and Collett TS (1999) Summary. In: Dallimore SR, Uchida T, and Collett TS (eds.) *Scientific Results from JAPEX/JNOC/GSC Mallik 2L-38 Gas Hydrate Research Well, Mackenzie Delta, Northwest Territories, Geological Survey of Canada Bulletin,* vol. 544, pp. 1–10. St. John's, NL: Geological Survey of Newfoundland and Labrador.

Ginsburg GD and Soloviev VA (1998) *Submarine Gas Hydrates*. St. Petersburg, Russia: VNIIOkeangoeolia.

Henriet J-P and Mienert J (1998) *Gas Hydrates: Relevance to World Margin Stability and Climate Change. Geological Society Special Publication No. 137*. London: Geological Society of London.

Hovland M, Gallagher JW, Clennell MB, and Lekvam K (1997) Gas hydrate and free gas volumes in marine sediments: Example from the Niger Delta front. *Marine and Petroleum Geology* 14(3): 245–255.

Hovland M and Gudmestad OT (2001) Potential influence of gas hydrates on seabed installations. In: Paull CK and Dillon WP (eds.) *Natural Gas Hydrates, American*

Geophysical Union Monograph, vol. 124, pp. 300–309. Washington, DC: American Geophysical Union.
Kennett J, Cannariato KG, Hendy IL, and Behl RJ (2003) Methane hydrates in Quaternary climate change: The clathrate gun hypothesis. *American Geophysical Union Special Publication 54.* Washington, DC: American Geophysical Union.
Kleinberg RL and Brewer PG (2001) Probing gas hydrate deposits. *American Scientist* 89: 244–251.
Kvenvolden KA and Lorenson TD (2001) Introduction. In: Paull CK and Dillon WP (eds.) *Natural Gas Hydrates, American Geophysical Union Monograph,* vol. 124, pp. 5–7. Washington, DC: American Geophysical Union.
Lorenson TD, Winters WJ, Hart PE, and Paull CK (2002) Gas hydrate occurrence in the northern Gulf of Mexico studied with giant piston cores. *EOS, American Geophysical Union Transactions* 83(51): 601, 607.
Mathews M (1986) Logging characteristics of methane hydrate. *The Log Analyst* May–June: 26–63.
Milkov AV and Sassen R (2002) Economic geology of offshore gas hydrate accumulations and provinces. *Marine Petroleum Geolology* 19: 1–11.
Nisbet EG (2002) Have sudden large releases of methane from geological reservoirs occurred since the Last Glacial Maximum, and could such releases occur again? *Philosophical Transactions of the Royal Society of London, Series A* 360: 581–607.
Sloan ED (1998) *Clathrate Hydrates of Natural Gas,* 2nd edn. New York: Marcel Dekker.
Soloviev VA (2002) Global estimation of gas content in submarine gas hydrate accumulations. *Russian Geology and Geophysics* 43: 609–624.
Suess E, Bohrmann G, Greinert J, and Lausch E (1999) Flammable ice. *Scientific American* Nov.: 52–59.
Wood WT, Gettrust JF, Chapman NR, Spence GD, and Hyndman RD (2002) Decreased stability of methane hydrates in marine sediments owing to phase-boundary roughness. *Nature* 420: 656–660.

The Petroleum System

C Cornford, Integrated Geochemical Interpretation Ltd, Bideford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction and Definitions

The term 'petroleum system' implies an understanding of the generation of oil and gas within a source rock, followed by their expulsion, migration, entrapment, and survival to the present day. The petroleum system approach focuses on the 'genesis' of petroleum and hence the 'source rock', rather than merely the current location of oil and gas and hence the 'trap'. This differentiates the petroleum system approach of the 1990s from the 'play' concept of the 1970s and 1980s (Figure 1).

A simple definition of a petroleum system might be 'all commercial oil and gas deriving from a single pod of mature source rock'. With a better understanding of efficiencies and losses in the processes, this definition has developed to 'a petroleum system accounts, in terms of *volumes*, *composition*, and *process efficiencies*, for all commercial oil and gas expelled from a single pod of mature source rock'. What constitutes a 'commercial accumulation' varies with circumstances, such as oil composition, drilling costs, oil prices, political stability, and proximity to market.

Use of the petroleum system concept accepts that commercial oil and gas is generated from organic matter (kerogen) in sedimentary rocks. These rocks are at or have been buried to temperatures in the range 80–165°C (oil) or 145–220°C (gas), and the petroleum is expelled from the source rock, migrates, and then accumulates in traps. The term is inapplicable if an inorganic origin of petroleum is adopted and when exploring for biogenic methane, and unhelpful when exploring for coal-bed methane and shale-gas (gas trapped intimately within seam coal and shales, respectively).

A discussion of petroleum systems can be broken down into the contributing processes, viz:

- Source rock deposition and kerogen accumulation;
- Maturation of source rock upon burial;
- Generation of hydrocarbons in the source rock;
- Expulsion of oil and gas out of the source rock;
- Migration of oil and gas to the reservoir; and
- Alteration of reservoired hydrocarbons during survival to the present day.

These processes are illustrated gastronomically in Figure 2 and discussed in more detail in the subsequent sections.

In mapped terms (Figure 3), the prospect or trap is surrounded by a drainage area (polygon), from which oil generated within and expelled from mature source rock is focussed during migration towards the proposed accumulation. As illustrated, the actual amount of oil or gas generated will be determined by the richness of the source rock (type of organic matter, e.g., marine or terrigenous kerogen) and the maturity level reached (e.g., oil- or gas-mature). To undertake a full volumetric calculation, the source rock yield and thickness is also required.

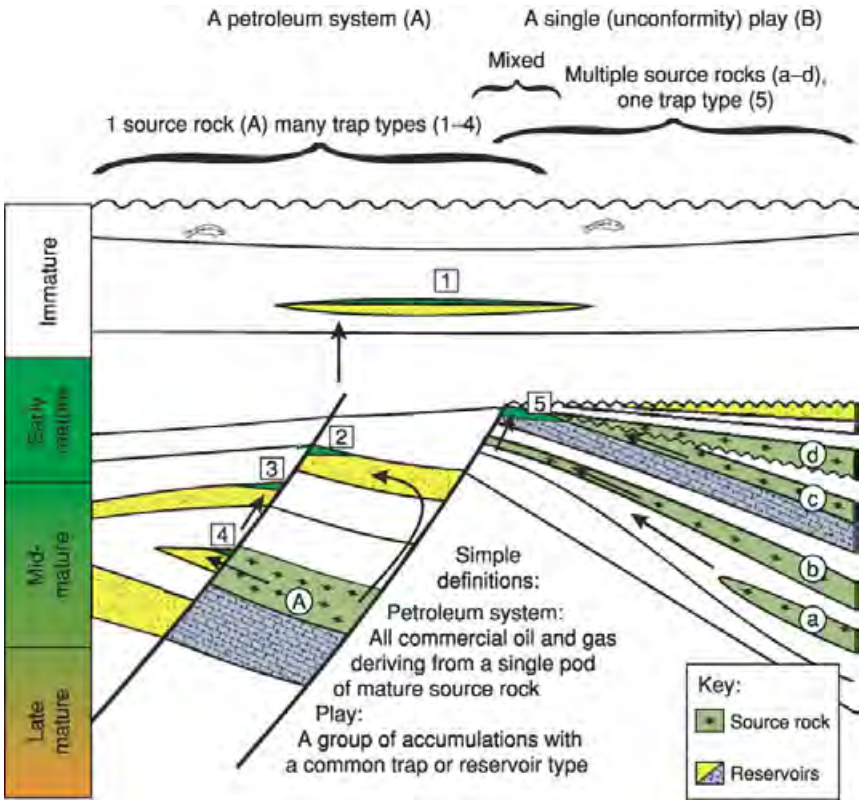


Figure 1 Contrasting the petroleum system with the play concept used for hydrocarbon exploration.

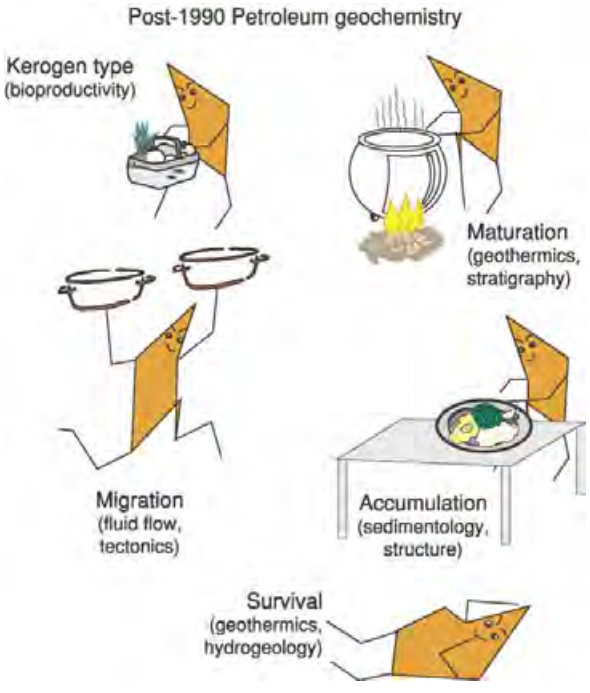


Figure 2 The petroleum system seen as a series of processes starting with (sedimentary) organic matter and resulting in a satisfyingly long-lasting accumulation! Note the proceedings can be seen in terms of geological processes (text labels) or molecular activities (cartoon figures).

Source Rocks

Oil and gas are generated in organic-rich sediments, the source rock. As sediments accumulate, the soft body parts of organisms settle out with the mineral grains and carbonate or siliceous skeletal debris. The soft body parts are dominantly protein, carbohydrate, and fats, and in the case of the higher land plants, lignin. On bacterial attack and initial burial, these tissues condense to form kerogen. Most aquatic or littoral sedimentary environments have an abundant supply of organic matter, but petroleum source rocks require, in addition, a depositional environment promoting the preservation of that organic matter.

Source Rock Deposition

Depositional environments favouring the accumulation of petroleum source rocks are illustrated in Figure 4 in the context of a section through a tectonic plate. The amount of kerogen present in sediments is a balance between bioproductivity, survival, and dilution by inorganic grains. To take some extremes:

Low productivity—Aeolian (wind-blown) desert sandstones where land plant growth is restricted due to the extremely arid environment.

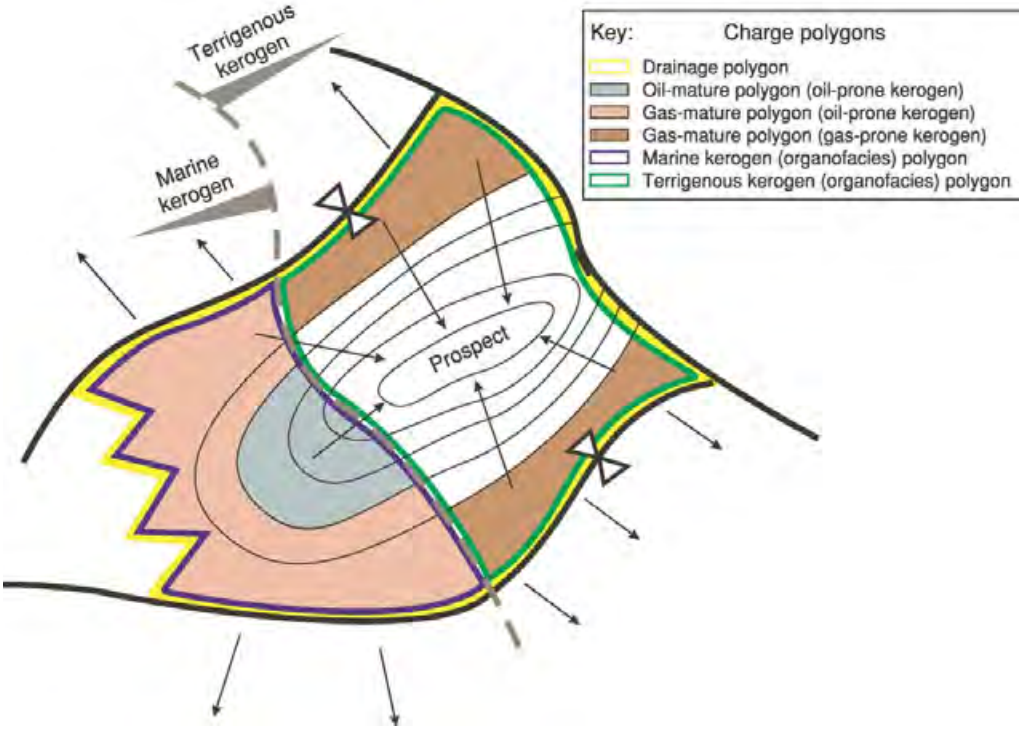


Figure 3 Mapping a trapping structure and the surrounding source rock drainage polygon supplying the oil and gas. The overall drainage area is bounded by faults and synclinal and distal boundaries and contains maturity kitchen- and kerogen-type polygons.

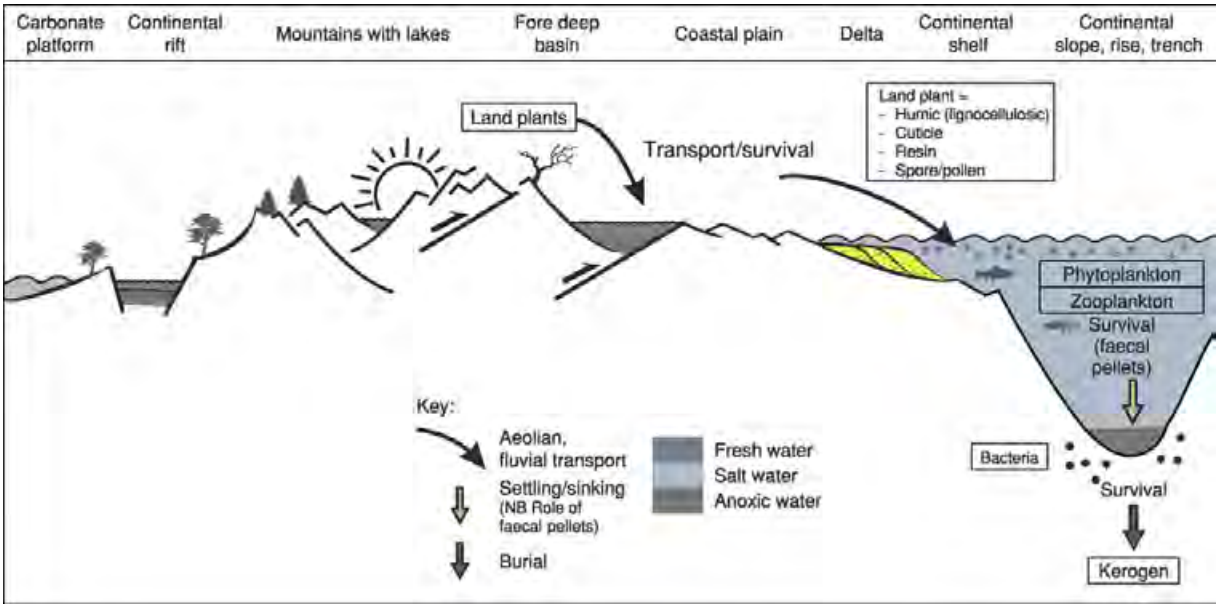


Figure 4 Environments where bioproductivity and depositional environment favour the accumulation of organic-rich sediments.

High productivity—Areas of ocean upwelling where the enhanced supply of nutrients fuels explosive growth of phyto- and zooplankton.

Poor survival—Chalk composed almost exclusively of coccolith skeletal debris (high bioproductivity) but where all the coccolith body tissues are destroyed by bacterial activity under strongly oxic open marine conditions.

Excellent survival—Early rifting phases of oceans where both optimum sedimentation rates and anoxic conditions of the lakes and enclosed seaways promote high rates of organic matter preservation.

Strong dilution—The prolific sediment supply of major deltas produces organic lean delta-front and pro-delta sediments despite high bioproductivity.

Minimal dilution—Coals where the lignocellulosic and other tissues of high-productivity land plants are well preserved in a delta-top environment starved of (or bypassed by) mineral grains.

As illustrated above, the aspects of the depositional environment favouring organic preservation are anoxia and elevated sedimentation rates, though excessive sedimentation rates will eventually lead to dilution.

As implied in Figure 4, organic matter falling to the ocean or lake floor has to pass through various zones of bacterial degradation (Figure 5). The normal open-water oxic bacterial community is highly effective at destroying oil-prone organic matter, so time spent in this zone has disastrous effects on potential oil source rocks. These conditions are found in the water column and surface sediments in the open ocean, but in strongly stratified basins (e.g., the present-day Black Sea) only the top of the water column may be oxic.

Thus to summarize (Figure 6), the optimum oil-prone source rocks are deposited where anoxia develops in an aqueous environment enjoying high rates of sedimentation. The combination of high sedimentation rates and oxic environments favours the accumulation of gas-prone source rocks such as coals. The combination of depositional conditions and kerogen amount and type is termed 'organofacies', as discussed in the next section.

Source Rock Characterization

For a sediment to be considered a potential petroleum source rock it has to contain or have contained an appropriate amount and type of organic matter

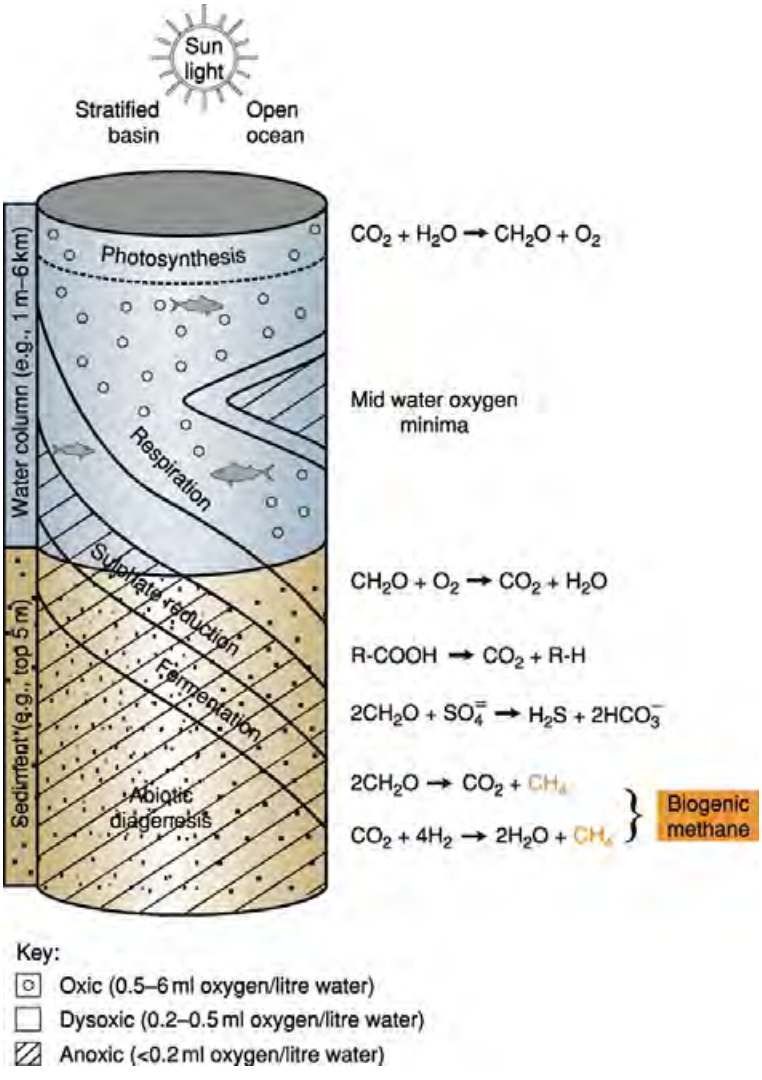


Figure 5 Processes affecting organic matter deposition and preservation.

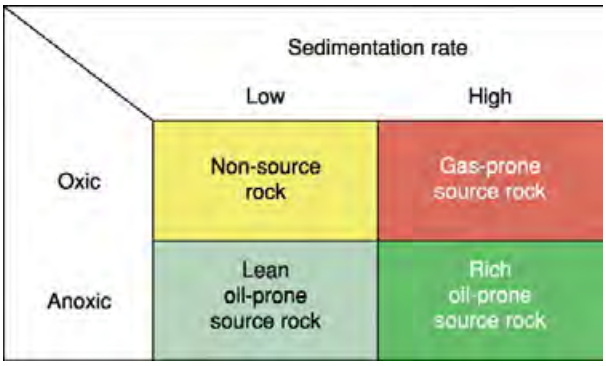


Figure 6 Summary of major processes controlling petroleum source rock accumulation (red = gas-prone; green = oil-prone).

(kerogen). In the same way that the mineral grains of a sediment can be linked to the conditions of deposition via the 'lithofacies', kerogens can be linked to depositional conditions via the 'organofacies'. An organofacies is characterized by a specific combination of organic matter abundance and type.

Abundance of organic matter The amount of kerogen is determined as the total organic carbon (TOC), measured as the weight percentage of the rock (Figure 7). The amount of organic matter can then be described as fair, good, or very good, as suggested in Figure 8. Experience has indicated that carbonate source rocks require lower TOC values to reach a certain level of source potential, probably due to more efficient expulsion of any generated hydrocarbon.

Type of kerogen—optical methods The most graphic way to determine the types of organic matter comprising kerogen is to view kerogen using optical microscopy. This technique is termed 'organic petrography' and is a combination of coal petrographic and palynological techniques and nomenclature. For a full organic petrographic characterization, three types of microscopy are needed:

- Reflected light (also used for vitrinite reflectance);
- Transmitted white light (also used for palynology); and
- Ultraviolet (UV) excitation fluorescence.

The reflected light technique is also the basis for the common maturity parameter, vitrinite reflectance, whereas kerogen colour in transmitted light and fluorescence colour can also contribute to maturity estimations (see next section).

A view of the same area of a bituminous coal thin section using all three methods is shown in Figure 9 with the logic for classification shown in Figure 10.

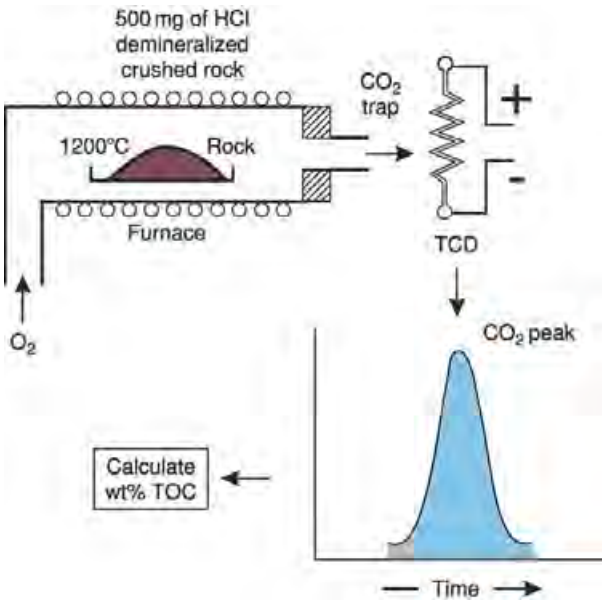


Figure 7 Schematic of a carbon analyser used to determine total organic carbon (TOC wt%) by burning acid-demineralized rock powder in oxygen in a furnace at 1200°C, and determining the amount of carbon dioxide (CO_2) liberated.

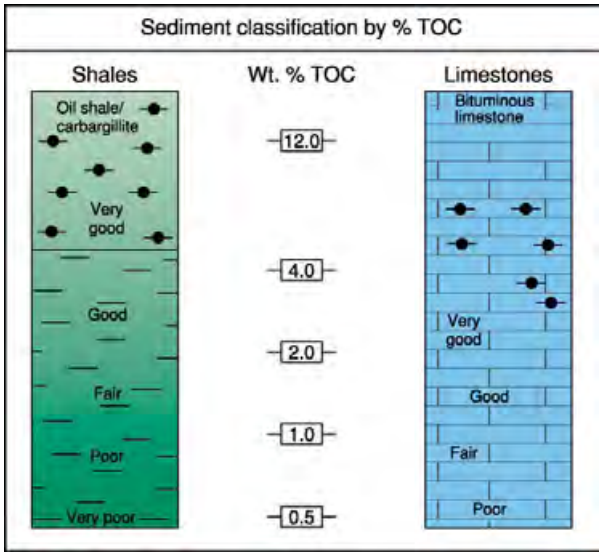


Figure 8 Quality of petroleum source rocks based on amount (but not type) of organic matter.

Discrimination using fluorescence can only be made for relatively shallowly buried sediments (e.g., <4 km burial) where the kerogen is in the immature to mid oil-mature levels of maturation. From Figure 9 it can be seen that this coal contains oil-prone spores and cuticle, gas-prone lignocellulosic vitrinite, together with inertinite (fungal sclerotinite and semifusinite)

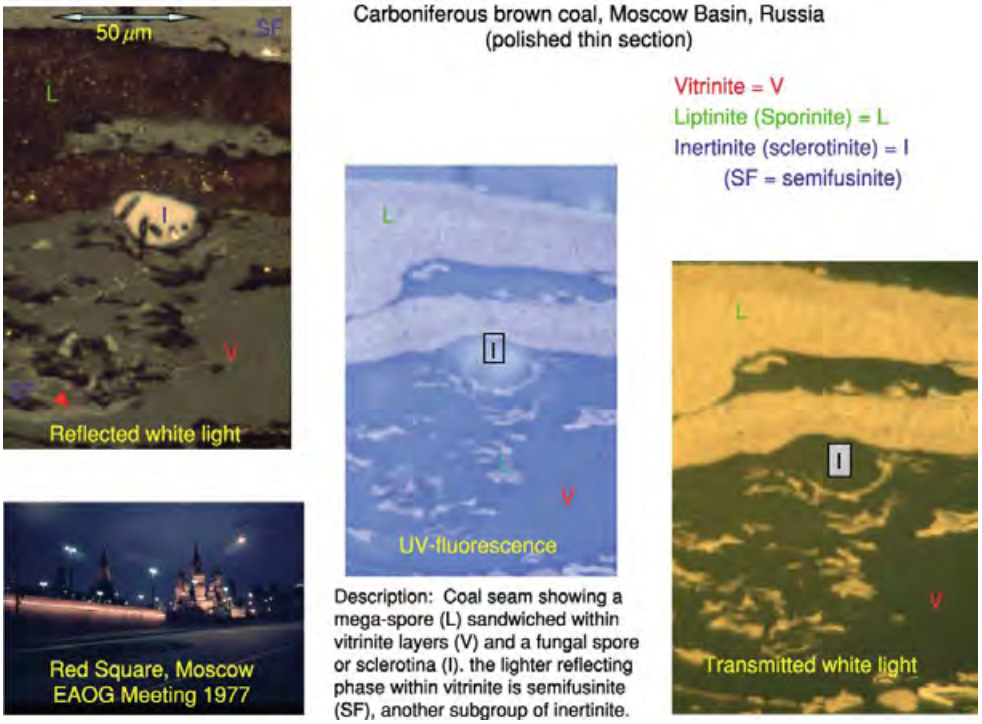


Figure 9 Reflected, transmitted, and fluorescent light views of a polished thin section of Russian carboniferous coal showing oil-prone lipinite, gas-prone vitrinite, together with inert kerogen macerals (inertinites).

with little or no hydrocarbon potential. The inert material is the result of oxidizing land plant material, this mainly being fossil charcoal resulting from natural forest fires and spontaneous combustion of humus.

The combination of these microscopy techniques allows the visible particulate organic matter to be classified into three groups:

- Oil-prone liptinites (fluorescent at low maturity and low reflectance);
- Gas-prone vitrinites (nonfluorescent and medium reflectance); and
- Inert or dead carbon (nonfluorescent and high reflectance at all levels of maturity).

The decision tree used by the organic petrographer is illustrated in Figure 10, where the initial distinction between fluorescent oil-prone liptinites and the nonfluorescent components is followed by high and low reflectance to split the latter into gas-prone and inert particles. Occasionally some vitrinites can exhibit low levels of flourescence. Some further breakdown of the three primary groups is also shown.

Visual descriptions (or a photograph) of a kerogen provides an excellent overview of the large variety of plant and animal tissues preserved in the sediment. However, this diversity of kerogen types also gives rise to a bewildering range of descriptive terms (Figure 11). Some of the commonest general terms are equivalences with respect to provenance (see comments in right column) and oil and gas potential (left column).

The coal micrographs shown in Figure 9 are unusual in that a polished thin section has been made to allow viewing in reflected and transmitted light together with fluorescence. As illustrated in Figure 12, visual characterization of the kerogen in sedimentary rocks may be undertaken on polished rock surfaces (reflected light and fluorescence) or on kerogen concentrates. The kerogen concentrate is made by removing carbonate with hydrochloric acid (HCl) and quartz and silicates (mica, etc.) with hydrofluoric acid (HF). The kerogen concentrate may still contain heavy minerals and pyrite, which can be removed by centrifuging in a high-density liquid. Some organic matter can be lost, so if a representative kerogen is required this step may be omitted.

Whole rock and kerogen concentrates are contrasted in Figure 13. The advantage of the 'whole rock' method is that the sedimentary context of the kerogen is preserved and the tedious acid digestion process—with associated health and safety issues—is omitted. The disadvantage is the dispersed nature of the kerogen makes some identifications difficult and quantification impossible. In contrast, viewing a representative kerogen concentrate allows for semiquantitative compositional estimates (%liptinite,

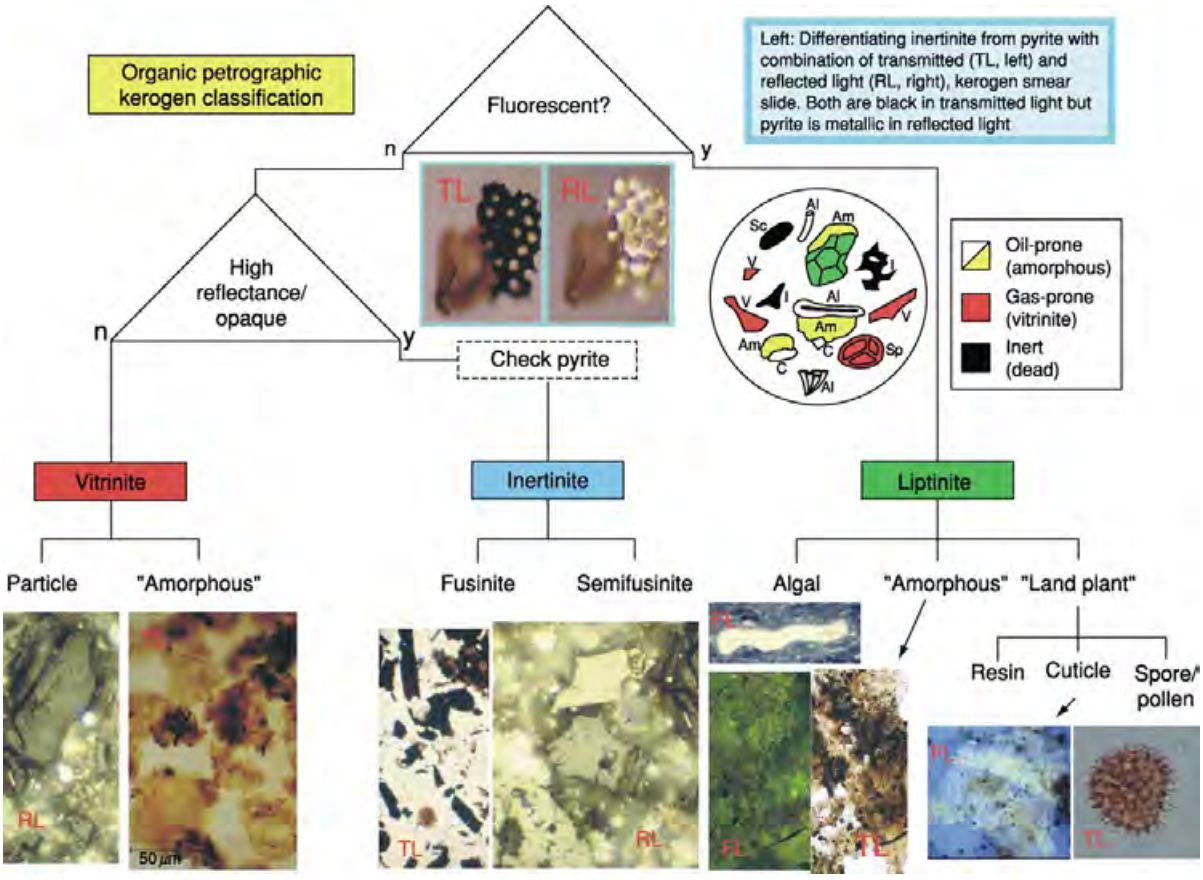


Figure 10 Classification of immature and mature kerogens using reflected (RL), transmitted (TL), and fluorescent light (FL) microscopy (y = yes, n = no).

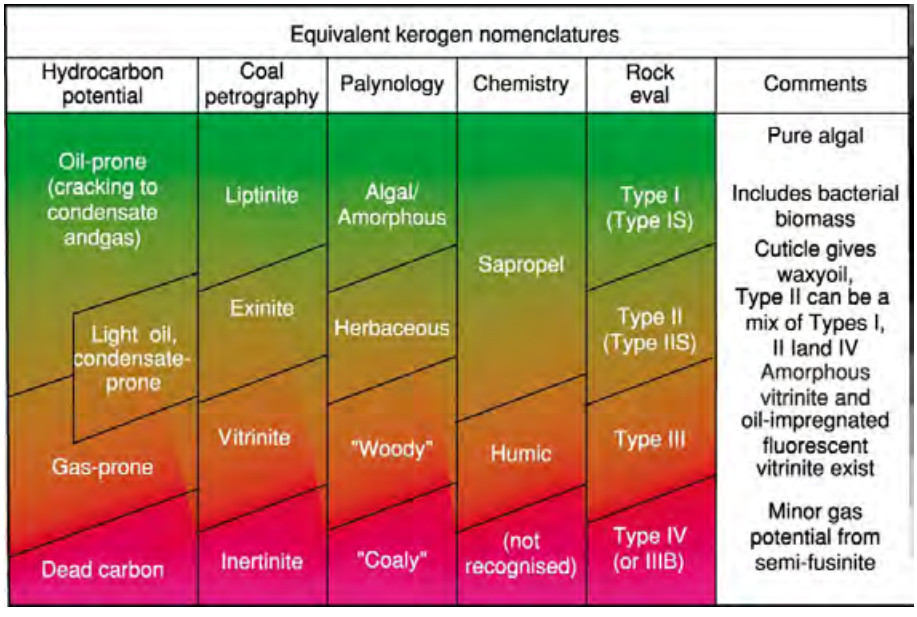


Figure 11 Approximate equivalent terms used for the optical description of kerogen.

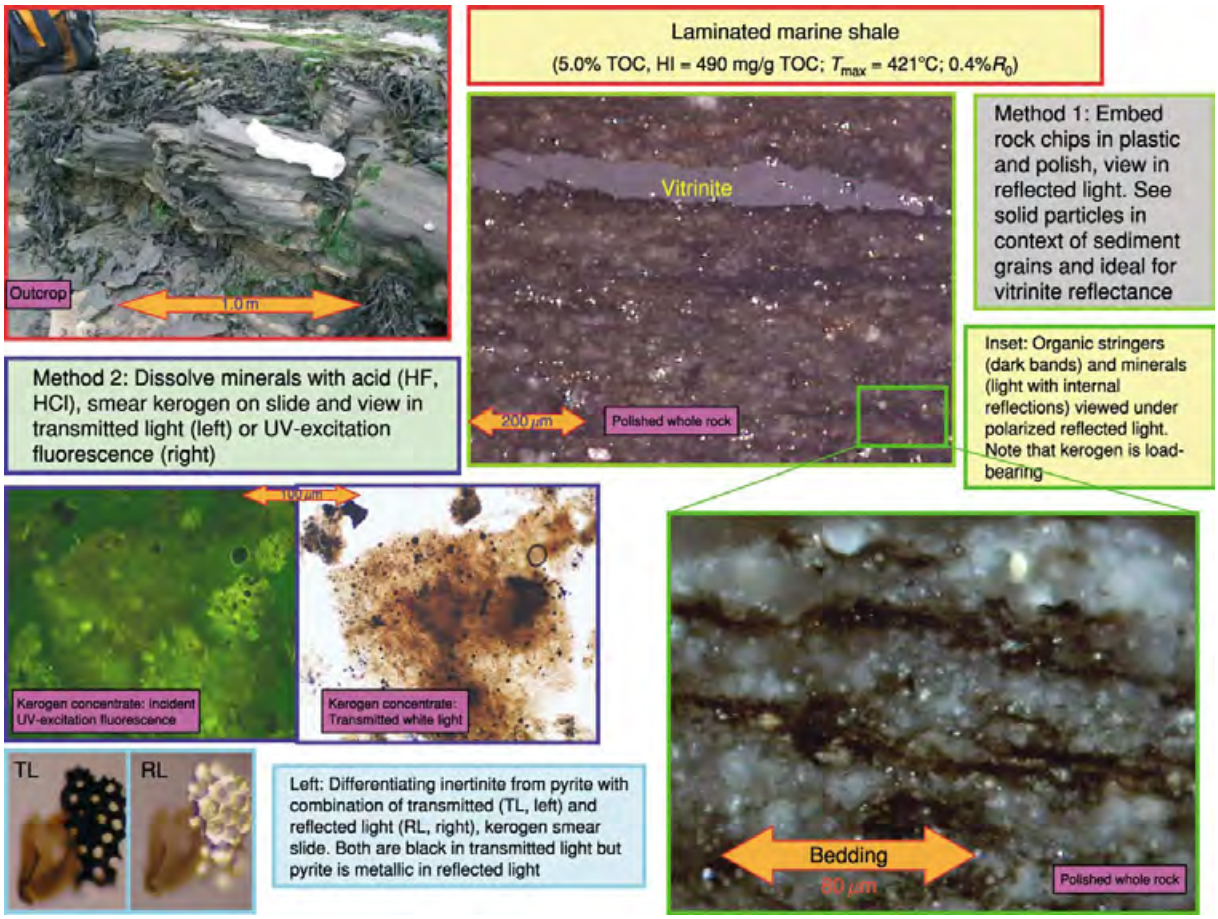


Figure 12 Organic petrographic methods applied to a sampled outcrop (top left) with rock chips polished and viewed in reflected light (right) and kerogen concentrates viewed in transmitted light together with UV-excitation fluorescence (bottom left). Lower Lias oil shales, Kilve, Somerset, UK.

%vitrinite, %inertinite) as the particles are closely packed and may be viewed in all three microscopy modes (Figure 14).

Type of kerogen—pyrolysis methods Pyrolysis, particularly the widespread commercial Rock-Eval family of analysers, rapidly yields 'cheap and cheerful' information concerning kerogen type, maturity, potential yield of hydrocarbon, and the presence of migrated hydrocarbons. Developed by the Belgians and popularized by the French, Rock-Eval pyrolysis has become the most widespread apparatus for geochemical screening of drill samples. Its popularity derives from its modest sample requirements together with its relative cheapness and simplicity (though not necessarily reliability) of interpretation. For most applications, TOC data are required for the interpretation of Rock-Eval data. The recent versions of the Rock-Eval machine do, however, produce an 'organic carbon equivalent' value, arguably dispensing with the need for separate TOC data (*see* **Analytical Methods:** Geochemical Analysis (Including X-Ray)).

The conventional Rock-Eval machine progressively heats (pyrolyses) powdered rock in an inert atmosphere to produce the following sequence of events (Figure 15):

- Release of the 'free' hydrocarbons to give an S_1 peak (to about 300°C).
- Simulation of maturation and generation of new hydrocarbons from the kerogen (300–550°C) to yield a second peak, S_2, with maximum at temperature T_{max} (°C).
- Release of organically bound CO_2 over the temperature range 300–550°C to give an S_3 peak.

The interpretation of pyrolysis results in general and Rock-Eval in particular covers many aspects of petroleum geochemistry. Interpretation is normally undertaken using industry-standard cross-plots. The simplest, termed the pseudo-van Krevelen diagram, plots the oxygen index against the hydrogen index.

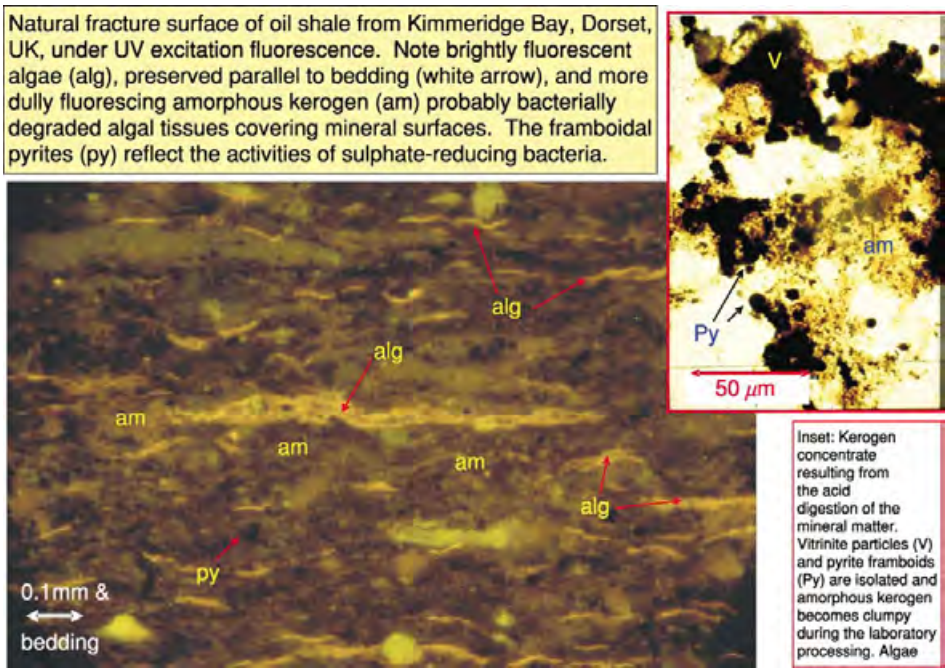


Figure 13 Visual kerogen descriptions can derive from whole rock surfaces viewed in reflected light or fluorescence (main picture) or as a kerogen concentrate in transmitted white light (inset).

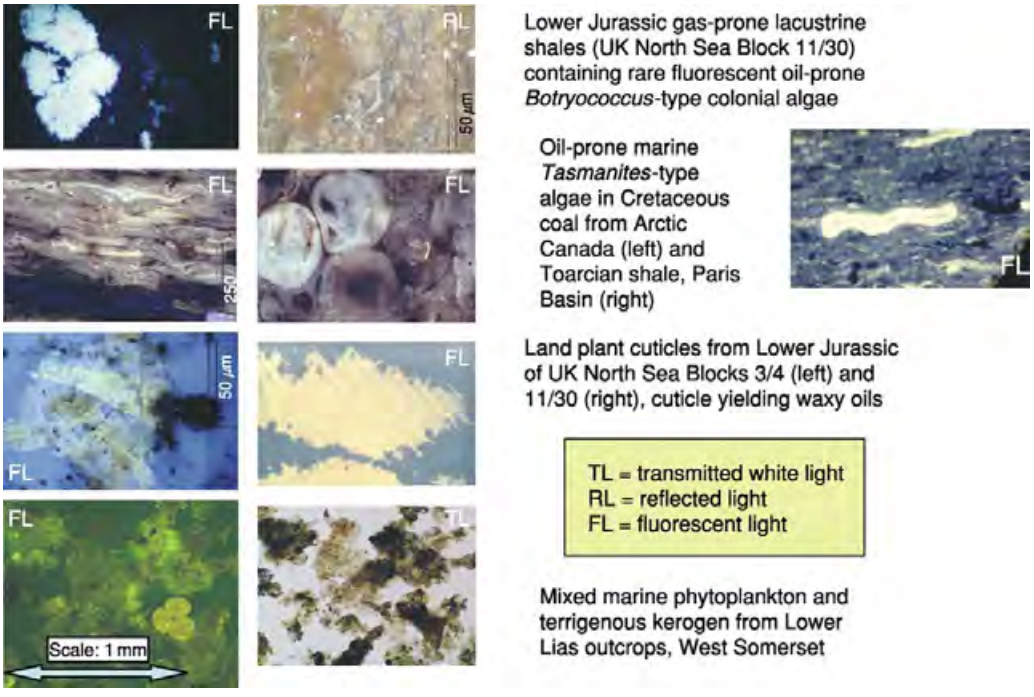


Figure 14 Some typical oil- and gas-prone kerogen associations illustrated as organic petrographic micrographs.

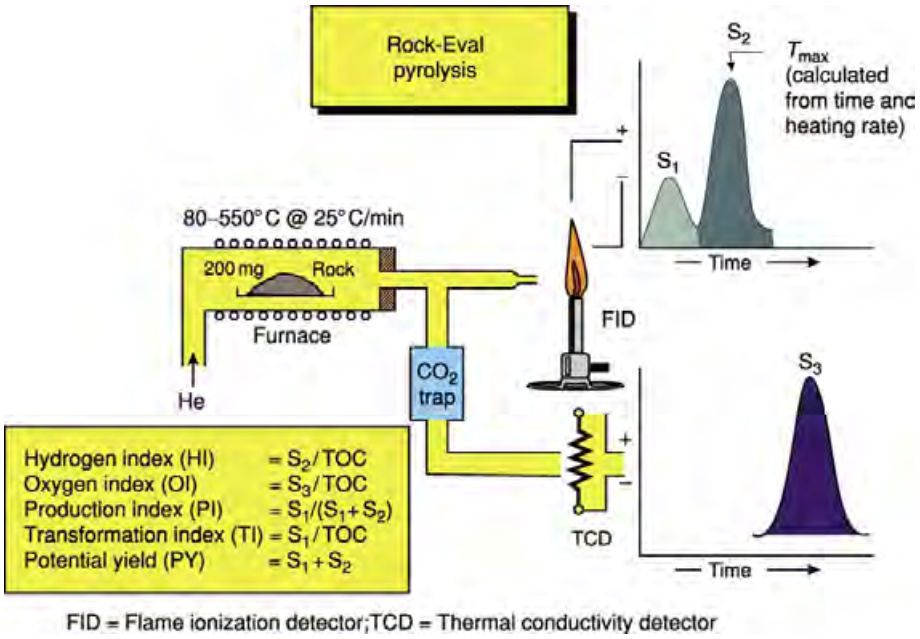


Figure 15 Schematic of the Rock-Eval pyrolysis apparatus, output, and (with TOC) derived parameters.

This is shown in Figure 16, where four main kerogen types can be identified:

- Type I kerogen of algal origin, which is oil-prone;
- Type II kerogen deriving from a number of origins though all are oil-prone:
  - Land plant spores, exines, and resins;
  - Bacterially degraded algal (originally Type I) kerogen; and
  - A mixture of Type I (algal) and Type III (humic) particles.
- Type III kerogen of lignocellulosic or humic origin which is gas-prone;
- Type IIIb or IV kerogen defined as altered (oxidized) humic material or fossil charcoal: it has no oil or gas potential.

The reliability of oxygen indices has been questioned since some carbonates (e.g., iron carbonates) decompose at the temperatures where the 'organic CO_2' is being collected (see Figure 15). High oxygen indices are suspect if deriving from rocks with high carbonate contents.

A more detailed approach to kerogen identification is shown in Figure 17, where a plot of the Rock-Eval S_2 yield versus TOC is used to identify the average percentage of dead carbon (%DC) as the intercept of the best-fit data trend on the TOC axis. An incremental part of this intercept has also been interpreted as resulting from the adsorption of S_2 organics on active mineral surfaces.

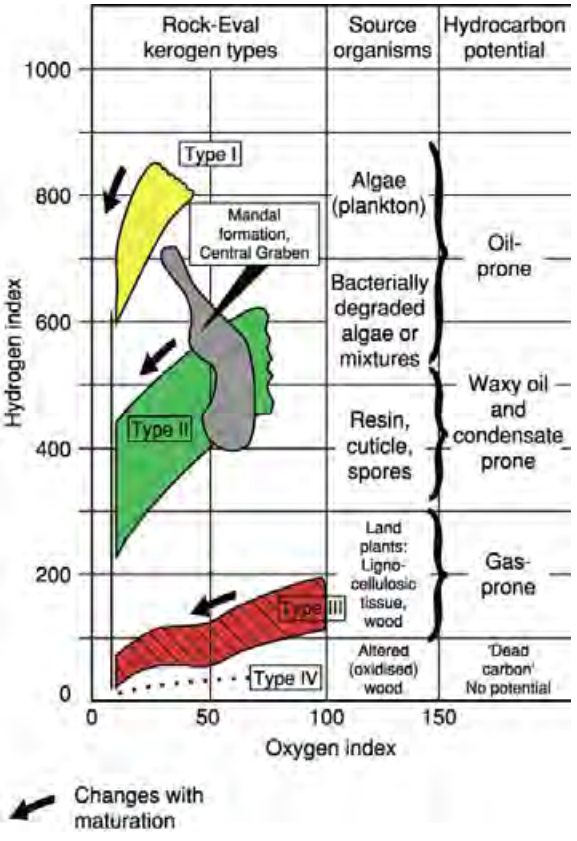


Figure 16 Kerogen type from Rock-Eval hydrogen indices (mg Pyrolate/g TOC) and oxygen indices (mg CO_2/g TOC), with maturity trends indicated by the heavy black arrows.

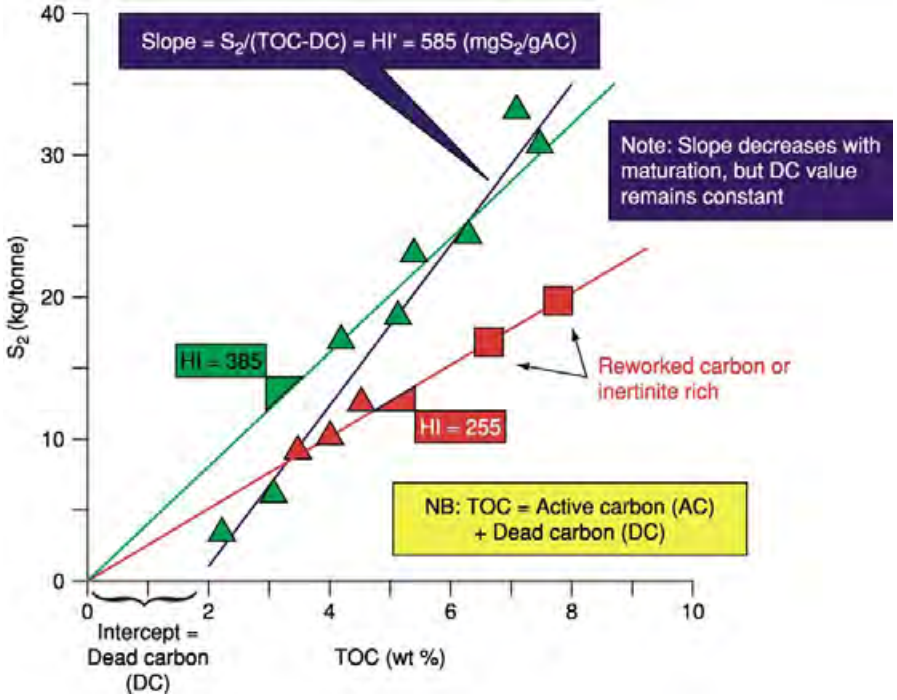


Figure 17 A more detailed determination of kerogen type in immature and early mature kerogens using Rock-Eval S_2 and TOC data where the gradient is hydrogen index (S_2/TOC in units of mg Pyrolysate/g TOC).

Green and red trends (data plus lines) pass through the origin of the graph and define two types of kerogen with HI of 385 and 255 mg g TOC^{-1}, respectively. Ignoring two points, a single blue trend better fits all the values. The TOC can then be considered as the sum of dead carbon (DC) and active carbon (AC):

$$\%\text{TOC} = \%\text{DC} + \%\text{AC}$$

Note %TOC represents a percentage while TOC refers to a ratio.
It follows that the slope of the blue trend in Figure 17 is a modified hydrogen index (HI′), calculated on a dead-carbon-free basis (HI′ = 585 mg S_2/g AC^{-1}):

$$\text{HI}' = 100 \times S_2/(\%\text{TOC} - \%\text{DC}) = 100 \times S_2/\%\text{AC}$$

The above analysis is really only applicable to immature and early oil-mature kerogens (see next section). The determination of kerogen type for mid-, late-, and post-mature kerogens can be addressed with Figure 18, where Rock-Eval T_{max} is the maturity parameter. The typical bacterial-algal Type II kerogen shows a small increase in hydrogen index in the immature stage (up to $T_{max} = 430°C$), followed by a linear decrease, which reflects oil generation up to about $T_{max} = 465°C$. The gas-prone Type III kerogen with a lower initial hydrogen index falls slowly over a larger T_{max} maturity range.

For the highly oil-prone Type I algal kerogen the immature/early mature boundary can be as high as $T_{max} = 440°C$, but once reached the rapid generation of oil produces a swift decrease in hydrogen index. A singular kerogen type is likely to require its own T_{max} calibration: for example, generation from a sulphur-rich Type IIS kerogen produces a reduction in hydrogen index at an anomalously low maturity level. Thus, using the Rock-Eval pyrolysis approach to kerogen typing requires knowledge of source rock maturity.

The pyrolysis S_2 peak can be trapped and passed to a gas chromatograph (GC) to produce a characteristic fingerprint (Figure 19). This pyrolysis-GC fingerprint differentiates oil-prone kerogen dominated by the alkane–alkene doublets, from gas-prone kerogens where the aromatic compounds such as toluene, xylenes, and low-molecular-weight alkanes dominate.

Kerogen type can also be determined somewhat more indirectly from a number of other kerogen properties, such as elemental analysis, stable carbon, and hydrogen isotopes, and the molecular fingerprints from biomarkers, such as steranes and terpanes.

Using these more sophisticated properties a source rock kerogen can be attributed to an organofacies based on water chemistry (fresh, saline, hypersaline),

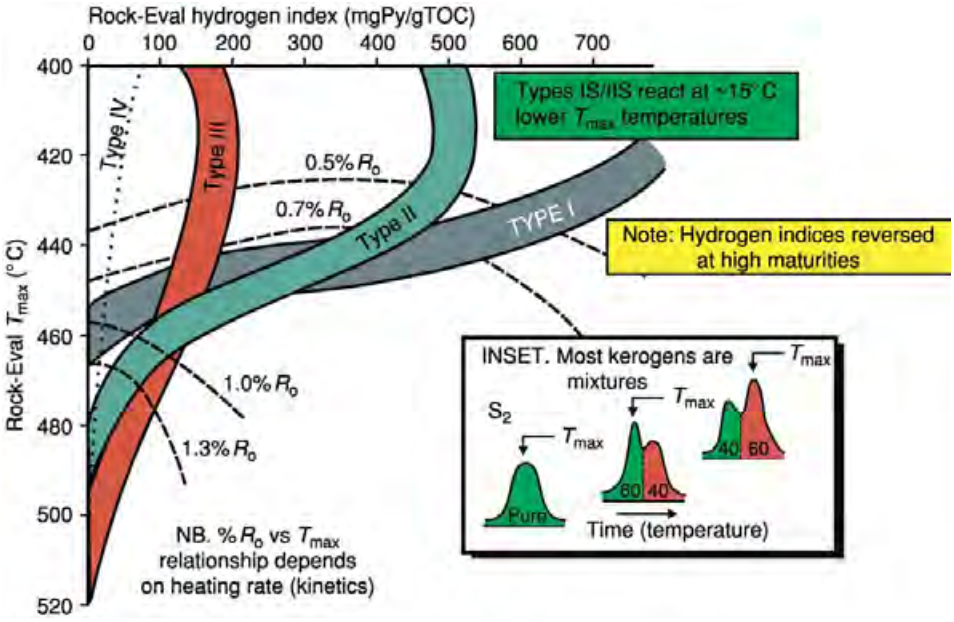


Figure 18 The effect of maturity changes on hydrogen index values for the three major kerogen types.

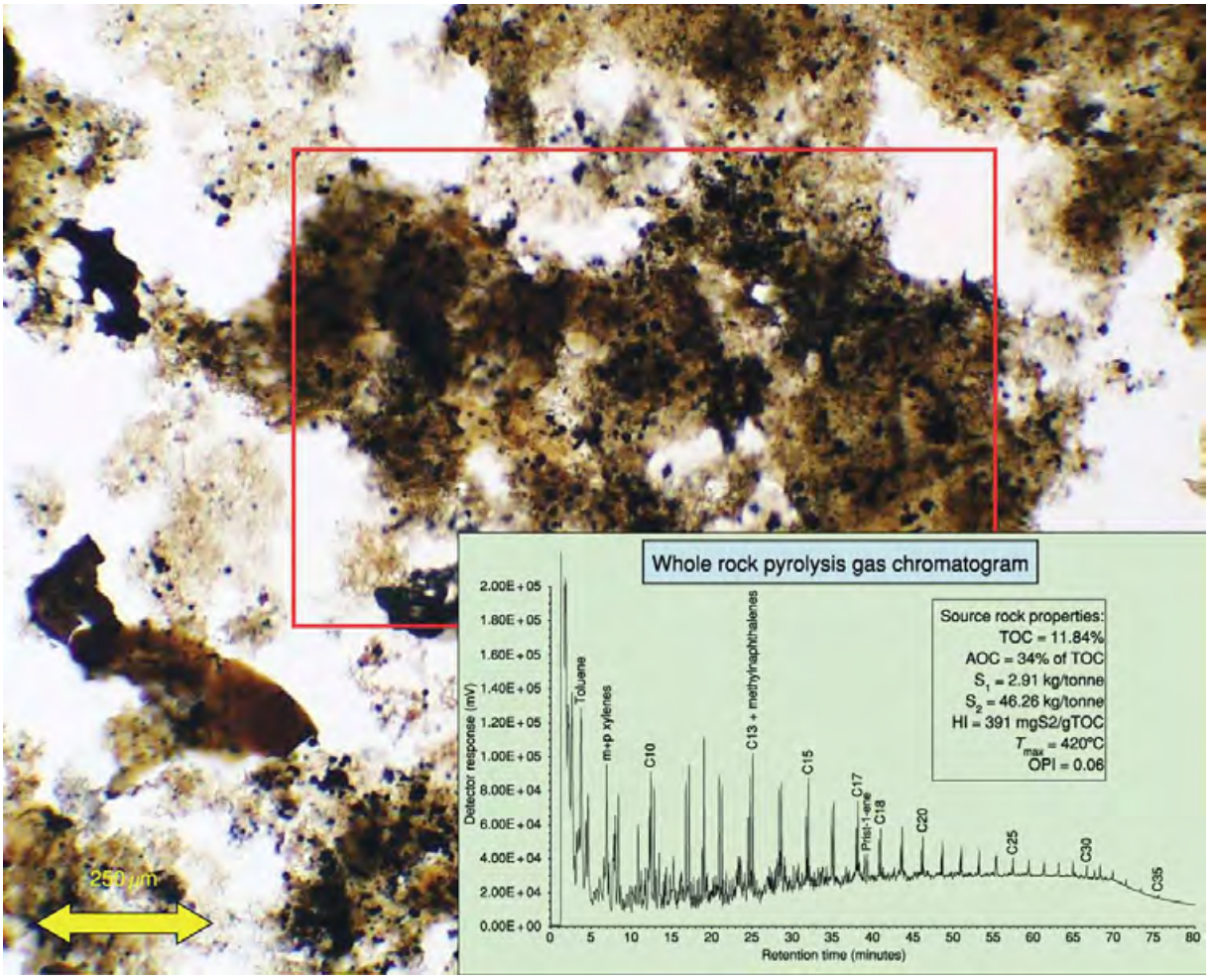


Figure 19 Kerogen typing using pyrolysis-GC fingerprints to confirm the presence of oil-prone kerogen (Lias shale from a well in the UK Western Approaches).

mineralogy (clastics, carbonates), and dissolved oxygen levels (anoxic, dysoxic, oxic) at deposition. In addition the molecular and isotopic signature can place limits on the stratigraphic age and types of organism contributing to the kerogen. The importance of these molecular and isotopic properties is that they are carried with the oil to the trap or surface seep. This allows the organofacies of the source to be estimated from migrated oil and hence address the question of oil–source rock correlation.

Maturation

As any sediment is buried the contained organic matter is matured as a function primarily of increasing temperature, with ancillary effects of time and possibly pressure. Maturity can be related to the stages of hydrocarbon generation and can be used to constrain geological models. In this section maturity is defined in terms of temperature and time, using measured indicators such as Rock-Eval T_{max}, vitrinite reflectance, and spore colour indices. Some common maturity parameters are related to oil generation levels in Table 1 and to gas generation levels in Table 2.

These equivalencies shown in Table 1 and Table 2 are subject to a number of caveats:

- Different responses to temperature and time for different kerogens (kinetics);
- Differences between sampling and laboratory analytical procedures; and
- Different concepts of what constitutes early, mid-, and late maturity.

The recommended procedure is to validate these equivalencies using locally generated data from within the project area.

The most common technique for measuring source rock maturation is vitrinite reflectance, developed for studies of coals in the 1960s and 1970s. It has subsequently been increasingly applied to sediments and derived kerogen concentrates. The reflectance of a material ($\%R$) is a function of its refractive and absorptive indices together with the refractive index of the observational medium (air, water, oil, etc.), as expressed by Beer's equation:

$$\%R = [(n - N)^2 + n^2k^2]/[(n + N)^2 + n^2k^2]$$

where n = the refractive index of the material, N = the refractive index of the measuring medium, and k = the absorptive index of the material.

The basic vitrinite reflectance technique is shown in Figure 20, where the upper left panel illustrates the reflected light polarizing microscope fitted with a photomultiplier detector. The reflectance is measured relative to a glass or crystal standard:

$$\text{vitrinite reflectance} = (\text{PMR of vitrinite}/\text{PMR of standard}) \times \text{reflectance of standard}$$

where PMR = photomultiplier response.

Reflectance, determined orthogonal to a polished surface, is expressed as a percentage of the incident light. It may be measured in air ($\%R_a$, typical of Russian data) or using oil-immersion objectives ($\%R_o$, typical in the West).

The first step for determining vitrinite reflectance is to identify vitrinite in general and the 'right type' of vitrinite (termed telocollinite) in particular (Figure 20A, inset). If all organic particles are measured, the low reflecting liptinite (L) may be differentiated from vitrinites (V) and the high-reflecting inertinites (I) (Figure 20B, upper histogram). This is rarely undertaken. In practice only particles believed

Table 1 Approximate[a] equivalence of maturity parameters with respect to oil generation

Generation stage (oil)	*%R_o (%)*	*T_{max} (°C)*	*SCI (1–10)*	*TAI (1–5)*	*LOM (1–20)*	*User*[a]
Immature						
	0.50	432	4	2.25	7.8	
Early mature						
	0.70	442	5	2.30	8.0	
Mid-mature						
	1.00	455	9	2.60	10.5	
Late mature						
	1.30	465	>9	3.50	12	
Post-mature						

$\%R_o$ = vitrinite reflectance; T_{max} = Rock-Eval T_{max}; SCI = spore colour index; TAI = thermal alteration index; LOM = Shell's level of organic maturation.

[a]To determine your own equivalences, produce your own cross-plots.

Table 2 Approximate[a] equivalence of maturity parameters with respect to gas generation

Generation stage (gas)	*%R_o (%)*	*T_{max} (°C)*	*SCI (1–10)*	*TAI (1–5)*	*LOM (1–20)*	*User*[a]
Immature						
	1.00	455	9	2.6	10.5	
Early mature						
	1.30	465	>9	3.50	12	
Mid-mature						
	2.20	525	10+	4.00	14.5	
Late mature						
	>3.00	>575	10+	5.00	18	
Post-mature (onset greenschist facies metamorphism)						

$\%R_o$ = vitrinite reflectance; T_{max} = Rock-Eval T_{max}; SCI = spore colour index; TAI = thermal alteration index; LOM = Shell's level of organic maturation.

[a]To determine your own equivalencies, produce your own cross-plots.

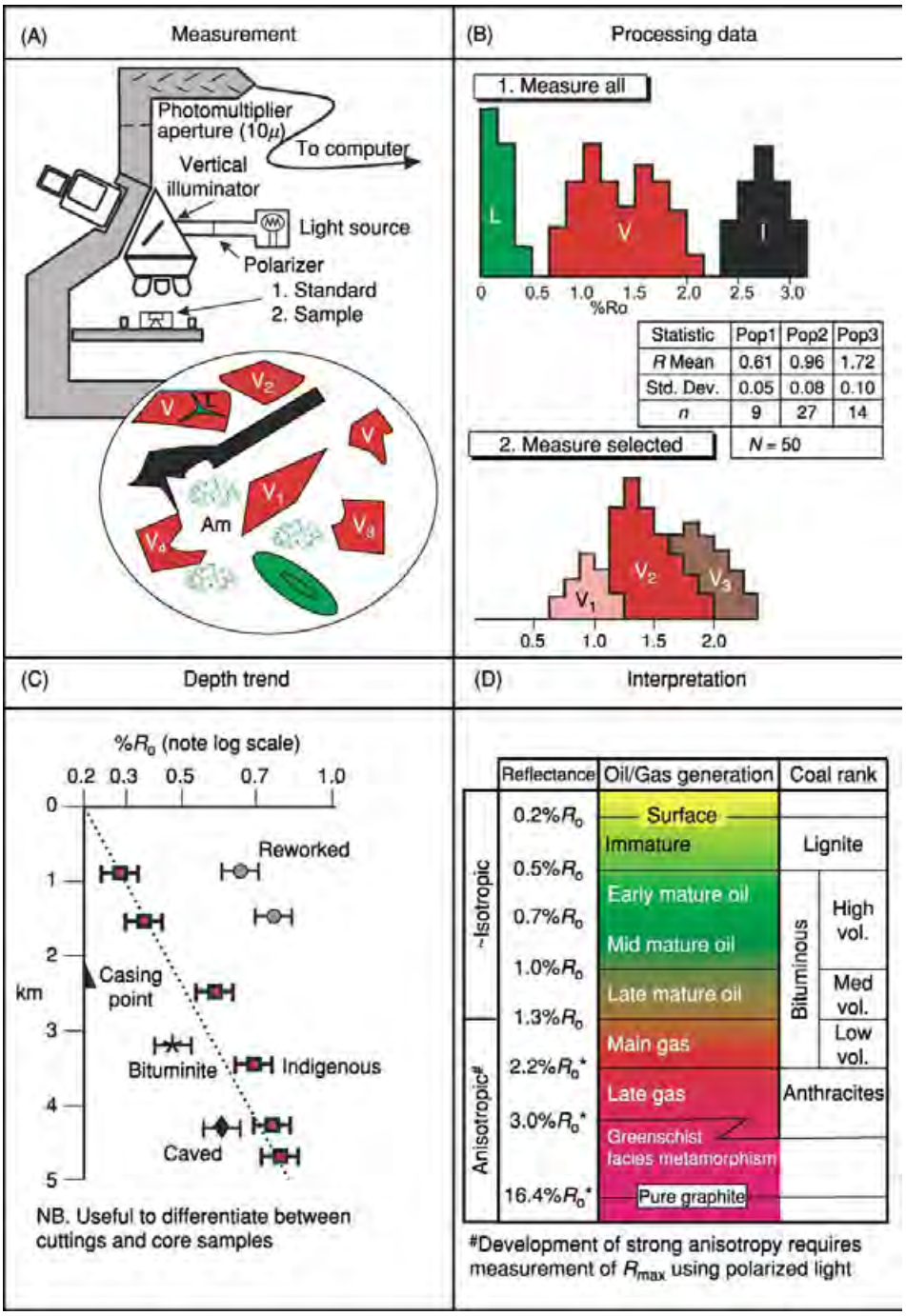


Figure 20 The measurement and interpretation of vitrinite reflectance for determining the maturity of sedimentary rocks in general and oil and gas source rocks in particular.

by the petrographer to be vitrinite are measured (Figure 20B, lower histogram): typically between 20 and 50 particles of vitrinite are measured, and an arithmetic mean value calculated. If multiple populations of vitrinite are recognized, then individual mean values are calculated for each population (Figure 20B, inset table).

If linearly polarized light is used (Figure 20A, see 'polarizer'), rotating the microscope stage and hence the sample with produce a maximum ($\%R_{o,\,max}$) and a minimum ($\%R_{o,min}$) if the material is optically anisotropic. As a result of maturing under overburden pressure, vitrinite develops bedding-parallel anisotropy at the late oil and early gas

generation levels ($\sim 1.0\%R_o$). If the vitrinite is anisotropic and maximum and minimum reflectances are determined, then the mean of the maximum values may be reported ($\%R_{o,\ \text{mean-max}}$). Theoretically the maximum $\%R_{o,\ \text{max}}$ is the most significant indicator of maturity.

The interpretation of the mean vitrinite reflectance values is normally undertaken by plotting $\%R_o$ on a log-scale against sample depth (Figure 20C), where anomalies may be recognized as plotting away from the linear trend. The interpretation then uses the industry-standard relationship with both oil and gas generation windows and coal rank (Figure 20D). Discontinuities in the maturity trend may be seen at unconformities (where the amount of uplift may be estimated from the offset), and excursions may relate to intrusions or hydrothermal flow. As well as defining the maturity of the sampled section, the reflectance trend can be projected to predict—in the absence of major unconformities—maturity levels ahead of the drilling bit.

In addition to vitrinite reflectance two other maturity parameters are commonly reported, Rock-Eval T_{max} (Figure 15), and kerogen or spore colour estimated on TAI (1–5) and SCI (1–10) scales, respectively. Acronyms are defined and equivalences detailed for oil and gas generation in Table 1 and Table 2, respectively. Although the T_{max} values are often available in large numbers, in a single well they often show substantial scatter with respect to depth. In contrast, fewer visual estimates of kerogen colour are determined, but they often produce a better maturity trend with depth. Spore colour estimates are particularly reliable since reworked or caved spores and spores introduced with drilling mud or other contamination can be readily identified as being out of stratigraphic sequence.

Finally the maturity involved in a petroleum system will normally be interpreted in the context of basin modelling and a burial history plot (Figure 21). The burial history plot (time and hence stratigraphy versus depth) traces the burial depth (and temperature) for each modelled strata (black lines) from deposition to the present day, including periods of uplift reflecting erosion at the palaeosurface. Compaction is modelled by various methods as the sediments are progressively buried, and the temperature grid (red lines) calculated from geothermal gradients or mantle heat flow. The heat flow calculation requires the knowledge of the thermal conductivity of each sediment unit—a function of mineralogy and porosity (compaction):

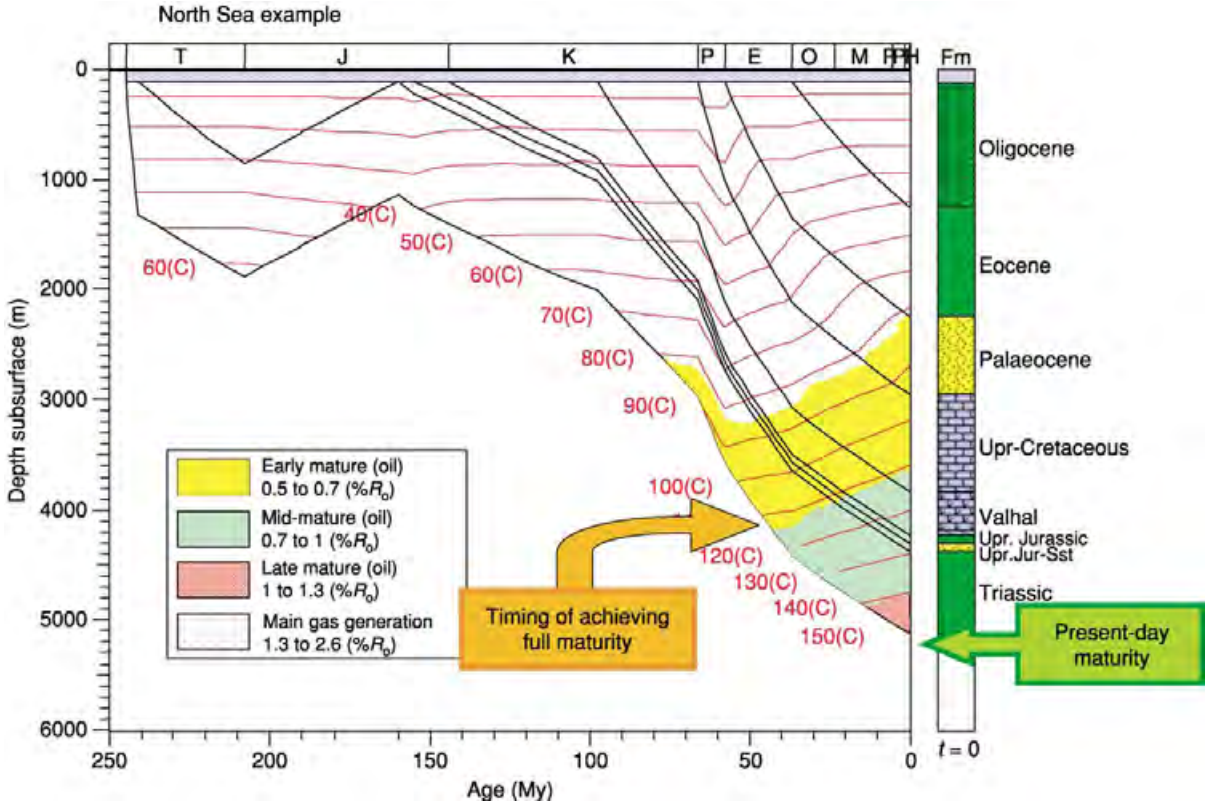


Figure 21 Burial history model showing present-day and palaeotemperature and early-, mid-, and late-mature oil windows defined by vitrinite reflectance.

$$\text{Heat flow}(\text{mW m}^{-2}) = \text{Geothermal gradient}\ (^{\circ}\text{C km}^{-1}) \times \text{Thermal conductivity}\ (\text{W m}^{-1}\,^{\circ}\text{C}^{-1})$$

The calibration step then minimizes the difference between measured and modelled temperature (by adjusting heat flow) and maturity parameters as given in Table 1 (by adjusting uplift, etc.). From knowledge of the temperature history for each stratum, the kinetics of decomposition of the kerogens (i.e., the generation of oil and gas) are modelled using the Arrhenius equation. Once calibrated, the model is used to predict the amount, composition, and timing of hydrocarbon generated as discussed below.

Generation and Expulsion

The process of generation describes the conversion of kerogen to petroleum (oil and gas):

$$\text{Kerogen} \xrightarrow{TR} \text{oil} + \text{gas} + \text{residue}$$

Generation, as opposed to maturation, is measured in terms of the extent of kerogen conversion (transformation ratio, TR) and is used to define the oil and gas windows (shown as yellow, green, and red areas in Figure 21).

The generation process can be monitored by laboratory measurement of the volumes (or masses) of generated oil or gas extracted from source rocks (Figure 22), or liberated during pyrolysis (Figure 23). Although the former is constructed from solvent extraction of the materials within source rocks as a fraction of the TOC (units of milligrams of extract per gram of TOC), the latter is reported as the Rock-Eval pyrolysis production index [$PI = S_1/(S_1 + S_2)$]. The PI is the ratio of the free hydrocarbon (S_1, kg t^{-1}) as a function of the remaining pyrolysate (S_2, kg t^{-1}). In both cases, the onset of generation (early generation) is indicated by an increase in yield. While the early generation phase starts the build-up of petroleum in the source rock, the continuation of the process saturates the source rock, leading to expulsion and eventually the cracking of oil to gas.

Assuming a typical geological heating rate (e.g., 1°C My^{-1}), the following maturity zones can be defined with respect to oil generation, expulsion, and cracking from Type II oil-prone kerogens:

Immature, <80°C (<175°F): No significant oil generation in the source rock: biogenic methane can be generated in recent and shallowly buried sediments.

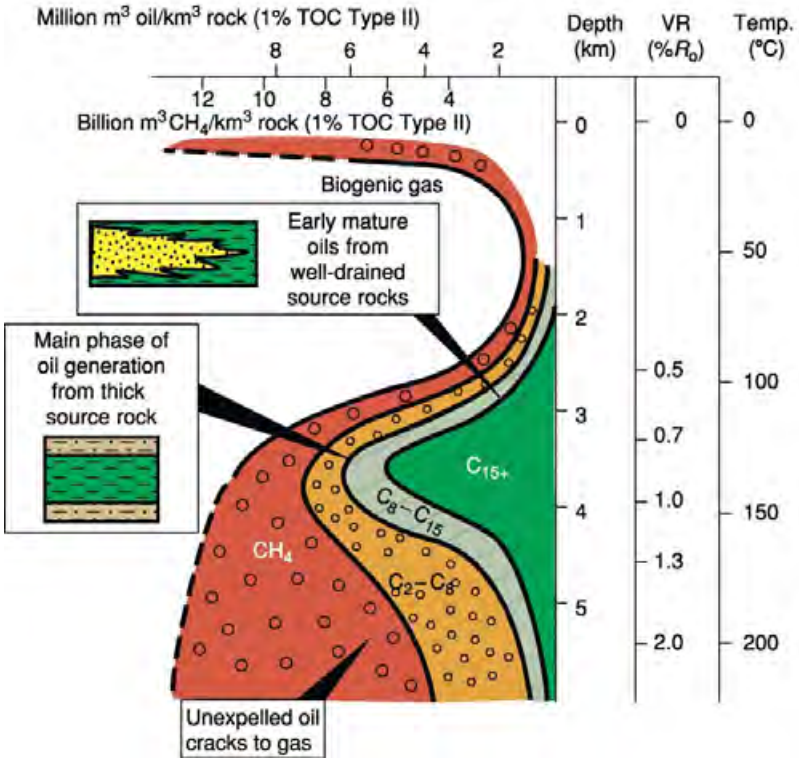


Figure 22 Illustration of generation based on a summation of methane, gas–gasoline (C_2–C_8), light oil (C_8–C_{15}), and C_{15}+ fractions measured in marine source rocks containing 1% TOC of Type II kerogen at different levels of maturity.

Early mature, 80–115°C (175–240°F): Oil starts to generate in and saturate the source rock, hydrocarbons only being expelled in situations promoting highly efficient primary migration (e.g., where fault scarp fan sands, turbidites, or deltaic sands interdigitate with a source rock claystone or coals, or with fractured cherts and carbonates).

Mid-mature, 115–145°C (240–295°F): This is the main phase of oil generation and expulsion. Oil is expelled from 'massive' source rock units with efficiencies eventually in the range 60–80%.

Late mature, 145–160°C (295–330°F): Waning generation associated with cracking of unexpelled oil to give progressively lighter oils and condensate. Continued generation/expulsion phase particularly in poorly drained source rocks.

Post-mature, >165°C (>330°F): No remaining oil generative capacity exists in the source rock kerogen, and unexpelled or reservoired oil is cracking down to condensate and wet gas.

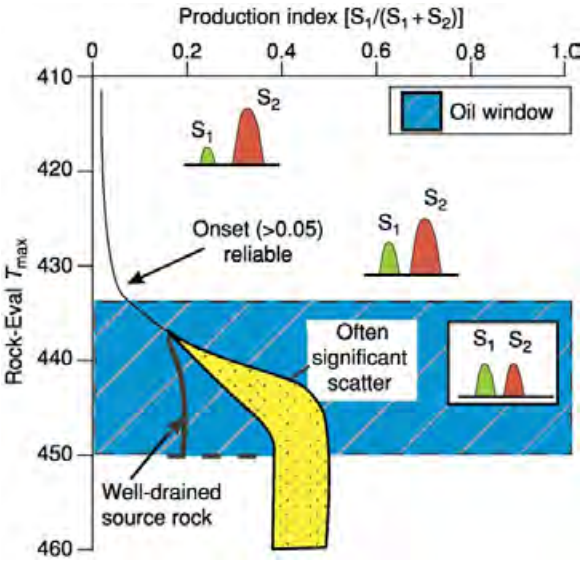


Figure 23 Oil generation as indicated by the Rock-Eval production index (PI) where the maturity trend depends on both generation and expulsion (retention).

The temperatures given above relate to corrected wire-line log temperatures and Tertiary effective burial, as in the North Sea. Studies related to DST temperatures quote higher temperatures (up to 12°C higher) for these boundaries, while absence of temperature correction can produce substantially lower temperature boundaries. Lower temperature (e.g., 10°C lower overall) may be expected where a Mesozoic burial event is the controlling factor, while higher temperatures (up to 25°C higher) are reported where rapid Neogene–Quaternary burial controls generation.

Both extract yields and production index discussed above are measured parameters. The theoretical and hence modellable measure of kerogen degradation is the transformation ratio on a scale of $0 \rightarrow 1$:

$$\text{TR} = \text{Generated petroleum}/\text{Original petroleum potential}$$

where:

$$\text{Generated petroleum} = \text{Original petroleum potential} - \text{Residual petroleum potential}$$

The original petroleum potential is determined from immature source rocks of the same organofacies, whereas the residual petroleum potential is determined from the source rocks as it is buried through the oil and gas window. The equivalence of the transformation ratio as a measure of generation to the maturity parameters discussed in the previous section (Tables 1 and 2) is given in Table 3 for oil generation from Type II kerogen and Table 4 for gas generation from Type III kerogen.

Generation is a kinetically controlled process, being a function of the effects of both temperature and time on the breakage of chemical bonds in the kerogen present in the source rock. It can be modelled using simple reactions or networks of competing or sequential reactions, expressed in terms of the Arrhenius equation. The Arrhenius equation effectively expresses the ease of degradation of the kerogen to petroleum as a distribution of chemical bond strengths reflecting the C–C, C–H, C–O–C, C–S–C, and C–N–S bonds holding the kerogen network

Table 3 The generation-defined oil window approximately related to some common maturity parameters

Parameters (oil & gas)	*Temperature (°C)*	*Transformation ratio[a] (0–1)*	*Vitrinite reflectance (%)*	*Spore colour index (1–10)*	*Rock-Eval T_{max}(°C)*
Early mature	80–115	0.05–0.15	0.5–0.7	4.0–5.0	432–442
Mid-mature	115–145	0.15–0.65	0.7–1.0	5.0–7.0	442–455
Late mature	145–165	0.65–0.95	1.0–1.3	7.0–9.0	455–465
Post-mature	>165	>0.95	>1.3	>9.0	>465

[a]Transformation ratio of Type II kerogen (LLNL).

together. Each of these chemical bonds has a unique strength (activation energy), depending on the adjacent atoms and functional groups, together with stereochemistry. Using the information shown in a burial history plot (Figure 21) to solve the Arrhenius equation, the extent of kerogen degradation can be modelled.

The understanding of generation in both space and time is at the centre of the use of the petroleum system concept to explore for oil and gas, and this requires the use of computer calculation and modelling. The calculation of a theoretical transformation ratio using the Arrhenius equation can be undertaken using industry standard kinetics and basin modelling software, and can be used to calculate generation as a function of depth (Figure 24) and geological time (Figure 25).

The generation versus depth plot (Figure 24) is a simplification in that it assumes that all units within the modelled section contain the same kerogen type

Table 4 The generation-defined gas window approximately related to some common maturity parameters

Parameters (gas)	*Temperature (°C)*	*Transformation ratio[a] (0–1)*	*Vitrinite reflectance (%)*	*Spore colour index (1–10)*	*Rock-Eval T_{max}(°C)*
Early mature	115–145	0.05–0.45	0.7–1.0	5.0–7.0	455–465
Mid-mature	145–220	0.45–0.85	1.0–2.2	7.0–9.0	465–525
Late mature	>220	0.85–0.95	2.2–>3.0	10	525–575
Post-mature	??	>0.95	>3.0	10	>575

[a]Transformation ratio of Type III kerogen (LLNL).

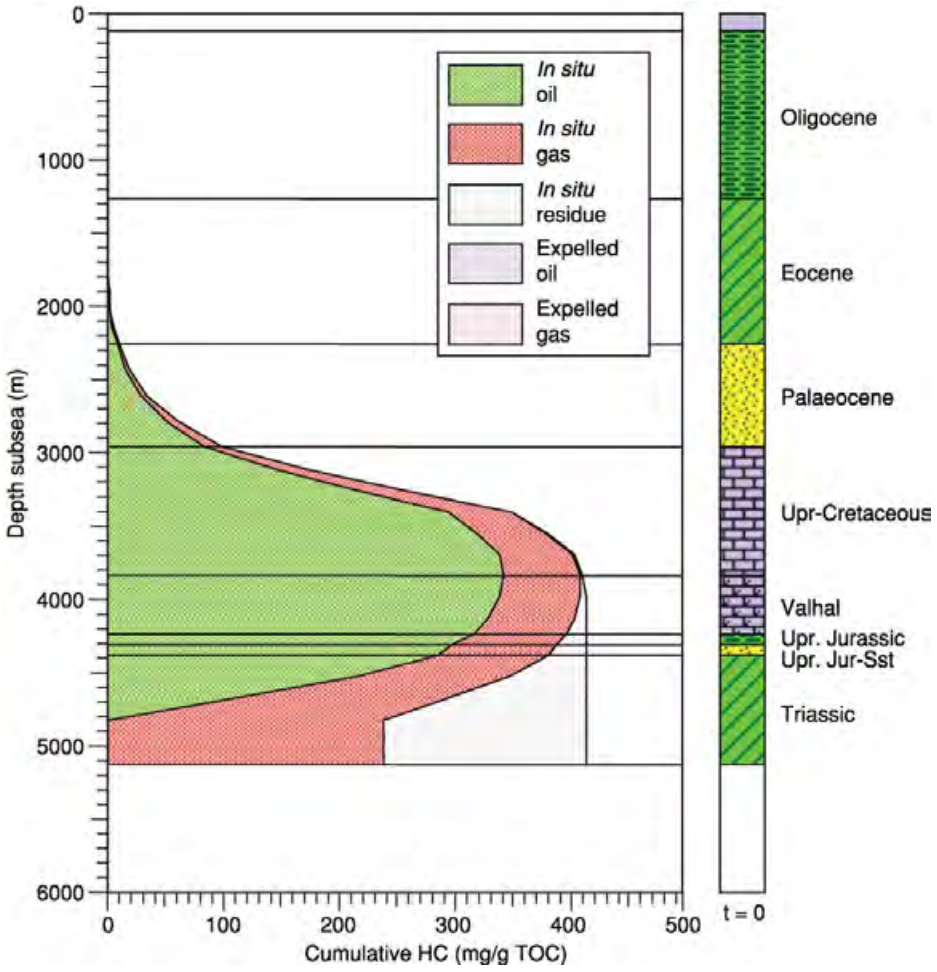


Figure 24 Computer-modelled depth trend for the potential generation of oil and gas, assuming all rock units contain a common kerogen type (compare with Figure 22 based on measured data).

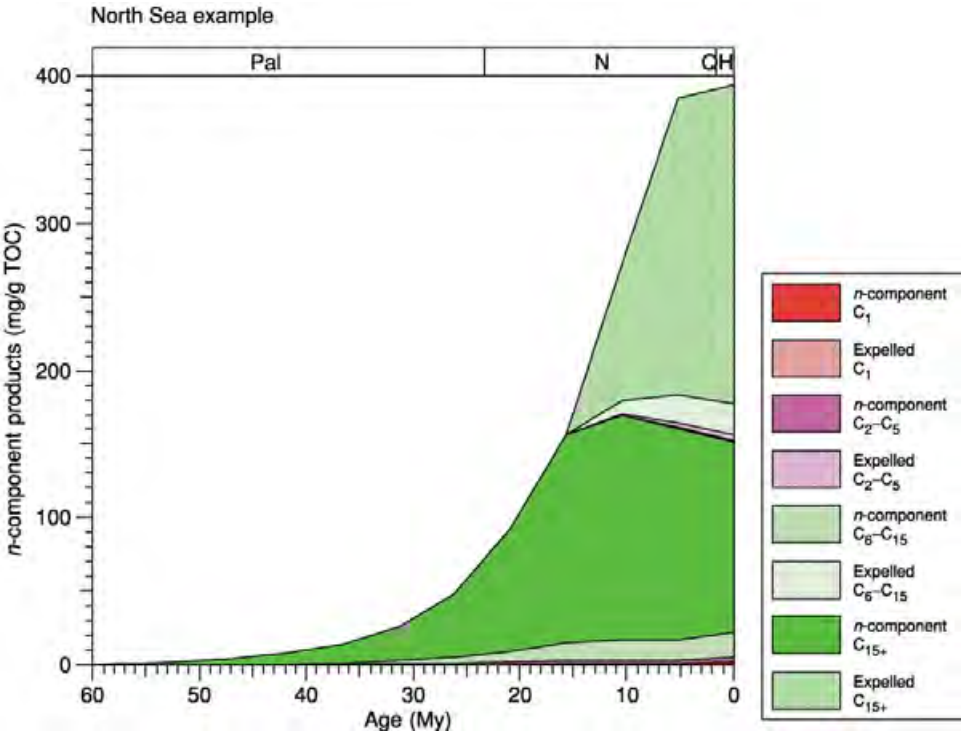


Figure 25 Computer-modelled generation and expulsion of oil and gas through time from a single source rock unit containing kerogen.

(unlikely). In this case the kerogen is modelled to be broken down to oil and then oil potentially cracked to gas in a two-step procedure:

$$\text{Kerogen} \rightarrow \text{Oil} + \text{Primary gas}$$

$$\text{Oil} \rightarrow \text{Secondary gas}$$

Considering generation in terms of geological time, a single source rock unit will progressively generate a range of products, grouped somewhat arbitrarily as methane plus wet gas, light oil, and main oil ranges of molecules in Figure 25. The subdivision of oil into these molecular ranges, though subjective, is useful for estimating compositional risk in oil exploration.

In the source rock, these molecules form a single petroleum phase (monophasic mix), which can be modelled as being expelled from the source rock once the pore space in the source rock is saturated. Expulsion is poorly understood in terms of process, but well understood in terms of efficiency. Possible mechanisms for petroleum expulsion include movement:

1. From high to low pressure potential—with pressure also causing rock fracturing;
2. From high to low concentrations (diffusion);
3. Under capillary forces induced at mineral pore throats;
4. Under capillary forces through a continuous kerogen 'wick'; and
5. In aqueous solution during decompaction and clay dehydration.

A combination of these mechanisms possibly contributes to expulsion from the source rock, the relative contributions being the basis for the current debate. The favoured view envisages the dominant process as being based on pressure, where compaction coupled with the generation of hydrocarbon fluids from solid kerogen and temperature-induced expansion of pore fluids produces overpressure in the source rock. The overpressure eventually fractures the source rock, sporadically releasing the fluids into the migration pathway.

The expelled hydrocarbon yields shown in Figure 25 are calculated from the total generated less that required to saturate (a fraction of) the porosity of the source rock. The modelled porosity can be calculated using one of a number of possible compaction equations.

A quantitative understanding of expulsion efficiency is essential for successful prospect evaluation. Expulsion efficiency (EE) is defined in mass units as:

$$\text{EE} = \text{Petroleum expelled/Petroleum generated}$$

where

$$\text{Petroleum expelled} = \text{Petroleum generated} - \text{Petroleum retained}$$

and

$$\text{Petroleum generated} = \text{Initial potential} \times \text{Transformation ratio}$$

Within a basin such as the North Sea, which contains a uniform oil-prone source rock organofacies in the Upper Jurassic Kimmeridge Clay Formation, the initial potential is determined by analysis of immature source rock samples from the rim of the basin. The retained (i.e., generated but not expelled) petroleum can be determined by solvent extraction or pyrolysis (S_1 peak—see Figure 15). The transformation ratio can be determined from modelling, equated with vitrinite reflectance or pyrolysis T_{max} (e.g., Table 3) or from ratios of the initial to measured hydrogen index (e.g. initial HI = 600, present HI = 300, therefore TR = 0.5). Examples of the range of expulsion efficiencies calculated from North Sea source rock data are shown in Figure 26, where values rise from about 20% at 3.0–3.5 km to about 50% at 4.0 km with the deepest samples exhibiting values in the 75–95% range.

Considering dry gas (methane) generation in, and expulsion from seam coals requires an understanding of additional factors, in particular the very high

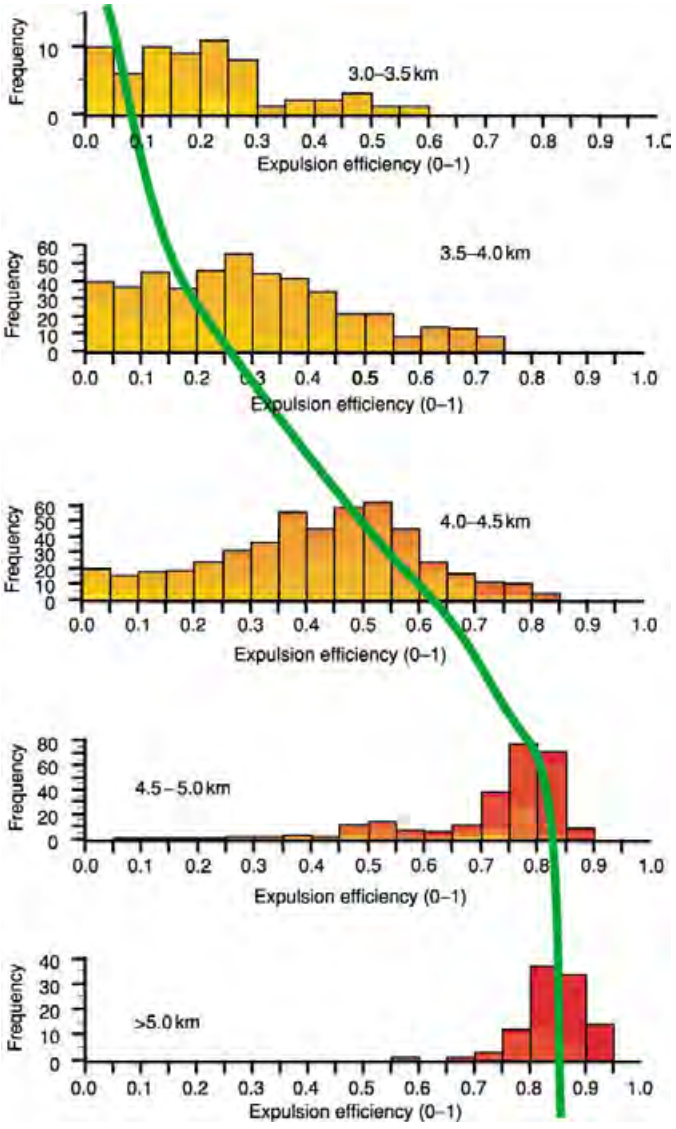


Figure 26 Expulsion efficiency as a function of depth in the North Sea based on solvent extraction (green trend line, from Larter (1988)) and Rock-Eval pyrolysis data (histograms).

surface area (macro-, meso-, and microporosity) of coal in general and the vitrinite maceral in particular. The developing understanding seems to suggest a three-step process:

1. Generation of the methane from humic kerogen;
2. Adsorption of the methane on the abundant porosity of vitrinite; and
3. Expulsion of the methane at higher maturity as the porosity collapses.

In all cases the expelled monophasic petroleum leaves the source rock, but will only form commercial accumulations if migration provides a focus towards a trap. The processes controlling migration are discussed in the next section.

Migration

How petroleum travels long distances in the subsurface is a matter for speculation: facts are hard to come by.

Certainly oil being more buoyant than water, and gas being more buoyant than both, there is a natural tendency for the upward movement of separate gas and oil phases in a water-saturated permeable sediment. The fluids are essentially separating out within the porous media of sedimentary rocks. Attempting to understand the processes contributing to migration requires some knowledge over a large range of scales of the properties of the solid rock and the fluid phases flowing therein (Figure 27). The scales are not only

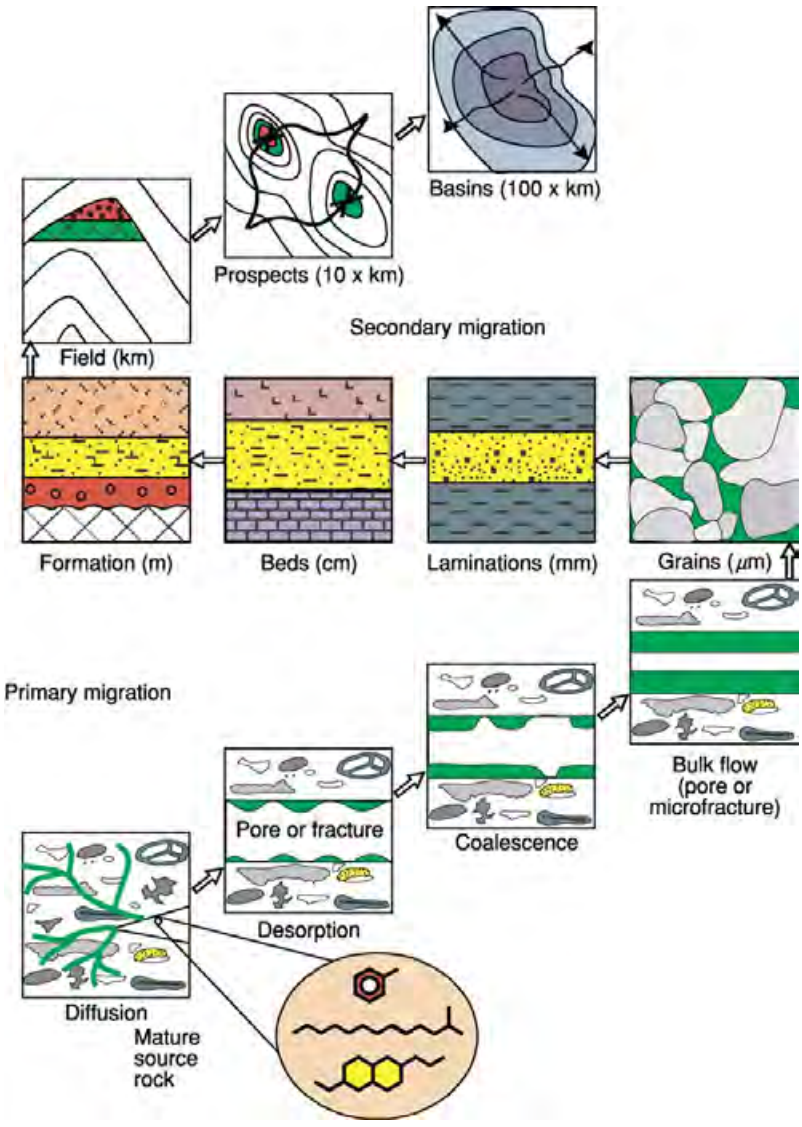


Figure 27 Our understanding of how oil and gas move through water-saturated rock requires a study of flux over a large range of scales. Modified after Mann *et al.* (1991).

those from sedimentary grains to basins, but also from oil molecules to accumulations containing billions of barrels of oil.

At deposition, sedimentary rock is initially porous, but compaction and diagenesis rapidly reduces the porosity. Once sediments are indurated (compacted and cemented), fracture porosity may take over as the dominant conduit allowing fluid flow. Thus fluids can move through both intergranular or fracture porosity.

Once expelled into a more porous conduit such as a sandstone bed, a fault plane, or diagenetically altered carbonate, petroleum will start to move upwards under the force of buoyancy (Figure 28, inset). The buoyancy derives from the lower density of oil (typically $0.86\,g\,cc^{-1}$ or 36°API) relative to formation water (typically $1.03\,g\,cc^{-1}$), as in the kitchen where fat rises to the surface of milk. Once within the migration conduit the petroleum will rise to the roof of the rock layer and if the layers are tilted, the 'river of oil' will move up-dip and along zones of highest permeability from high point to high point. When seen at outcrop, this produces a braided network of oil-saturated rock, running under the top-seal of the migration conduit (Figure 28).

Migration can also be influenced by any water flux sharing the same migration route (hydrodynamic control) or by pore pressure gradients associated with the progressive increase in pore throat diameters (permeability control). The pore throat diameters together with wettability essentially control the permeability of the rock with respect to oil or gas.

Migration distances can vary widely. In situations where mature source rock is interbedded with reservoir sands the migration path may only be a number of metres. Large accumulations, however, demand the gathering of oil and gas over a wide area, and hence long-distance migration pathways. In the North Sea migration distances are mainly in the range 10–20 km with 65 km as a maximum. In large basins such as the West Canada Basin and West Siberian Basin migration pathways of many hundreds of kilometres have been proposed.

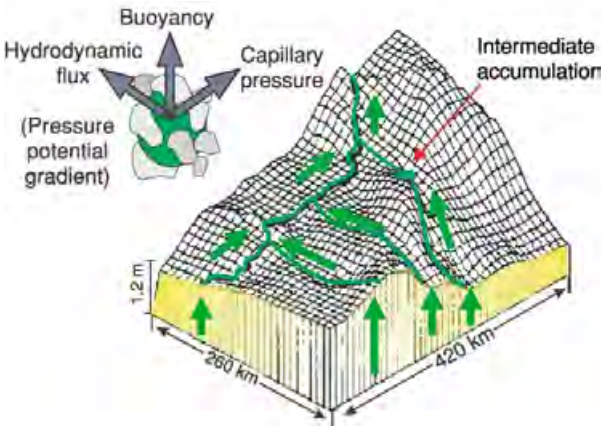


Figure 28 Migration occurs as 'rivers of oil' running up-dip mainly under the influence of density-driven buoyancy, though hydrodynamic flux and capillary pressure may exert ancillary controls. Modified after IES (1985).

As the expelled monophasic petroleum moves upwards under buoyancy, the pressure and temperature regime will change and the bubble-point or dew-point may be reached (Figure 29). Separation into a liquid (oil) phase and a gas phase will result, and these may migrate separately due to differential buoyancy and relative permeabilities.

Studying a petroleum system demands an understanding of the efficiency of the migration process. This is probably the least well-quantified element of a petroleum system, and is normally determined by difference:

$$\text{Accumulation} = \text{Generated} \times \text{Expulsion efficiency} \times \text{Migration efficiency}$$

Therefore

$$\text{Migration efficiency} = \text{Accumulated}/(\text{Generated} \times \text{Expulsion efficiency})$$

The above equation can be used to determine the migration efficiency if the size of an accumulation, the amount generated, and expulsion efficiency are all known. Calibrated against known accumulations, the migration efficiency can then be used to predict the petroleum charge to undrilled prospects. Migration efficiencies are not widely discussed or reported, but values seem to fall in the range of 50% (maximum) to 0 (i.e., no oil or gas reached the drilled prospect). It may be helpful to consider the inefficiency of migration as the residual oil left behind along the migration path. The recognition that oil migration is a focussed process affecting only a small percentage of the rock between kitchen and accumulation fits well with the calculated migration efficiencies.

Accumulation and Survival

Once charged to the trap, the survival of the petroleum depends on a good seal and avoiding the effects of various oil alteration processes. An approach to use bulk, molecular, and isotopic characterization of oil and gas to identify such reservoir processes is shown in Figure 30. As emphasized in the figure, the analytical approach to differentiating between oil alteration processes requires knowledge of the geometry and history of the trap.

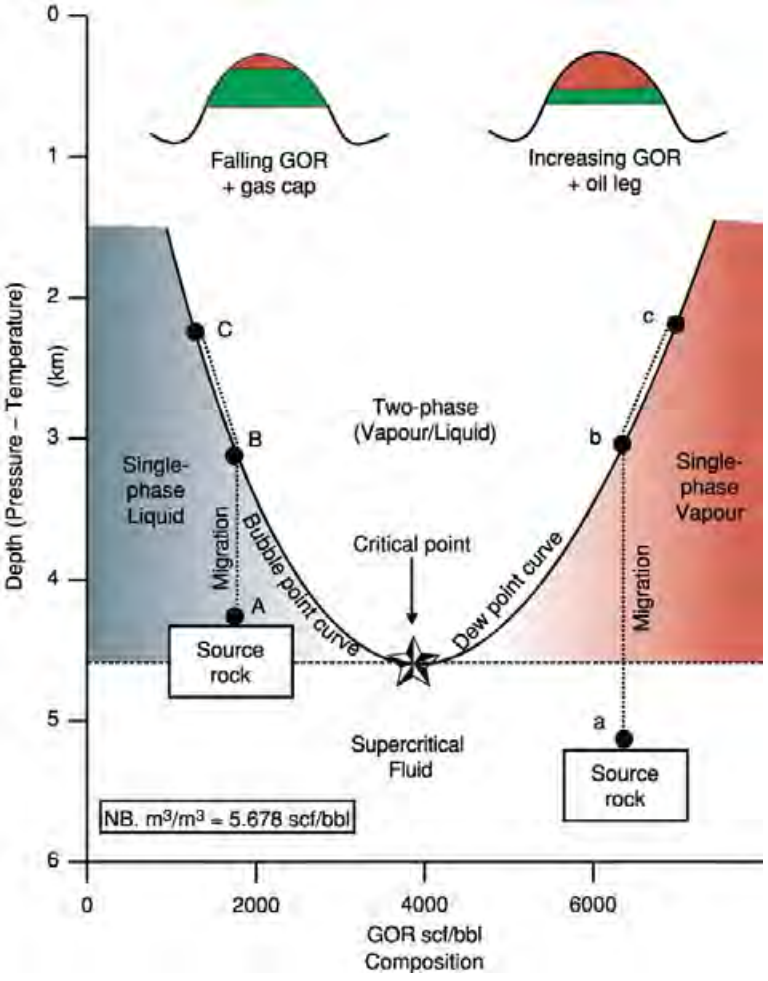


Figure 29 Phase separation of upward-migrating petroleum depending on the initial gas/oil ratio (GOR) and reductions in temperature and pressure.

In terms of a seal, efficiencies range from near perfect in the case of intact salt (Halite) to 'highly leaky' in the case of thin and poorly compacted mudstone containing a significant portion silt and fine sand. Diagenetic cement can greatly enhance the sealing efficiency of more porous strata, even when relatively shallow. In essence the efficiency of a seal must be measured against the charge rate if generation continues, or the time since charge ceased if in the past. A more specific and dramatic way of breaching the seal of the trap is faulting.

For structural traps, the prediction of whether a fault is a seal or a migration conduit is complex, the controlling factors being only poorly understood at present. Some unusual seals comprise a hydrodynamic trap, where a counterflow of formation water opposes the buoyancy of the oil, and shallow seals in the Arctic involving the permafrost layer or gas hydrate (*see* **Petroleum Geology:** Gas Hydrates).

The concept of the 'half-life' of an oil or gas field has been proposed, this being defined as the time (millions of years) required for half of the charge volume to escape through the seal. Gas sealed by mudstone has a half-life in the range 30–60 My, and oil about an order of magnitude greater. A continuous and stable halite seal can be considered near perfect with an infinite half-life.

A number of intrareservoir processes are listed towards the base of Figure 30, and of these the most common are bacterial degradation and thermal cracking. These occur when the reservoir is unusually cool (<65°C) and unusually hot (>165°C), respectively.

Of the listed intrareservoir processes, bacterial degradation is arguably the most destructive. In essence it occurs when petroleum accumulates in a reservoir that is at a temperature allowing bacterial degradation of the oil or gas. As reported, the maximum

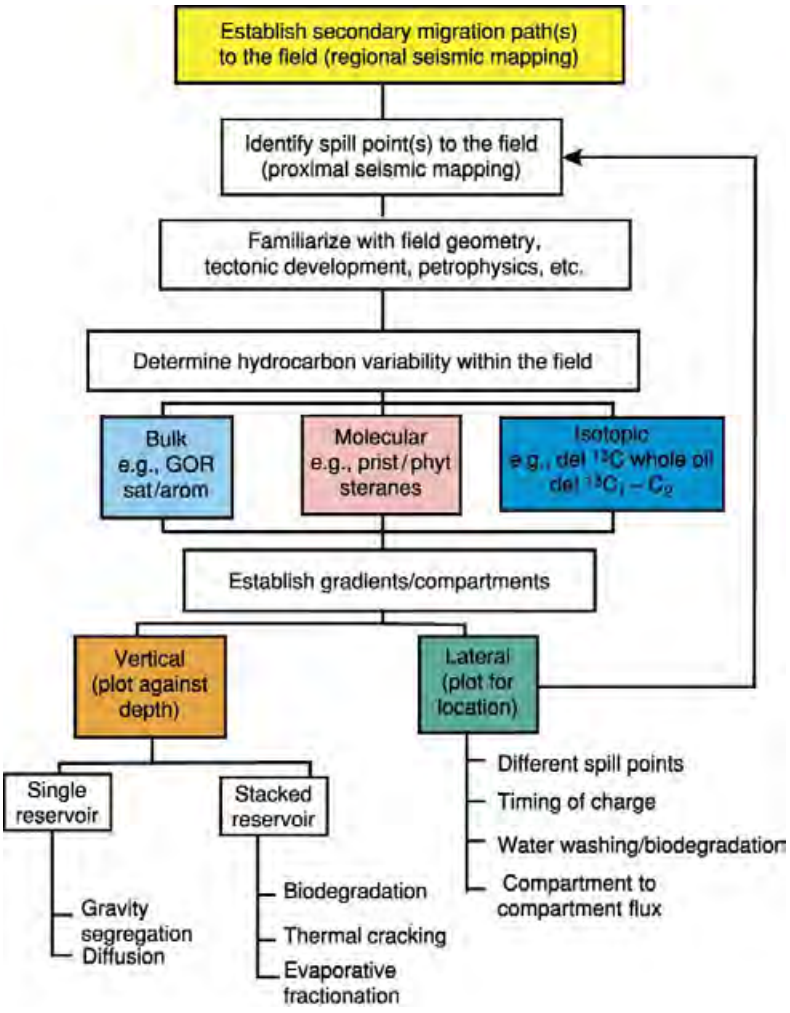


Figure 30 Steps required to identify and interpret vertical and lateral compositional gradients resulting from intrareservoir processes.

temperature boundary is in the range 65–75°C, the range reflecting different oil chemistries, formation waters, and hence active bacterial communities, together with different methods of measuring reservoir temperatures. In nature, viable bacteria will be preserved if the reservoir has never been buried to temperatures hotter than about 90°C. If hotter and then uplifted to a cooler regime, meteoric water influx into the reservoir is required to reintroduce bacteria.

The bacterial alteration of the oil progresses with the selective removal of various families of molecules (Table 5). The sequence of removal is broadly *n*-alkanes > branches alkanes > cyclic alkanes -> aromatics, this reflecting the 'digestibility' of hydrocarbon molecules by the bacteria.

Bacterial degradation of a gas accumulation removes the wet gas components (C_2–C_5) and in particular propane (C_3) with preferential removal of molecules with the ^{12}C isotope. Thus an initial wet gas is reduced to a dry gas accumulation of >95% methane with isotopically heavy residual propane. Bacterial activity is often confirmed by the thin rim of biodegraded heavy oil below the dry gas.

Most reservoirs are cooler than the source rocks from which they were filled. If the petroleum-filled reservoir is buried to higher temperatures, the oil will crack to condensate and gas (Table 6). This process of intrareservoir cracking starts at about 160°C and black oil is normally destroyed by 175°C. This has been termed the base of the oil preservation (as opposed to oil generation) window. An oil trap that has suffered intrareservoir cracking will be recognized as containing gas, with black flecks of pyrobitumen seen occupying some of the remaining porosity:

$$2 \times \text{oil}(CH_2) \rightarrow \text{pyrobitumen}(C) + \text{methane}(CH_4)$$

Table 5 The degree of bacterial degradation of oil on a 1–10 scale based on progressive loss of specific chemical groups

Degree of bacterial degradation	*Loss of named molecules indicates stage of biodegradation*			
1	$C_{15}+$ fraction	C_{15}–C_7	Gasoline (C_4–C_7)	Gas (C_1–C_4)
			n-Alkanes	*n*-Propane[b]
2	$<C_{15}$ *n*-alkanes	Alkyl aromatics		
3	Total loss of *n*-alkanes, appearance of 10-demethyl hopanes			Ethane to *n*-butane . . . iso-butane
4	C_{15} to C_{20} acyclic isoprenoids		iso-alkanes	
5				
6	Regular steranes	Uncertain	Cycloalkanes	Dry gas only survives
7	Hopanes and			
8	diasteranes		Aromatics[a]	
9	Aromatized biomarkers			
10	Total loss of resolvable peaks ('hump' of UCM left as residue)			

UCM-unresolved complex mixture.
[a]Earlier loss of aromatics (benzene and toluene) may result from concurrent water washing.
[b]Early loss of propane produces stable carbon isotope anomaly (isotopically heavy C_3).

Table 6 Intrareservoir cracking of oil: The relationship of transformation ratio of the kinetic cracking reaction to reservoir temperature based on kinetics from Platte River's BasinMod with heating rate of 1.25°C/My

Temperature (°C)	*Transformation ratio*[a]	*Comments*
110	0.00	Temperature of onset of oil window
130	0.00	Peak oil generation temperature
140	0.00	
145	0.01	Temperature of end of main oil window
150	0.03	Primary generation in source rock ceased, oil stable in reservoir
160	0.07	
165	0.16	Onset significant oil cracking
170	0.34	Main phase of oil cracking
175	0.65	
180	1.00	End of modelled black oil survival

[a]Transformation ratio of the oil → gas + residue cracking reaction.

System Efficiencies

Based on the definition of a petroleum system given at the start of this article a gross system efficiency may be defined as:

$$\text{System efficiency} = \text{Discovered petroleum} / \text{Generated petroleum}$$

Expressed as a percentage, reported system efficiencies fall in the range 0.5 to 30%. This range includes both natural variations attributable to the geology and errors in our understanding thereof. Based on the previous discussion and as illustrated in Figure 3, the simple equation given above can be expanded to give a more complete description of the controls:

$$\text{Discovered petroleum} = (\text{area of field} \times \text{thickness of reservoir} \times \text{porosity}) - \text{losses}$$

$$\text{Generated petroleum} = \text{source rock yield} \times \text{thickness} \times \text{area} \times \text{transformation ratio}$$

In addition to technical uncertainty in the above measurements, the reported values for 'discovered petroleum' may also be influenced by commercial and political considerations.

Uncertainty in the 'generated petroleum' value will include variations in total organic carbon, kerogen type, and its variation with source rock thickness: the transformation ratios (Table 3) for discrete areas of mature source rock will be determined from measured and modelled maturity data. These areas will be measured by planimeter on basin-scale mapping derived from the interpretation of a regional seismic grid. If the generated and discovered petroleum values are precisely defined, then the efficiency of the petroleum system is essentially the expulsion efficiency multiplied by the migration efficiency.

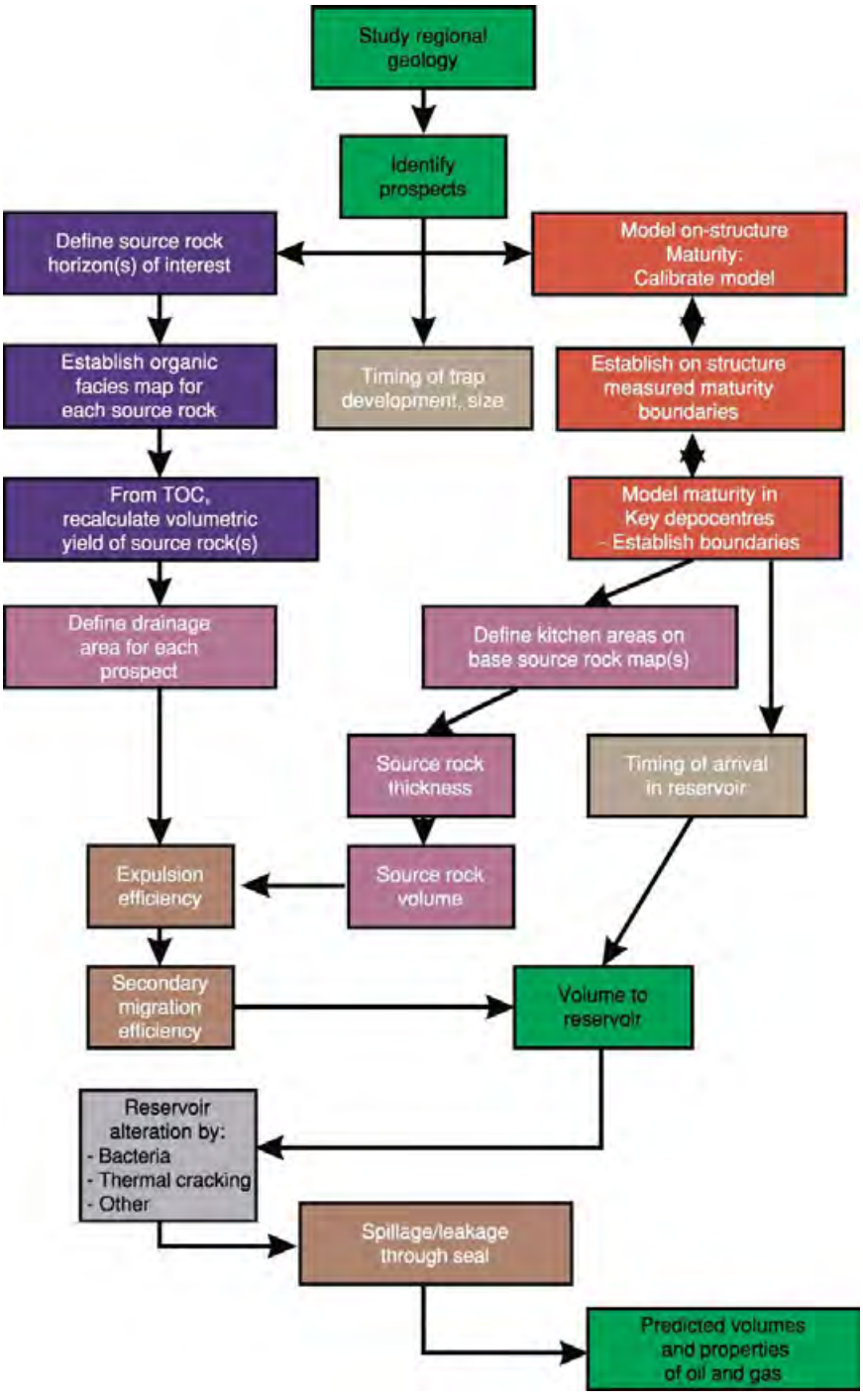


Figure 31 A summary of the steps leading to the determination of the efficiency of a petroleum system.

A possible workflow tying all these factors together is given in Figure 31 by way of a summary.

See Also

Analytical Methods: Geochemical Analysis (Including X-Ray). **Petroleum Geology:** Overview; Chemical and Physical Properties; Gas Hydrates; Exploration; Production.

Further Reading

Biteau J-J, de Janvry GC, and Perrodon A (2003) The petroleum system: a fundamental tool. *Oil and Gas Journal* 101(30): 34–39.

Biteau J-J, Janvry GC, and Perrodon A (2003) How the petroleum system relates to the petroleum province. *Oil and Gas Journal* 101(31): 46–49.

Cornford C (1994) Mandal-Ekofisk (!) petroleum system in the Central Graben of the North Sea. In: Magoon LB and Dow WG (eds.) *The Petroleum System – from Source to Trap,* AAPG Memoir No. 60. AAPG., pp. 537–571. Oklahoma: Tulsa.

Demaison G and Huizinga BJ (1991) Genetic classification of petroleum systems. *American Association of Petroleum Geologists Bulletin* 75(10): 1626–1643.

Magoon LB and Dow WG (eds.) (1994) *The Petroleum System – from Source to Trap,* AAPG Memoir No. 60. AAPG., p. 655. Oklahoma: Tulsa.

Selley RC (1996) *Elements of Petroleum Geology,* 2nd Edition. San Diego, California: Academic Press.

General Geochemistry

Cornford C (1998) Source rocks and hydrocarbons of the North Sea. In: Glennie KW (ed.) *Petroleum Geology of the North Sea,* pp. 376–462. Oxford: Blackwell Science.

Engel MH and Macko SA (1993) *Organic Geochemistry: Principles and Applications,* p. 861. New York: Plenum Press.

Hunt JM (1996) *Petroleum Geochemistry and Geology,* p. 743. New York: Freeman and Company.

Killops SD and Killops VJ (1993) *An Introduction to Organic Geochemistry,* p. 265. Harlow, Essex: Longman Scientific and Technical.

Peters KE and Moldowan JM (1993) *The Biomarker Guide,* p. 363. Eaglewood Cliffs, New Jersey: Prentice-Hall Inc..

Tissot BP and Welte DH (1984) *Petroleum Formation and Occurrence*. Berlin: Springer-Verlag.

Tyson RV (1995) *Sedimentary Organic Matter: Organic Facies and Palynofacies,* p. 615. London: Chapman and Hall.

Waples DW (1985) *Geochemistry in Petroleum Exploration,* Geological Sciences Series. p. 232. Dordrecht, NL: D. Reidel Publishing Company (Kluwer Academic Publishers).

Basin Modelling

Beardsmore GR and Cull JP (2001) *Crustal Heat Flow. A guide to measurement and modelling,* p. 324. Cambridge: Cambridge University Press.

Duppenbecker SJ and Iliffe JE (eds.) (1998) *Basin Modelling: Practice and Progress,* Geol. Soc. Sp. Publ. No. 141. p. 245. London: The Geological Society.

Pepper AS and Corvi PJ (1995) Simple kinetic models of petroleum formation Part 1: oil and gas generation from kerogen. *Marine and Petroleum Geology* 12(3): 291–321.

Pepper AS and Corvi PJ (1995) Simple kinetic models of petroleum formation Part III: modelling an open system. *Marine and Petroleum Geology* 12(4): 417–452.

Welte DH, Horsfield B, and Baker DR (eds.) (1997) *Petroleum and Basin Evolution,* p. 524. Berlin: Springer-Verlag.

Generation, Migration, Alteration and Volumetrics

Ahsan A, Karlsen DA, and Patience RL (1997) Petroleum biodegradation in the Tertiary reservoirs of the North Sea. *Marine and Petroleum Geology* 14(1): 55–64.

Cooles GP, Mackenzie AS, and Quigley TM (1986) Calculation of petroleum masses generated and expelled from source rocks. *Organic Geochemistry* 10(1–3): 235–245.

George S, Boreham CJ, Minifie S, and Teerman SC (2002) The effect of minor to moderate biodegradation on C_5 to C_9 hydrocarbons in crude oils. *Organic Geochemistry* 33(12): 1293–1318.

Johannesen J, *et al.* (2002) 3-D oil migration modelling of the Jurassic petroleum system of the Statfjord area, Norwegian North Sea. *Petroleum Geoscience* 8(1): 37–50.

Lewan MD, Henry ME, Higley DK, and Pitman JK (2002) Material-balance assessment of the New Albany-Chesterian petroleum system of the Illinois basin. *American Association of Petroleum Geologists Bulletin* 86(5): 745–777.

Leythaeuser D, Schwark L, and Keuser C (2000) Geological conditions and geochemical effects of secondary petroleum migration and accumulation. *Marine and Petroleum Geology* 17(7): 857–859.

Larter SR (1988) Some pragmatic perspectives in source rock geochemistry. *Marine and Petroleum Geology* 5(3): 194–204.

Mann J, Duppenbecker SJ, Langen A, Ropertz B, and Welte DH (1991) Pore network evolution of the Lower Toarcian Posidonia Shale during petroleum generation and expulsion – a multidisciplinary approach. *Zentralblatt fur Geologie und Palaeontologie* 1(8): 1051–1071.

Volkman JK, Alexander R, Kagi RI, Rowland SJ, and Sheppard PN (1984) Biodegradation of aromatic hydrocarbons in crude oils from the Barrow sub-basin of Western Australia. In: Schenck PA, Leeuw JWD, and Lijmbach GWM (eds.) *Advances in Organic Geochemistry 1983,* pp. 619–632.

Exploration

J R Parker, Formerly Shell EP International, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

All commercial accumulations of hydrocarbons are found within, or in close proximity to, sedimentary basins, but not all sedimentary basins contain oil or gas. The presence of large quantities of hydrocarbons in a sedimentary basin requires (Figure 1)

- a source rock rich in organic carbon,
- sufficient heat over long periods of time to convert this organic carbon into hydrocarbons,
- migration pathways that enable the hydrocarbons to migrate away from the source rock,
- a suitable reservoir rock that is sufficiently porous for the hydrocarbons to accumulate and sufficiently permeable to allow the hydrocarbons to be produced at economic rates,
- an effective seal to prevent leakage from this reservoir, and
- a closed structure (a geometrical disposition of the reservoir and seal) to prevent further migration of the hydrocarbons.

Petroleum exploration depends primarily on the scientific understanding of these factors (*see* **Petroleum Geology:** Overview). However, exploration remains a risky business with an uncertain outcome, both in terms of the degree of confidence of realizing a successful outcome (probability of success) and in terms of the amount of hydrocarbons to be found in the case of success (success volume).

Thus, to justify an exploration programme, the expected benefits must exceed the expected costs. The principal benefit is the value of any oil or gas found, while the costs are those incurred in carrying out the programme. Hence, it is an important task of the petroleum geologist to develop quantitative estimates of the probability of success and of the likely recoverable volumes that may be present. In many countries the petroleum laws provide for governments and exploration companies to negotiate contracts that, typically, oblige a company to carry out an exploration programme of agreed scope and cost and entitle it to participate in the development and production of any oil and gas discovered in return for an agreed share of the rewards. Thus, even before exploration starts, a careful assessment of the risk and possible benefits is required.

The basis of this assessment is the so-called expected-monetary-value (EMV) calculation, in which the risk-weighted value of the potential reward should exceed (or at least equal) the risk-weighted cost of the potential loss, as risk without benefit is unacceptable:

$$\mathrm{NPV} \times \mathrm{POS} > \mathrm{AEC} \times (1 - \mathrm{POS})$$

where NPV is the net present value (monetary return in the case of a successful project), POS is the probability of success (chance that the project will be successful), AEC is the abortive exploration cost (the lost investment in the case of failure), and (1 – POS) is the probability of failure (complement of the probability of success).

An exploration programme is, therefore, a carefully planned series of activities aimed at discovering whether oil and gas fields are present in a particular area and, if present, whether they are sufficiently large and productive to be economical. Based on an analysis of the geological information, often involving sophisticated computer modelling of geological processes and an analysis of analogous information from petroleum provinces around the world, the probabilities of finding reserves of certain magnitudes can be assessed. In a well-defined area with fully appraised discoveries, for instance, there may be a high degree of certainty about the current reserves but little chance of finding major additions. In a speculative venture in a little-known area, on the other hand, the chances of finding any hydrocarbons may be low but there is an outside chance of making a very large discovery. However, even with the most sophisticated analysis, the discovery of any quantity of oil and gas cannot be anticipated with certainty, and petroleum exploration remains a classic example of 'decision-making under uncertainty'.

Exploration Methods

Exploration methods are the techniques employed in the search for oil and gas. Their primary purpose is not directly to find oil and gas but to provide evidence about the geology of the subsurface, the interpretation of which may eventually lead to drilling and the discovery of hydrocarbons.

Geological Analysis

The first stage in any exploration programme is to define a 'play' concept to explain how oil and gas may have accumulated in the basin. A play is defined

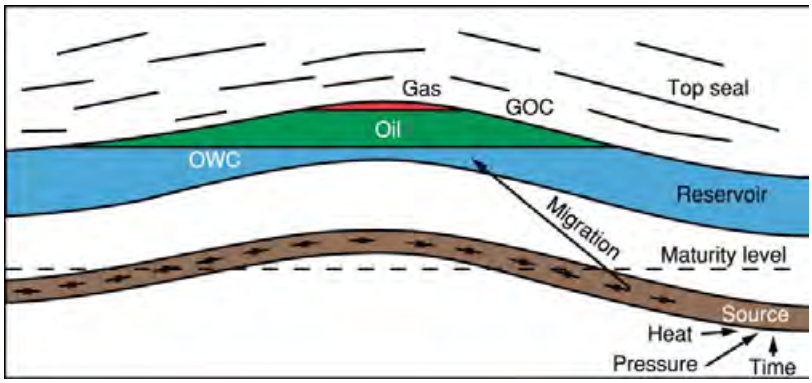


Figure 1 Geological conditions for hydrocarbon accumulation, showing the migration of oil and gas from mature source rock into an initially water-filled reservoir. GOC, gas–oil contact; OWC, oil–water contact. Reproduced with permission of Nautilus Limited.

as a group of prospects (sites of potential hydrocarbon accumulation) and any related fields having common oil or gas sources, migration relationships, reservoir formations, seals, and trap types, thus sharing common elements of geological risk (Figure 2).

The analysis of hydrocarbon plays involves the synthesis and mapping of all the key geological parameters controlling the occurrence of oil and gas, which are principally source, maturation and migration, reservoir, trap, timing, seal, preservation, and recovery (Figure 3).

Geochemical techniques can help to establish the presence of a suitable source rock, while sedimentary geological models can predict the distribution of reservoir and seal rocks. Data are derived from previously drilled wells in the basin (although not necessarily testing an identical play), from traditional field geology (examining surface outcrops within or on the margins of the basin), and from remote sensing (Figure 4) (which may elucidate the geological structure of the area). For previously unexplored areas lacking in 'hard' data, concepts can be tested against analogous basins elsewhere in the world. Play analysis helps to estimate the risks and potential volumes of prospects, as the absence of any one of the factors listed above invariably means that there will be no hydrocarbons in the prospect; a poor development of source rock will mean a limited generation of hydrocarbons, and a poor development of reservoir rock will mean only limited opportunities for entrapment.

Geophysical Analysis

Several geophysical techniques are used to enhance the geological understanding of a basin. Gravimetric (*see* **Analytical Methods:** Gravity) and magnetic surveys (both surface and airborne) (Figure 5) can be used in a reconnaissance mode to delineate the deeper parts of the basin and the major highs before undertaking much more expensive seismic surveys (*see* **Seismic Surveys**). However, seismic surveys are the most important geophysical tool for obtaining a detailed understanding of the subsurface structure, including identifying likely migration paths, inferring the relative timing of trap formation and charge, delineating the geometry and size of potential traps, and even establishing the presence of hydrocarbons themselves.

Seismic surveys involve recording artificially generated shock waves that are reflected or refracted from the different rock strata, and they require a suitable sound source and an appropriate set of detectors. Dynamite has largely been replaced on land by special trucks called vibrators, which have metal plates on their undersides to shake the ground in a controlled manner (Figure 6). At sea, powerful air or water guns are the main energy sources. The reflections are picked up by receivers called geophones on land and hydrophones offshore (Figure 7), which are laid out in such a way as to form directional antennae, termed arrays. Geophones and hydrophones work in a similar way to seismographs, which register earthquakes, but all data are recorded in digital rather than analogue form.

When gathering data in new or little-explored areas, two-dimensional seismic surveys are generally acquired first; with lines spaced as much as 20 km apart, these can cover a great deal of ground at comparatively low cost. If these surveys show interesting structures, then more closely spaced two-dimensional lines can be shot to provide a more detailed picture. However, for an enhanced understanding of the structural and reservoir detail of a prospect, three-dimensional seismic surveys employing a dense grid of lines are commonly used, even in the exploration phase before any wells have been drilled or hydrocarbons discovered. This is

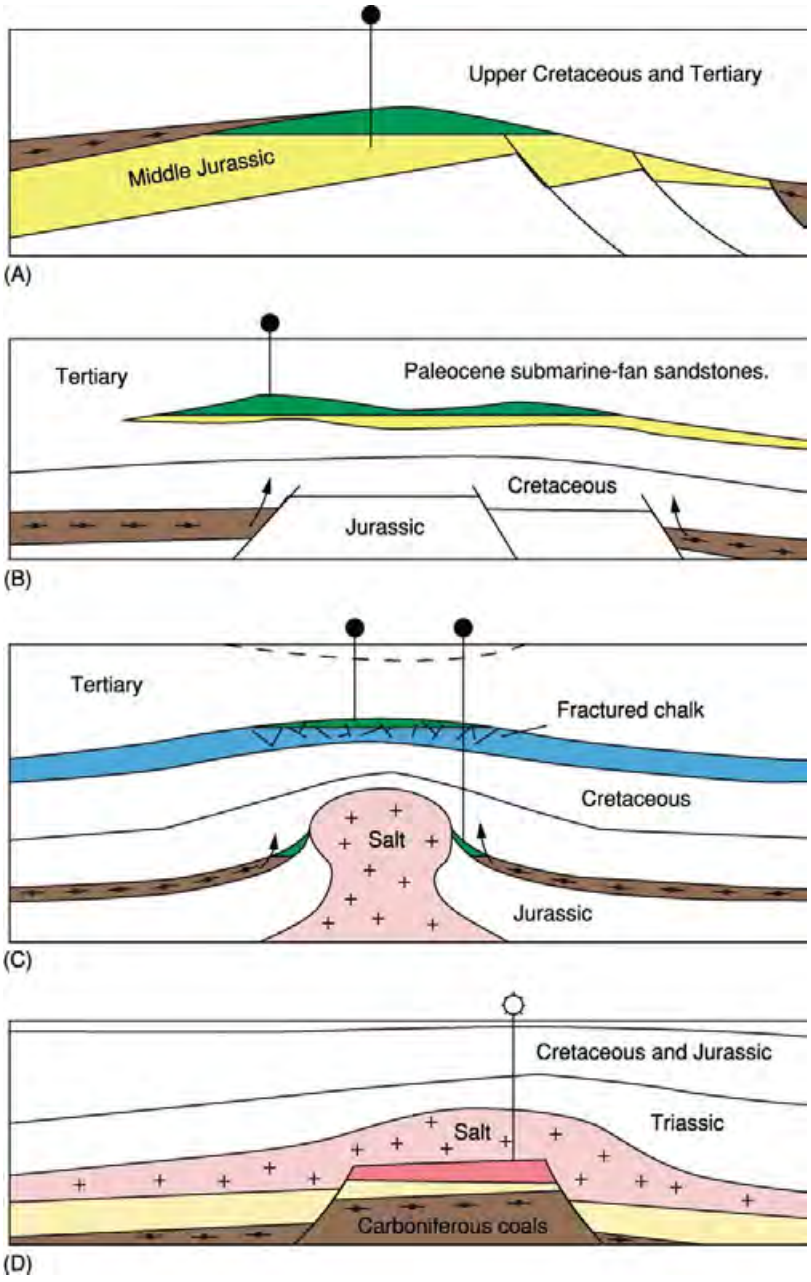


Figure 2 North Sea play types: (A) a truncated tilted fault block, typical of the northern North Sea; (B) a drape or pinchout trap, typical of the central graben; (C) a salt-induced structure typical of the central graben; and (D) a horst block capped by salt, typical of the southern North Sea. Yellow, sandstone reservoirs; brown, with —• symbol, source rock; pink, salt; blue, chalk reservoir; red, gas; green, oil. Reproduced with permission of Nautilus Limited.

particularly true in geologically complex areas, where the cost of a three-dimensional survey can be considerably less than the cost of a mislocated well.

A traditional two-dimensional seismic survey is shot along single lines to provide vertical slices through the Earth's crust; from a series of parallel and intersecting lines, the three-dimensional structure of the subsurface can be interpreted (Figure 8). However, the assumption is that the reflections seen on each line originate within the plane of the vertical slice, whereas in reality many of the reflections originate from inclined rock layers that may be hundreds of metres outside the plane of the seismic line; also, the gaps between the lines may hide faults and other geological complexities. These limitations have been overcome by the use of three-dimensional seismic surveys, which employ multiple sound sources and simultaneous recording by lines of detectors laid out in a dense grid to

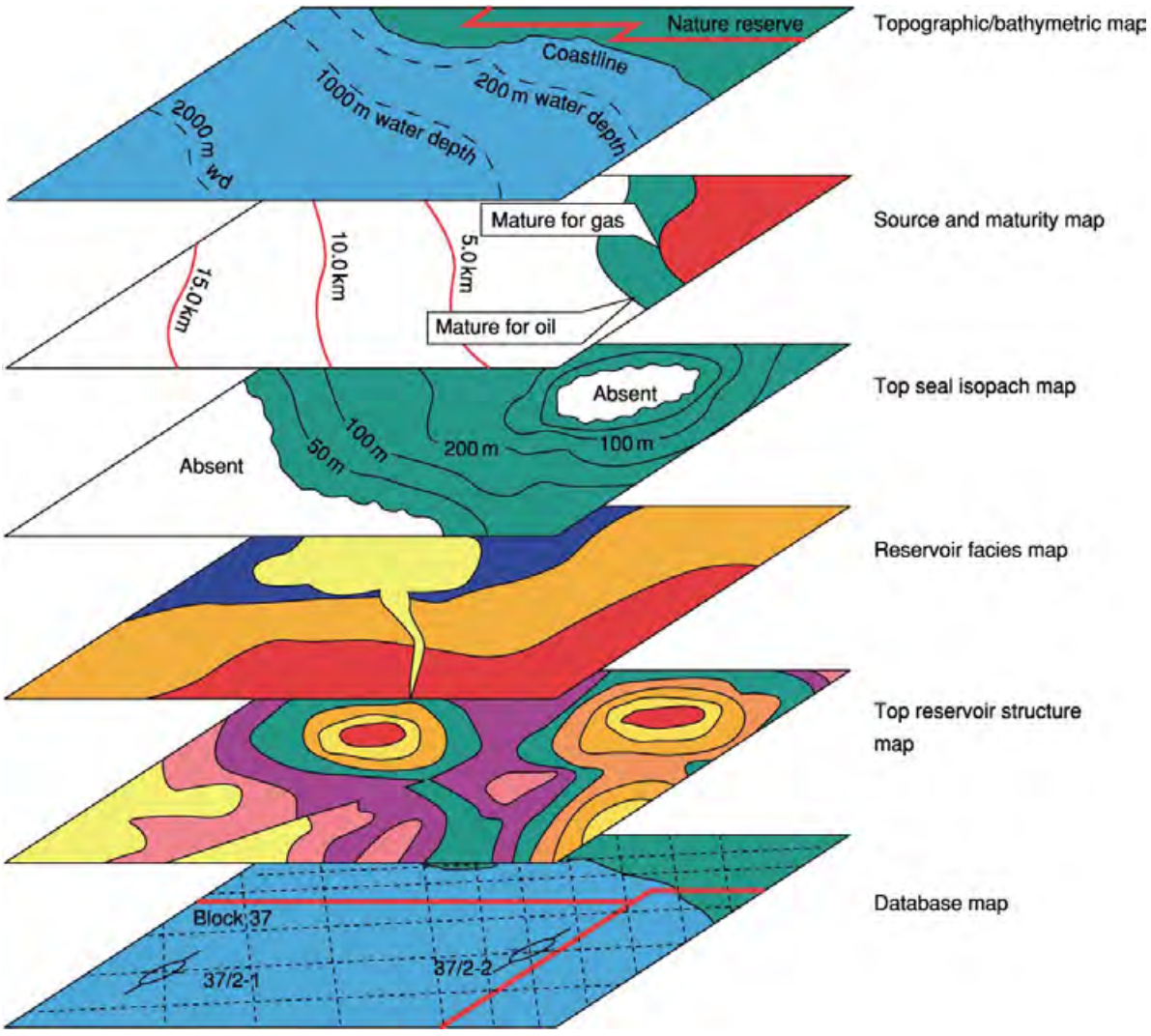


Figure 3 Elements of hydrocarbon-play analysis, combining topographic/bathymetric data, source rock maturity, reservoir distribution and structure, and seal. Reproduced with permission of Nautilus Limited.

Figure 4 Landsat image of south-eastern Iran, showing large, eroded anticlinal structures. The area shown is 45 km × 30 km. Image from Landsat 7, data available from US Geological Survey, EROS Data Center, Sioux Falls, SD, USA.

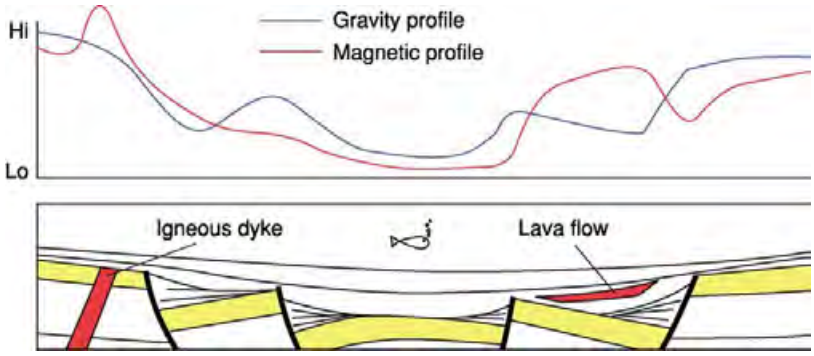


Figure 5 Gravimetric and magnetic profiles (above) which mirror the deeper parts of the basin and the major highs shown on the geological section (below). Note the effect of iron-rich rocks, such as lava flows and igneous dykes on the magnetic profile. Yellow, (sandstone) reservoir. Reproduced with permission of Nautilus Limited.

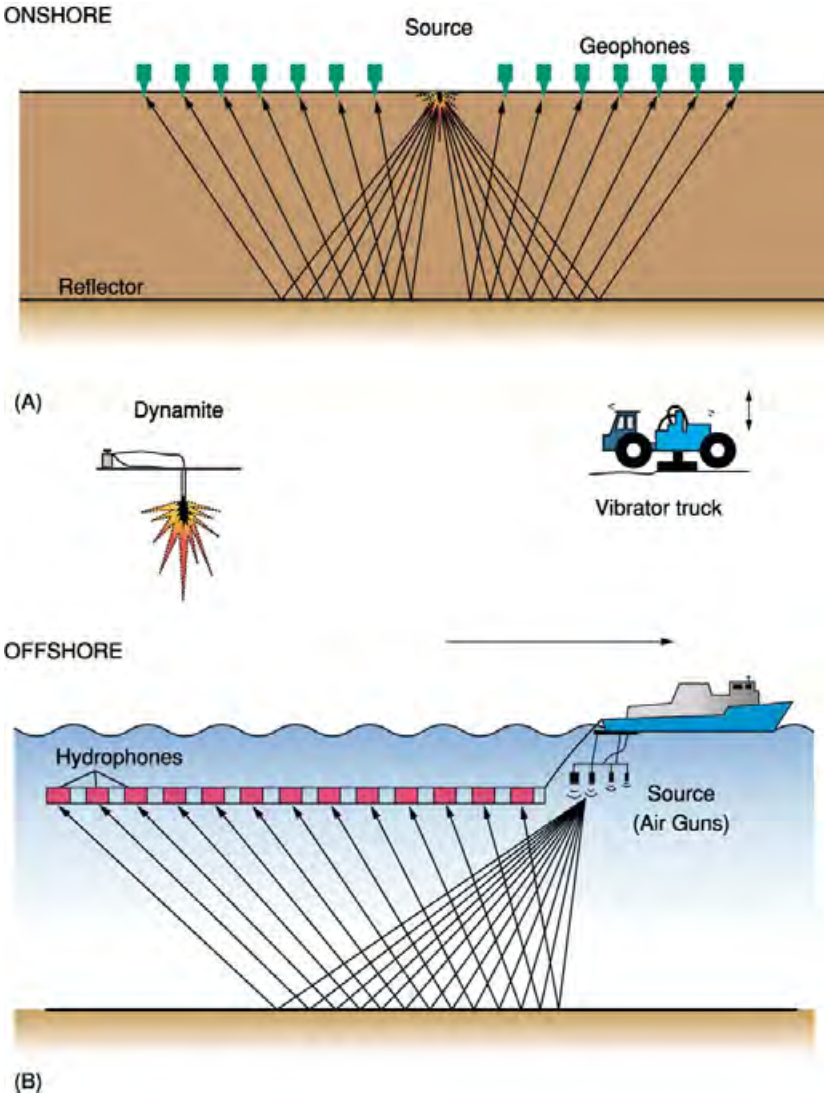


Figure 6 Principles of seismic surveying: (A) onshore, with diagrams showing the various methods of producing the signal, and (B) offshore. Reproduced with permission of Nautilus Limited.

Figure 7 Aerial view of an offshore seismic survey, showing the array of hydrophones being towed behind the ship. Reproduced with permission of Millenium Atlas Company Limited.

acquire a true three-dimensional picture (Figure 9). This development has been made possible by an almost exponential increase in computer power over the years coupled with a concomitant decrease in cost, which has enabled the vast amount of data acquired by three-dimensional surveys to be quickly processed, separating true signals from interference and providing a clearer picture by accurately locating all reflections in their true subsurface positions.

Increasing sophistication in both seismic acquisition and seismic processing techniques allows the direct identification of hydrocarbons in the subsurface. Each interface between separate rock layers will reflect some of the sound waves back towards the surface, and the amplitude of this reflection depends on the nature of that interface. For a porous rock, such as a sandstone, the reflectivity will depend in part on the fluid present in the rock pores, and from this it may be possible to identify oil, gas, and water in the subsurface (Figure 10).

An extension of this concept that is being increasingly applied is the so-called four-dimensional or time-lapse seismic survey. A three-dimensional seismic survey is re-acquired at regular intervals (say, every 3–4 years) over a producing field. The changes in the positions of the reflections indicating the positions of the oil–water contact and the gas–oil contact are used

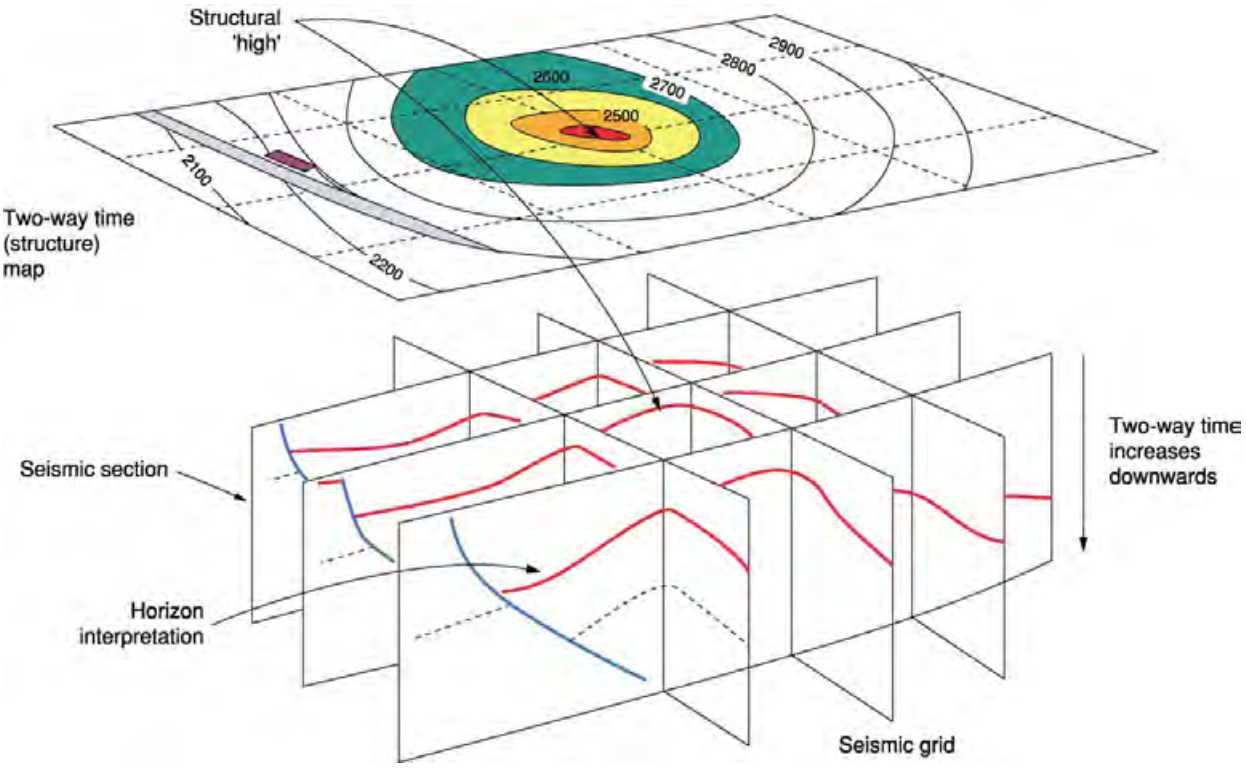


Figure 8 Interpretation of a two-dimensional seismic survey. A geological horizon is traced around all the lines of the grid and then plotted in map view. Reproduced with permission of Nautilus Limited.

to monitor the efficiency with which oil and gas are being produced from the field and can identify areas that are not being drained for future infill drilling (Figure 11).

The increase in computer power that has allowed the acquisition and processing of the vast amounts of data resulting from three-dimensional seismic surveys has also revolutionized the interpretation of these data. Computer-based interpretation systems, termed workstations, allow the profiles to be viewed on colour screens and key geological horizons to be identified (Figure 12).

Once they have been identified through the use of appropriate algorithms, these horizons can be automatically 'picked' by the computer by recognizing their 'character' throughout the rest of the dataset. Information from wells and on the regional geological setting can also be integrated. The resulting interpretation can then be viewed and manipulated in a variety of ways: vertical sections can be viewed in any direction; time and horizon slices can be generated that can be horizontal, dipping, or undulating; maps of horizons can be produced; and the frequency and amplitude of the seismic reflectors can be analysed to give clues as to the nature of the lithology and the fluid content of the subsurface layers. The three-dimensional survey can be considered as a cube of data, and computer visualization techniques allow the data to be viewed from any direction and, indeed, to be rotated on screen (Figure 13). The latest development is the use of a so-called immersive 'visionarium' environment, in which observers are surrounded by the data, viewing it in actual three dimensions using special spectacles (Figure 14).

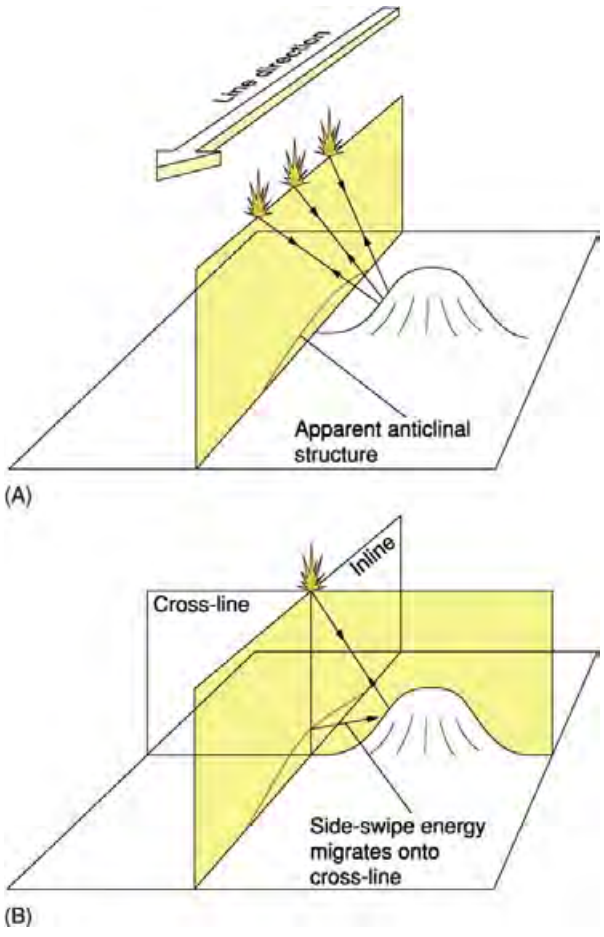


Figure 9 Correct positioning of reflections by the use of three-dimensional seismic surveys. (A) A two-dimensional seismic survey may contain spurious information generated from out of the plane of the profile being acquired. (B) A three-dimensional seismic survey and its subsequent processing, enables all information recorded to be correctly located in the subsurface. Reproduced with permission of Nautilus Limited.

Prospect Appraisal Systems

A number of methods have been developed to attempt to translate complex geological data, derived from the geological and geophysical studies described above, into the set of numbers needed for decision-making. The particular appraisal method used depends on the size of the unit to be assessed (from global to part of a structure) and the amount and type of information available (from sparse regional data to detailed drilling results, or data of a more statistical nature such as the number of wells drilled and the amount of reserves found). The methods can be broadly grouped under three main headings: subjective methods, statistical methods, and deterministic models.

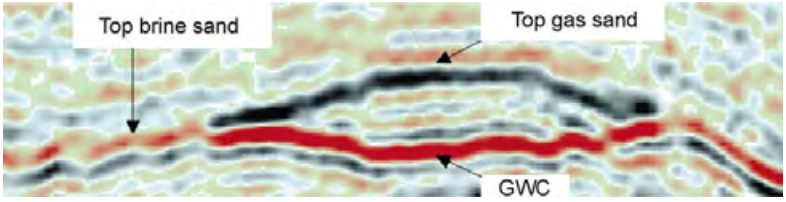


Figure 10 Gas effect seen on seismic showing the change in amplitude of the signal (so-called 'bright spot'). There is a phase reversal at the gas-water contact (GWC) and a time sag. Reproduced with permission of BG Group PLC.

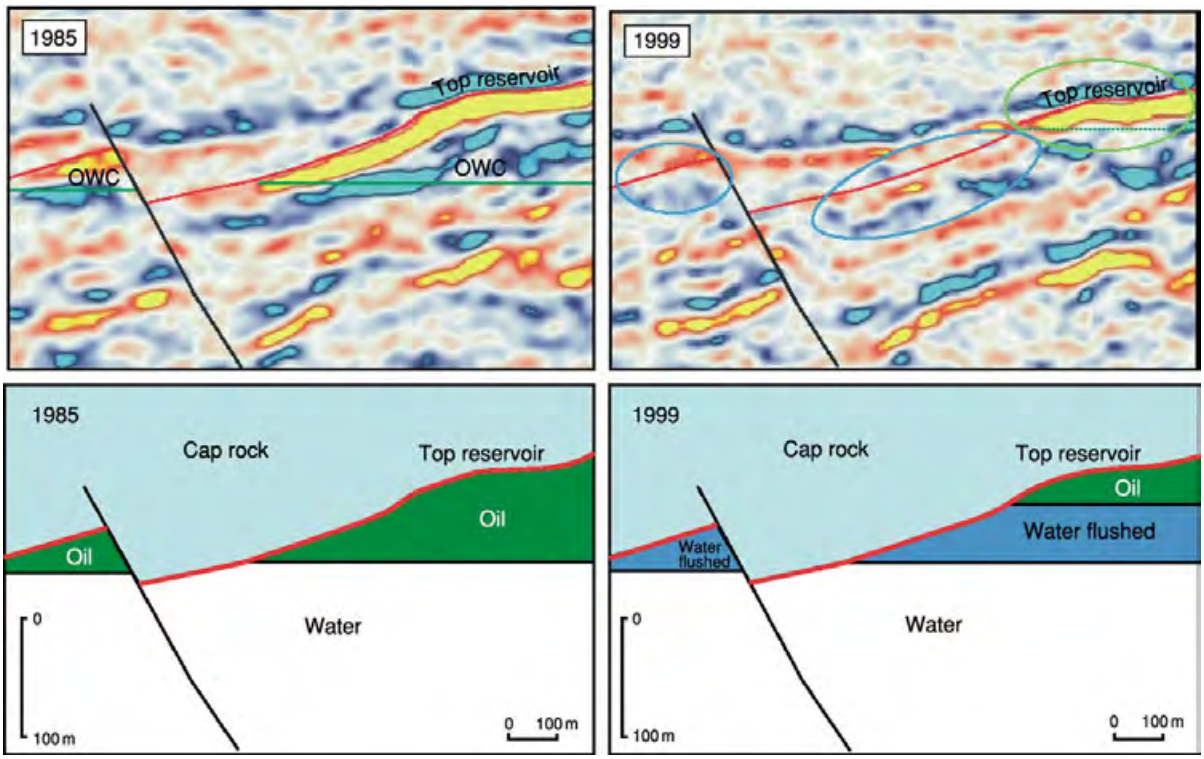


Figure 11 Time-lapse (four dimensional) seismic survey over a North Sea field; the changes in amplitude result from a significant depletion of the oil due to extraction. OWC, oil–water contact. Reproduced with permission of Millenium Atlas Company Limited.

Subjective Methods

Subjective methods result from the implicit intuitive thought processes of a single individual or group of individuals and can be formidable tools when a substantial body of expertise is available. However, there is the problem of bias: usually people are more certain about the unknown than they can afford to be on the basis of the information available, and some people are consistently pessimistic, while others are optimistic. There is also the problem of consistency of knowledge: local knowledge can override possibly more relevant but less well-known worldwide information.

Statistical Methods

Statistical methods attempt to extrapolate past experience using a variety of statistical techniques. Some depend on the validity of an essentially linear extrapolation, for example a hydrocarbon richness per unit volume of sediments in explored basins, which can then be applied to unexplored basins. Other methods attempt to improve this extrapolation by considering past exploration results, such as the decline of success ratio through time and the decline of field size as a function of exploration effort. In the example shown in Figure 15, the larger fields are discovered first (the so-called 'creaming' effect), but, although the average size of the fields found is decreasing, the number found is still increasing.

One of the best known of such studies is that of M King Hubbert, who in 1956 predicted that US oil production would peak in the early 1970s; this proved to be correct (the so-called 'Hubbert's Peak'; Figure 16).

Disadvantages of these methods are that a certain degree of exploration maturity is required for the necessary historical data to be available, the results may be strictly applicable only to the particular basin or province analysed, and the interrelationships of the variables chosen are not always known and correlations may be fortuitous.

Deterministic Models

Deterministic models attempt to integrate all the relevant geological concepts and research findings on the processes of hydrocarbon generation, migration, accumulation, and retention into a coherent model that will provide quantitative estimates of the hydrocarbon volumes likely to be present and the risks involved. These methods are now widely employed in the industry. The problem is that the information available is usually insufficient for precise estimates of the required variables to be made, and, in such

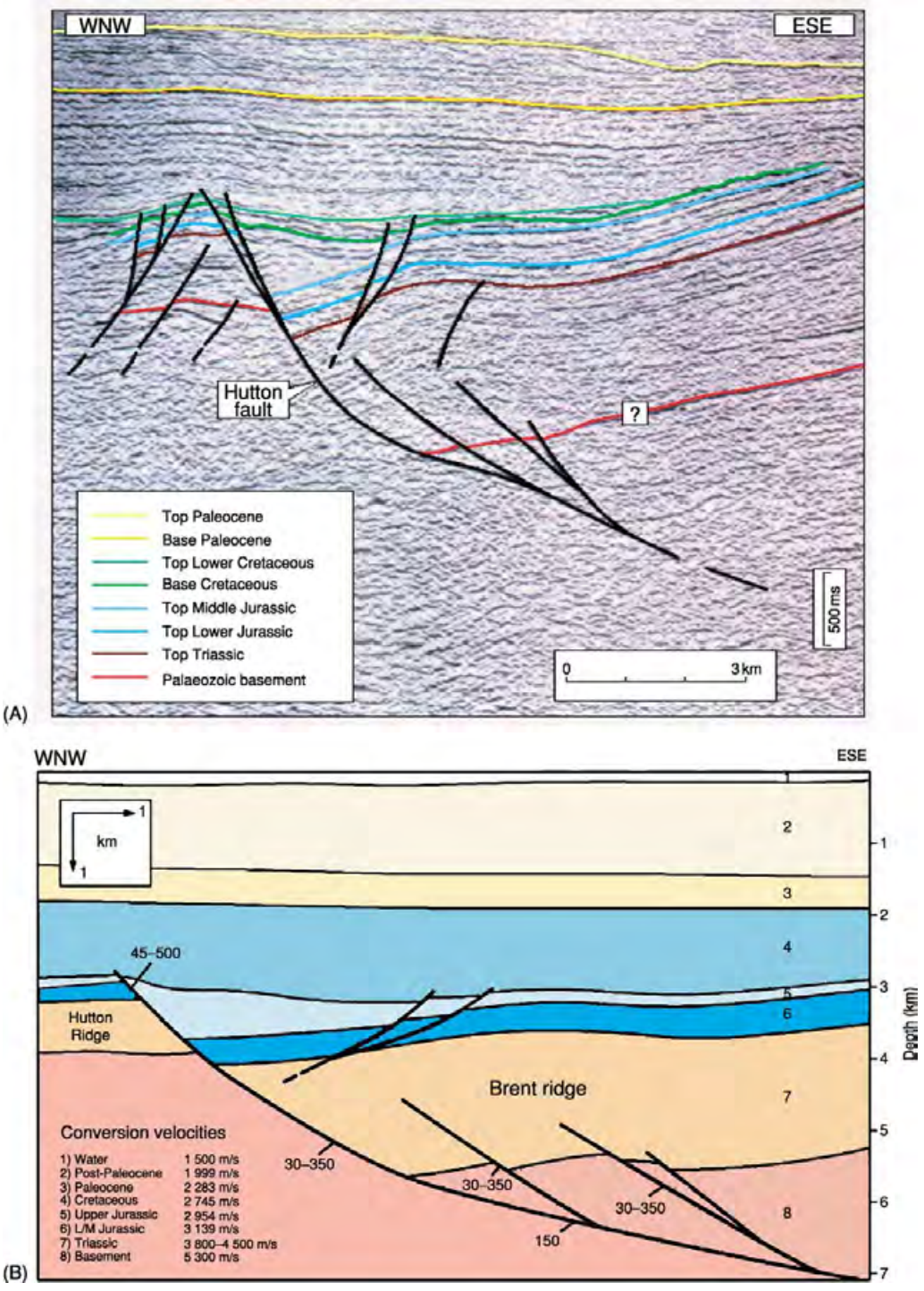


Figure 12 (A) Geological interpretation of a seismic line, and (B) depth converted to show the true fault geometry. Reproduced with permission of Millenium Atlas Company Limited.

cases, a single numerical estimate is often more misleading than helpful. Hence, the inputs will almost always be in the form of probability distributions, and the use of repetitive calculations, each time using an equally likely choice (the so-called 'Monte Carlo' technique; Figure 17) for the input parameters, results in a series of equally likely outcomes, reflecting the uncertainties of both the inputs and the applied formulae. In addition, despite many advances in the understanding of the geological processes involved, any model, although logically consistent and based on the best of present knowledge, is

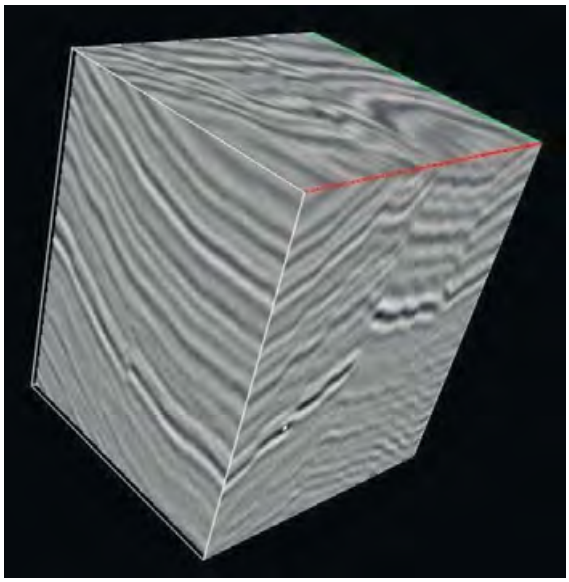


Figure 13 Three-dimensional seismic cube which can be reviewed from any direction, or rotation, on the computer workstation screen. Reproduced with permission of Ikon Science Limited.

likely to be only partially correct because there are variables and processes that are unknown, cannot be measured with sufficient precision, or cannot be accommodated in the model. Hence, it is important to calibrate the output from these models against real-life situations.

Exploration Drilling

Assuming that the geological and geophysical work outlined above has identified a 'prospect' – a possible hydrocarbon-bearing structure – it has then to be decided whether to drill an exploratory well, which is the only way to prove the actual presence of hydrocarbons. Factors involved in this decision are the likelihood of hydrocarbons being present in the prospect, the extent – and hence economic viability – of any hydrocarbons found, the cost of development if hydrocarbons are found, and political and economic considerations relating to the concession and the country in which the prospect is situated.

Figure 14 Three-dimensional seismic interpretation system (so-called 'visionarium').

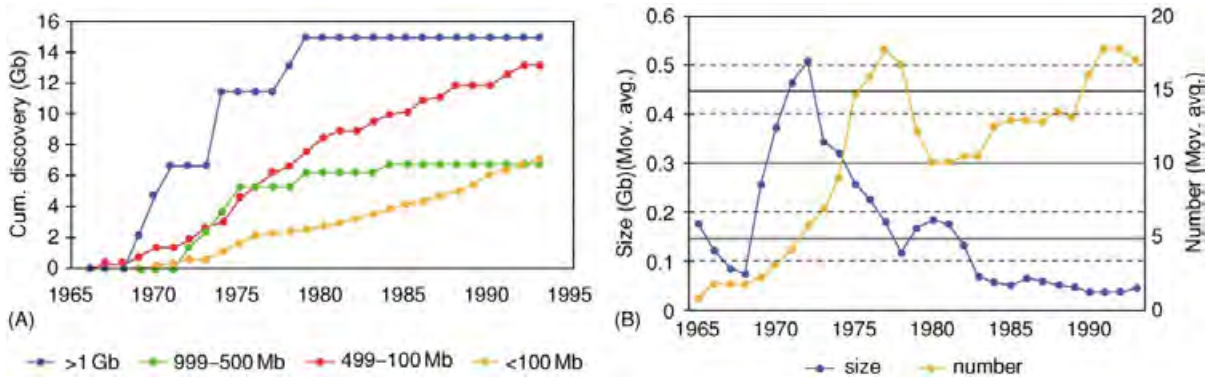


Figure 15 Discovery by (A) field-size classes, showing that the larger fields are discovered first (so-called 'creaming effect'), and (B) average size and number of fields showing that although the size of fields found is still decreasing. North Sea data, both plotted against time.

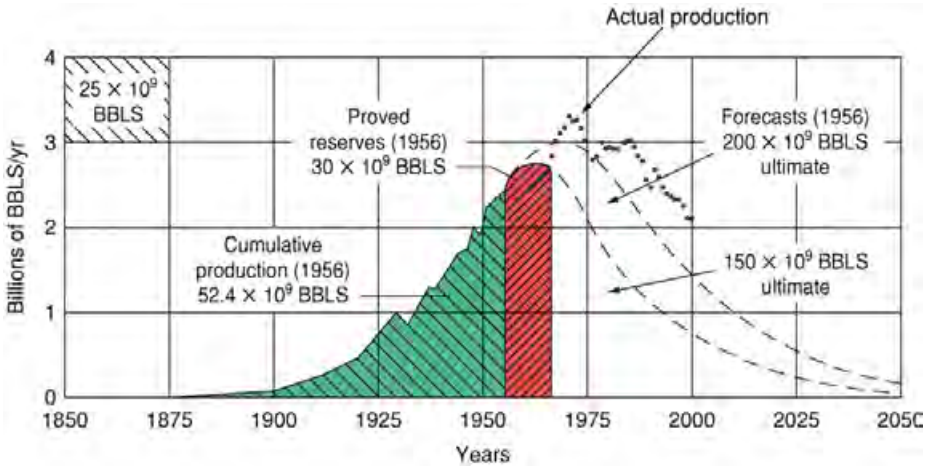


Figure 16 Hubbert's Peak: predicted and actual US oil production.

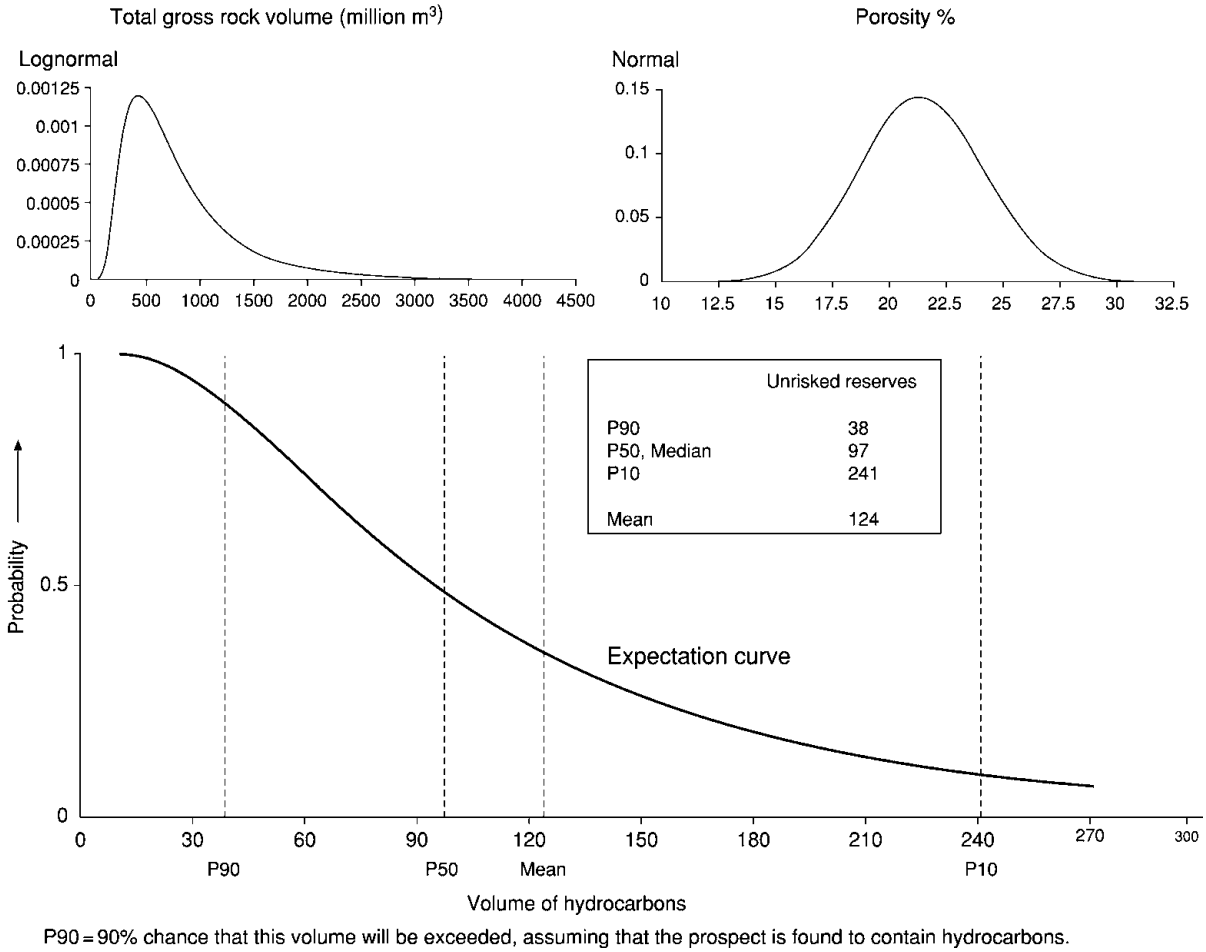


Figure 17 Monte Carlo technique, combining input probability distributions describing total gross rock volume and porosity (illustrated in the top diagrams) together with those describing net-to-gross ratio, hydrocarbon saturation, formation volume factor and recovery factor (not illustrated) to give an expectation curve (cumulative frequency distribution; bottom diagram). Reproduced with permission of Nautilus Limited.

In addition to confirming the presence of hydrocarbons, an exploration well provides information from which the size of the accumulation can be assessed, the development of the field planned, and further exploration in the area undertaken. Rock cuttings brought to the surface by the drilling mud and specially taken core samples contribute to an understanding of the geological history of the sedimentary basin and, specifically, of the nature and characteristics of the reservoir rocks. Wireline logging (Figure 18), using special sondes to measure the electrical, acoustic, and radioactive properties of the rocks, allows key physical properties of the rocks, such as porosity and the fluid content – of oil, gas, or water – within the pore space, to be determined. Other logging tools can detect the presence of fractures (which can be important for the flow characteristics of a reservoir), the dip (or inclination) of the rock layers, and sedimentary features (such as cross-stratification). If this information indicates a potentially commercial accumulation, then a flow test can be undertaken, in which the oil or gas is flowed to the surface: from flow and pressure measurements, further insights into the porosity and, particularly, permeability of the reservoir and the likely performance of wells under operational conditions can be gained.

Further surveys and additional drilling are then undertaken to appraise the size of the accumulation and reduce the geological uncertainties; this information is used to plan the development of the field.

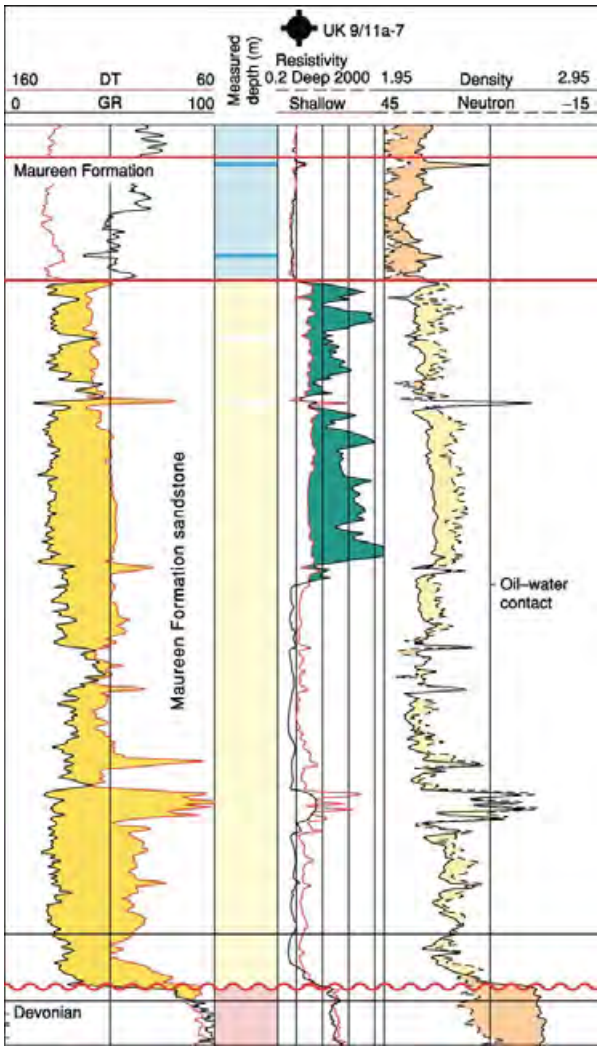


Figure 18 Wireline log response of a hydrocarbon-bearing sandstone reservoir, with a shale seal above and water-saturated reservoir below the oil. DT, interval transit time; GR, gamma ray. Reproduced with permission of Millenium Atlas Company Limited.

Exploration Costs

Onshore, the exploration costs depend very much of the nature of the terrain: operations in remote mountainous or jungle areas are much more expensive than those in flat desert country. Operations in urban and agricultural areas and in areas of environmental sensitivity are expensive, as access can be difficult and strict regulations relating to noise, pollution, and environmental protection must be adhered to. Offshore, costs are less variable, though remote and less climatically favourable locations inevitably attract higher costs (Table 1).

Petroleum Agreements

Apart from the onshore USA, where individuals can have title to the mineral rights, in most countries these rights are vested in the state, and exploration for oil and gas can be carried out only under licences or contracts granted by the state. Petroleum agreements can be divided into two main types: licences (including leases, permits, and concessions) and production-sharing contracts.

Under a licence agreement, a company is granted exclusive rights to explore particular areas. The company finances the exploration campaign and, if the exploration is successful, the development cost and in return is entitled to dispose of the production, sometimes subject to a deduction of a royalty in kind and an obligation to supply the domestic market. The state obtains revenues from the royalty and from taxation, often through special petroleum taxes. A variation on this type of agreement is a joint-venture agreement, where the state shares with the company the risk and expenses of the development phase; generally the company will carry the project at its own cost through the exploration phase.

Table 1 Indicative costs of exploration (for the year 2003) (all values are in US dollars)

Onshore jungle terrain	
Seismic cost per two-dimensional kilometre	7 000
Seismic cost per three-dimensional square kilometre	12 000 (but can be significantly higher)
Well cost (to 3500 m)	3–5 million
Onshore desert terrain	
Seismic cost per two-dimensional kilometre	2 000
Seismic cost per three-dimensional square kilometre	4 000
Well cost (to 3500 m)	1.5–2 million
Onshore urban or agricultural environment	
Seismic cost per three-dimensional square kilometre	12 000
Offshore	
Seismic cost per two-dimensional kilometre	600
Seismic cost per three-dimensional square kilometre	5 000
Well cost (to 3500 m)	6–8 million (but can be significantly higher for deep wells encountering high pressure and high temperature, or in remote locations; up to 25 million)

In a production-sharing agreement, the company carries the exploration risk and funds all the development and operating costs. These costs are then recovered from part of the production, known as 'cost oil'. The remaining 'profit oil' is then split in a pre-determined manner between the state and the company.

A further type of agreement is the service contract, in which the state contracts for a service from the company for which the company receives a fee; this can be, for example, a set fee per barrel produced or a percentage of the hydrocarbons produced while providing the service.

Petroleum agreements take many forms, are increasingly complex, and are often a combination of elements from the different types. The main considerations for the company are the investment at risk, its ability to manage the operations, access to oil in the event of success, and the economic return on the investment.

Postscript

The development of new exploration techniques continues to improve geologists' and geophysicists' understanding of petroleum geology and to increase the efficiency of exploration by allowing wells to be sited more accurately and with a greater chance of success. However, even if the geological conditions for the presence of hydrocarbons are promising, exploration remains a high-risk business and investments are made in exploration many years before there is any prospect of producing the oil: those ventures that are successful must generate sufficient profit to pay for the unsuccessful ventures, both past and future. Thus, the fiscal framework established by states is vital to the commercial success of any exploration venture. The interests of governments wishing to develop their petroleum resources and the interests of companies as risk takers have much in common; petroleum-exploration strategies must take these mutual interests into account.

See Also

Analytical Methods: Gravity. **Petroleum Geology:** Overview; Chemical and Physical Properties; The Petroleum System; Production; Reserves. **Seismic Surveys**.

Further Reading

Deffeyes KS (2001) *Hubbert's Peak: the Impending World Oil Shortage*. Princeton: Princeton University Press.

Gluyas JG and Swarbrick RE (2004) Petroleum Geoscience. Oxford: Blackwell.

Johnson HD and Fisher MJ (1998) North Sea plays: geological controls on hydrocarbon distribution. In: Glennie KW (ed.) *Petroleum Geology of the North Sea: Basic Concepts and Recent Advances,* pp. 463–547. Oxford: Blackwell.

Poelchau HS, Baker DR, Hantschel Th, Horsfield B, and Wygrala B (1997) Basin simulation and the design of the conceptual basin model. In: Welte DH, Horsfield B, and Baker DR (eds.) *Petroleum and Basin Evolution,* pp. 5–70. Berlin: Springer-Verlag.

Selley RC (1996) *Elements of Petroleum Geology,* 2nd edn. San Diego: Academic Press.

Shell Briefing Service (1994) *Upstream Essentials*. London: Shell International Petroleum Company Limited.

Steinmetz R (ed.) (1992) *The Business of Petroleum Exploration*. Treatise of Petroleum Geology. Handbook of Petroleum Geology. Tulsa: American Association of Petroleum Geologists.

Production

K J Weber, Technical University, Delft, The Netherlands
L C van Geuns, Clingendael International Energy Programme, The Hague, The Netherlands

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The geological activities following the discovery of a petroleum accumulation commence with a thorough analysis of the find in terms of structural definition, reservoir rock characteristics, and accumulation conditions. This is followed by the selection of appraisal well locations and the study of the results. If considered commercially viable, a multidisciplinary team, including specialists of all disciplines in petroleum engineering, designs an initial development plan. At this stage, the volume of petroleum-in-place is estimated and a dynamic model is built to test the production capacity and producible reserves of the field, which may be generated by a series of alternative development plans.

After executing the selected initial development plan, the reservoir model is updated with new data. The increasing information on field performance is used to calibrate the model by history matching. In certain favourable cases, fluid movements can be observed on additional three-dimensional (3D) seismic data (*see* **Seismic Surveys**). Further development activities, such as infill drilling, side tracking, recompletions, or fluid injection schemes, are planned on the basis of the updated model. Detailed field studies by integrated teams of petroleum engineers and production geologists are carried out periodically until the ultimate stages of the field life (Figure 1).

Reservoir performance is a function of the reservoir characteristics and the petroleum properties. The study of reservoir characteristics is the primary task of the production geologist. An overview is given of the influence of various reservoir heterogeneities on fluid flow. To analyse these characteristics, and to quantify the results, requires a wide range of techniques. These cover such a large field of subjects that the work has to be carried out by a group of specialists in seismology, sedimentology, structural geology, geostatistics, and petrophysics. In addition, a good understanding of the principles of reservoir and production engineering is required to be able to function successfully in a multidisciplinary team.

The production geological activities at the various stages of field development have a number of different purposes which all require geological input of some kind (Figure 2). This is illustrated by the different techniques and modelling methods that are employed at each stage (Figure 3). The pace of development of new techniques, in particular integrated computer systems, is very fast. As a result, the production geological activities are centred around workstations being served by a database containing all relevant information (Figure 4).

Historical Development of Production Geology

Production geology as a separate petroleum engineering discipline was only generally recognized in the 1950s. Early field development, with shallow, closely spaced wells and only cuttings and surface outcrop observations, provided little scope for reservoir studies. In the USA and Russia, where grid drilling was and still is in fashion, most companies did very little production geology work in the appraisal stage. In Indonesia, where complicated structures were encountered, the need was felt for more geologist involvement at an early stage.

Around 1915, the Bataafse Petroleum Company (Royal Dutch) started to engage mining engineers with the particular task to improve cooperation between production and geology to achieve a more efficient exploitation of the field. The reports from that time show that structural interpretation and detailed mapping of the producing horizons constituted the main tasks. In these pre-logging and pre-seismic days, correlations were carried out with the aid of cuttings glued to narrow planks.

The wireline logs developed by the Schlumberger brothers in 1927 progressively replaced these early methods. Just before World War II, effective reflection seismic methods came into use, although any sophisticated processing had to wait for the availability of more powerful computers after 1960. This is also the time that production geology, also called 'development geology', began to be more generally employed, although, in most companies, the geologists were still firmly attached to the exploration departments.

In the meantime, sedimentology had made great strides forwards and more attention was given to the internal architecture of the reservoir. Core descriptions and analysis were improved. Research studies of outcrops and recent sedimentation of both

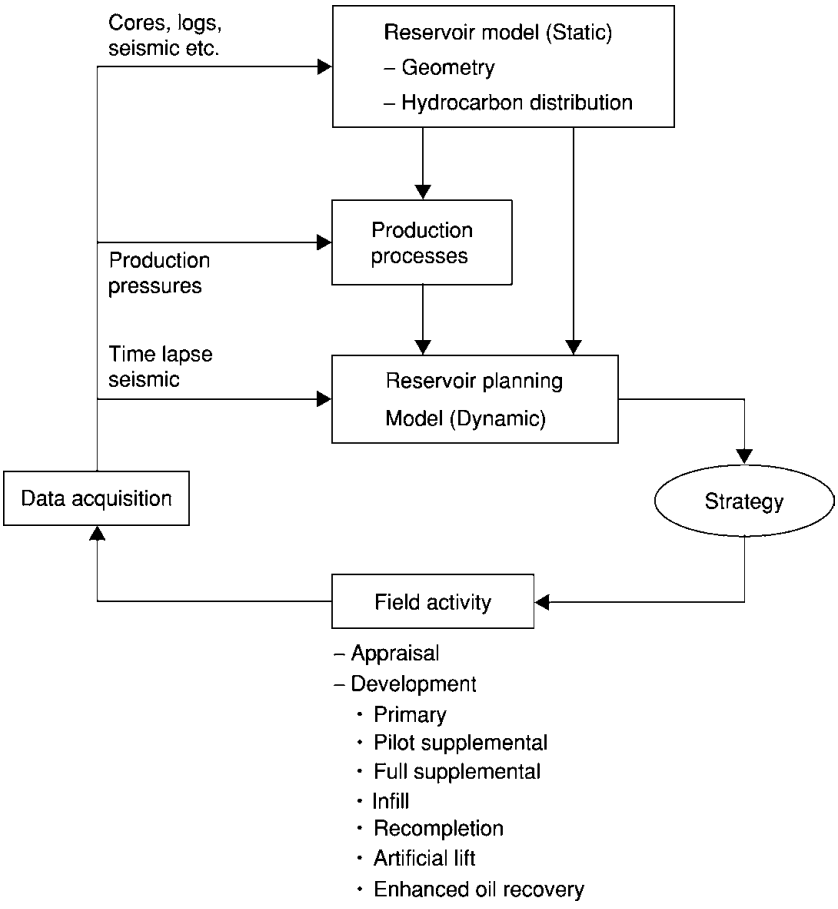


Figure 1 Field study cycles.

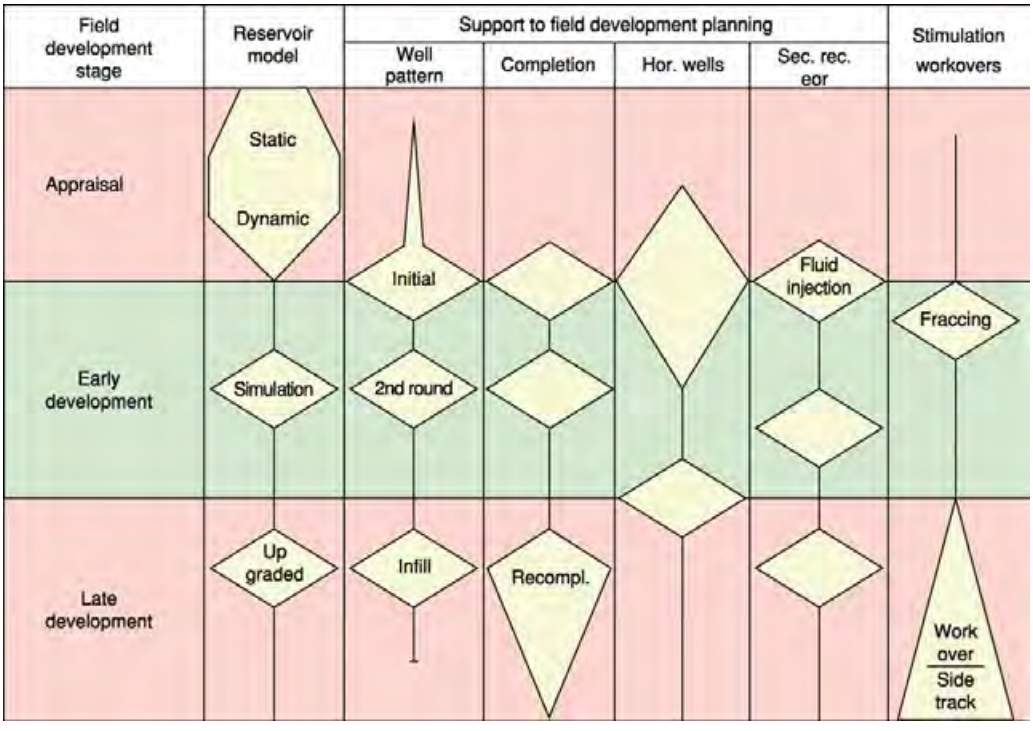


Figure 2 Purposes of reservoir characterization. Courtesy of Shell.

Field development stage	Seismic techniques	Core/log techniques	Modelling	
			2D	3D
Appraisal	Structure Fault sealing Hydrocarbon indications Sequence stratigraphy	Facies Poro/Perm Pore geometry Lithoclass (Cores) Lithoclass (Logs)	Stratigraphy Facies distribution	Probabilistic → Sensitivity → Connectivity
Early development	Reservoir architecture Lithology Porosity	Fracture types + orientation Borehole imaging	Well-to-Well Correlations	Hybrid model → Simulation input
Late development	Time lapse Well-to-Well Tomography	Sand body geometry Fault sealing	Distribution of saturation	Deterministic model → Remaining oil distribution

Figure 3 Reservoir modelling techniques during successive development stages. Reproduced with permission from Weber KJ (1995) Visions in reservoir management – what next? In: *Reservoir Characterisation: Integration of Geology, Geophysics and Reservoir Engineering, The Third JNOC-TRC International Symposium, February 2–23, Technical Research Centre, Chiba, Japan*, pp. 1–15. Chiba: Japan National Oil Corporation.

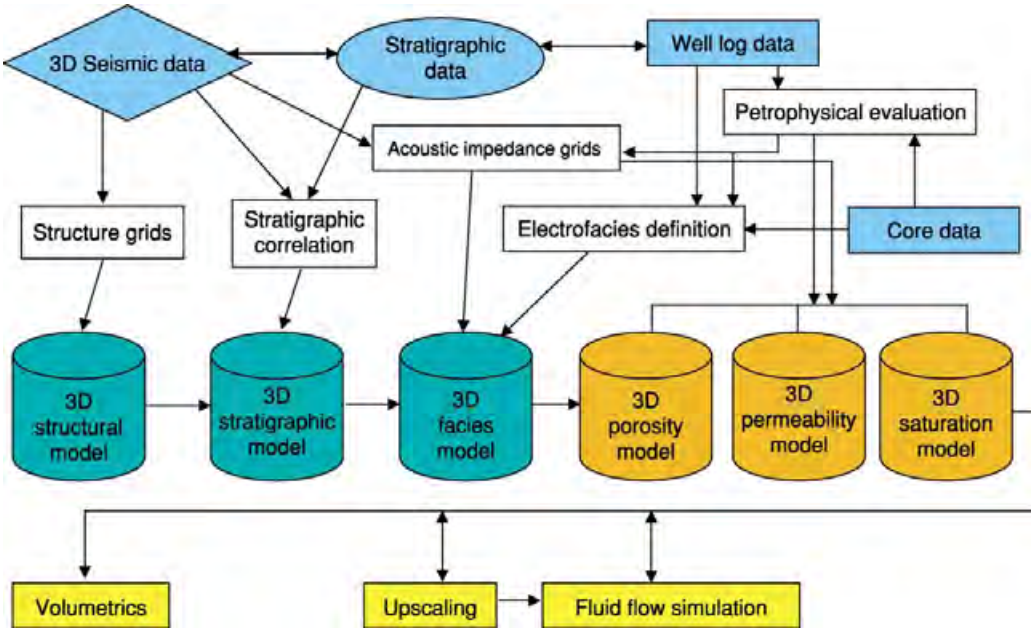


Figure 4 High-level workflow for computerized reservoir modelling. Courtesy of Shell.

clastics and carbonates were carried out by several major oil companies, in combination with laboratory studies of diagenetic processes. Rock classification systems such as the 'Archie classification' appeared, providing the link between rock texture and permeability. The introduction of induction-, latero-, formation density- and dip-meter logs greatly improved the quantitative assessment of hydrocarbon volumes

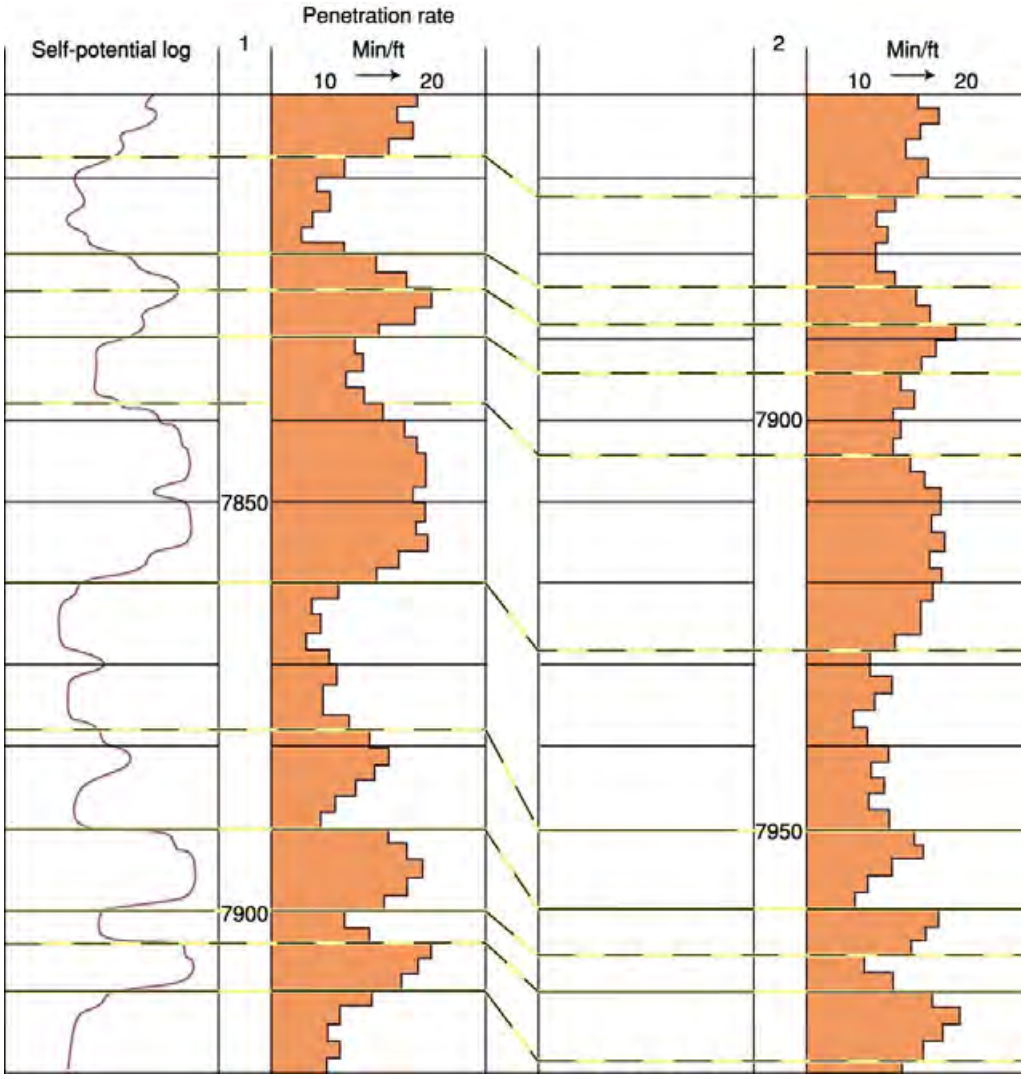


Figure 5 Penetration log correlation.

and distributions. Between 1960 and 1980, most of the presently used production geological methods were developed.

As a typical example of a classical production geological method, Figure 5 shows a penetration log correlation. The different rates of penetration of sandstones and shales (whilst drilling with constant speed and weight) can often be related to spontaneous potential (SP) or gamma-ray (GR) log curves, and thus the progress of wells could be monitored prior to the introduction of modern logging-while-drilling (LWD) much later. Further refinements in structural and reservoir imaging came with the spectacular advance of seismic acquisition, processing, and interpretation. Seismic attribute analysis is now one of the major production geological techniques. Sophisticated borehole imaging logs formed another new source of detailed information. The increase in computer capacity also led to reservoir simulation based on complex heterogeneous reservoir models comprising very large numbers of grid cells, which, in turn, required realistic well-calibrated geological models. This promoted the use of geostatistical methods and stochastic modelling. The present trend is to use a combination of all relevant reservoir data in integrated databases from which visualization can be generated, simulation models can be constructed, and well trajectories can be planned.

Reservoir Characteristics Influencing Fluid Flow and Storage Capacity

Controlling Reservoir Characteristics

The quality of a reservoir is directly related to the available pore space (porosity) and the capacity to

conduct fluid flow through the pores (permeability). Porosity depends on texture (grain size distribution, grain shape, and packing) and the degree of diagenesis (compactions, leaching, cementation). Permeability is a function of the size, shape, and distribution of pore spaces and their interconnectedness. Rocks with a similar texture that have been subjected to similar diagenetic processes will generally show a good correlation between porosity and permeability. However, for the same porosity, significantly different permeabilities may be found in the same reservoir because of facies-related textural differences and/or diagenetic variations.

There are a series of geological features that influence fluid flow and hydrocarbon recovery (Figure 6). Faults, fractures, reservoir architecture, internal zonation of genetic units, low permeability intercalations, sedimentary structures, and pore shape all have an impact on a different scale. The drive mechanism can also be influenced by these factors because they control the pressure support from the aquifer surrounding the reservoir. An accurate evaluation of a reservoir has to take into account all of the relevant heterogeneities at the right scale. The resulting 3D reservoir model forms the basis for volumetric estimates of the trapped hydrocarbons and simulations of

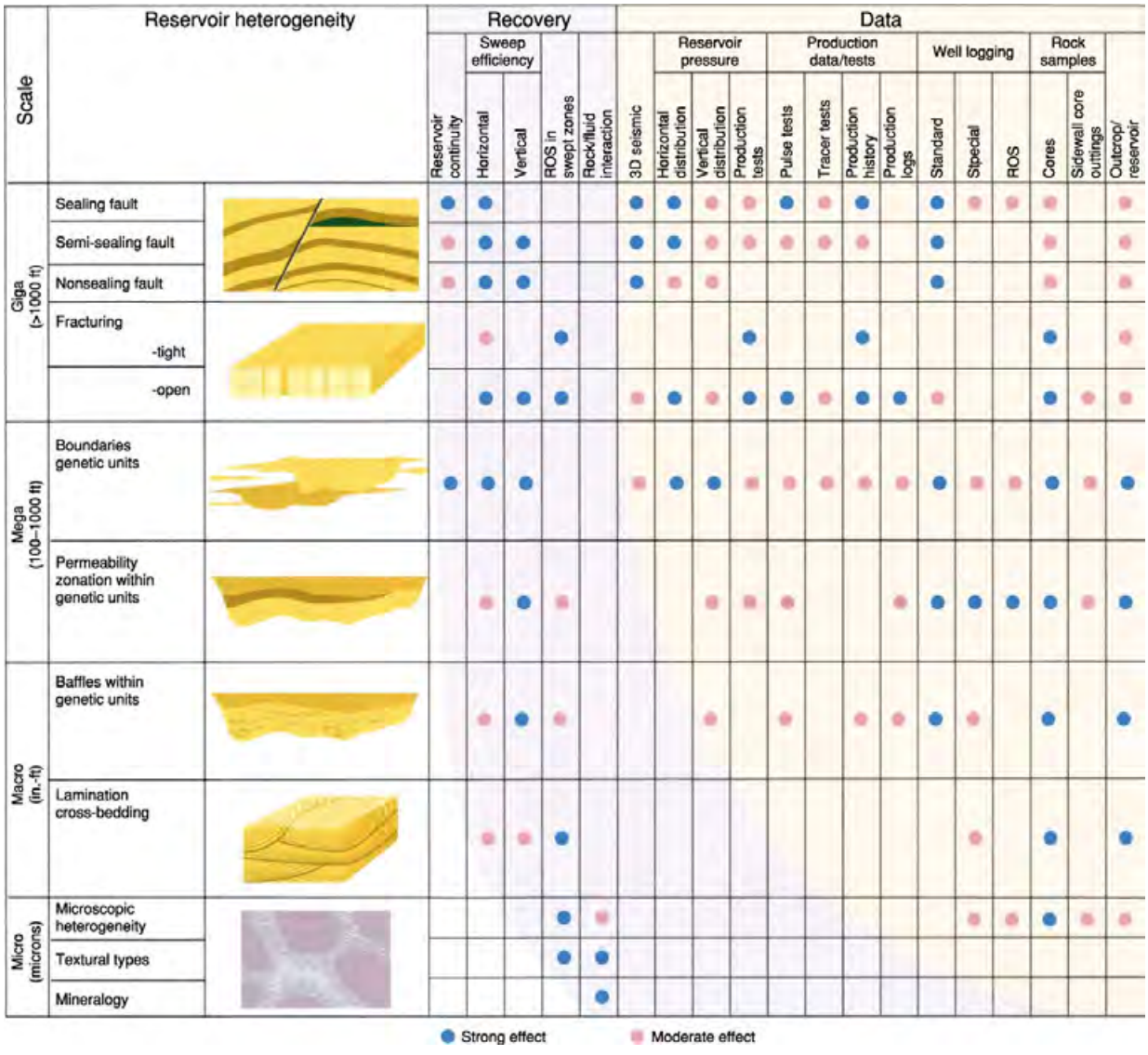


Figure 6 Reservoir heterogeneity influencing fluid flow and hydrocarbon recovery, and relevant data for analysis. ROS, Residual Oil Saturation. Modified with permission from Weber KJ and van Geuns LC (1990) Framework for constructing clastic reservoir simulation models. *Journal of Petroleum Technology* 42(10): 1248–1253, 1296–1297. © Society of Petroleum Engineers.

reservoir performance. In this way, the appraisal campaign can be planned and the optimum development scheme can be determined.

Faults and Fractures

Faults have a significant influence on hydrocarbon migration and trapping. The sealing capacity of a fault can be controlled by cataclasis and diagenesis in the fault zone, juxtaposition against a tight layer, or smearing of clay along the fault. In deltaic settings, where multiple reservoirs separated by shales are common, the determination of the clay smear potential along faults is particularly useful.

Fractures can have a positive or a negative effect on fluid flow. Cemented fractures compartmentalize the reservoir and reduce recovery. Swarms of shear fractures are common along many faults and can significantly reduce productivity over a wide zone. Open fractures, which are mainly associated with folded carbonate reservoirs, can increase productivity enormously, although they may also be the cause of early water breakthrough to the wells.

Clastic Reservoirs

Clastic reservoirs occur in a wide variety of sedimentary settings, ranging from almost homogeneous beds to an intricate 3D network of sinuous sand bodies. From an observation of the existing reservoir configurations, it appears that most clastic reservoirs can be classified into one of three basic architectural types – layer-cake, jigsaw, or labyrinth – which represent a decreasing order of connectivity (Figure 7). Correlation is comparatively easy in layer-cake reservoirs, such as barrier bars or shallow marine sheet sands. Thus, such reservoirs can already be modelled quite realistically in the appraisal stage. At the other end of the scale are the labyrinth-type reservoirs, such as low-sinuosity distributary channel complexes, which require a dense well spacing to achieve a deterministic correlation.

If we add common internal permeability distributions and discontinuity patterns, the impact of the reservoir types on sweep efficiency can be shown (Figure 8). In layer-cake reservoirs, not only the sands but also the shales can be very continuous, preventing vertical flow. High permeability at the base of point bars can lead to early water breakthrough, while the reverse, as in barrier bars, can have the opposite effect. Jigsaw reservoirs, with interlocking sand bodies with contrasting properties and local baffles to vertical and horizontal flow, often show uneven drainage and erratic flood fronts. Labyrinth reservoirs reflect their discontinuous nature in very uneven sweep efficiency and bypassing of major oil volumes.

Carbonate Reservoirs

Carbonate reservoirs are generally more heterogeneous than clastic reservoirs, because the original organic sediments can differ widely in their grain size and composition, but also because of the large influence of diagenesis. The original sedimentary setting can be split into four main types: the deeper chalks and turbidites; the shallow-water build-ups; platform/shoal sediments; and ramp/nearshore packages (Figure 9). These reservoirs are sometimes left with rock characteristics closely related to their original properties, but often considerable changes are effected by the various diagenetic processes. However, in the majority of cases, the original material with a similar texture and composition will be transformed into the same rock type. There may be a complete reversal of the original porosity and permeability distribution. A porous reef may become tight and a low-permeability, fine-grained back-reef may change into a permeable sucrose dolomite (Figure 9).

Several diagenetic processes can increase the heterogeneity markedly. Karstic leaching can be pervasive or restricted to certain zones. Some carbonate reservoirs consist of virtually tight rock in which leached fractures and voids comprise the only effective pore space. Another complicated type, which is very difficult to evaluate, is formed by reservoirs submitted to a mixed process of leaching and dolomitization, forming a network of moldic sucrose dolomite nodules and tunnels through a tight matrix.

For nearly all carbonate reservoirs, detailed microscopic studies are required to unravel their diagenetic history and to identify facies and composing organisms. The rock types have to be classified to be able to analyse and predict the permeability distribution, which, in carbonates, can cover an extremely large range even in a single reservoir (Figure 10).

Small-Scale Heterogeneity

Although the larger scale heterogeneities usually control the production performance, the sedimentary structures and rock texture can cause anisotropy of the reservoir and have a strong influence on residual oil distribution. In cross-bedded reservoirs which have been subject to severe diagenesis, it may be necessary to measure the permeability contrast between the laminae and to take the resulting anisotropy into account in the dynamic fluid flow model. It has also been shown that, even with permeability contrasts between the laminae as low as two, capillary forces can trap of the order of 10% additional oil in comparison with more homogeneous rock. Detailed studies of the capillary pressure curves and relative

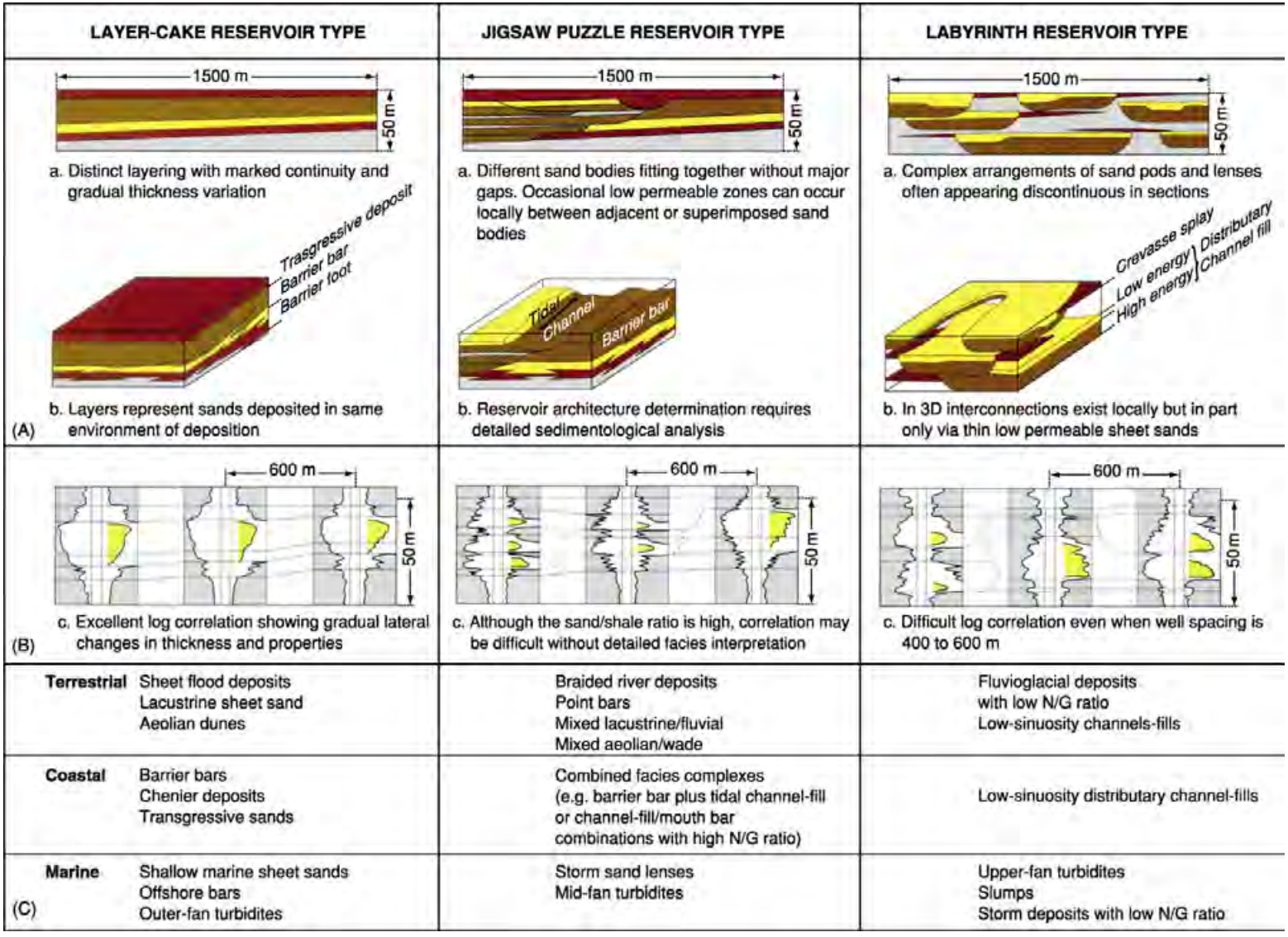


Figure 7 Clastic reservoirs classified into three architectural types: layer-cake, jigsaw, and labyrinth. N/G ratio, Net/Gross. Modified with permission from Weber KJ and van Geuns LC (1990) Framework for constructing clastic reservoir simulation models. *Journal of Petroleum Technology* 42(10): 1248–1253, 1296–1297. © Society of Petroleum Engineers.

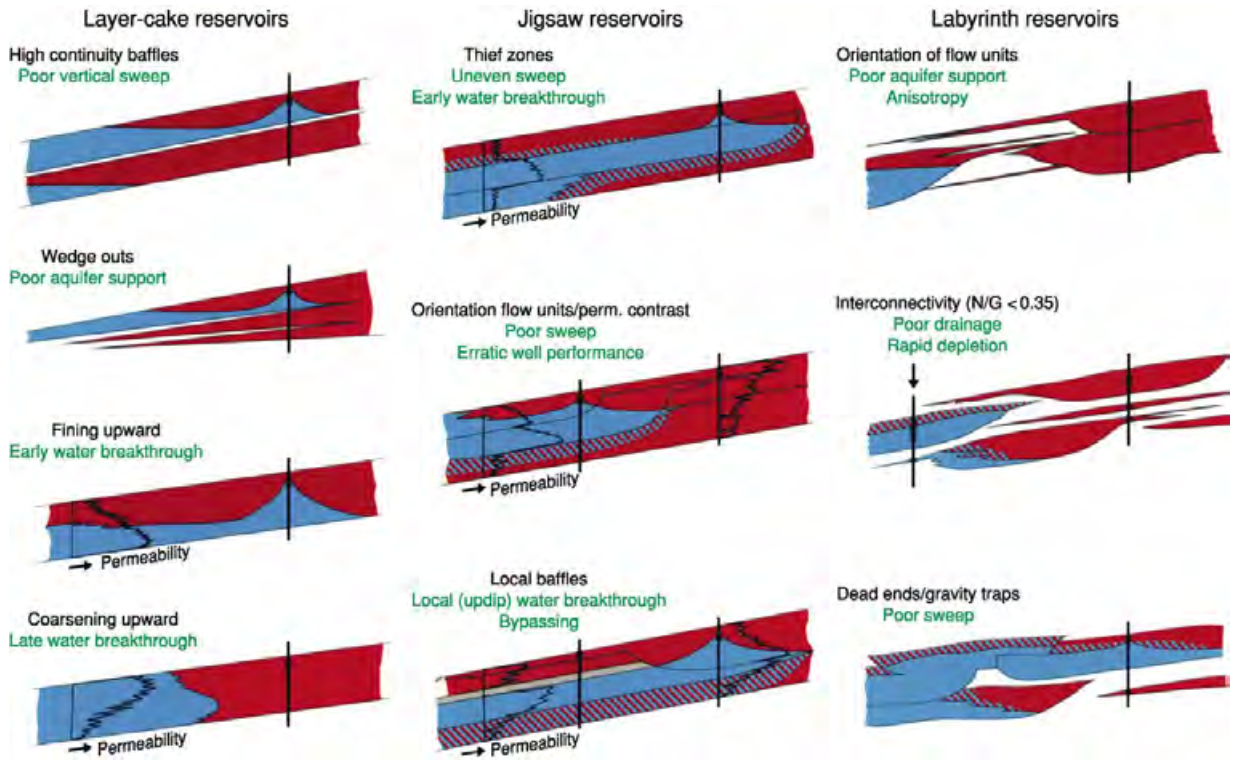


Figure 8 Connected permeable pathways through the major reservoir types. Courtesy of Shell.

permeability to oil and gas for each rock type are necessary. Major lateral differences in apparent oil–water contact can result from variations in rock type.

Techniques

Seismic Methods

The range of methods, techniques, and tools used in production geology has become very large. Seismic techniques are of primary importance in the appraisal stage, but are also used for reservoir monitoring (Figure 11). The traditional use of seismic techniques for structural definition has been much improved by 3D surveys and better processing techniques. Where seismic resolution is good and acoustic impedance contrasts are sufficiently large, amplitude measurements along the beds can reveal depositional patterns as well as oil and gas accumulations (Figure 12).

Besides acoustic impedance mapping, amplitude versus offset (AVO) techniques and shear-wave analysis are now being employed. These techniques are powerful methods to determine lithology, while shear waves are affected by open fractures. Repeat 3D surveys (so-called 4D seismic) can reveal movements of fluid interfaces and saturation changes (Figure 12). Analysis of changes in acoustic impedance and velocity can also be used to estimate pressure changes in the reservoir resulting from hydrocarbon production. This can reveal compartmentalization caused by sealing faults and uneven aquifer drive. Rock stresses can be derived from measurements of shear-wave polarization.

In favourable cases, porosity and pore-fill analysis is carried out via acoustic impedance measurements, calibrated with well data. In this way, volumetric estimates of hydrocarbon-in-place may be made. Of particular importance is to relate the 3D seismic reservoir models to models made from well data correlations with geostatistical methods. Improved calibration of seismic data, attributed using neural network techniques, can provide estimates of facies types and related properties throughout the reservoir. Seismic methods are frequently used for planning well trajectories (Figure 12). Three-dimensional displays and interactive virtual reality techniques are employed for this purpose. The developments in seismic methods for reservoir analysis have been so spectacular in the recent past that further research is likely to result in important advances.

Core Description and Analysis

Facies and rock types have to be identified and characterized from cores and wireline logs. Systematic

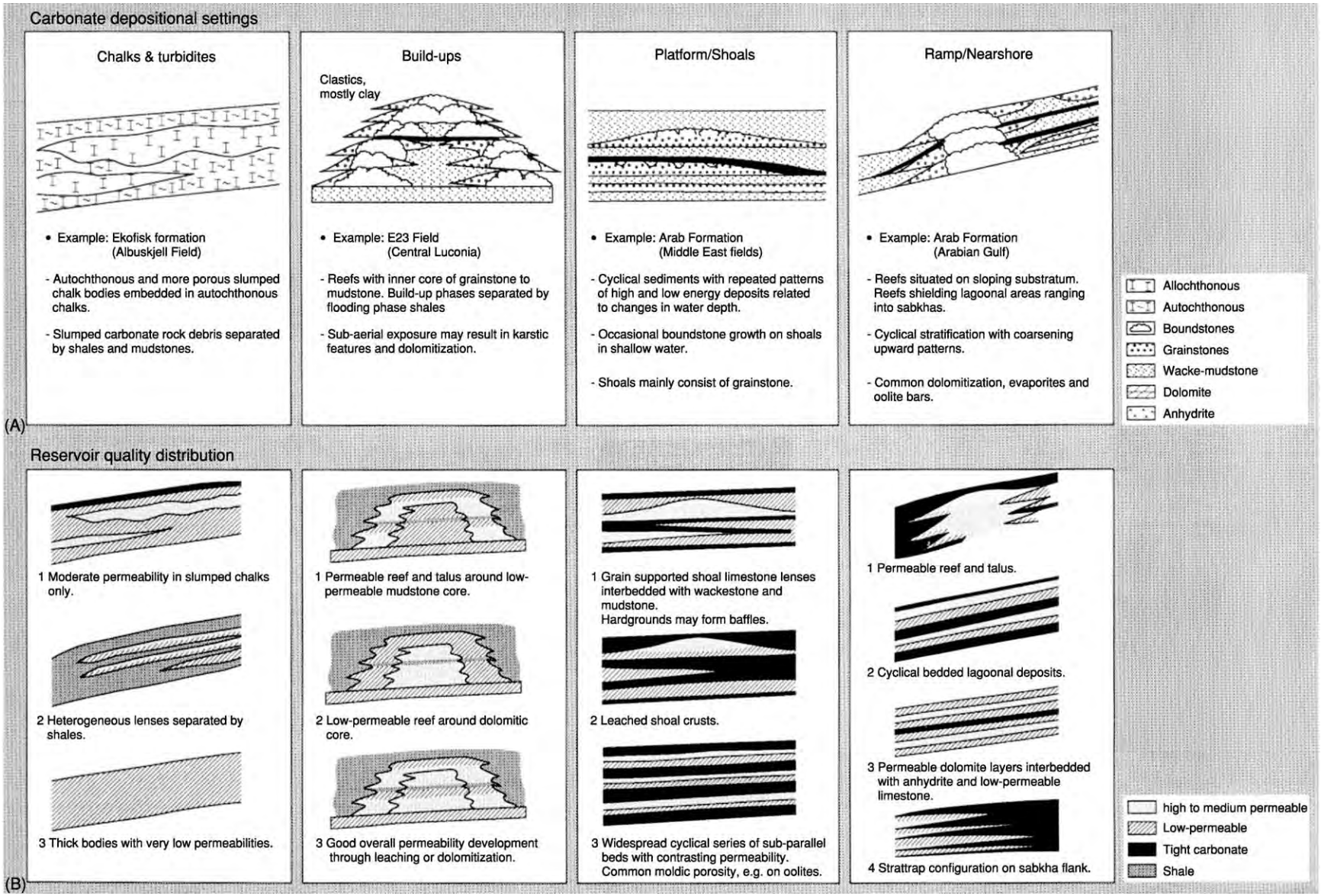


Figure 9 Scheme of the main carbonate depositional settings and common pattern of permeability distribution. Courtesy of Shell.

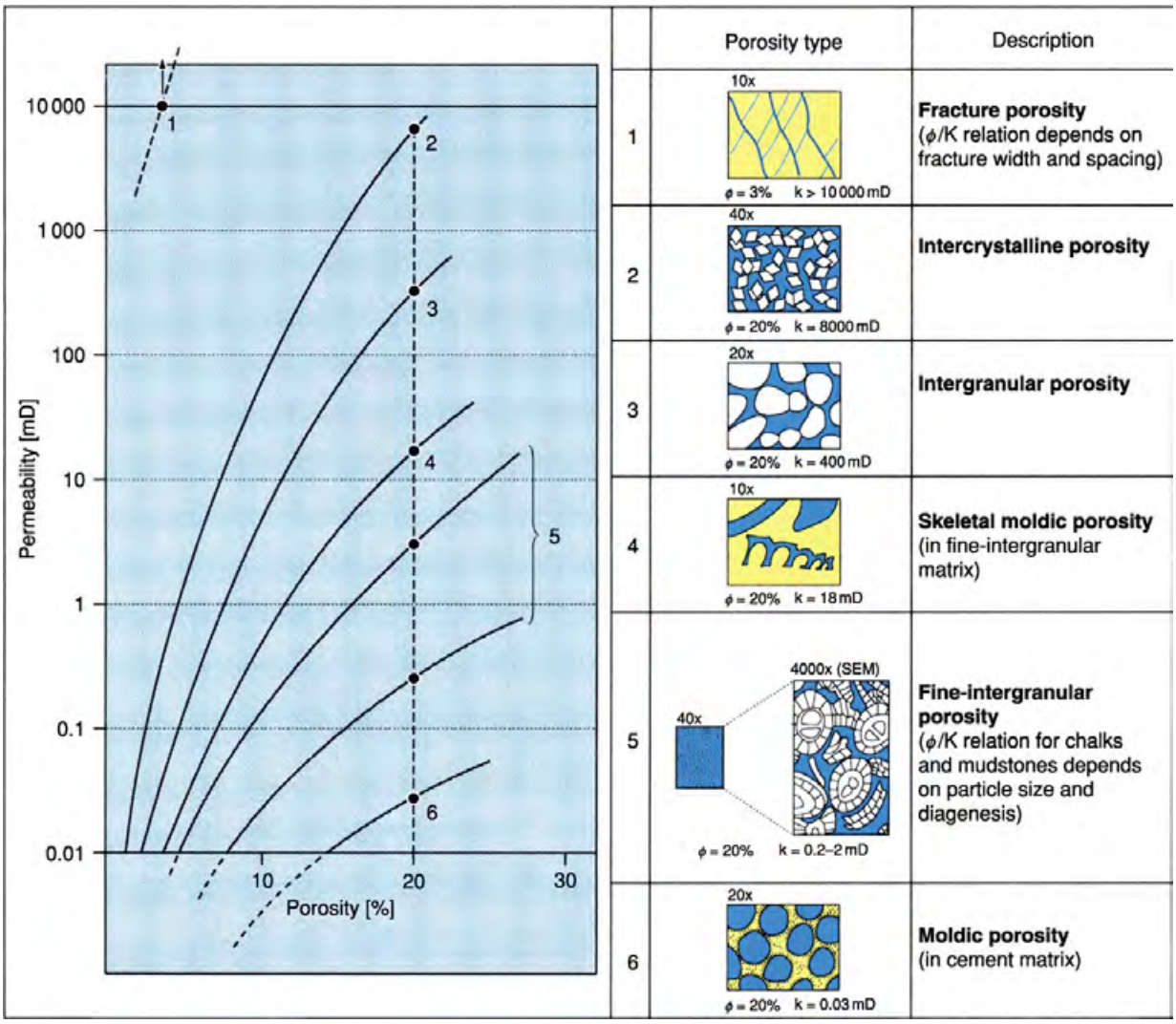


Figure 10 Generalized trends for porosity versus permeability of the main primary and secondary carbonate pore types. Courtesy of Shell.

geologically steered core analysis is the key to the proper understanding of the reservoir composition and the basis for log calibration, static and dynamic modelling, detailed seismic interpretation, and the identification of key heterogeneities. An overview of core analysis methods and applications is given in Figure 13. Core observations under normal light can be augmented by ultraviolet (UV) and infrared photography to reveal remaining oil and sedimentary structures. Standard petrophysical analysis should be carried out with a sample choice and spacing reflecting the reservoir heterogeneity. Probe permeability measurements or cube-shaped samples may be required to analyse permeability anisotropy (Figure 14). The influence of *in situ* stress on reservoir properties and rock compressibility has to be measured.

The core data are used to establish rock types with similar characteristics, such as porosity/permeability relationship, capillary pressure, and relative permeability curves. Each rock type usually represents a specific facies type having a similar lithology and texture. The core data are correlated with the logs to identify typical characteristics that allow the recognition of rock types in uncored wells. This is necessary to delineate facies distribution for well-to-well correlation, as well as to estimate permeability via porosity–permeability transforms. High-resolution borehole imaging logs are very useful for the identification of structural features, sedimentary structures and dips, and trends of individual sand bodies (Figure 15).

A series of electric logs is used to identify rock types in uncored wells. Linear multiple regression of, for example, GR, neutron and density log readings of

Field development stage	Modelling objectives	Seismic attributes	Techniques
Appraisal	– Structure	Reflection time/continuity acoustic velocities	3D survey dip/az plotting shaded relief
	– HC distribution	Acoustic impedance	Mapping impedance, avo
	– Reservoir architecture	Reflection cont./strength	Seismic horizon mapping
	– Porosity distribution	Acoustic IMP/velocities	Acoustic IMP, modelling
	– Volumetric estimates	Acoustic impedance	Finite element impedance distribution determination
	– Fault sealing capacity	Acoustic impedance along fault planes	Plotting impedance slices along faults
Early development	– High precision structure	Detailed acoustic velocity model	Pre-stack migration
	– Fracture distribution	Shear-wave polarization	Measuring shear-wave splitting and attenuation
	– Lithology	Poisson's ratio, velocities acoustic impedance	Shear and compressional wave analysis, avo
	– Facies identification	Acoustic impedance well data	Segmentation, neural network methods
	– Constraining stochastic reservoir models	Impedance amplitude patterns, well data	Geostatistical approach
	– Steering wells	Reflection time/continuity	Wsp (walkaway survey) look-ahead vsp, lwd
Late development	– Movements of HC contacts	Acoustic impedance	Time-lapse seismic, avo
	– Pressure distribution	Acoustic IMP/velocities	Time-lapse seismic
	– EOR steam drive monitoring CO2	Acoustic IMP/velocities poisson's ratio	Cross-well seismic tomography
	– Stress distribution	Shear-wave polarization	Shear-wave polarization analysis, neural networks

Figure 11 Overview of seismic techniques for reservoir delineation, analysis, and monitoring. AVO, amplitude versus offset; EOR, enhanced oil recovery; LWD, logging-while-drilling; VSP, Vertical Seismic Profile. Modified with permission from Weber KJ (1995) Visions in reservoir management – what next? In: *Reservoir Characterisation: Integration of Geology, Geophysics and Reservoir Engineering, The Third JNOC-TRC International Symposium, February 2–23, Technical Research Centre, Chiba, Japan*, pp. 1–15. Chiba: Japan National Oil Corporation.

known rock types in cored well is used to set up a diagnostic system for uncored wells. Alternatively, neural networks can be trained to carry out this identification. The advantage of the neural network techniques is that the influence of the input parameters is non-linear.

Permeability Distribution

Permeability measurements are amongst the most difficult to derive from well data and yet comprise one of the most useful and relevant measurements in production geology. The permeability distribution in the well is derived from cores and via rock type identification in uncored wells. However, this only provides local horizontal permeability data around boreholes. Cores and logs in horizontal holes can provide lateral permeability development, but only along narrow zones. There are a large number of techniques to augment the permeability database both horizontally and vertically on a range of scales (Figure 16). Production geologists must play an active role in identifying opportunities to measure the permeability, which provides an important control on fluid flow in the reservoir. Vertical permeability on the medium and large scale is particularly difficult to estimate from cores and logs. For this purpose, multiprobe well testers, pulse testing, production logging, and pressure build-up analysis are used.

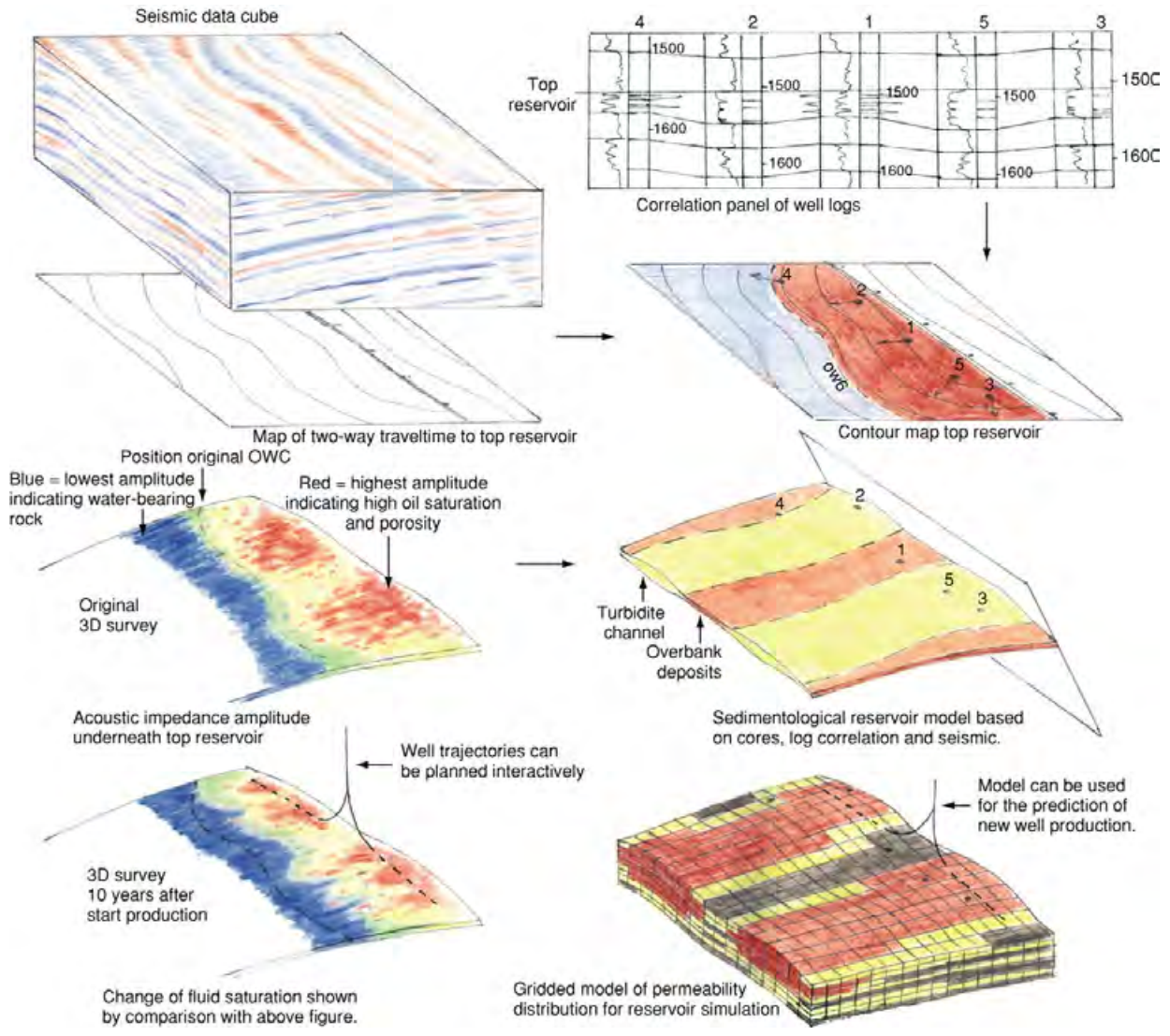


Figure 12 Integrated use of three-dimensional (3D) seismic measurements for static modelling purposes.

Fault and Fracture Analysis

Faults are mainly analysed from seismic data, although well-to-well correlation can reveal fault cut-outs. Dip-meter logs can also indicate fault positions by showing drag patterns. The fault types have to be analysed from the deformation pattern seen on seismic measurements and the structural history of the sedimentary basin. The associated sealing capacity of the faults is closely related to the type of fault, fault throw, reservoir lithology, and depth. Cataclastic fault zones tend to be cemented and impermeable after burial to a depth of 1–2 km.

Clay smearing along normal faults is an important process in many deltaic reservoirs. Estimates of the sealing potential can be based on the shale thickness that passed along a fault plane (Figure 17). This can be achieved by a log analysis of wells near the fault and the fault throw derived from the cut-out observed in a well intersecting the fault. If good quality seismic data are available, however, the fault throw distribution can be established. With 3D seismic data, it is possible to sample the acoustic impedance values along both flanks of a fault and to translate these data into synthetic lithological logs. By comparing and correlating the hanging- and footwall stratigraphy, both the 3D distribution of the fault throw and the thickness of the shales that moved across any point of the fault surface can be obtained. The method requires local calibration because the ductility of the shales at the time of faulting can differ markedly.

Fractures are seen on borehole imaging logs. By the use of Stoneley waves (shear waves), open fractures can be identified and the fracture width can be estimated. This provides estimates of the usually very low

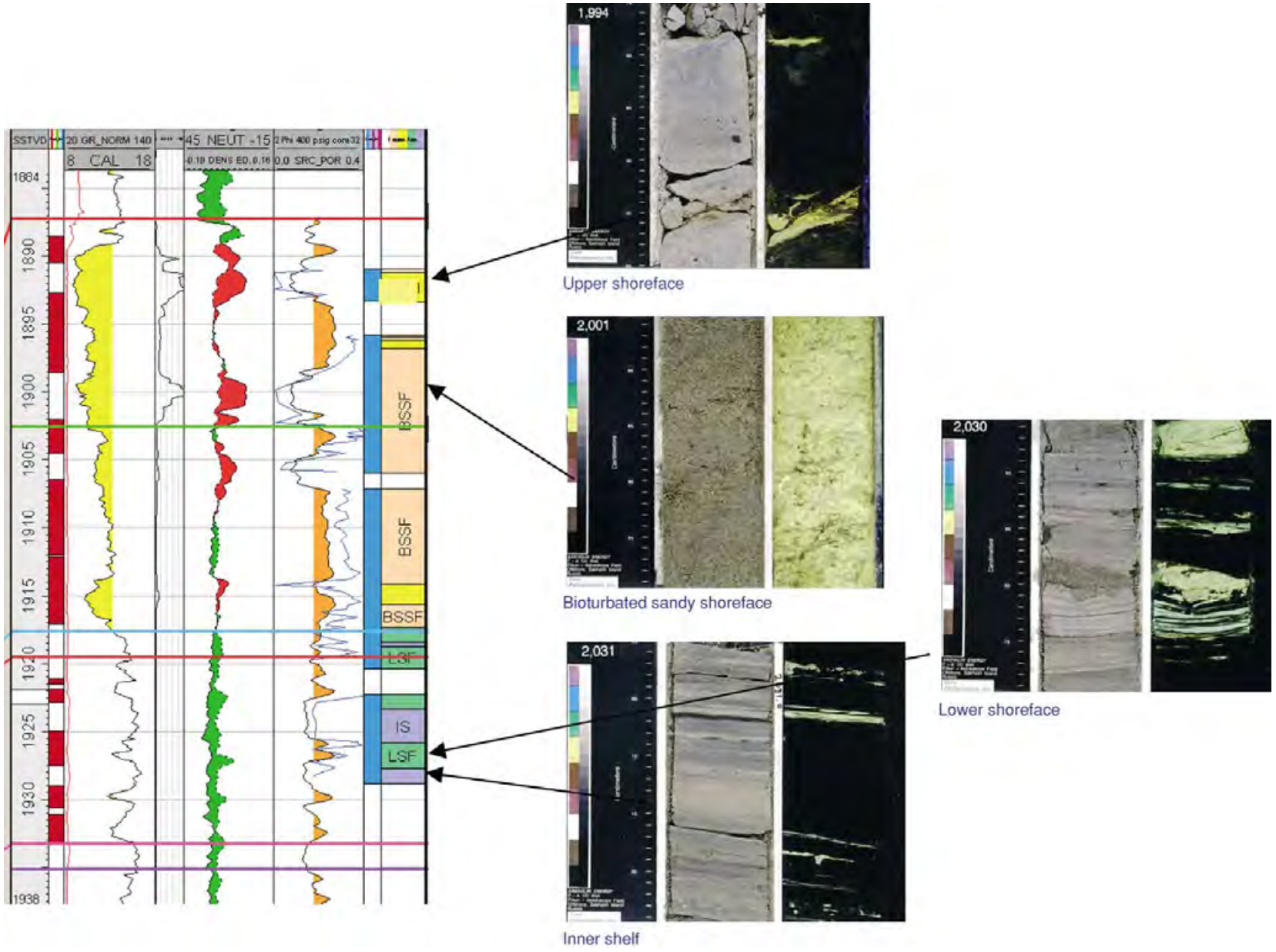


Figure 13 Facies definition from cores and logs. Courtesy of Shell.

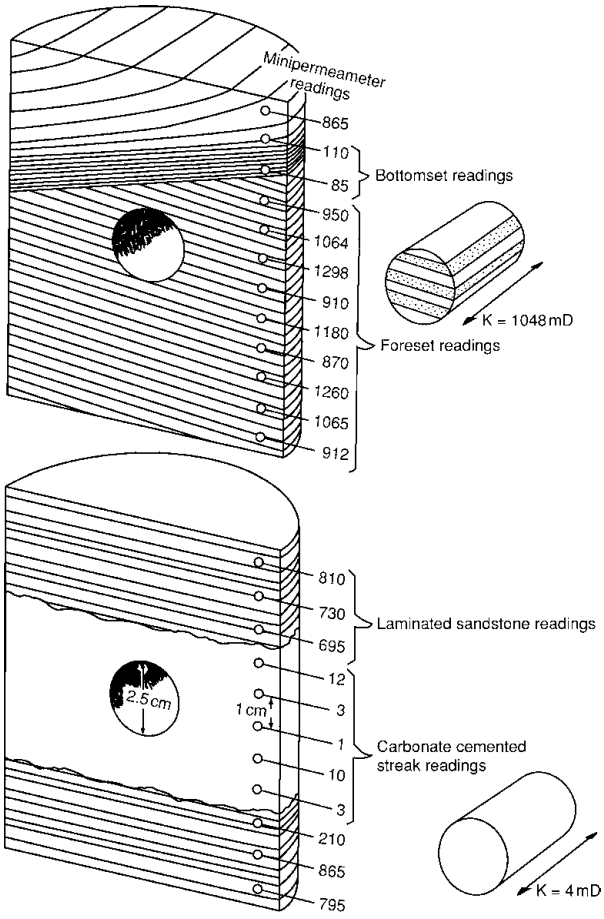


Figure 14 Probe permeability measurements showing small-scale heterogeneities. Reproduced with permission from Weber KJ and van Geuns LC (1990) Framework for constructing clastic reservoir simulation models. *Journal of Petroleum Technology* 42(10): 1248–1253, 1296–1297. © Society of Petroleum Engineers.

fracture porosities. Flowmeter-type production logs can also show open fractures. The shear zones around faults can be analysed with borehole imaging logs in horizontal wells. Fracture density, as a result of tectonic deformation, can be estimated by determining the rate of change of the dip. In fault and fracture analysis, extensive outcrop studies have added enormously to the knowledge of fault zone compositions and fracture distributions (Figure 18).

Correlation and Use of Analogues

The correlation from well to well requires a detailed knowledge of the facies distribution and expectation values of the typical shape and dimensional relationships of genetic reservoir bodies. For this purpose, there is a need for a data bank of such information based on outcrop and field data, complete with their sedimentological setting, including type of basin and climate, to be able to choose a proper analogue example. Such data banks have been compiled by several oil companies, while many universities are engaged in reservoir analogue outcrop studies. The stratigraphy forming the major framework of the reservoir can be interpreted with sequence stratigraphical principles. These principles and, especially, the analysis of sea-level fluctuations can be used to delineate the sedimentary successions, and are a great help in correlating the wells (Figure 19). Sea-level changes have a large influence on the sedimentary processes along time slices of the reservoir and, in carbonates, they also have a strong impact on diagenesis (e.g., leaching and dolomitization).

The well-known techniques of comparing the floral and faunal contents of discrete intervals remain very useful, not only for stratigraphical purposes, but also for facies analysis. The palaeontological and palynological studies can be carried out on drill cuttings whilst drilling, and thus can provide important indications of drilling progress through a known stratigraphical sequence. In combination with LWD techniques, the correct casing setting depth can be determined with sufficient safety margins.

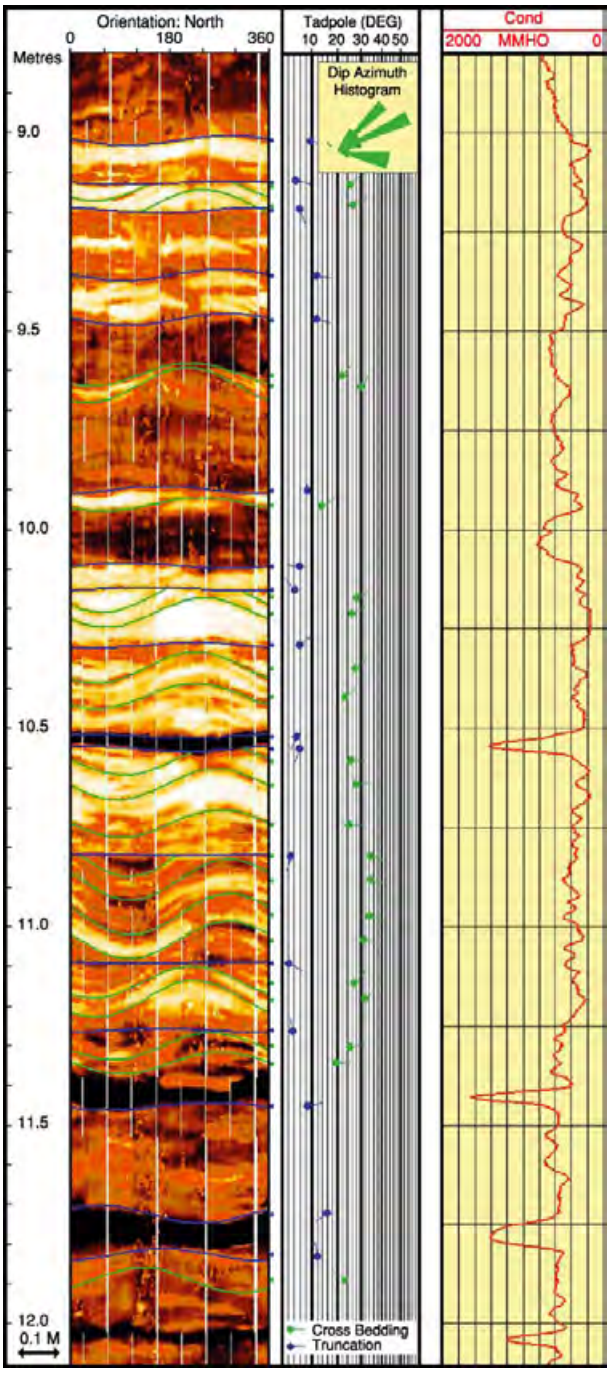


Figure 15 High-resolution acoustic borehole images compared with core slab showing aeolian cross-bed set. Reproduced with permission from Frikken HW (1999) *Reservoir-Geological Aspects of Productivity and Connectivity of Gasfields in the Netherlands*. PhD thesis, TU Delft.

Another useful technique is oil typing via gas chromatography. Detailed comparison of oil samples has proven to be a sensitive indicator of reservoir continuity. Oils that are not in communication either vertically or laterally will show distinctive differences in their chemical fingerprints.

Volumetric Estimates

The static modelling carried out as the basis for volumetric estimates and dynamic modelling requires 3D infilling of the space between the wells. In an early stage, hydrocarbon-in-place estimates are made by

Objective	Scale (m)	Type of measurement
Distribution of horizontal permeability along borehole	0.001–0.01	Probe permeameter on core slabs
	0.01–0.05	Core plugs
	0.05	Grain-size distribution from sidewall samples combined with porosity from logs
	1.2	Nuclear magnetic resonance (NMR) log
	0.5–5	Rock typing in non-cored well and use of porosity/permeability relationship
	1.0–5.0	Pressure build-up from Repeat Formation Tester (RFT) combined with production log
Permeability within well drainage radius	100–500	Well test pressure build-up combined with geological model
	100–500	History matching of well performance after significant production
Vertical permeability of intervals	0.001–0.05	Cube-shaped samples and probe-permeameter
	0.1–0.7	Multi-probe formation tester
	5.0–50	Vertical pulse test
		Computation with geological model using outcrop shale continuity statistics
Lateral and vertical permeability continuity	1.0–50	RFT or multi-probe tester in infill wells
	100–300	Well-test analysis combined with geological model and vertical seismic profile
	200–1000	Horizontal pulse test
Permeability distribution around horizontal wells	0.001–1.0	Core plugs, probe-permeameter, log-derived estimates, production logging
	100–300	Well-test analysis combined with geological model and production logging results

Figure 16 Overview of methods to determine or estimate permeability distribution on a range of scales. Modified with permission from Weber KJ (1995) Visions in reservoir management – what next? In: *Reservoir Characterisation: Integration of Geology, Geophysics and Reservoir Engineering, The Third JNOC-TRC International Symposium, February 2–23, Technical Research Centre, Chiba, Japan*, pp. 1–15. Chiba: Japan National Oil Corporation.

combining the tentative probability distribution of the controlling parameters: reservoir volume, net-to-gross, porosity, petroleum saturation, and formation volume factor (FVF), the ratio between the net petroleum volume *in situ* and under standard conditions. The cumulative probability distribution curve or expectation curve for the petroleum-in-place is generated with Monte Carlo techniques. The probability distributions of the parameters are randomly sampled and the resulting petroleum-in-place volumes are plotted until the graph hardly changes by additional values (Figure 20).

As an interim step towards 3D modelling, Net-Oil-Sand (NOS) maps can be constructed, which yield better estimates than combining the field averages of all parameters. For thin reservoirs, such maps almost represent 3D models, if the properties of the individual sand bodies can also be plotted (Figure 21).

Geostatistics

Classic statistical methods for the analysis of numerical data sets to determine distributions, standard deviations, and correlations have been used in production geology for a long time. There are, however, a number of problems for which these methods are not suitable. Firstly, one is frequently confronted with sparse data sets which do not allow reliable contouring or correlation. Secondly, there is a need to quantify uncertainty in spatial distributions and also to estimate the possible reduction in uncertainty resulting from drilling an additional well in a given place. Thirdly, one wants to use spatial relationships established for specific types of reservoir body to constrain the models based on well-to-well correlations.

The method which has been developed to handle these problems is called 'kriging', after the originator Daniel Krige, a South African mining engineer. It is based on the reasonable assumption that the unknown spatial distribution of a geological property can be predicted on the basis of the spatial distribution of the measurement of that property. The main tool in this method is the 'variogram', which quantifies the spatial continuity of a property. It is a graph of the variance of the difference of two measurements as a function of their spacing. One can use variograms made on the basis of a data set. If the number of data points is insufficient to define a variogram, one can select a best-fitting variogram or use a variogram that is typical for a modelled property. Known trends can be honoured by using different variograms in different directions.

An example of the above is shown in Figure 22. A map is made of the thickness variation of a reservoir using a best-fitting variogram. Next, the residual

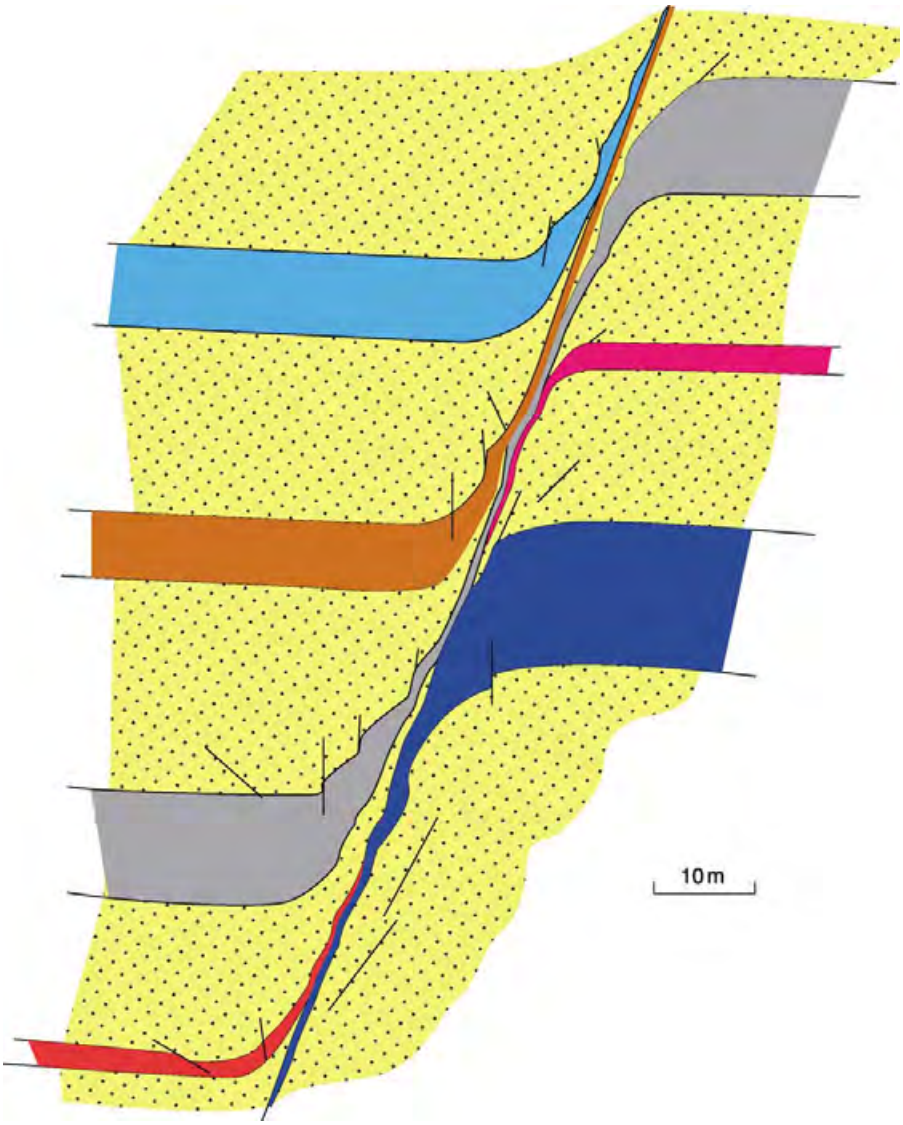


Figure 17 Clay smearing along fault as seen in outcrop. Note discontinuous clay smear of thin (red) shale.

variation between the variogram and the thickness map is computed to show the uncertainty at any point of the map. Finally, one can place a grid over the map and compute the reduction of uncertainty resulting from drilling a fictitious well in any of the grid points. This method can be used to select a next appraisal well.

Further applications of 'kriging' methods include the combination of data from different sources and with different accuracy to estimate a property. In this way, reservoir depth data from wells and from seismic measurements can be merged into one map. Probabilistic reservoir modelling also makes use of variograms to constrain the extent of the bodies within the reservoir. A series of equally probable models can be generated. Here lies one of the most useful applications of analogue data derived from outcrops and densely drilled fields.

Static Modelling

The particular technique used for 3D reservoir modelling depends on the reservoir type. For layer-cake reservoirs, deterministic models can often be made directly from the well-to-well correlations and the rock properties can be interpolated. For labyrinth reservoirs, on the other hand, deterministic modelling is rarely possible without a very close well spacing. Computer systems have been developed to generate a series of equally probable 3D models through 'probabilistic' modelling techniques. Most systems work in three steps. Firstly, correlatable reservoir bodies are determined and incorporated in the

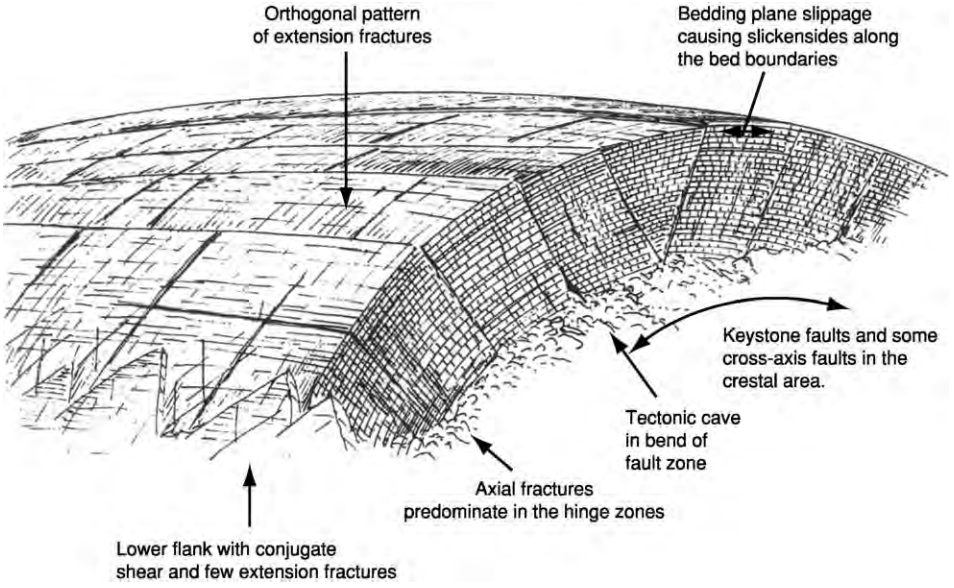


Figure 18 Faults and fractures in anticlinal structure of a thick limestone formation.

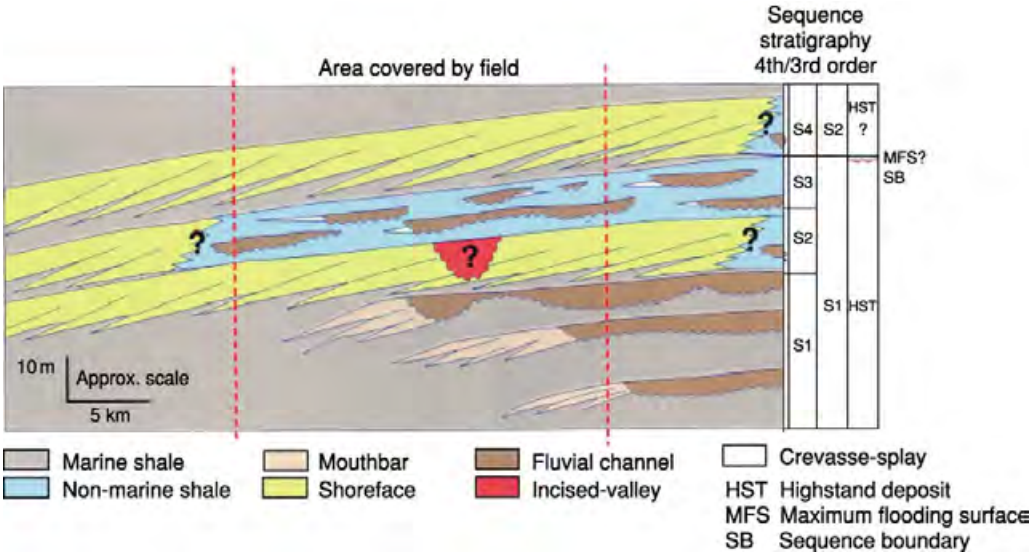


Figure 19 Sequence stratigraphical correlation in reservoir modelling. Courtesy of Shell.

model. Where extrapolations beyond control points are necessary, use is made of an analogue database for the relevant genetic type. Secondly, the uncorrelatable bodies are considered, for which the dimensions are also derived from the database, while body orientations are taken from borehole imaging logs or estimated on the basis of the general geological model. Characteristic variograms for the thickness distribution of genetic sand body types in different orientations relative to their expected trend are also used. Thirdly, especially when the well spacing is rather large, there are smaller or narrower reservoir bodies not penetrated by wells. Such bodies are added using statistical estimates of their occurrence from the wells. Their position and dimensions are conditioned by geological modelling rules and the analogue database (Figure 23). Jigsaw-type reservoirs are also difficult to correlate in the appraisal stage and probabilistic modelling is required. In a later stage, a large part of the architecture may be determined and only limited recourse has to be taken to probabilistic techniques.

The resulting architectural models can be compared with the seismic models and, in favourable

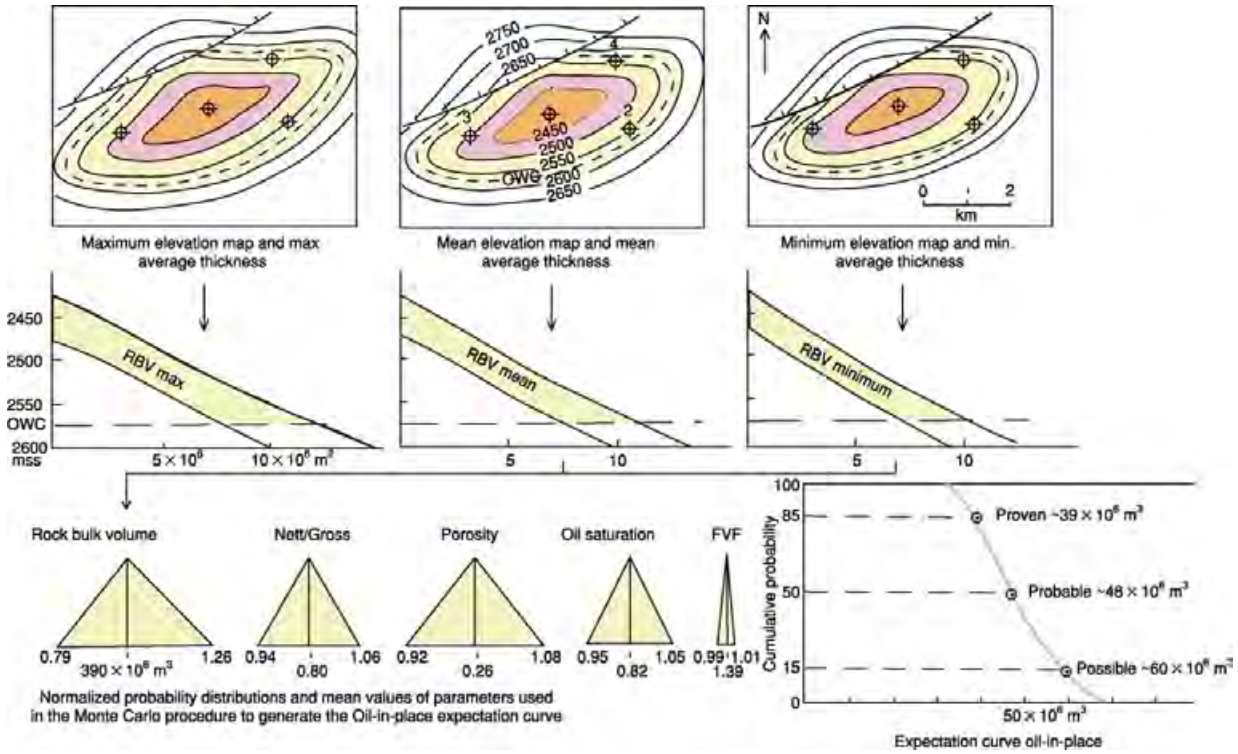


Figure 20 Approach to determine the cumulative probability (expectation) curve for the oil-in-place volume. RBV, Rock Bulk Volume.

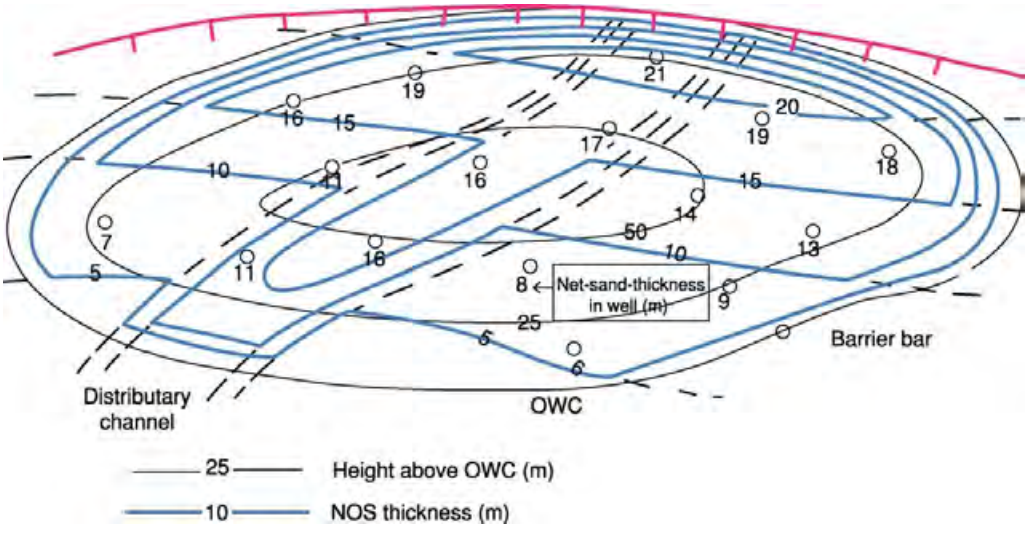


Figure 21 Net-Oil-Sand (NOS) map of deltaic reservoir.

cases, quantitative correlations can be made. This reduces the spread of possible model configurations. In any case, the permutation of the probabilistic models generated has to be carefully ranked in order of probability by screening them with regard to the presence of geological anomalies. A limited number of 3D models are selected as representing the likely range of variation in the reservoir with respect to hc-volume (hydrocarbon volume), connectivity, architecture, and permeability distribution.

Dynamic Modelling

The profitability of a hydrocarbon reservoir depends crucially on how its development is planned in

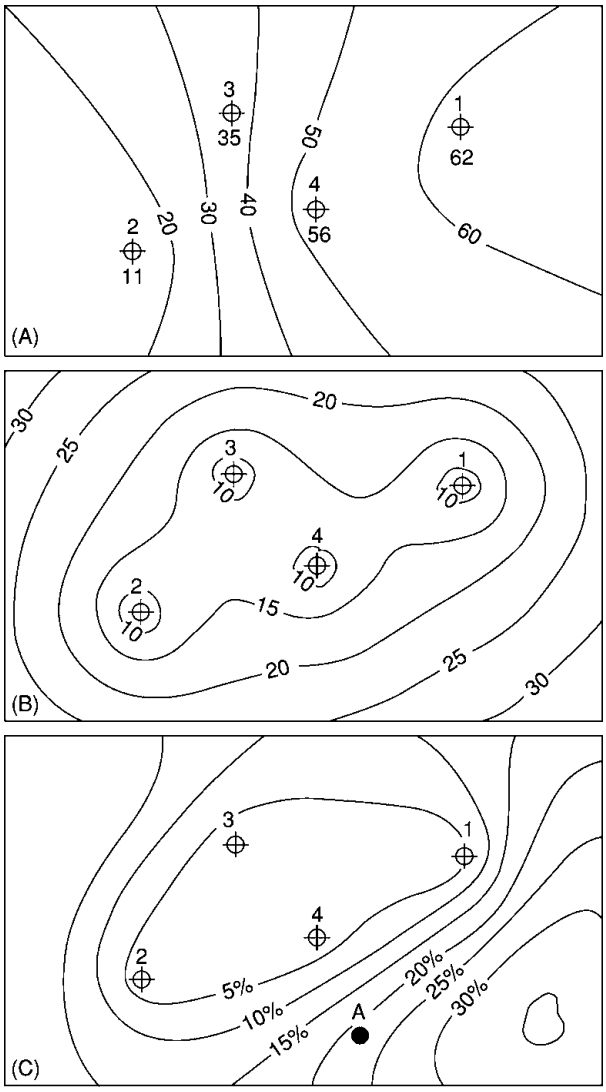


Figure 22 Geostatistical methods in mapping and well planning. (A) Estimated thickness map made with the best-fitting variogram. (B) Thickness uncertainty (66% confidence limit) computed from the residual variation between the variogram and thickness map. (C) Expected reduction in estimation variance of the mean thickness after drilling well in location A.

the context of recovery mechanism(s) and infrastructure environment. A key element from the outset of planning is a conceptual model of the hydrocarbon reservoir (ranging from a simple analytical model to a multitude of equiprobable realizations). The reservoir model allows the evaluation of alternatives by providing an estimated range of hydrocarbon volumes that are potentially recoverable under various development schemes.

The majority of the field development plans are based on reservoir simulation models with varying degrees of complexity. Integrated dynamic reservoir models are constructed by reservoir engineers and require input on geological structure and architecture, rock and fluid properties and their distribution, displacement characteristics, and description of the wells and the surface facilities. Static and dynamic modelling programs differ from each other in terms of the number of volume elements they can handle. Like the 3D static modelling packages, the dynamic modelling programs require the reservoir to be segmented into unit cells, each of which has uniform properties, before it can perform computations. Static modelling computations involve the categorization, sorting, and counting of voxel cells; dynamic modelling computations involve the balancing of the fluid masses and pressures within and between 'gridblock' cells. Upscaling programs enable one to aggregate the geological detail (=3D static model) into the bigger elements of a reservoir engineering model (=3D dynamic simulation model) by using a method of permeability averaging (e.g., harmonic, arithmetic, geometric). These programs also produce a corresponding gridblock model of porosity through conventional averaging of voxel porosity values.

It is important that the 3D modeller identifies the key geological characteristics that should not be lost during upscaling (e.g., thin-bedded shales that may act as permeability barriers or baffles) (Figure 24). Not all geological architecture will have equal weighting in terms of its influence on reservoir performance, and therefore the geologist (with the help of a reservoir engineer) must filter through the geological detail in order to determine what is/is not important for reservoir simulation. Iteration between the static and dynamic models is recommended in order to ensure that the relevant detail (flow units and barriers) is not compromised during the upscaling procedure.

Conclusions

The scope of production geological activities and methods has become very wide, requiring an equally diversified group of specialists to cover all topics. Advanced seismic techniques are employed at every stage, from discovery to additional development planning late in the life of a field. The backbone of the profession is formed by expert sedimentologists who can understand and model the architecture of reservoirs, including the features that influence fluid flow. There is clearly a division in clastic and carbonate reservoir expertise because of the difference in sedimentological and diagenetic processes and relevant analysis techniques.

Structural geology is another speciality that should be handled by experts. Again, there are large differences between faults and fractures in clastic and carbonate reservoirs. The advances in seismic resolution

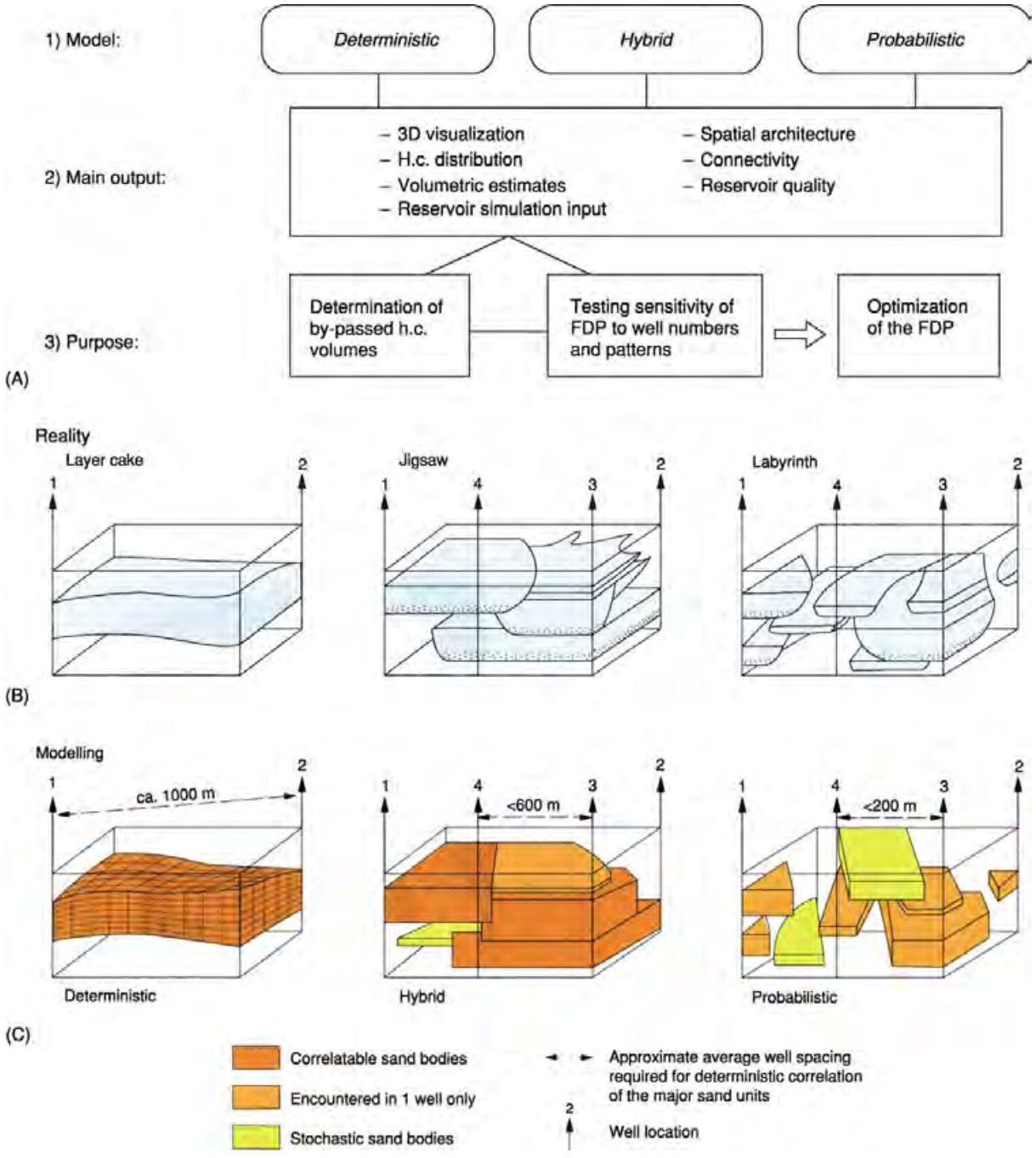


Figure 23 Summary of three-dimensional (3D) reservoir modelling. Courtesy of Shell.

on the large scale and borehole imaging logs on the small scale have greatly improved the analysis of faults and fractures. Geostatistical methods are widely used in constructing maps and reservoir models, which reflect the level of uncertainty, but in which analogue information can be incorporated.

The present approach to production geology is heavily dependent on computer systems, and much of the time is spent behind a workstation or personal computer. Modern interactive computer techniques and integrated data systems are used to design models and well trajectories. Visualization plays an important part in this work and virtual reality techniques are already being used. Typical of the modern reservoir management approach is the integrated team composed of various production geology specialists, together with petrophysicists, reservoir engineers, and production technologists.

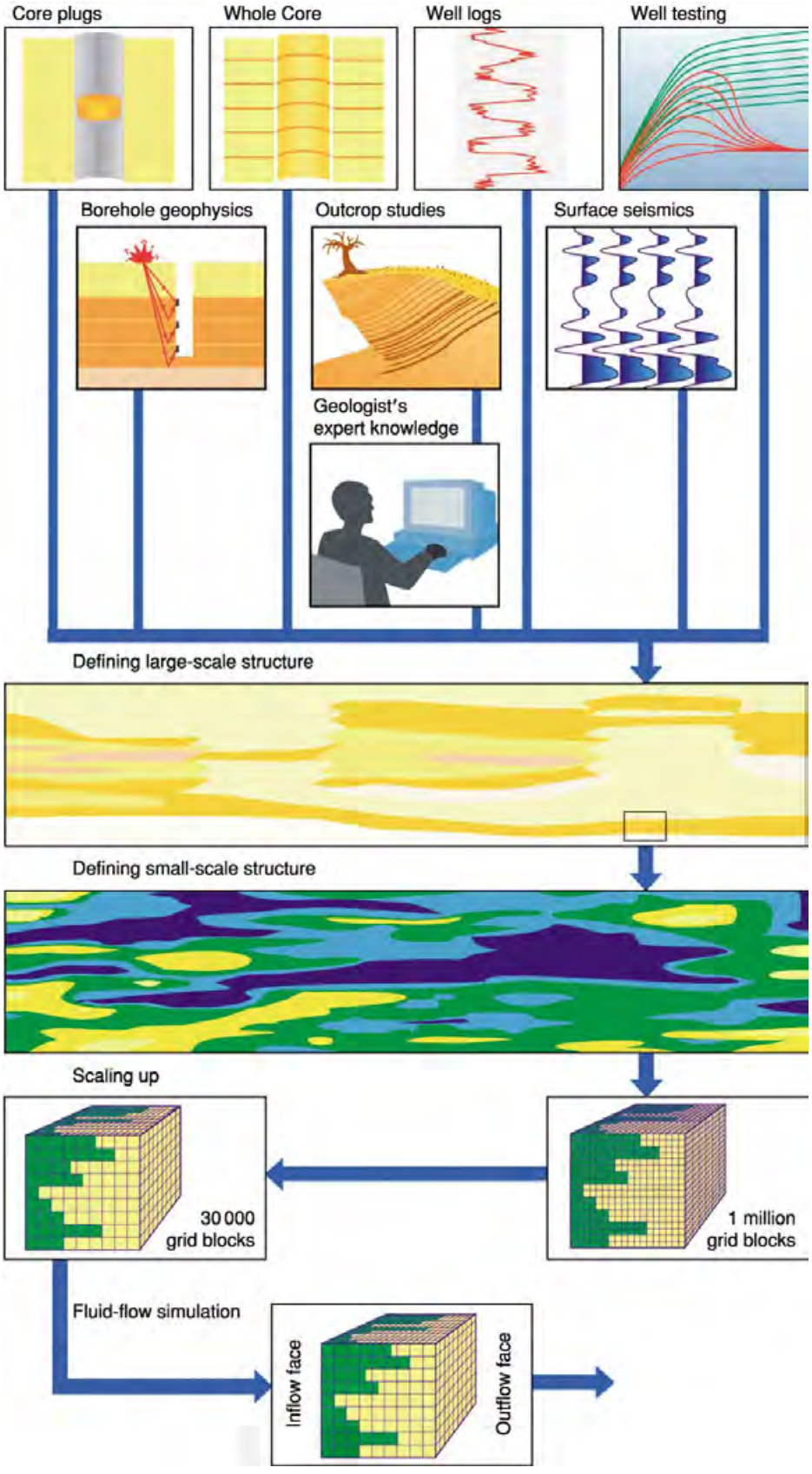


Figure 24 Building a reservoir model with measured data, expert knowledge, and statistics. © Schlumberger Limited, used with permission.

The future will undoubtedly bring important further improvements in seismic techniques. There is a need for more analogue data and sedimentological knowledge to aid in reservoir modelling. The reservoir models will benefit more and more from integration with seismic information, both with respect to reservoir architecture and to fluid content and movements. The high costs of deep-water field appraisal and development will place a premium on the timely construction of realistic reservoir models. On land, there is a large scope for enhanced oil recovery that equally requires reliable detailed reservoir models. Thus, it may be expected that there will be a significant further development in production geological techniques.

See Also

Diagenesis, Overview. **Petroleum Geology:** Overview; Chemical and Physical Properties; Exploration; Reserves. **Sedimentary Environments:** Depositional Systems and Facies. **Sedimentary Rocks:** Dolomites; Limestones; Sandstones, Diagenesis and Porosity Evolution. **Seismic Surveys**.

Further Reading

American Association of Petroleum Geologists (AAPG) (1993) *Development Geology Reference Manual. AAPG Methods Series No. 10*. Tulsa, OK: American Association of Petroleum Geologists.

Barwis JH, McPherson JG, and Studlick JRJ (eds.) (1990) *Sandstone Petroleum Reservoirs*. New York: Springer Verlag.

Bishop CM (1995) *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.

Brown AR (1999) *Interpretation of Three-Dimensional Seismic Data. AAPG Memoir 42*. Tulsa, OK: American Association of Petroleum Geologists.

Dickey PA (1986) *Petroleum Development Geology,* 3rd edn. Tulsa, OK: PennWell Publishing Company.

Dikkers AJ (1985) *Geology in Petroleum Production. Developments in Petroleum Science 20*. Amsterdam New York: Elsevier Science.

Flint SS and Bryant ID (eds.) (1993) *The Geological Modelling of Hydrocarbon Reservoirs and Outcrop Analogues. Special Publication No. 15 of the IAS (International Association of Sedimentologists)*. Oxford: Blackwell Scientific Publications.

Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Jahn F, Graham M, and Cook M (1998) *Hydrocarbon Exploration and Production. Developments in Petroleum Science 46*. Amsterdam New York: Elsevier Science.

Jensen JL, Lake LW, Corbett PWM, and Goggin DJ (1997) *Statistics for Petroleum Engineers and Geoscientists*. Upper Saddle River, NJ: Prentice Hall.

Laudon R (1996) *Principles of Petroleum Development Geology. Petroleum Engineering Series*. Upper Saddle River, NJ: Prentice Hall.

Lowell JD (1985) *Structural Styles in Petroleum Exploration*. Tulsa, OK: OGCI Publications, Oil and Gas Consultants Inc.

Moller-Pedersen P and Koestler AG (eds.) (1997) *Hydrocarbon Seals – Importance for Exploration and Production. Norsk Petroleum – Forening/NPF Special Publication No. 7*. Singapore: Elsevier Science.

Nelson RA (1985) *Geological Analysis of Naturally Fractured Reservoirs*. Houston, TX: Gulf Publishing Company.

Reading HG (ed.) (1996) *Sedimentary Environments and Facies,* 3rd edn. Oxford: Blackwell Scientific Publications.

Van Wagoner JC, Mitchum RM, Campion KM, and Rahmanian VD (1990) *Siliciclastic Sequence Stratigraphy in Well Logs, Core and Outcrops for High-Resolution Correlation of Time and Facies. AAPG Methods in Exploration Series No. 7*. Tulsa, OK: American Association of Petroleum Geologists.

Weber KJ (1995) Visions in reservoir management – what next? In: *Reservoir Characterisation: Integration of Geology, Geophysics and Reservoir Engineering, The Third JNOC-TRC International Symposium, February 2–23, Technical Research Centre, Chiba, Japan*, pp. 1–15. Chiba: Japan National Oil Corporation.

Weber KJ and van Geuns LC (1990) Framework for constructing clastic reservoir simulation models. *Journal of Petroleum Technology* 42(10): 1248–1297.

Reserves

R Arnott, Oxford Institute for Energy Studies, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Fossil fuels can be broadly categorized as either resources or reserves. Resources include all fuels, both those identified and those as yet unknown. Reserves are that portion of the identified resources which can be economically extracted and exploited using current technology. Petroleum reserves can be labelled under a wide variety of physical, chemical, and geological circumstances. For example, the boundaries between crude oil as a liquid and condensates have long been the subject of controversy. In addition, there are issues of definition as to what to include or exclude from a particular production forecast, as there are as to what can and cannot be reported as reserves because of legal and political considerations. The only near-certainty on the supply side is the actual volume of oil that has been produced. This is because the definitions of petroleum reserves often include assumptions with regard to existing technology and present economic conditions. However, there is no uniformity or stated policy as to the time period over which the existing technology and present economic conditions are anticipated to prevail. As a consequence, there is often fierce debate about how long existing petroleum reserves are likely to last and the economic consequences if future production cannot meet demand.

Definitions of Reserves

Because of the considerable uncertainty surrounding the definitions of reserves, most authors prefer a probabilistic rather than a deterministic approach (Figure 1). An initial declaration of recoverable oil that can, with reasonable certainty, be recovered in the future under existing economic and operating conditions is the usual definition of the 'proven' reserves. However, all fields will also be declared as having additional volumes of 'probable' and 'possible' reserves. The definition of 'proven', 'probable', and 'possible' reserves varies from country to country. 'Proven' reserves are usually defined as being P90 reserves, indicating that there is a greater than 90% chance that the actual proven reserves base will be higher, and a 10% chance that it will be less. Similarly, 'probable' and 'possible' reserves can be defined as P50 and P10 reserves, respectively. The important point here is that, when oil reserve numbers are quoted in the literature, it must be realized that the numbers are probabilistic and, in the case of proven reserves, are more than likely to be exceeded. Therefore, it is wrong for commentators to argue that the volumes of discovered reserves are a fixed entity as, in the case of proven reserves, there will be a 90% chance that the initial reserve number will be exceeded.

Declarations of proven reserves are often only for a specific reservoir. Therefore, when such information is summed, eventually, to the regional and national level, the simple arithmetical addition of a large number of independent values, each representing the 90% probability of a specified volume in a specific reservoir, produces a higher joint probability of the total. It is for this reason that field growth and rates of field growth are well documented in practice.

Resources can be defined as existing reserves plus all of the accumulations of reserves that may eventually become available. These additional reserves might have already been discovered but be uneconomic, or may not yet have been discovered. The distribution of fossil fuels can therefore be viewed as a pyramid, with a small amount of higher quality resource at the top, but with increasingly large amounts of lower grade resource as we move down the pyramid (Figure 2). The costs of retrieving the resource increase lower down in the pyramid, making a larger amount of the resource available at higher prices. The issue as to what defines the total volume of petroleum resources depends on where the pyramid is sliced, and this is a very subjective decision.

There is often a curious circularity in the estimates of undiscovered and ultimate reserves. Undiscovered reserves are the difference between estimates of the discovered and ultimate reserves. However, ultimate resource estimates obviously depend on an estimate of undiscovered reserves. Although there are a number of ways in which to try and surmount this difficulty, the most commonly described method involves an examination of the discovery pattern. In general, there is some regularity to this pattern, with the largest fields being discovered first with attention then switching to the smaller fields. Therefore, in any one basin, the curve (commonly referred to as the creaming curve) relating cumulative discoveries to wildcat wells is usually hyperbolic with the asymptote revealing the level of ultimate reserves. However, this method is not immune to error, especially if larger

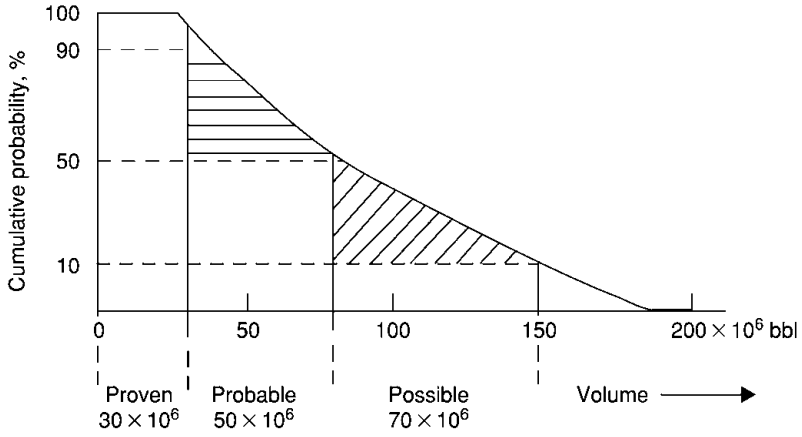


Figure 1 Expectation curve used to find approximate values of proven, probable, and possible reserves. bbl, barrels.

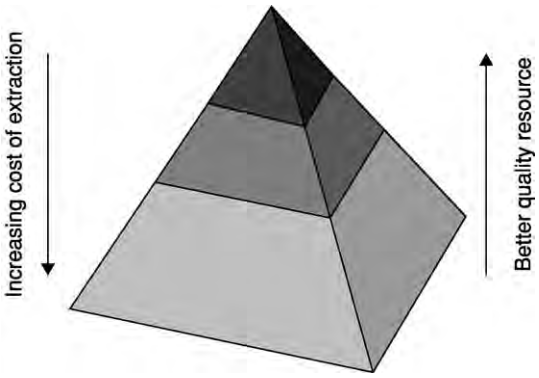


Figure 2 Resources visualized as a pyramid, with a small volume of prime resources that are of high quality and easy to extract, and a large volume of resources of lesser quality that are more difficult or expensive to extract.

fields are discovered late in the life of a particular basin. In the case of the UK sector of the North Sea (Figure 3), the asymptote has moved over time as the older fields have grown in size through better recovery techniques and reservoir definition. As a consequence, estimates of the total resource base have almost doubled in size from the initial values.

Predictions of Ultimate Recoverable Reserves

Numerous authors have attempted to predict the magnitude of the ultimate recoverable resource base. What is interesting about these estimates is that they all essentially identify those crude oil reserves that will be produced over the next 40–50 years. In other words, every assessment tends to slice lower through the resource pyramid (Figure 2). Estimates of the ultimate resource vary significantly depending upon the relative optimism or pessimism of the subsurface geotechnical assessment, as well as the definition of what is the minimum size of economic importance.

Recent estimates of the ultimate recoverable reserves of oil (Figure 4) vary between 1.5 and 3.8 trillion barrels. It is interesting to note that, despite the range of estimates, the mean estimate has remained stable at 2 trillion barrels for at least the past 20 years, suggesting that technology has reached an asymptote. Of course, not all of the estimates are based on the same set of data, as different estimates slice the resource pyramid at different levels. The most pessimistic estimate only includes conventional oil reserves and excludes all other oil types (for example, heavy oil, tar sands, secondary recovery). In contrast, the optimistic estimate by the United States Geological Survey (USGS) only excludes from its estimates oil from tar sands, oil from polar regions, and oil shales. However, nearly all estimates rely on existing technology and do not attempt to predict the impact of future technological developments. As a result, the estimates can only be regarded as static estimates at a fixed point in time under a particular set of technological and economic circumstances.

The actual volume recovered will depend on a number of factors, not least whether the reserves are economic to extract. Political, economic, and technological constraints all play a role in deferring or accelerating production. In addition, there is interplay with other sorts of fuel, which is likely to have a significant impact on determining the rate of extraction.

Historical evidence from producing fields highlights the fact that reserve estimates are dynamic as circumstances change and knowledge and technology progress. An understanding of the potential importance of the growth of reserves can be gained from considering the current estimate of the global oil resource base and the current average recovery factor. Using consensus estimates that the global average

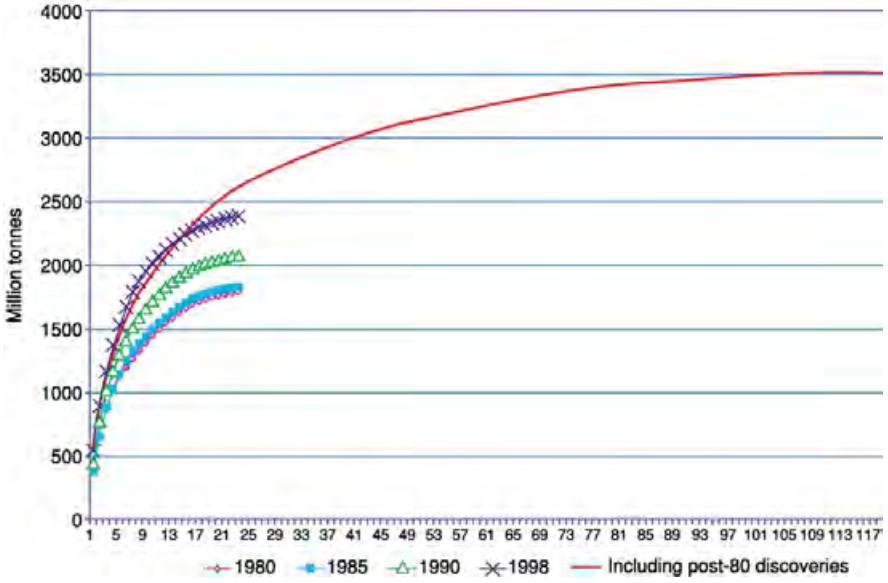


Figure 3 Creaming curves for the UK sector of the North Sea showing the impact of reserves creep on the older fields. The upper curve includes all fields that had been discovered and were producing by 1998. The symbols represent estimates of the same 25 fields made at different times (1980, 1985, 1990, and 1998). The figure shows the change in estimated size of these discoveries. The full line shows the addition of a further 100 discoveries made up to 1998 and which are currently in production.

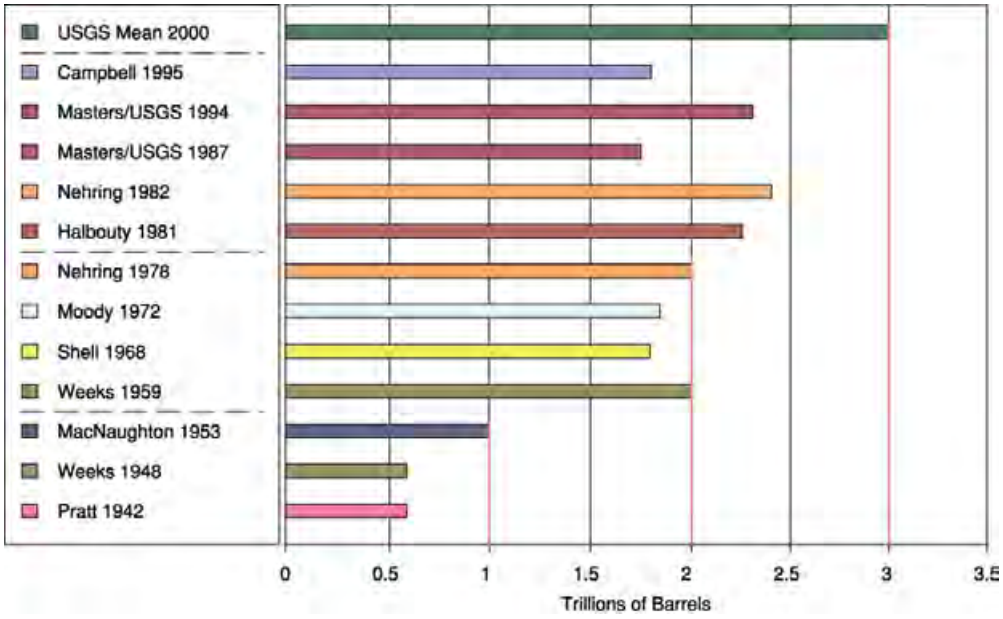


Figure 4 Published estimates of world ultimate oil recovery (trillions of barrels). This chart shows the various published estimates of ultimate oil recovery through time. Earliest estimates made in the 1940s suggested an ultimate recovery of around 0.5 trillion barrels. However, the average estimate since Halbouty (1981) has been around 2.0 trillion barrels. USGS, United States Geological Survey.

oil recovery factor is around 35% (Figure 5), and that the current resource base of reserves is around 2 trillion barrels, original oil in place would be nearly 6 trillion barrels. Therefore, every 1% increase in global oil recovery would lead to an additional 55 billion barrels of oil being produced. Moving the average global recovery factor up to 45%, a not unrealistic target, could increase global recoverable reserves by around 550 billion barrels.

Over the past 50 years, estimates of global recoverable oil reserves have risen annually, except for two small falls in the late 1970s, despite annual increases in oil production (Figure 6). This has led to a great deal of attention being paid to reserve reporting by

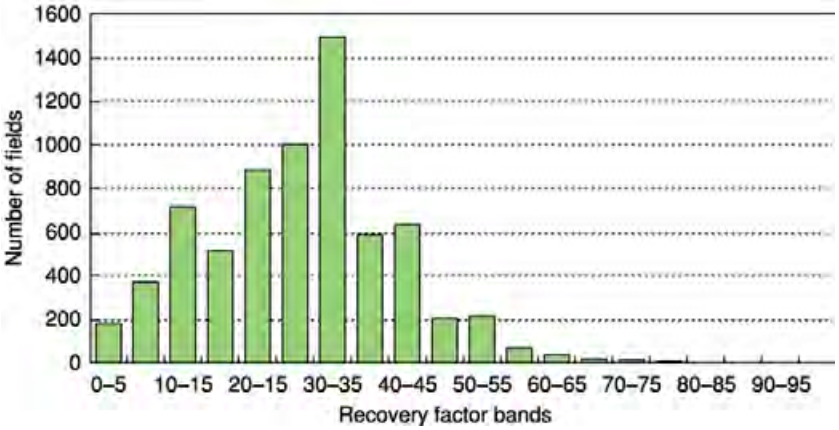


Figure 5 Distribution of recovery factors in all currently producing oil fields.

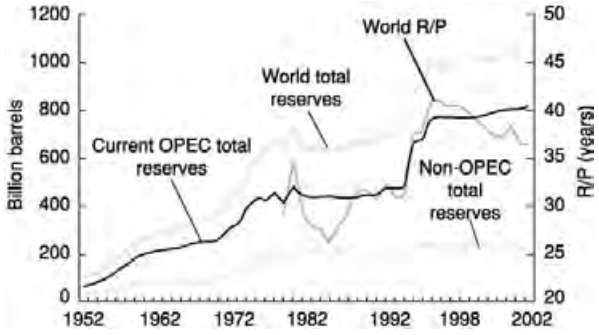


Figure 6 Global estimated reserves and reserves to production (R/P) ratios by year. OPEC, Organization of Petroleum Exporting Countries. R/P, reserves to production ratio.

companies and countries. Individual Organization of Petroleum Exporting Countries (OPEC) members have been accused of inflating reports of reserves because export quotas were originally based on estimates of recoverable reserves. The huge reserve revisions reported by many OPEC companies in 1987 have been used by some to undermine confidence in published numbers. In reality, it is possible that previous OPEC estimates were too conservative, having been inherited from private companies before nationalization.

Although the absolute level of reserves has risen only slightly over the past 5 years, within that total, upward revisions to existing fields and discoveries have exceeded 150 billion barrels, 30% more than actual consumption. The USGS recently published a mean estimate of 612 billion barrels from increased recovery, significantly increasing their estimate of the world's ultimate recoverable reserves. In the UK sector of the North Sea, reserves have grown by an average of 35% from the time of field approval (Annex B) to the present day or abandonment (Figure 7).

Estimates of the total resource base remaining depend not just on new discoveries being made and brought on stream, but also on the amount of reserves used due to depletion of existing fields. Over 70% of current world oil supply comes from oil fields that were discovered prior to 1970, and 14 of these fields produce over 20% of the world's total supply. Even though data for many of these fields are very hard to come by, the decline rates for these fields are likely to become of increasing importance in the future. However, the effect of depletion is often overstated as forecasters often assume a depletion rate of 10–20%, and no further investment. As an example, estimates of non-OPEC production have always tended to indicate that production is at or near its peak, and yet overestimation of depletion rates has led to successive estimates being proven to be incorrect (Figure 8).

The Peak Oil and Depletion Debate

Ever since oil was first produced in significant quantities over 140 years ago, debate has taken place on whether petroleum will run out within the foreseeable future. Historically, near-term supply concerns arise when the relative rate of production capacity growth falls short of expected rates of demand growth. Through time, however, changing attitudes to oil supply and demand have led to varying perceptions of whether the resource is in short supply or not. It is also noticeable that changes in perception can take place over a very short period of time depending on a particular event on either the supply side or the demand side. The discovery of super giant oil fields in Texas in the 1930s gave rise to perceptions of a huge glut in oil. This perception was sustained by the discovery of major reserves in the Middle East. However, by the early 1970s, there was a clear paradigm

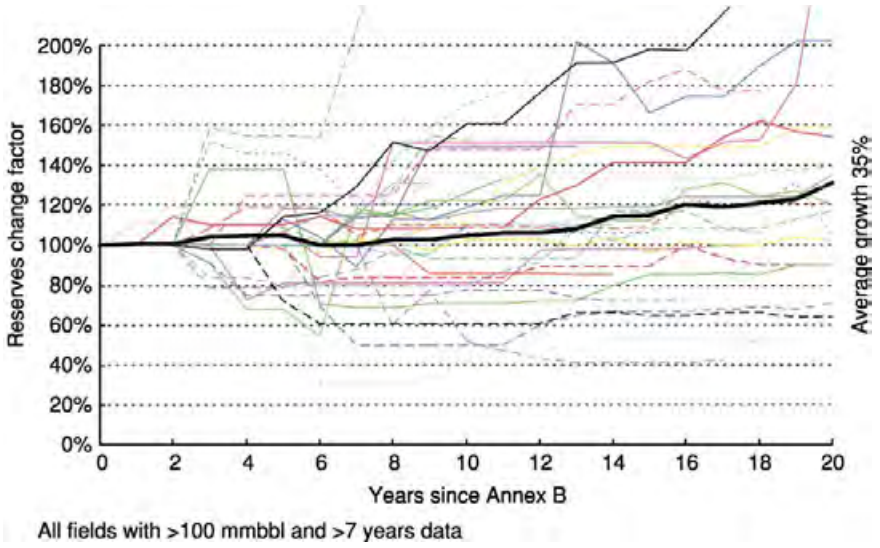


Figure 7 Growth of reserves in UK oil fields. This chart shows the change in reserve estimates of UK oil fields with time, since their 'Annex B' approval. Some field estimates have decreased, but most have increased very substantially. The average of all the fields (full line) has grown by 35% over 20 years.

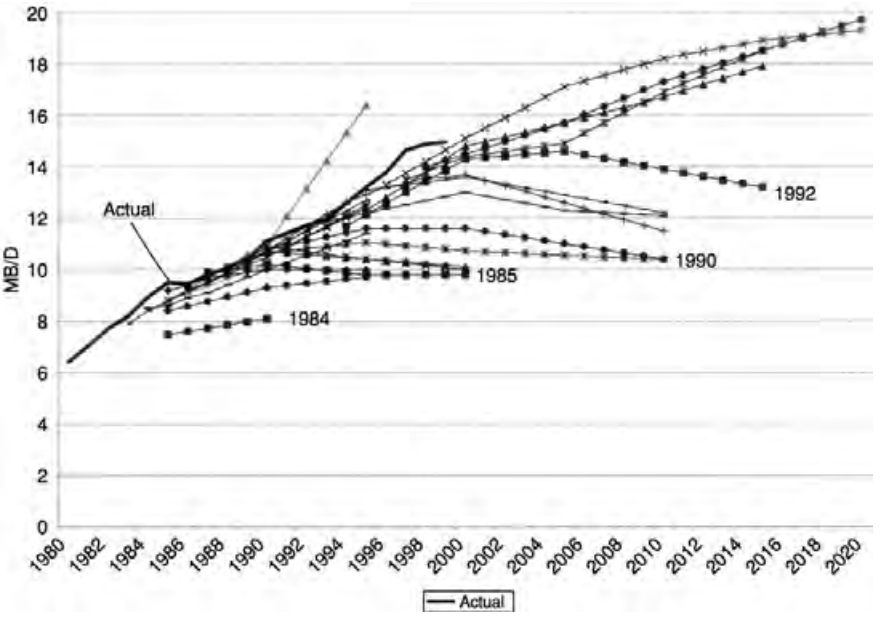


Figure 8 Forecast of non-Organization of Petroleum Exporting Countries (OPEC) Third World oil production over time made by the United States Department of Energy. The chart shows that estimates of non-OPEC production have consistently proved to be conservative through time. For example, the forecast made in 1990 for the year 2000 is now some 4 million barrels per day lower than actual production. MB/D, million barrels per day.

shift as part of a trend towards pessimism about resource availability as studies began to emerge that oil production outside the Middle East had peaked. This was triggered when US production levels started to fall from peak levels by the end of the 1980s. In the 1990s, the introduction of new development techniques to boost recovery rates, coupled with the revolution in computers which transformed seismic acquisition and processing, the technical innovations that enabled deep-water development, and the opening up of new petroleum provinces that were previously closed due to political constraints, all led to a period of optimism that weakened concerns about the exhaustibility of reserves. More recently, these

concerns have been raised again as the oil industry worldwide struggles to replace its reserves and achieve targeted rates of production growth.

Much of the published literature on the depletion debate falls into two camps. On the one hand, there are those who warn of the imminent danger of a collapse in oil supply and the economic consequences of that occurring. On the other hand, there are those who argue to varying degrees that there is no limit to oil supply in the near future. What is common to both camps is the extensive reference to the work of King Hubbert.

Hubbert gained prominence in the American geological community because he anticipated the peak and subsequent fall of US oil and gas production. At the heart of his prediction was the assumption that all resources are finite, and that eventually all resources will be depleted and exhausted over a period of time that is determined by the rate of production.

Hubbert argued that the complete cycle of exploitation must have the following characteristics. Beginning at zero, the rate of production tends initially to increase exponentially. Then, as difficulties of discovery and extraction increase, the production rate slows in its growth, passes one or more maxima and, as the resource is progressively depleted, declines eventually back to zero. He argued that this cycle of production would be bell shaped when plotted against time. The approximate rule here is that peak production will correspond to the mid-point of depletion and that this usually occurs about 20–25 years after the discovery of the mid-point. Although some production curves do have a shape that approximates such a 'normal' distribution, there is no inherent reason why production should follow such a pattern, and some production curves show very strong asymmetry. For example, when production is dominated by a small number of large fields, peak output tends to precede the depletion mid-point. Where the discovery pattern is more dispersed, or where offshore fields are significant, the production peak usually comes after the mid-point.

Hubbert's basic assumption was that, if known past and prospective rates of production are combined with a reasonable estimate of the amount of fuel initially present, one can calculate the probable length of time that the fuel can be exploited. In other words, the area under the complete production curve would be equal to the size of the resource. However, estimating the amount of oil and gas that will ultimately be discovered and produced in a given area is full of uncertainty, as we have seen. The Hubbert curve might have fitted US production, but, on a global scale, the modelled curve no longer fits the historical data (Figure 9). Up until the mid-1970s, the modelled production matched the actual production, but, after the supply shocks, which led to sharp oil price rises, annual increases in demand for oil fell from around 7% to just 2% per annum.

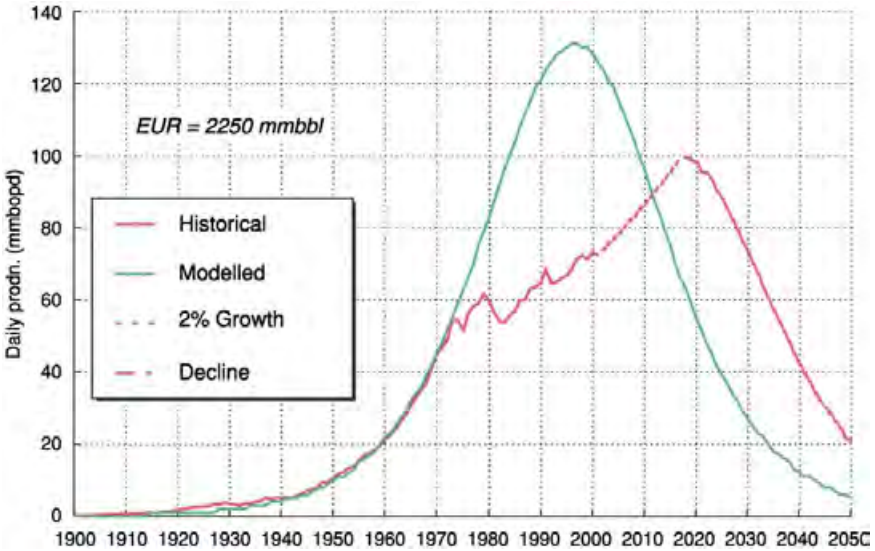


Figure 9 World production and the Hubbert peak. This chart shows a theoretical Hubbert curve based on production up to the oil price shocks of the 1970s. The resulting high oil prices led to a decrease in demand for oil. As a result, the slope of the oil supply curve departs from the predicted Hubbert curve after 1973. Assuming oil demand growth of 2% from 2003, and an ultimate reserve base of 2.25 trillion barrels, the peak in oil production could be reached in 2018. The chart highlights the impact that changes in demand can have on predictions of when global oil production will peak. mmbbl, billion barrels; mmbbpd, million barrels per day.

Models of Resource Depletion

There are, in essence, two main models that have been applied to the estimations of the volume of remaining reserves and the rate of depletion. On the one hand, there is the model for an open market and, on the other, the model for a closed market. The closed market model, in essence, treats the global oil reserves as a finite or 'fixed stock', whereby market depletion would ultimately result in higher prices. In the open market model, the cost of new energy sources is forecast to fall because of technological innovation, economies of scale, and the development of efficient systems of transportation. With lower costs, new energy sources become competitive with existing energy sources and ultimately lead to their replacement. The open market model could ultimately lead to lower energy costs through time through the potential substitution of fuels, in particular natural gas.

The open market model has stronger affinities to actual market examples. The US energy market is an open market, which has benefited from a long-term reduction in energy prices despite rising demand. Coal substituted for firewood; petroleum substituted for whale oil for illumination purposes, and eventually coal for transportation; gas substituted for coal in domestic heating; and gas and coal are currently competing in the electrical generation sector (Figure 10).

The End of the Petroleum Era?

With any finite resource, there will come a time when unfavourable economics will not permit further extraction. Worries concerning the long-term supply of petroleum began shortly after production started in 1859. The huge growth in energy markets after World War II also led to renewed fears that fossil fuels would be exhausted. In the 1970s, many people argued that the growth in world population and industrial production would lead to the total depletion of oil within 50 years of the forecast. The fact that the deadlines for these scenarios have now passed, and oil production and reserves continue to rise, shows that these forecasts were all overly pessimistic.

The resource exhaustion spectre was initially raised largely on the basis of a comparison of the current year's oil production with the so-called proven reserves of oil. However, this type of evaluation is inadequate as proven reserves represent nothing but the working stock of the oil industry. For example, the 'optimum' reserves life of 10 years is merely an expression of confidence from private and state oil companies as to where oil will come from in the future in order to fund investment in new infrastructure. Therefore, the current reserves life of oil is deemed to be irrelevant, especially in the context of a global reserves life that has risen to around 40 years from 1940 to 1975 despite the significant increase in global production.

The debate has now moved on from reserves life to a focus on the point in time when global production starts to decline. Supporters of this 'peak oil' school of thought forecast serious economic and social consequences in the aftermath of this occurring.

They highlight the fact that most of the world's major oil discoveries were made during the 1960s and 1970s and that, since that time, the discovery rate has declined (Figure 11). Using examples from various basins around the world, not least the USA, it is argued that peak oil production usually occurs around 25–30 years after peak exploration success. Using this as a proxy, it is predicted that oil production will peak within the next 10 years and, given that there are no near-term substitutes to oil, an economic crisis will ensue.

The Economic Viewpoint

One key assumption made by the supporters of 'peak oil' is that the resource base is finite. However, the amount of oil that can be recovered depends not only

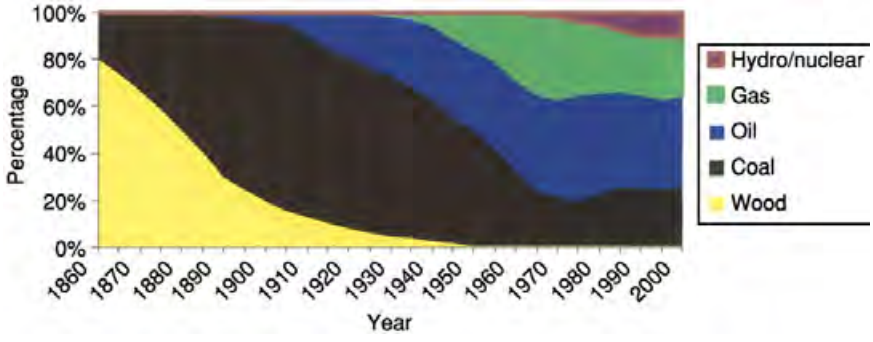


Figure 10 United States primary energy consumption by fuel, 1860–2000.

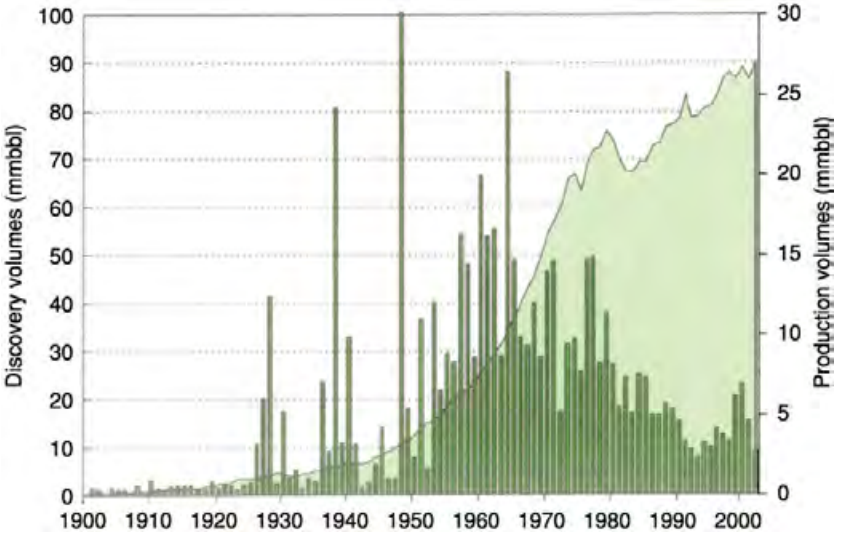


Figure 11 Discovery volumes and world oil production, 1900–2000. mmbbl, billion barrels.

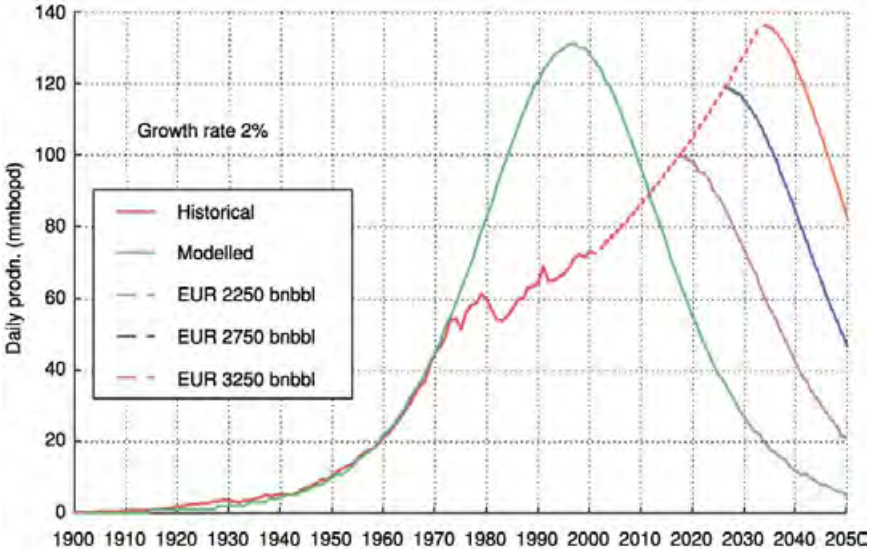


Figure 12 The effect of changing the resource base on the timing of peak oil. bnbbl, billion barrels; mmbopd, million barrels oil per day.

on the total resource base of oil, but also on dynamic variables, such as price, infrastructure, and technology. The level of peak production will also increase if recoverable reserves rise and, as a consequence, the timing of peak oil will be pushed further into the future (Figure 12).

Economists argue that oil reserves are not a 'fixed stock' and that energy has not been getting scarcer in basic economic terms, but rather has been getting more plentiful. The fall in the long run cost of oil production has been used in support of this argument. It is also argued that there is no such thing as an exhaustible natural resource and that the total petroleum reserves in the Earth are an irrelevant non-binding constraint. If oil becomes uneconomic to produce, the industry will disappear and whatever is left in the ground will remain unknown: a geological fact of no economic interest.

On the demand side, it has been proven wrong to project current rates of growth well into the future. For example, in the late 1970s, the history of nearly 30 years of an average 7% per annum rate of growth in oil demand still played a powerful role in determining attitudes to the longevity of oil resources. This was at a time when growth rates had already fallen to around 1% per annum and when most

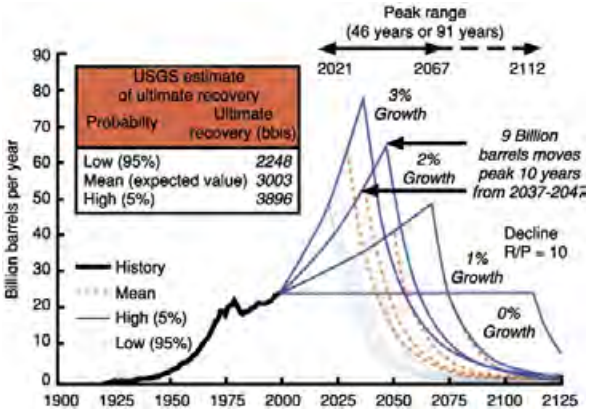


Figure 13 Conventional oil supply model showing the timing of peak oil with four rates of oil demand growth. R/P, reserves to production ratio; USGS, United States Geological Survey.

economists were forecasting a maximum growth rate of 3%. Changes in forecast to oil demand therefore have a major impact on the timing of peak oil; a difference of just 3% in average annual demand growth for oil gives a 46 year range in the timing of peak oil (Figure 13). Similar arguments can be applied to changes in the level of investment and the rate of depletion.

The Real Debate

The amount of fossil fuel on our planet is finite, certainly within the time-frame of civilization. However, does this necessarily mean that we will run out of petroleum reserves imminently or at all? Or will we develop alternative sources of energy long before the planet's petroleum reserves have been extracted? The changing pattern of world oil supply has meant that many of the predictions that oil would run out by the turn of the twentieth century have been proven wrong. Fears of impending scarcity of non-renewable energy have all been proven to be groundless as important changes on both the supply and demand sides have occurred.

The history of those who predict future trends in the depletion of petroleum reserves is one of persistent tension between those who assert that the best decisions are based on quantification and numbers, determined by the patterns of the past, and those who base their decisions on more subjective beliefs about the uncertain future. Ultimately, oil is a finite resource, but the real debate is not so much whether we are likely to significantly diminish the petroleum reserve base within the next few decades, but whether prices will fluctuate sufficiently to enable energy substitution. Volatility in the price of oil will sharpen the debate and could actually cause near-term unjustified changes in investment patterns by industry and governments alike.

See Also

Geological Surveys. **Petroleum Geology:** Overview; Chemical and Physical Properties; The Petroleum System; Exploration; Production. **Seismic Surveys**.

Further Reading

Adelman MA (1990) Mineral depletion, with special reference to petroleum. *The Review of Economics and Statistics* LXXII(1): 1–10.

BP (2003) *BP Statistical Review of World Energy 2003*. London: BP plc.

Campbell CJ (1997) *The Coming Oil Crisis*. Brentwood: Multi-Science Publishing Company and Petroconsultants.

Campbell CJ and Laherrere JH (1998) The end of cheap oil. *Scientific American* March: 80–86.

Deffeyes KS (2001) *Hubbert's Peak: The Impending World Oil Shortage*. Princeton, NJ: Princeton University Press.

Hubbert MK (1956) Nuclear energy and the fossil fuels. *Drilling and Production Practice*: 7–25.

Hubbert MK (1971) The energy resources of the Earth. *Scientific American* February: 31–40.

Laherrere JH (1999) World oil supply – what goes up must come down, but when will it peak? *Oil and Gas Journal* 97: 57–64.

Lynch MC (1996) The analysis and forecasting of petroleum supply: sources of error and bias. In: El Mallakh DH (ed.) *Energy Watchers VII*, pp. 51–71. Colorado: International Research Center for Energy and Economic Development.

Lynch MC (2001) Forecasting oil supply: theory and practice. *Quarterly Review of Economics and Finance* July: 1–28.

McCabe PJ (1998) Energy resources – cornucopia or empty barrel? *American Association of Petroleum Geologists Bulletin* 82(11): 2110–2134.

Odell PR (2001) *Oil and Gas: Crises and Controversies 1961–2000*. Brentwood: Multi-Science Publishing Company Ltd.

Simmons MR (2001) *The World's Giant Oilfields. Energy Service Industry Research*. Houston: Simmons and Company.

Smith AJ and Lidsky BJ (1993) King Hubbert's analysis revisited: update of the lower 48 oil and gas resource base. *The Leading Edge* 12: 1082–1086.

Weeks LG (1958) Fuel reserves of the future. *American Association of Petroleum Geologists Bulletin* 42: 431–438.

PLATE TECTONICS

R C Searle, University of Durham, Durham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

In the theory of plate tectonics, it is assumed that the Earth has a thin, brittle, outer layer, or 'lithosphere', which is broken up into a small number of tectonic 'plates'. These plates are further assumed to be rigid and undeformable, but capable of independent motion relative to each other on the surface of the Earth. The different relative motions of the plates are considered to be responsible for much of the varying geological activity on the Earth. The plate motions are describable in simple geometric terms and, once they have been determined, usually from a limited number of observations; they can be used to predict the type, magnitude, and direction of relative motions across plate boundaries anywhere in the world, both present and past. The power of plate tectonics theory arises both from its unifying view of global tectonics and from its powerful predictive ability. The subject of plate tectonics, *sensu lato*, is sometimes divided into 'plate kinematics', which deals with plate motions and their geometry, and plate tectonics *sensu stricto*, which deals with the geological consequences of the motions.

Tectonic Plates, Lithosphere, and Asthenosphere

The fundamental assumption underlying plate tectonics is that the outer part of the Earth consists of a small number of thin, rigid plates which are curved to fit the spherical shape of the Earth. The plates are considered to suffer no internal deformation, and only to deform and interact with each other along their boundaries, which are considered to be very narrow compared with the plate dimensions.

This view arose in the 1960s, largely as a result of the newly observed pattern of global seismicity provided by the World Wide Standard Seismograph Network. This showed that, over much of the world, earthquakes occur in narrow bands no more than a few tens of kilometres wide, whilst the large areas between these bands are essentially aseismic (Figure 1). The assumption was made that the aseismic areas are largely devoid of deformation and behave rigidly, and the bands of seismicity mark the plate boundaries. In simple plate tectonics theory, the plate boundaries are considered to have negligible width, although in practice they are associated with narrow but finite bands of geological activity (*see* **History of Geology Since 1962**).

Early workers on plate tectonics theory divided up the Earth's surface into about 12 major plates (Figure 2), although since then considerable numbers of smaller 'microplates' have been recognized.

The rigid outer layer of the Earth, which comprises the plates, is known as the 'lithosphere', from the Greek '*lithos*' meaning stone. Below this is a weaker layer, the 'asthenosphere' (Greek '*asthenia*', weakness), which deforms relatively easily and allows the plates to slide around. It is important to understand that these are terms describing the mechanical behaviour of the Earth and not its composition. In particular, the lithosphere does not normally correspond to the crust, but comprises both the crust and the uppermost part of the upper mantle, which usually behave together in a brittle manner (*see* **Earth:** Mantle; Crust). The lithosphere ranges in thickness from a few kilometres at mid-ocean ridge axes (where it may be entirely crust) to around a hundred kilometres in old ocean basins, and perhaps several hundred kilometres under continental cratons (where it is predominantly mantle). The asthenosphere constitutes a relatively weak layer of mantle below the lithosphere, capable of plastic deformation. The thickness of the asthenosphere is less well defined than that of the lithosphere, but may range from around 50 km in some places to several hundred kilometres, and may even encompass the whole upper mantle. Furthermore, except in the case of some subduction zones (see below), tectonic plate boundaries do not follow the boundaries between continents and oceans. An individual plate typically comprises regions of both continental and oceanic crust. For example, the South American plate consists of the continent of South America plus the western half of the South Atlantic Ocean (Figure 2).

The different mechanical behaviours of the lithosphere and asthenosphere are explained by the different deformation mechanisms that occur at different temperatures and pressures. At relatively low temperatures, most rocks deform elastically and suffer brittle failure (faulting) when their elastic limit or 'strength' is exceeded. This strength increases with pressure and hence depth (Figure 3). However, as the temperature increases, solid-state creep becomes increasingly important. Deformation by this mechanism becomes easier at higher temperatures, and the rock weakens exponentially (Figure 3). The depth at which brittle deformation gives way to ductile creep

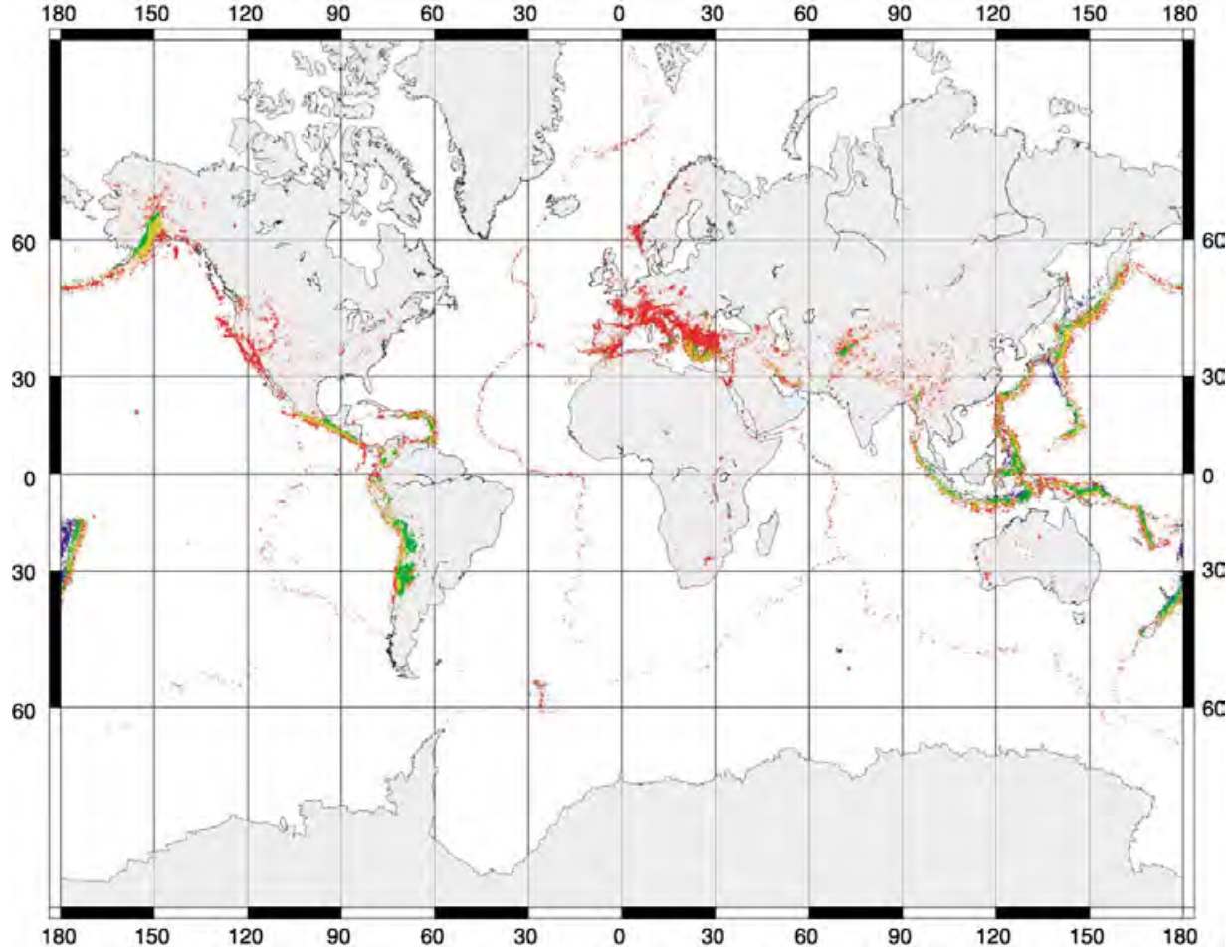


Figure 1 Distribution of shallow (0–33 km, red), intermediate (33–70 km, yellow; 70–300 km, green), and deep (300–700 km, blue) earthquakes of magnitude 4.0 or greater recorded between 1990 and 1996. Reproduced from D. Sawyer, http://www.geophysics.rice.edu/plateboundary/intro.html.

depends mainly on the composition of the rock, the rate of strain, and the geothermal gradient. Laboratory measurements of the mechanical properties of rocks and minerals predict depths of the brittle–plastic transition in the range suggested for the base of the lithosphere by observations of the Earth.

Evidence for the existence of the lithosphere and asthenosphere comes from a variety of sources. Seismic waves generated by earthquakes, particularly shear or S waves, show very little attenuation of their energy whilst travelling through the lithosphere, but are attenuated somewhat more in the asthenosphere (*see* **Tectonics:** Earthquakes). The velocities of these waves through the Earth show a marked decrease in the region of the asthenosphere. The effective viscosity of the Earth's mantle, as inferred from, for example, the rate of uplift of continental areas following the retreat of glaciers, shows a decrease in viscosity in the asthenosphere.

In general, crustal rocks are weaker than mantle, and, because oceanic crust is much thinner (around 6 km) than continental crust (approximately 35 km), oceanic lithosphere has a much greater component of mantle rocks and is correspondingly stronger. For this reason, oceanic areas tend to behave in a more rigid way than continents. In particular, continental mountain belts often display deformation over quite broad zones compared with oceanic plate boundaries (Figure 1).

Plate Tectonics

Although plate boundaries are primarily defined by the earthquakes along their boundaries, they are also marked by distinctive morphological features, narrow bands of tectonic activity, and, especially at transform boundaries (see below), by individual major faults. The active plate boundary may be only

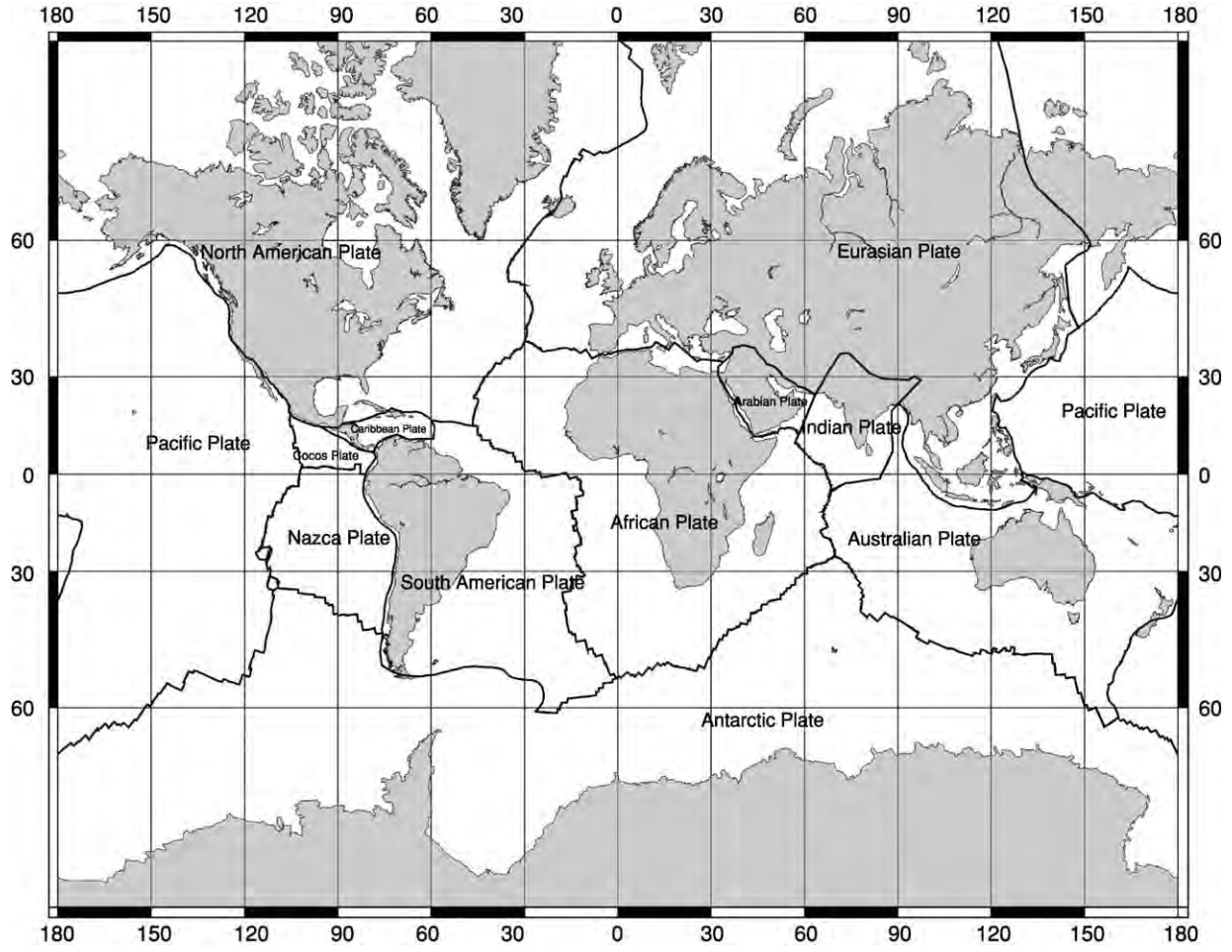


Figure 2 Major plates of the world and their boundaries. Reproduced from D. Sawyer, http://www.geophysics.rice.edu/plateboundary/intro.html, after D. Mueller, University of Sydney, Sydney, Australia.

a few kilometres wide, and rarely more than a few tens of kilometres.

Three types of plate boundary are recognized, corresponding to the three types of relative motion (Figure 4): 'divergent' (also 'ridge' or 'accretionary') plate boundaries, 'transform' (or conservative) plate boundaries, and 'convergent' ('subduction' or destructive) plate boundaries. In general, any type of plate boundary can join any other type.

Divergent (Ridge) Boundaries

Divergent boundaries occur only at mid-ocean ridges, and are the sites of creation of new lithosphere by seafloor spreading (*see* **Tectonics:** Mid-Ocean Ridges). The Mid-Atlantic Ridge and East Pacific Rise are typical examples. These boundaries are characterized by extensional tectonics dominated by normal faulting and extensional earthquake focal mechanisms. Generally, a divergent plate boundary is approximately orthogonal to the plate separation or spreading direction, but this is not essential, and ridges with greater or lesser extents of obliquity are fairly common.

The mid-ocean ridges associated with these plate boundaries arise because new plate is created at the boundary; thus, the plate gets older with increasing distance from the plate boundary and, as it does so, it cools, becomes denser, and thermally contracts. The plate boundary is thus the youngest, hottest, and shallowest region, forming the crest of the mid-ocean ridge. The detailed morphology of the plate boundary appears to depend on the temperature of the underlying mantle, which is mainly controlled by the spreading rate. For 'slow spreading ridges' (with a plate separation rate of less than about 60 km per million years), there is usually a 'median valley', a few kilometres deep and a few tens of kilometres wide, formed by rifting (Figure 5). At 'fast spreading ridges' (separation faster than 70 km per million years), the plate is too hot and weak to support rifting, and instead there is a volcanically built axial high.

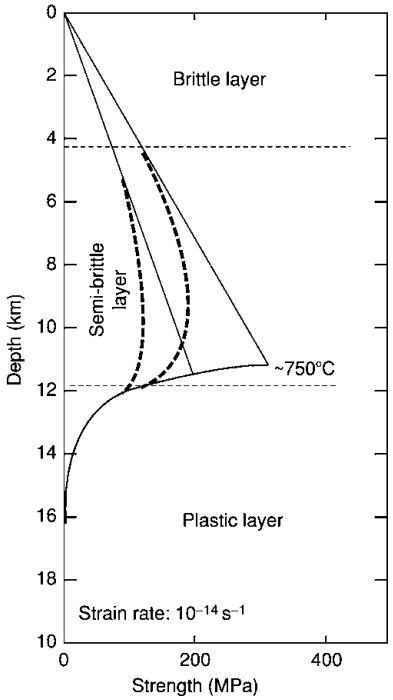


Figure 3 Calculated strength envelopes for material subject to brittle, semi-brittle, and plastic deformation with increasing depth. Reproduced from Searle RC and Escartín J (2004) The rheology and morphology of oceanic lithosphere and mid-ocean ridges. In: German C, Lin J, and Parson L (eds.) *Thermal Regime of Ocean Ridges and Dynamics of Hydrothermal Circulation. American Geophysical Union, Geophysical Monograph*, Fig. 8. Washington, DC: American Geophysical Union.

Ridges are formed by melting of dry mantle at depths of ~60 km, producing basaltic magma. Basaltic volcanism is widespread at mid-ocean ridge plate boundaries.

Plate tectonics theory allows for asymmetrical spreading (one plate accreting faster than the other). However, although temporary asymmetrical spreading is common, the net effect averaged over millions of years is usually approximately symmetrical. The precise reasons for this are not fully understood.

Transform Boundaries

Transform boundaries are those in which plates slide past one another with essentially no convergence or divergence; they thus conserve the areas of the adjacent plates. Because of this, they have the important property that they are exactly parallel to the direction of relative plate motion, and can be used for its estimation. Transform boundaries are characterized by strike-slip faulting and earthquakes with strike-slip mechanisms. A typical example of a transform boundary is the San Andreas Fault Zone in California, USA (Figure 6).

In simple plate tectonics theory, a transform boundary consists of a single fault called a 'transform fault', although, in practice, there is often a zone of such faulting some tens of kilometres wide. Because there is no large-scale convergence or divergence at these boundaries, there is also relatively little vertical relief,

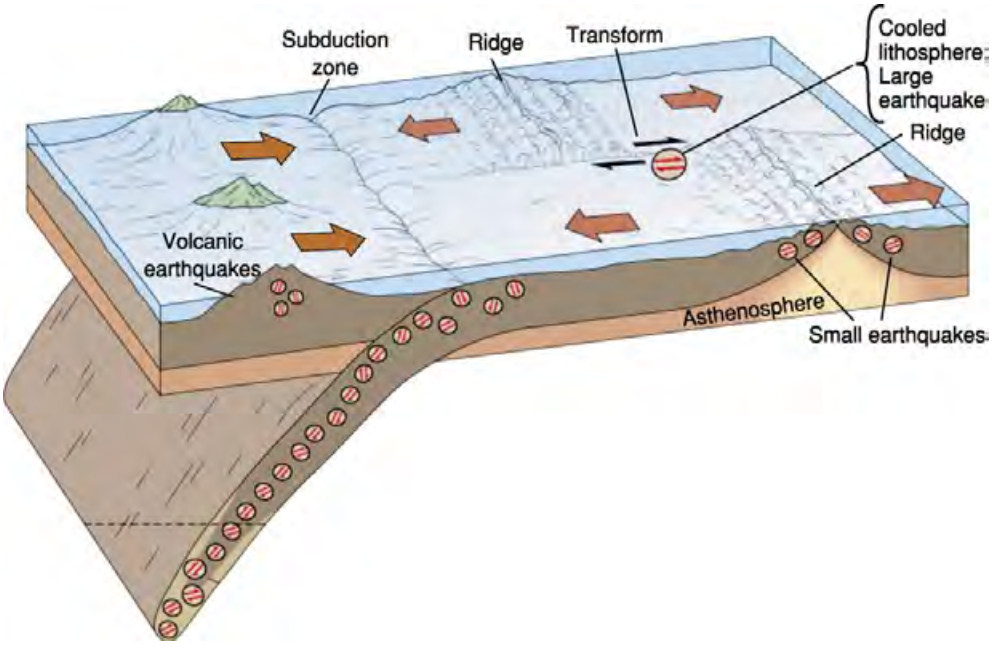


Figure 4 Diagram illustrating the three principal types of plate boundary (ridge, transform, and subduction), their general morphology, and the distribution of earthquakes. Large arrows show relative plate motion. Circles with double arrows indicate the general distribution of earthquakes. Modified from Davidson JP, Reed W, and Davis PM (2002) *Exploring Earth*, Fig. 8.13. Upper Saddle River, NJ: Prentice Hall.

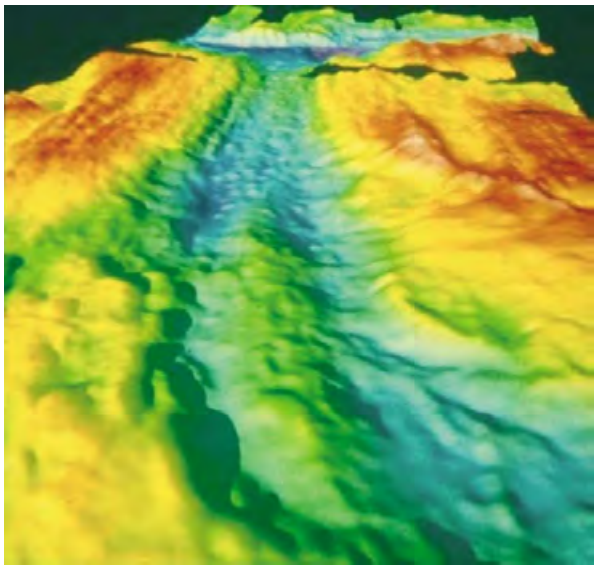


Figure 5 Oblique view of the topography of the Mid-Atlantic Ridge Median Valley near 24° N, based on multibeam echosounder measurements. The view is towards the south, with the cross-cutting Kane Transform Fault in the distance. The width of the distant view is approximately 50 km, width in the foreground about 20 km, and distance from the foreground to the horizon about 100 km. Depths less than 2500 m, red; greater than 4200 m, blue. The median valley is approximately 30 km wide. There is evidence of fault scarps on the valley walls. The ridge in the centre foreground is a linear volcano sitting above the plate boundary and representing the current axis of plate accretion. Image produced by M. Jones, University of Durham, Durham, UK, using data from Purdy GM, Sempere JC, Schouten H, Dubois DL, and Goldsmith R. (1990) Bathymetry of the Mid-Atlantic Ridge, 24° N–31° N Amap series. *Marine Geophysical Researches* 12: 247–252; and Pockalny RA, *et al.* (1988) *Journal of Geophysical Research* 93: 3179–3193.

although small valleys and scarps usually mark the trace of the transform fault itself. Transforms link other plate boundaries, and can thus 'transform' one type of tectonic boundary (e.g., extensional) to another (e.g., compressional), hence the name. Many transform faults offset mid-ocean ridges, producing (in map view) a staircase pattern of alternating ridge and transform boundaries (Figure 2). Such offsets range from a few tens of kilometres to over a thousand kilometres in length. Transform boundaries occur in both continental and oceanic lithosphere.

Convergent Boundaries

Convergent plate boundaries in the strict plate tectonics sense occur only at the deep-sea trenches and related 'subduction zones' (*see* **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges). Areas of active mountain building on continents (e.g., the Alpine–Himalayan zone) are zones of plate convergence, but are characterized by broad and almost continuous zones of deformation. This is partly because continental lithosphere is weaker and more easily deformed than oceanic lithosphere, and partly because it is less dense and so is not readily removed by subduction (see below). Plate tectonics does not provide a very useful description of such broad continental convergence zones.

Where one of the converging plates consists of oceanic lithosphere, it will be overridden by the other and pushed down into the asthenosphere in a process known as subduction (Figure 7). This can occur because oceanic lithosphere (unlike continental lithosphere) has a similar density to the underlying asthenosphere. Where both plates are oceanic, either one may be subducted under the other, and sometimes, depending on local conditions, the 'polarity' of subduction may reverse (i.e., the subducted plate breaks off and subsequently becomes the overriding one, whilst that which was overriding begins to subduct).

The plate boundary at a subduction zone is marked by a deep trench, the deepest and probably most famous of which is the Mariana Trench in the western Pacific, whose base is over 11 km below sea-level (*see* **Tectonics:** Ocean Trenches). As the slab subducts, it is deformed and generates earthquakes. These occur in a narrow band, known as the Wadati–Benioff zone, that follows the position of the subducting slab, and constitutes some of the best evidence for the existence of subduction zones. Subduction zones are very complex, and contain local regions of both compressional deformation (mainly at the actual plate boundary and in the leading edge of the overriding plate) and extensional deformation (mainly where the subducting plate bends to begin its descent). Earthquake mechanisms reflect this, and also show variation in mechanism with depth in the subducting plate. Some of the world's largest earthquakes are associated with subduction zones.

As the slab descends, it heats up and gives off water trapped in the crustal rocks. The presence of this water lowers the melting point of the surrounding mantle and, at a depth of some 100 km, magma is generated that rises through the overriding plate to create a volcanic arc, a hundred or so kilometres behind the trench (depending on the angle at which the slab is descending).

As with ridges, oblique subduction is allowed by plate tectonics theory, and is relatively common.

Plate Kinematics

Rotation Poles

The great strength of plate kinematics is that it can describe plate motions in terms of simple geometry,

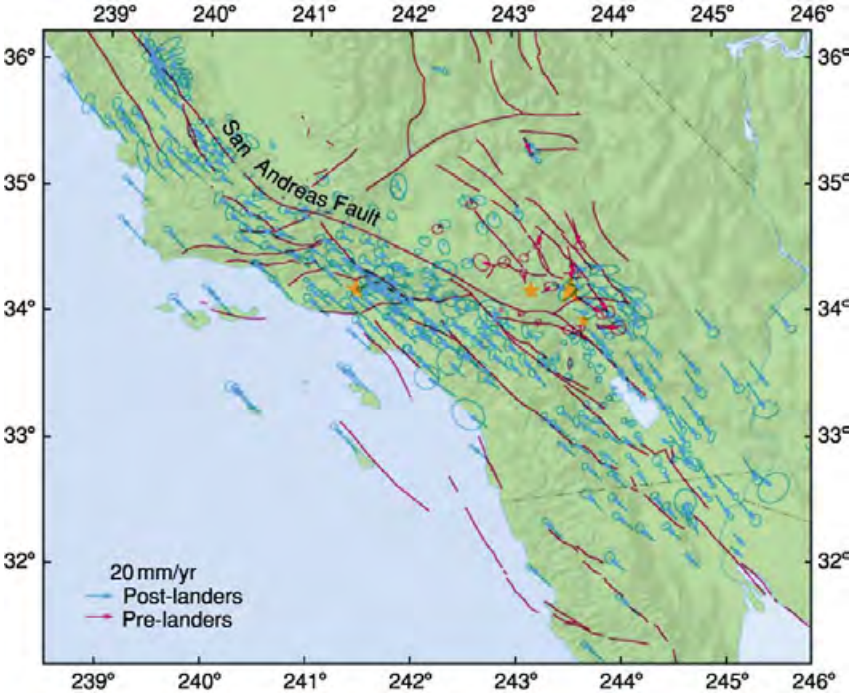


Figure 6 Map of major faults (red) and movement measured over 1 year by a Global Positioning System (GPS) (arrows) over part of the San Andreas Fault Zone in California, USA. Reproduced from Davidson JP, Reed W, and Davis PM (2002) *Exploring Earth*, Fig. 8.24. Upper Saddle River, NJ: Prentice Hall.

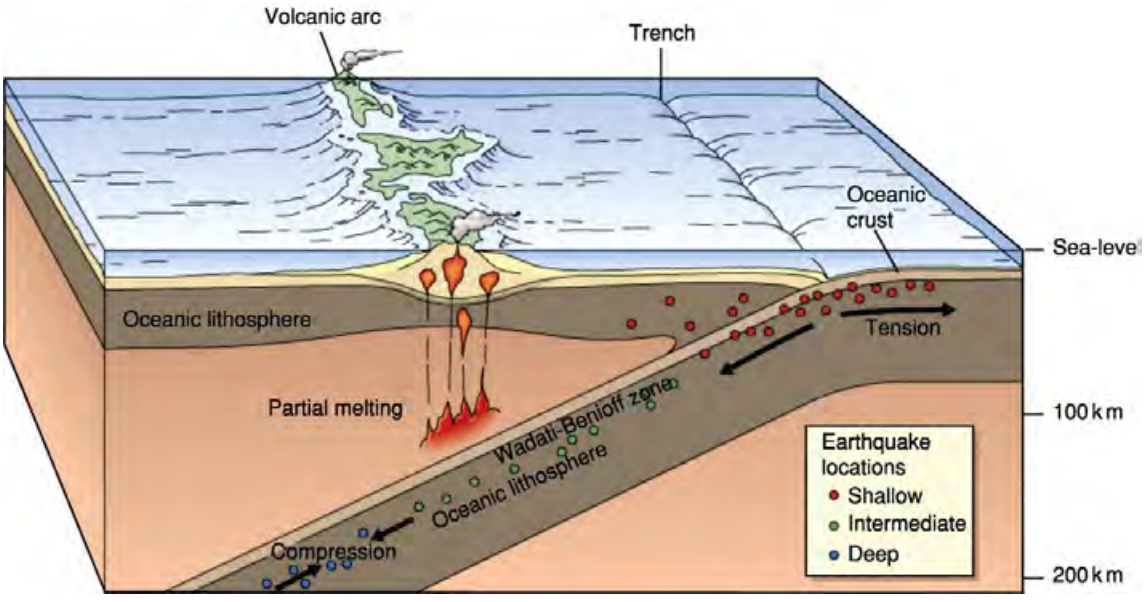


Figure 7 Diagram of a typical ocean–ocean subduction zone. Reproduced from Davidson JP, Reed W, and Davis PM (2002) *Exploring Earth*, Fig. 10.3a. Upper Saddle River, NJ: Prentice Hall.

and hence make precise predictions of relative motions anywhere on the globe. At the heart of this geometry is a concept called Euler's theorem, which states that any displacement of a rigid body on the surface of a sphere can be described in terms of a single rotation about a specified axis (Figure 8). Such axes cut the Earth's surface at pairs of points called 'rotation poles' (or 'Euler poles'). Once the poles and the angular rotations are specified, the whole motion is completely determined. Thus, the

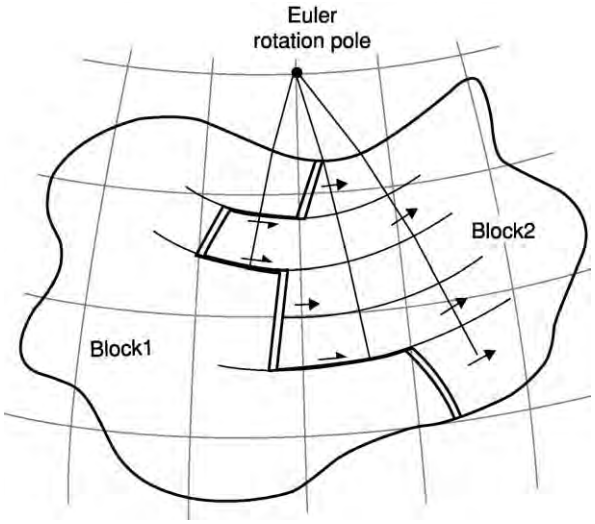


Figure 8 A Euler rotation pole describing the motion between Block 1 and Block 2. Plate boundary shown by double lines (mid-ocean ridge) and bold single lines (transform faults). Circles (medium lines through transforms) and great circles drawn orthogonal to them and passing through the pole are analogous to latitude and longitude in the geographic coordinate system. Light lines show the geographic coordinate system for comparison. Reproduced from Lowrie W (1997) *Fundamentals of Geophysics*, Fig. 6.30. Cambridge: Cambridge University Press.

motion of a given plate is specified in terms of its Euler pole and a corresponding angular rotation rate.

In practice, determining the motions of individual plates relative to some common reference frame is not easy, but determining 'relative' motions between pairs of plates is quite straightforward. These relative motions can also be described by rotation poles, although then the pole position must be specified relative to one or other of the pair of plates. Such relative rotation poles are the basis of most descriptions of plate kinematics.

As Euler poles define the 'directions' of the rotation axes, and the angles or rates of rotation define their 'magnitudes', rotations can also be described as vectors. Vector and matrix algebra can then be used to calculate plate motions, greatly simplifying and speeding up such calculations, which can be performed easily on computers.

It is important to distinguish between so-called 'instantaneous poles', which describe motion at an instant only, and 'finite rotation poles', which may describe the net result of motion over long periods of time. In descriptions of current plate motions, 'instantaneous' is usually taken to be about the past 1–3 million years. Motions over longer periods can be approximated by successions of so-called 'stage poles', each of which may describe the motion over a period of a few million years.

Like the geographic poles, we can imagine the Euler poles as centres of coordinate grids (Figure 8). 'Small circles' centred on the poles are the equivalent of lines of latitude, and plate relative motions are everywhere parallel to these small circles. Along any given small circle, the angular separation rate is constant, and increases with the distance from the rotation pole to the circle. 'Great circles' represent the shortest distance between two points on the surface of a sphere. Great circles that pass through the pole of rotation are equivalent to lines of longitude in the geographic system, and cut the small circles at right angles.

Measuring Plate Motions

Plates move at average rates of a few tens of millimetres per year. Thus, relatively indirect methods have usually been used to determine plate motions, although in recent years various geodetic measurements have been developed that are sufficiently precise to provide direct measurements of plate motions.

The most common way of determining plate separation rates is to use the linear magnetic anomalies (Vine–Matthews anomalies) produced during seafloor spreading to determine plate ages. These anomalies are produced as the Earth's magnetic field episodically reverses its direction, and the reversing field direction is recorded in the basalts of the oceanic crust as they cool after eruption at the mid-ocean ridge axis. This varying rock magnetization causes small variations, or 'anomalies', in the magnetic field above the seafloor, and these are readily observed by towing a magnetometer behind a ship or low-flying aircraft. Because the reversal process is irregular, the resultant magnetic anomalies have a characteristic pattern, which has been calibrated against crustal age through radiometric dating of samples of the magnetized rock. If the pattern of reversals can be recognized, the crust that is the source of the anomalies can thus be dated. The magnetic anomalies mark isochrons of crustal creation; therefore, by measuring the distance between a recent magnetic isochron on one plate and its conjugate on the other, the 'instantaneous' divergence rate, and hence the rotation rate, can be determined. (The commonly quoted 'spreading rate' is the rate at which a single plate accretes; for symmetrical spreading, it is half the divergence rate.) As the spreading or divergence rate varies with the distance from the Euler pole, in principle, the distance to the pole can be determined by measuring the spreading rate at several places along the plate boundary.

The best measure of the direction of relative plate motion is the azimuth of transform faults which, as stated above, are exactly parallel to the relative motion direction. Transform faults thus follow small

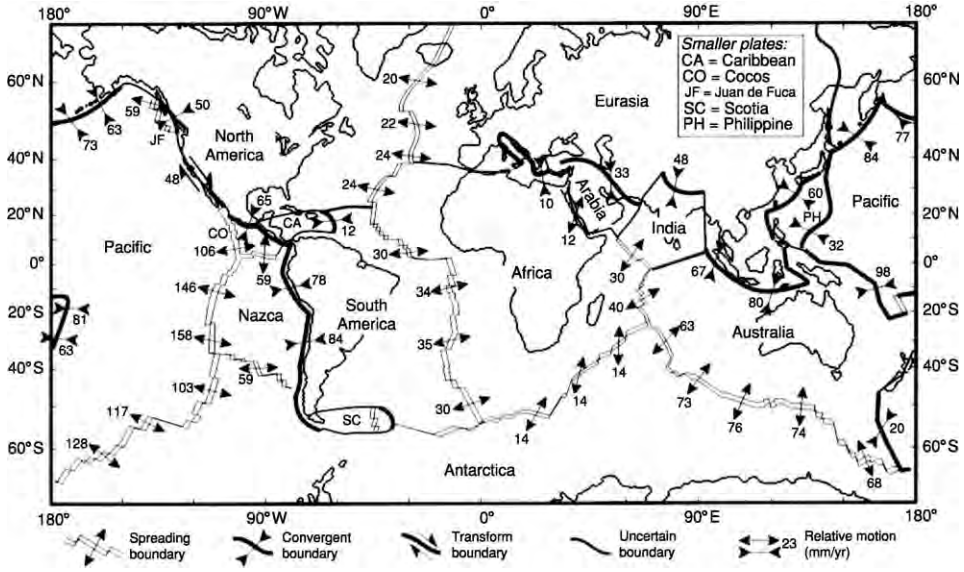


Figure 9 Map of plate boundaries showing the relative velocities across them from the data of DeMets C, Gordon RG, Argus DF, and Steins (1990). Reproduced from Lowrie W (1997) *Fundamentals of Geophysics*, Fig. 1.11. Cambridge: Cambridge University Press.

circles about Euler poles. A great circle at right angles to a small circle passes through the pole. Thus, if the azimuths of several transforms along a plate boundary are determined, great circles can be constructed normal to them, and should intersect at the Euler pole (Figure 8). Transform faults are readily recognized by their morphology (for example, a narrow linear valley for ridge–ridge transforms along mid-ocean ridges).

In practice, both spreading rates and transform azimuths are used together to solve for pole positions, sometimes supplemented by other data, such as earthquake focal mechanisms, to estimate relative motion directions. Euler poles may be calculated for individual plate pairs, for groups of plates, or for the global plate system (Figure 9). The most recent determination of global plate motions was performed by C. DeMets and others from Northwestern University, Illinois, USA, and provides a remarkably precise and self-consistent description of these motions (see 'Further Reading').

As stated above, an important attribute of plate kinematics is its ability to predict plate motions. An interesting example of this is the possibility of determining convergence rates across subduction zones. Even at ocean–ocean subduction zones, one plate is destroyed, together with the record of magnetic lineations carried on it. Thus, there was no direct way of measuring such motion until the recent development of sufficiently precise geodetic methods. However, the relative motions of the plate pair can be determined by global fits as described above, and then the motion at any point on the common plate boundary, including subduction zones, can be calculated from the Euler pole data.

In recent years, geodetic methods have been developed to the level at which they can begin to measure plate motions directly. Where plate boundaries exist on land (such as the Mid-Atlantic Ridge in Iceland or the San Andreas Fault in California, USA), standard geodetic methods, such as electronic distance measurement, can be used at a local scale (over ranges of a few kilometres). On a slightly larger scale of tens to hundreds of kilometres, precise relative position determinations (to precisions of a few millimetres) can be made by careful use of the Global Positioning System satellite network. Relative positions between widely separated continents can be determined by Very Long Baseline Interferometry, in which the variation in phase of radio signals from distant quasars is used. Repeat measurements by these methods over times of a few years can now resolve plate motions, and give results that, in general, agree well with the more traditional determinations.

The rotation rates of the major plates about their Euler poles range from about 2.1° per million years for the Cocos–Pacific pair, to about 0.1° per million years between Africa and Europe or Africa and Antarctica, and only 0.03° per million years for India–Arabia. Many of the minor plates (so-called 'microplates') rotate much faster than this, at tens of degrees per million years. In terms of linear rates, the fastest plate divergence rate at present is on the East Pacific Rise between the Pacific and Nazca plates, at

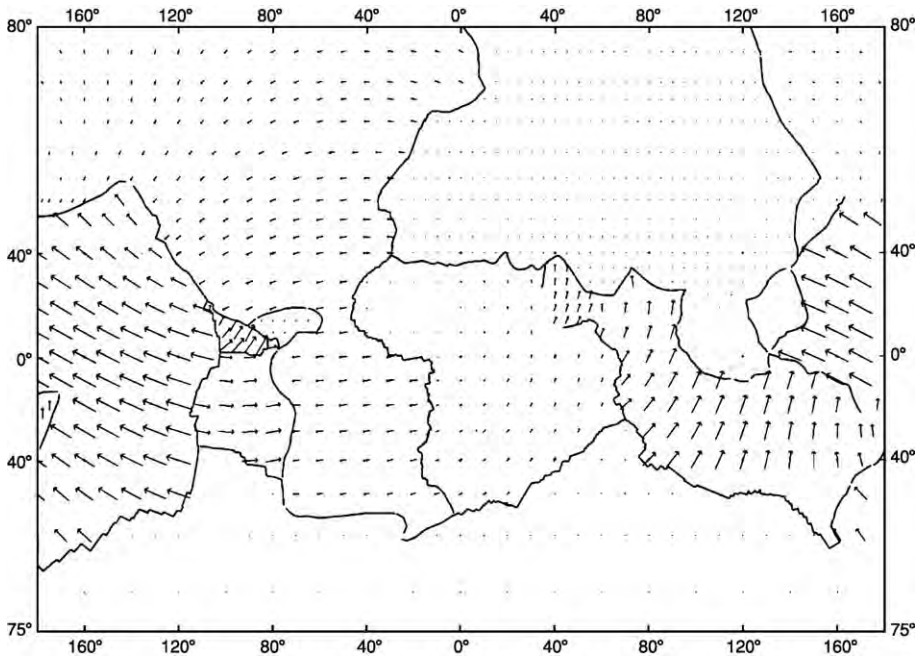


Figure 10 Map of plate boundaries and absolute motion vectors. Arrow lengths are proportional to the plate speed, with the longest (in the western Pacific) corresponding to speeds of about 100 mm per year. Reproduced from Fowler CMR (1990) *The Solid Earth: An Introduction to Global Geophysics*, Fig. 2.20. Cambridge: Cambridge University Press.

about 160 km per million years (or 160 mm per year). At the slow end, the North America–Eurasia plate boundary passes very close to its Euler pole in northern Siberia, where the relative motion becomes essentially zero.

Absolute Plate Motions

So far, we have dealt only with relative motions, which are fairly easy to determine. There is also interest in determining so-called 'absolute' plate motions, where the motions of all plates are related to some common reference frame. The possibility of doing this arises from the proposed existence of mantle plumes, which rise as narrow columns of relatively hot rock from deep in the mantle, possibly from the core–mantle boundary. They reach the Earth's surface in so-called 'hotspots' where they are manifest by clusters of intense volcanic and seismic activity. Well-known examples occur in Iceland and Hawaii, but there are thought to be many tens of such hotspots and associated plumes.

Hotspots leave clear trails on the Earth's surface, which comprise lines of volcanoes or volcanic seamounts and zones of thickened volcanically produced crust. The Hawaii–Emperor seamount chain, trending north-west from Hawaii, is an excellent example, but there are many other subparallel seamount trails in the south-western Pacific which are thought to have resulted from plumes. If points along the hotspot trails are dated (e.g., by radiometric dating of volcanic products), the relative motion between the plumes and the plate, and between given plumes, can be determined. We can also calculate the motions of individual plates relative to the average plume motion.

When this is performed, the relative motion between many of the hotspots turns out to be quite small, and significantly less than the average relative motions between plates. From this, and the fact that plumes are thought to rise through the mantle, it seems reasonable to assume that the average plume motion relative to the mantle is quite small, although it should be admitted that this has remained difficult to test in detail, and there are those who even question the existence of plumes. Nevertheless, if it is assumed that the average motion of hotspots relative to the mantle is zero, plate motions can then be given in the supposedly fixed mantle reference frame. These are referred to as absolute plate motions. They can also be described in terms of Euler poles, and are shown in Figure 10.

Mechanisms and Plate Driving Forces

Plates as Parts of the Mantle Convection Cycle

Plate motions are ultimately driven by the Earth's heat energy, and they are intimately related to the

mantle convection that is driven by this heat. One view of plates is that they simply represent the surficial parts of mantle convection cells: as hot, ductile mantle rises to the surface, it cools and becomes brittle – a plate – and then moves as a rigid block over the surface before being subducted, gaining temperature and becoming ductile again. Recent results from seismic tomography suggest that, around the rim of the Pacific, sheets of cold material descend below subduction zones deep into the lower mantle, implying a strong coupling of mantle motion and subducted plates.

However, the coupling is not perfect. There are some parts of the mid-ocean ridge (divergent plate boundaries) where it seems that the deeper mantle (below the asthenosphere) may be descending rather than rising. One such place is the so-called Australo-Antarctic Discordance south of Australia. Moreover, some plates, such as Africa, are almost entirely surrounded by ridges and have very few subduction zones on their boundaries. In such cases, a rigid coupling of plates to convection cells would imply the unusual scenario of upwelling along an expanding ring, with a downwelling column inside it. In fact, one of the advantages of plate tectonics is that it allows partial decoupling of plate motions from deeper mantle flow via the ductile asthenosphere.

The Forces Acting on Plates

Another way to look at the problem of the driving mechanism is to consider the forces acting directly on the plates. There are many possible forces, but amongst the most important are ridge push, slab pull, trench suction, and mantle drag. Ridge push arises from the tendency of the plates on a mid-ocean ridge flank to slide down the slopes of the wedge of thermally expanded asthenosphere that lies beneath the ridge. Slab pull arises from the negative buoyancy of the subducted plate which tends to drag the rest of the plate down with it. Trench suction is an additional force tending to pull plates together at subduction zones as a result of local convection driven by the subduction. Mantle drag is the frictional force between the base of the plate and the underlying asthenosphere.

It is possible to estimate some of these forces, at least approximately, and their relative importance can also be assessed by considering the observed stresses in plates and inferred absolute plate velocities. The latter is particularly instructive. Absolute velocities are largely independent of the total area of the plate, and so it is unlikely that mantle drag is an overall driving force, as was once thought (i.e., plates do not ride as passive passengers on top of mantle convection cells). However, plate velocities are inversely correlated with the area of continental lithosphere, suggesting that large areas of continent act as a brake (perhaps because such lithosphere is very thick or because subcontinental asthenosphere is cold and rather viscous). The fastest plates are those, mainly in the Pacific, which have large lengths of subduction zone along their boundaries, implying that slab pull and/or trench suction are important driving forces, as also suggested by calculation. There is a modest correlation with the effective length of ridge on the plate, indicating that ridge push is a driving force, but less strong than trench pull; this also tends to be backed up by calculations. Observations of intraplate stresses are also consistent with these conclusions.

Tests of Plate Tectonics

There have been numerous tests of plate tectonics theory. Its self-consistency, direct measurements of predicted plate motions, earthquake focal mechanisms, and distributions of earthquakes have all played their part in confirming the theory.

See Also

Earth: Mantle; Crust. **History of Geology Since 1962**. **Tectonics:** Convergent Plate Boundaries and Accretionary Wedges; Earthquakes; Mid-Ocean Ridges; Mountain Building and Orogeny; Ocean Trenches.

Further Reading

Cox A and Hart RB (1986) *Plate Tectonics – How It Works*. Oxford: Blackwell Scientific Publications.

DeMets C, Gordon RG, Argus DF, and Stein S (1990) Current Plate Motions. *Geophysical Journal International* 101: 425–478.

DeMets C, Gordon RG, Argus DF, and Stein S (1994) Effect of recent revisions to the geomagnetic reversal timescale on estimates of current plate motions. *Geophysical Research Letters* 21(20): 2191–2194.

Isacks B, Oliver J, and Sykes LR (1968) Seismology and the new global tectonics. *Journal of Geophysical Research* 73: 5855–5900.

Kearey P and Vine FJ (1996) *Global Tectonics,* 2nd edn. Oxford: Blackwell Science.

Le Pichon X (1968) Sea-floor spreading and continental drift. *Journal of Geophysical Research* 73: 3661–3697.

McKenzie DP and Parker RL (1967) The North Pacific: An example of tectonics on a sphere. *Nature* 224: 125–133.

Morgan WJ (1968) Rises, trenches, great faults and crustal blocks. *Journal of Geophysical Research* 73: 1959–1982.

Wilson JT (1965) A new class of faults and their bearing on continental drift. *Nature* 207: 343–347.

PRECAMBRIAN

Contents

Overview
Eukaryote Fossils
Prokaryote Fossils
Vendian and Ediacaran

Overview

L R M Cocks, The Natural History Museum, London, UK

Copyright 2005, Natural History Museum. All Rights Reserved.

Introduction

Since the Precambrian–Cambrian boundary can now be dated at 543 Ma ago, and since the Earth is estimated to have been formed at about 4500 Ma ago, it follows that the Precambrian represents about seven eighths or 88% of geological time. However, that fact was not at all obvious to early geologists who, 200 or so years ago, had no idea of the age of the Earth. Between about 1820 and 1845, Earth history was divided into various named systems, with the Cambrian the oldest, based on the successive assemblages of distinct fossils contained in the sedimentary rocks and on the Law of Superposition, which states that rocks are older than other rocks now above them (assuming that they have not been structurally overturned). Igneous rocks were either undated or dated relatively as younger than the sedimentary rocks through which they had been intruded. Since no unambiguous fossils were then known from rocks below the Cambrian, all such rocks were and are simply termed the Precambrian. It was not until the invention and progressive refinement of radioisotopic dating in the twentieth century that the true ages of both the Precambrian and the rocks above it began to be understood.

Divisions of the Precambrian

Unlike the Phanerozoic (Cambrian–Holocene), there are no formal subdivisions of the Precambrian. However, the Precambrian is normally divided into two: the earlier Archaean and the younger Proterozoic. The Proterozoic has itself been divided into three, and the Precambrian is thus divided as follows (the ages are obviously approximate):

Neoproterozoic – 1000 Ma–543 Ma,
Mesoproterozoic – 1600 Ma–1000 Ma,
Palaeoproterozoic – 2500 Ma–1600 Ma, and
Archaean – *ca*. 4000 Ma–2500 Ma

No Earthly rocks are known that are older than about 4000 Ma, so there is no universally used name for the period between the formation of the Earth at about 4500 Ma and the formation of the oldest known rocks of the Archaean, although the Hadean is a term used by some. The Neoproterozoic is divided into the Riphean (1000–600 Ma) and the Vendian (600–543 Ma), and the Vendian is sometimes termed the Ediacaran.

The Precambrian in this Encyclopaedia

A large number of articles in this encyclopaedia deal in various ways with aspects of the Precambrian, and thus this article is devised to help the reader to locate the appropriate entry through cross-referencing. Thus there follows a brief guide to the key aspects of the Precambrian, under the headings the origin of the Earth, major Precambrian outcrops, Precambrian sediments and climate, the origin of life and Precambrian fossils, Precambrian orogenies and Precambrian terranes and palaeogeography.

The Origin of the Earth

Our galaxy and the universe appear to have been formed about ten thousand million years (10 Ga) ago, but our star – the Sun – and its Solar System do not seem to have been formed for a long time after that, at about 4.5 Ga. The article on Earth structure and origins (*see* **Earth Structure and Origins**) describes how this process is thought to have occurred. The oldest known individual minerals are found within

zoned zircons from Western Australia, which give an age of nearly 4.4 Ga. However, the oldest known rocks are the Acasta Gneisses of Canada, which are about 4.0 Ga old; this age is closely followed in other areas, for example the 3.9 Ga Napier Complex of Antarctica. The oldest cratons so far identified form parts of the South African and West Australian shields (*see* **Shields**) and are dated at about 3.2 Ga.

Major Precambrian Outcrops

Although slivers of Precambrian rocks are found in many terranes consisting mainly of younger rocks (*see* **Terranes, Overview**), the great majority of Precambrian outcrops are found in the old shield areas of North America (principally the Canadian Shield, North American Precambrian Continental Nucleus (*see* **North America:** Precambrian Continental Nucleus; Continental Interior), South America (largely underlying Brazil and the adjacent areas), Africa (several areas) (*see* **Africa:** Pan-African Orogeny), Northern Europe (the East European Craton (*see* **Europe:** East European Craton)), Siberia (the Angara Craton, Central Asia (*see* **Asia:** Central), and Russia (*see* **Russia**)), India (*see* **Indian Subcontinent**), Antarctica (*see* **Antarctic**), and Australia (*see* **Australia:** Proterozoic). Several of these shield areas are made up of more than one shield unit; for example, three individual shields can be identified within Australia, and another three are found within Antarctica. From late in the Proterozoic and for all of the Palaeozoic, the South American, African, Indian, Antarctic, and Australian shields were grouped together to form the supercontinent of Gondwana (*see* **Gondwanaland and Gondwana**).

Precambrian Sediments and Climate

The Earth's atmosphere in Archaean and pre-Archaean times was a reducing one, consisting largely of carbon dioxide, nitrogen, water vapour, and inert gases, with subsidiary amounts of hydrogen, methane, and ammonia. A critical event in Earth history was the change just before about 2.2 Ga in the atmosphere from reducing to oxidizing, termed the Great Oxidation Event. The sediments reflect that profound change; for example, a characteristic sediment of many Archaean and Early Palaeoproterozoic shields is the banded iron formation (*see* **Sedimentary Rocks:** Banded Iron Formations). This sediment type is first known from the Isua Supracrustal rocks of Greenland, dated at 3.8 Ga, and these chemically precipitated deposits, which contain up to 35% iron, reach their maximum distribution between 2.8 Ga and 2.5 Ga, the end of the Archaean. They are often associated with greywackes deposited by turbidity currents in 'greenstone belts', which are extensive sedimentary mudstones and coarser rocks (usually metamorphosed) with high percentages of iron and other minerals. Following the change in the atmosphere from reducing to oxidizing, there are no true banded iron formations, although comparable granular iron formations are known from a few areas, notably Lake Superior, Canada (dated at about 1.9 Ga), and, much later (Late Neoproterozoic), the Yukon and Namibia.

Each grain of clastic sediment at the bottom of the oceans today has been through several (an estimated average of five) cycles of erosion and sedimentation during Earth history. Since there was no vegetation on the land and soils only developed progressively with time, Precambrian erosion rates were much higher than those of today. Nearly all Early Archaean sediments were volcanogenic, usually basic or intermediate in chemical composition, but today these account for only about 20% of sediments, most of the earlier ones having broken down into their individual mineral components. Since quartz is more resistant to weathering than most other minerals, the proportion of quartz in sedimentary rocks has progressively increased, reaching today's figure of approximately 50%. Limestones became common for the first time in the Proterozoic and often contain algal stromatolites. These Precambrian limestones had a higher magnesium content than average modern limestones, which resulted in a higher proportion of dolostones. A form of coal termed schungite is found for the first time in rocks aged about 2.0 Ga, and, although now metamorphosed, it probably formed initially from local accumulations of blue-green algae. Chert is known from Archaean rocks onwards and occurs in the form of layers within banded iron formations and as nodules within dolomites; however, these cherts are of inorganic or secondary origin, and primary cherts, formed from siliceous organisms such as radiolaria, are not known until the Late Neoproterozoic.

Precambrian climates have formed the topic of many research projects. As can be seen from the desert sediments that many Precambrian rocks contain, the climates were often very hot; however, at other times, glacigenic deposits were widespread. These Precambrian glaciations occurred in several phases, and some workers believe that the Earth was at one time completely covered in glaciers, the so-called 'snowball Earth', but that concept is controversial, although palaeomagnetic evidence indicates that some glacigenic rocks in Australia must have been close to the then equator.

The Origin of Life and Precambrian Fossils

It is not known when life originated here on Earth; however, as far as can be deduced, life originated only once. Prokaryotes, which include bacteria (*see* **Biosediments and Biofilms**), are apparently simpler than eukaryotes, since they lack a nucleus, and were probably the first organisms. Because early organisms were largely soft-bodied, the early fossil record is tantalizingly incomplete. As detailed in the article on Precambrian prokaryotes (*see* **Precambrian:** Prokaryote Fossils), the changes in overall carbon isotopes seen after 3.5 Ga indicate that biological systems were in place by that time. This was well before the Great Oxidation Event at about 2.2 Ga, and thus early life forms must have lived in a reducing environment.

Laminated structures known as stromatolites have been described from rocks as old as 3.4 Ga in Australia and South Africa; however, it appears likely that the earliest stromatolites were inorganic in origin. In contrast, stromatolites deposited after about 3.0 Ga were apparently organic in origin. Although the range and diversity of these complex organisms peaked in the Middle Mesoproterozoic and Neoproterozoic, their descendants are still to be found living today, most famously at Shark Bay, Western Australia. The role of stromatolites and other early organisms in taking carbon dioxide out of the atmosphere and replacing it with free oxygen played a major part in the change of the atmosphere during the Great Oxidation Event at about 2.2 Ga.

As the Precambrian progressed, life forms became more diverse (*see* **Precambrian:** Eukaryote Fossils), although they rarely exceeded 1 mm in size. A major evolutionary event took place in the later Proterozoic Vendian period (*see* **Precambrian:** Vendian and Ediacaran), when much larger macroscopic fossils, albeit still without hard parts, evolved; these are known as the Ediacara fauna. That fauna has been found on most of the larger Late Precambrian terranes, but apparently became extinct at or near the Precambrian–Cambrian boundary.

Precambrian Orogenies

In the Archaean there were many orogenies, but in most cases their detailed ages and lateral extents have not yet been fully determined, and long periods of orogeny are often conflated under a single name. For example, various orogenies have been recognized as occurring during the formation and subsequent fusion of the three cratons that make up present-day Australia. However, the 2.6–2.4 Ga Sleafordian Orogeny of Australia marked the assembly of the South Australia craton and spanned the Archaean–Palaeoproterozoic boundary (*see* **Australia:** Proterozoic). In central Canada the Trans-Hudson Orogeny extended from 2.2 Ga to 1.7 Ga, peaking at 1.9 Ga (*see* **North America:** Precambrian Continental Nucleus). Some workers correlate the 1.9–1.8 Ga Barramundi Orogeny of the North Australian craton with the Trans-Hudson Orogeny of North America, and the two together form a key element in the assembly of the Rodinia supercontinent (see below). In Antarctica there are at least three different Archaean–Palaeoproterozoic cratons that formed between 3 Ga and 1.6 Ga (*see* **Antarctic**).

The best-known Precambrian orogeny in the Mesoproterozoic is the Grenvillian Orogeny (*see* **Grenvillian Orogeny**), which occurred between 1.3 Ga and 1.0 Ga and probably resulted in the assembly of the superterrane of Rodinia. The Grenvillian Orogeny is principally represented in eastern North America and consists of an earlier accretionary stage between 1.3 Ga and 1.2 Ga, widespread magmatism from 1.2 Ga to 1.1 Ga, and a continent–continent collision at about 1.0 Ga. The 1.3–1.1 Ga Albany Fraser Orogeny of the South and West Australian cratons has been interpreted as part of the Grenvillian Orogeny.

During the final part of the Neoproterozoic, the Timanide Orogeny (*see* **Europe:** Timanides of Northern Russia) affected today's northern Europe and resulted in the accretion of a substantial number of terranes to the old craton of Baltica. In addition, the Cadomian Orogeny affected most of the terranes now in southern Europe; this continued until about 530 Ma, in the Early Cambrian.

Precambrian Terranes and Palaeogeography

It is not known exactly when continental and oceanic crust became differentiated on the Earth's surface, but the process is thought to have started before the Archaean. The oldest known cratons are dated at about 3.2 Ga, and it is not until after the formation of these cratons that we can be sure that wide shallow continental-shelf seas actually existed. Although there are glimpses of early terranes, for example between 3.0 Ga and 1.6 Ga (Archaean–Palaeoproterozoic) several Antarctic cratons were stabilized, there is no clear image of global palaeogeography prior to the Late Mesoproterozoic.

The existence of a Precambrian superterrane was suggested in the 1970s, as geologists noted a number of 1.3–1.0 Ga mountain belts now on different continents, and later the name Rodinia was adopted for

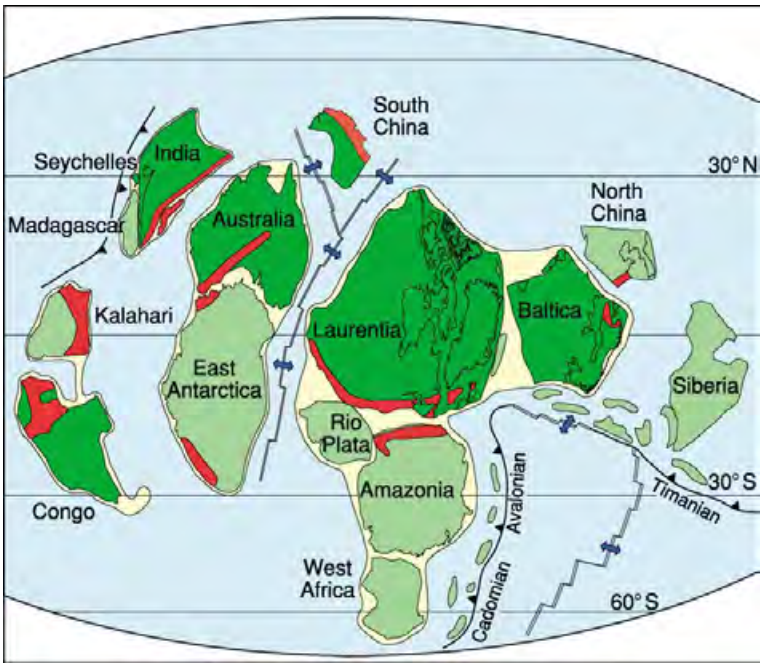


Figure 1 Palaeogeographical reconstruction at 750 Ma, just after the initial breakup of the Rodinia superterrane. The dark green shading represents terranes with good palaeomagnetic data, in contrast to the light green from with palaeomagnetic data is poor or absent. The red areas are the tectonic belts. Modified by Trond Torsvik, Trondheim, from Torsvik (2003).

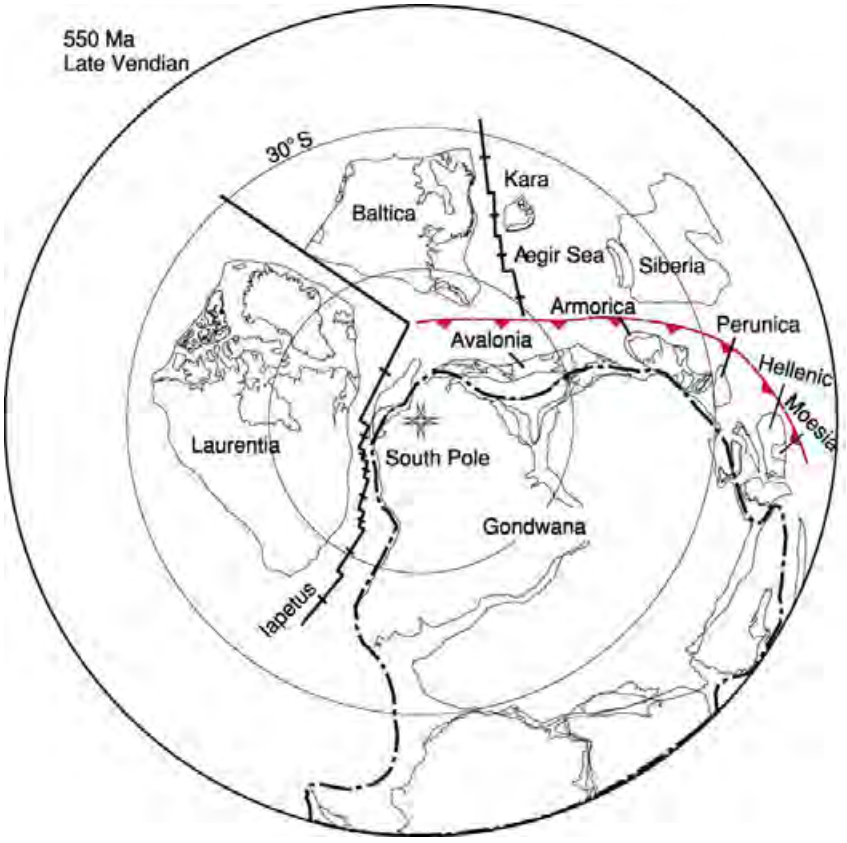


Figure 2 Palaeogeographical reconstruction at 550 Ma, near the end of the Neoproterozoic. New diagram by Trond Torsvik, Trondheim.

this terrane. There is now a progressively convincing scenario emerging of the slow assembly of the superterrane of Rodinia during the Mesoproterozoic between 1.3 Ga and 1.1 Ga. After relative stability for a long period, Rodinia started to break up at about 850 Ma. That breakup is documented by several events; for example, the 760 Ma Areyonga event in Australia may represent rifting between Australia and the new terrane of Laurentia. Similarly, Laurentia appears to have separated from Antarctica at about 750 Ma (Figure 1). However, South America, Africa, peninsular India, Australia, and Antarctica stayed together and, with further smaller accretions, remained together to form the superterrane of Gondwana in the Late Neoproterozoic, a process that was complete by the Early Cambrian. Between 580 Ma and 555 Ma there was massive plutonism in Scandinavia, and flood basalts in the Ukraine occurred at 576 Ma; these events may document the separation of Baltica from Laurentia. Gondwana remained a united superterrane until after the breakup of Pangaea in the Early Mesozoic.

Figure 2 shows a probable palaeogeography for the Late Neoproterozoic. The progressive rifting between Laurentia (North America) on the one hand and Gondwana and Baltica on the other resulted in the opening and steady widening of the Iapetus Ocean between the two sets of terranes, an opening that continued until about the end of the Cambrian. In addition the opening of the Ran Ocean marked the division between Baltica and Gondwana, and the Aegir Sea separated Baltica from Siberia. Most of the other side of the globe was occupied by the vast Panthalassic Ocean.

See Also

Africa: Pan-African Orogeny. **Antarctic**. **Asia:** Central. **Australia:** Proterozoic. **Biosediments and Biofilms**. **Earth Structure and Origins**. **Earth System Science**. **Europe:** East European Craton; Timanides of Northern Russia. **Gondwanaland and Gondwana**. **Grenvillian Orogeny**. **Indian Subcontinent**. **North America:** Continental Interior. **North America:** Precambrian Continental Nucleus. **Precambrian:** Eukaryote Fossils; Prokaryote Fossils; Vendian and Ediacaran. **Russia**. **Sedimentary Rocks:** Banded Iron Formations. **Shields**. **Terranes, Overview**.

Further Reading

Cocks LRM (ed.) (1981) *The Evolving Earth*. Cambridge: Cambridge University Press.

Edwards K and Rosen BR (2000) *From The Beginning*. London: The Natural History Museum.

Hancock PL and Skinner BJ (2000) *Oxford Companion to the Earth*. Oxford: Oxford University Press.

Hartz EH and Torsvik TH (2002) Baltica upside-down: a new plate tectonic model for Rodinia and the Iapetus Ocean. *Geology* 30: 255–258.

Torsvik TH (2003) The Rodinia jigsaw puzzle. *Science* 300: 1379–1381.

Windley BF (1995) *The Evolving Continents*, 3rd edn. New York: John Wiley.

Eukaryote Fossils

S Xiao, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Eukaryotes, archaeobacteria, and eubacteria are the three domains of living organisms (**Figure 1**). Most life forms that we can see with the naked eye are multicellular eukaryotes, which include animals, fungi, plants, and seaweeds. Many eukaryotes (for example, dinoflagellates, ciliates, and amoebae) are, however, single-celled and microscopic.

Eukaryotes are cytologically distinct from prokaryotes (archaeobacteria and eubacteria). A typical eukaryotic cell contains membrane-bound intracellular structures such as a nucleus, mitochondria, and, for photosynthetic eukaryotes, chloroplasts. Eukaryotes are also distinctively characterized by DNA-associated histone, eukaryotic gene regulation, and tubulin- and actin-based structures known as cytoskeletons, which help to maintain and manipulate the shape of the cell.

Nuclei, mitochondria, chloroplasts, histone, and cytoskeletons are rarely preserved in the fossil record, making it a serious challenge to recognize fossil eukaryotes, particularly single-celled ones. It has been suggested that cell size may be a useful guide: eukaryotes tend to have larger cells than prokaryotes. The size range of eukaryotic cells, however, overlaps considerably with that of prokaryotic cells. Cell size alone therefore is only indicative, not conclusive, evidence of eukaryotic affinity. Morphological complexity provides more reliable evidence. Because of the supportive function of cytoskeletons,

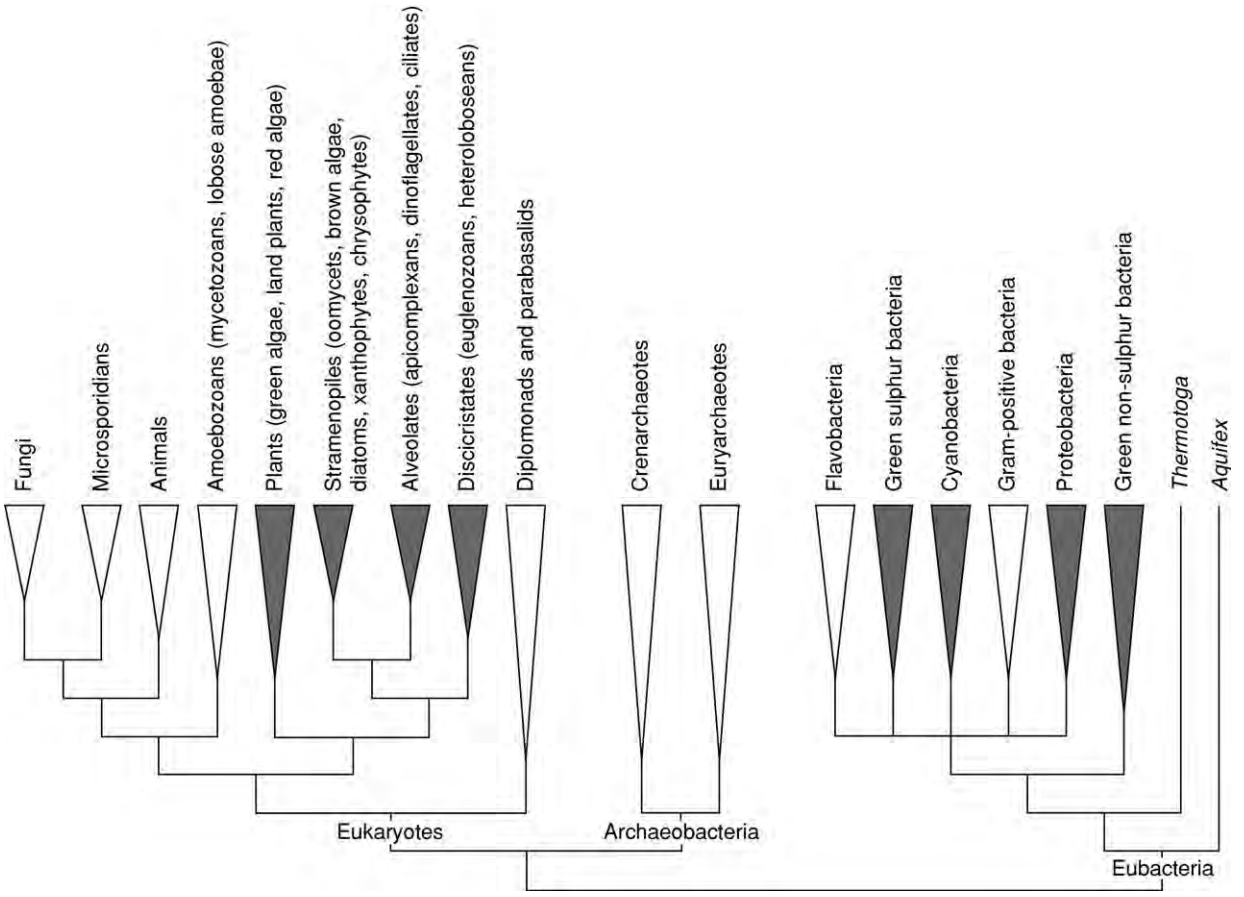


Figure 1 Phylogenetic relationships among the three major domains (eubacteria, archaeobacteria, and eukaryotes) and phylogenetic structure within the eukaryote domain. Shaded clades include photosynthetic members. Simplified from Woese CR, Kandler O, and Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the USA* 87: 4576–4579 and Baldauf SL, Roger AJ, Wenk-Siefert I, and Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972–977.

many eukaryotes maintain complex cell morphologies that can be preserved in the fossil record. For multicellular organisms, eukaryotic morphologies extend to cell organization and differentiation. Through intercellular interaction (e.g. via plasmodesmata and pit connections), controlled cell division, and genetic regulation, multicellular eukaryotes can have well-organized cell arrangements and some degree of cellular differentiation, which distinguishes them from prokaryotes. In addition, eukaryotic biochemistry may leave fingerprints in ancient rocks in the form of molecular fossils or biomarkers. Steranes, for example, are good eukaryote biomarkers. Steranes are derived from steroids, and the biosynthesis of steroids is almost exclusively a characteristic of eukaryotes.

Not only the size range but also the morphological disparity of eukaryotes overlaps with that of prokaryotes. Many eukaryotes, particularly single-celled ones, can be morphologically simple and volumetrically small. It can be very difficult to distinguish such eukaryotes from prokaryotes. This is a significant challenge because the majority of Precambrian microfossils are morphologically simple. Many of them are featureless filamentous or spheroidal structures, less than a few tens of micrometres in size, preserved in cherts or shales. The default interpretation of these microfossils is that they are prokaryotes. This is not an unreasonable interpretation for many, particularly if they can be shown to be microbial mat builders. For others, however, a eukaryote interpretation is equally plausible; there is simply not enough morphological detail to allow an unambiguous interpretation. These fossils are not the focus of this article. The discussion below takes a more conservative approach, focusing only on the Precambrian fossils that can be reasonably interpreted as eukaryotes. In addition, the discussion will concentrate on marine sediments; there are no known convincing eukaryote fossils in Precambrian terrestrial sediments.

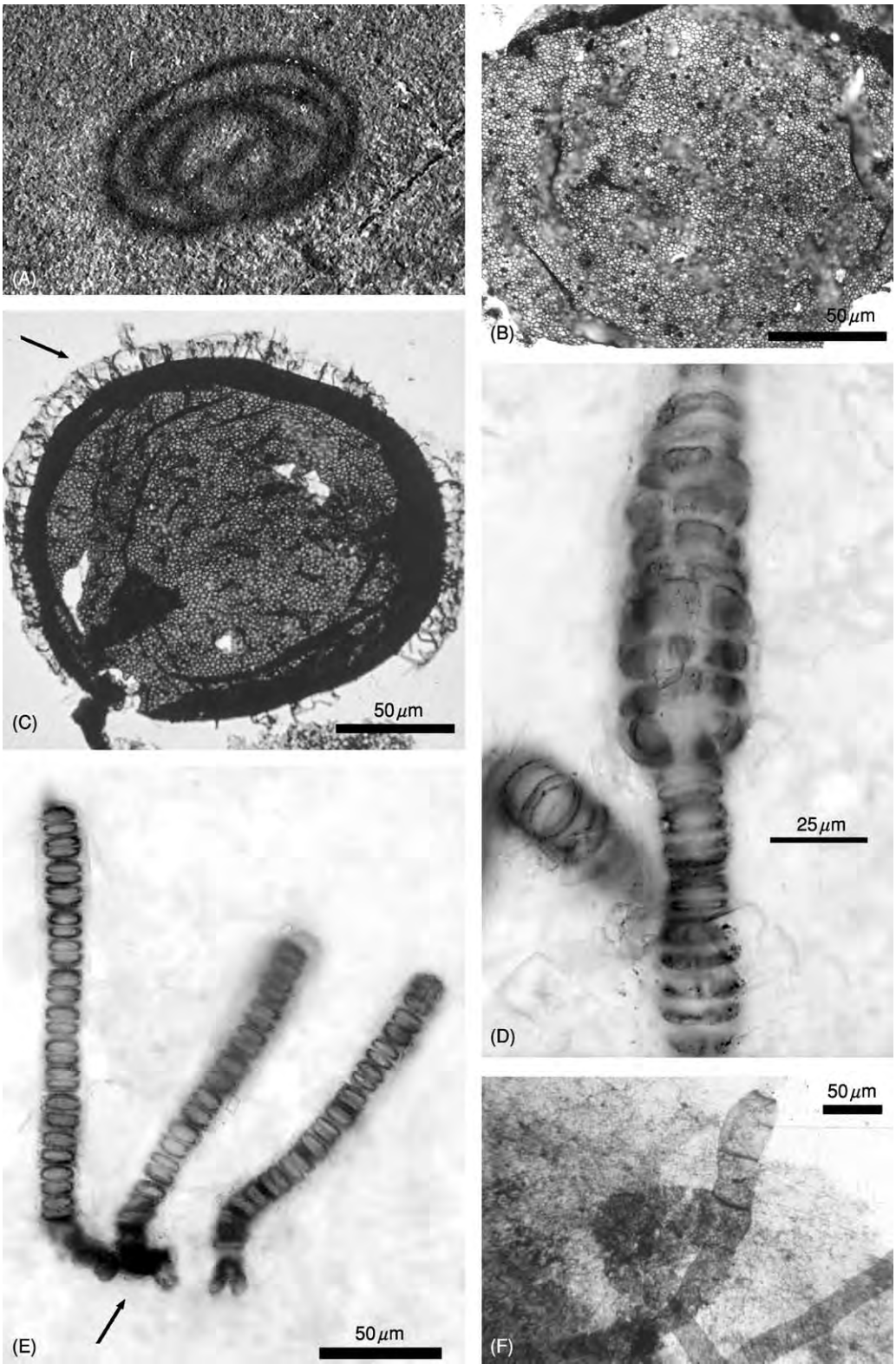


Figure 2 Palaeproterozoic (A) and Mesoproterozoic (B–F) eukaryote fossils. (A) Coiled ribbons from the Negaunee Iron Formation, Michigan, USA. Specimen is about 2 cm in maximum dimension. Photograph courtesy of Bruce Runnegar. (B) *Dictyosphaera delicate* from the Ruyang Group, northern China. Notice the polygonal pattern (*ca.* 1 μm in diameter) on the vesicle surface. (C) *Shuiyousphaeridium macroreticulatum* from the Ruyang Group, northern China. Arrow points to spines on the vesicle surface. This taxon also shows a polygonal pattern on the vesicle surface. Reproduced with permission from Xiao S, Knoll AH, Kaufman AJ, Yin L, and Zhang Y (1997) Neoproterozoic fossils in Mesoproterozoic rocks? Chemostratigraphic resolution of a biostratigraphic conundrum from the

Archaean (3800–2500 Ma) Eukaryotes

Phylogenetic relationships among the three domains (Figure 1) indicate that the eukaryotes may have a history that extends back as far as those of the other two domains. Since there are microbial fossils in the Archaean, the possibility of Archaean eukaryotes is intriguing. Indeed, we can follow the footprints of eukaryotes into the Archaean. Steranes, which are eukaryote biomarkers derived from steroids, occur abundantly in the ~2600 Ma Marra Mamba Formation and the ~2715 Ma Maddina Formation in northwestern Australia. But so far no eukaryote microfossils have been found in Archaean rocks.

The presence of steranes in Archaean rocks, however, does not necessarily imply that all of the features (mitochondria, cytoskeletons, and histone) that collectively define living eukaryotes evolved in the Archaean. It is possible that the biosynthesis of steroids appeared early in eukaryote history. These Archaean organisms may represent ancient branches that are more closely related to the eukaryotes than to the other two domains but that diverged before the last common ancestor of all living eukaryotes evolved. In other words, they may represent stem-group eukaryotes.

Palaeoproterozoic (2500–1600 Ma) Eukaryotes

Eukaryotic biomarkers continue to occur in Palaeoproterozoic and younger rocks, for example in the McArthur Group (approximately 1600–1700 Ma) in northern Australia. But it is in the Palaeoproterozoic rocks that the earliest morphological evidence for eukaryotes has been found. Such morphological evidence appears in two forms: macroscopic carbonaceous compressions and organic-walled microfossils.

Among Palaeoproterozoic carbonaceous compressions, the coiled ribbons (Figure 2A) of millimetric width and centimetric length from the *ca*. 2000 Ma Negaunee Iron Formation of Michigan are probably the most famous because of the ancient age of the formation. The Negaunee ribbons resemble the Mesoproterozoic fossil *Grypania spiralis*. Millimetre-sized discoidal to elliptical compressions resembling *Chuaria* and *Tawuia* are known from the *ca*. 1800–1900 Ma Changzhougou Formation and Chuanlinggou Formation in northern China. The slightly younger (*ca*. 1700 Ma) Tuanshanzi Formation in northern China contains millimetre- to centimetre-sized carbonaceous ribbons and blades. Although the cellular details of these fossils are not preserved, their macroscopic and stable morphologies suggest that they are probably the earliest eukaryotic fossils known so far.

The Changzhougou Formation and Chuanlinggou Formation in northern China also contain organic-walled microfossils (or acritarchs) and multicellular eukaryotes. The Chinese acritarchs are spherical vesicles with simple morphology but relatively large size (about 100 μm in diameter). They are commonly referred to the genus *Leiosphaeridia*. They are interpreted as single-celled resting cysts of ancient eukaryotes. The relatively thick and resistant vesicles of the resting cysts allowed their preservation in carbonaceous shales. The cellular details of another fossil, *Qingshania magnifica*, described from the Chuanlinggou Formation, are preserved, and the organism shows evidence of cellular differentiation – the expanded terminal cell at one end of the clavate filament was probably a reproductive cell. *Qingshania magnifica* may well be a multicellular eukaryote.

Palaeoproterozoic eukaryotic fossils do not have distinct morphological features that would allow them to be placed into extant eukaryotic clades. Like the Archean molecular fossils, the Palaeoproterozoic fossils may also reflect stem-group divergence in the early history of eukaryotes.

Mesoproterozoic (1600–1000 Ma) Eukaryotes

Morphologically simple acritarchs such as *Leiosphaeridia* continue to dominate Mesoproterozoic microfloral assemblages, and *Grypania spiralis*, *Chuaria*, and *Tawuia* are also known in Mesoproterozoic successions. However, several morphologically complex acritarchs first appeared in the Mesoproterozoic. In addition, the first crown-group eukaryotes (that is, fossils that can be phylogenetically resolved to an extant eukaryote clade) also occur in the Mesoproterozoic. This indicates that crown-group eukaryotes began to diversify in the Mesoproterozoic. By the

North China Platform. *Precambrian Research* 84: 197–220. (D, E) *Bangiomorpha pubescens*, interpreted as a bangiophyte red alga, from the Hunting Formation, arctic Canada. Arrow in (E) points to basal holdfasts. Specimen in (D) shows the transition from uniseriate to multiseriate growth, suggesting multiple ontogenetic phases and probably sexual reproduction in *Bangiomorpha pubescens*. Reproduced with permission from Butterfield NJ (2000). *Bangiomorpha pubescens* n.gen., n.sp.: Implications for the evolution of sex, multicellularity, and the Mesoproterosoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26: 386–404. (F) *Palaeovaucheria clavata*, interpreted as a xanthophyte alga, from the Lakhanda Group, south-eastern Siberia. Photograph courtesy of Andrew H. Knoll.

end of the Mesoproterozoic, several algal groups had already diverged.

Acanthomorphic acritarchs (or acritarchs bearing spines) and acritarchs with patterned vesicles are known in the 1400–1500 Ma Roper Group of northern Australia and the >1000 Ma Ruyang Group of northern China. The Roper and Ruyang assemblages include such acanthomorphs as *Tappania plana* and *Shuiyousphaeridium macroreticulatum* (Figure 2C), as well as acritarchs with polygonally patterned (*Dictyosphaera delicate* and *Shuiyousphaeridium macroreticulatum*; Figure 2B and 2C) or striated (*Valeria lophostriata*) vesicles. Some of these fossils have wide geographical and long stratigraphical ranges. For example, *Valeria lophostriata* and *Tappania* occur in Mesoproterozoic and Neoproterozoic rocks in Laurentia, Australia, India, and northern China. As no prokaryotes are known to have comparable levels of morphological complexity, these acritarchs are probably eukaryotic. Some of these acritarchs (for example *Dictyosphaera delicate*) preserve an organic $\delta^{13}C$ signature that is consistent with eukaryotic photosynthetic biochemistry, but it is unclear which algal group (e.g. chlorophytes, rhodophytes, or stramenopile algae – a group that includes chrysophytes, xanthophytes, diatoms, and brown algae; Figure 1) they belong to.

One of the earliest eukaryotic fossils that has been confidently attributed to a modern algal group is *Bangiomorpha pubescens* (Figures 2D and 2E) from the 1200 Ma Hunting Formation in arctic Canada. This is a multicellular filamentous fossil that shows evidence of holdfast differentiation and sexual reproduction. It is interpreted as a benthic bangiophyte red alga. Another phylogenetically resolved eukaryotic fossil is *Palaeovaucheria clavata* (Figure 2F), interpreted as a xanthophyte alga, from the upper Mesoproterozoic Lakhanda Group in south-eastern Siberia. Xanthophyte algae are members of the photosynthetic stramenopiles whose plastids were derived from a secondary endosymbiont (probably a red alga). The occurrence of *Bangiomorpha pubescens* and *Palaeovaucheria clavata* in Mesoproterozoic rocks suggests that not only must crown-group eukaryotes such as red algae have diverged but also the secondary endosymbiotic event leading to stramenopile algae must have occurred by the end of the Mesoproterozoic.

Neoproterozoic (1000–540 Ma) Eukaryotes

The Neoproterozoic era includes several major milestones in eukaryote evolution. The diversity and morphological complexity of eukaryotes increased appreciably in the Neoproterozoic, and several phylogenetically and ecologically important eukaryotic groups make their first appearance in the Neoproterozoic fossil record. These include heterotrophic protists, biomineralizing protists, and, towards the end of this era, animals. Molecular-clock estimates also indicate that land plants and fungi may have diverged in the Neoproterozoic, but so far this has not been confirmed by palaeontological evidence.

Major environmental crises occurred in the middle Neoproterozoic. Between about 720 and 600 Ma, the Earth experienced at least two global glaciations (also known as 'snowball Earth events'), during which glaciers reached the tropical oceans. More glaciations may have occurred in the Neoproterozoic, but these were not nearly as extreme. It is therefore convenient to divide the Neoproterozoic into three intervals: Early (1000–720 Ma), Middle (720–600 Ma), and Late (600–543 Ma) Neoproterozoic.

Early Neoproterozoic

A quick look at several fossiliferous units of Early Neoproterozoic age gives us a broad picture of eukaryote diversity at that time. The Early Neoproterozoic Little Dal Group (850–780 Ma) in northwestern Canada, the Chuar Group (>742 Ma) in the Grand Canyon, and the Huainan Group and Huaibei Group (*ca.* 740–900 Ma) in northern China contain some of the best-preserved carbonaceous compressions in the Early Neoproterozoic. *Chuaria* and *Tawuia* (Figures 3A and 3B) are abundant in these successions. In addition, *Longfengshania stipitata* – a benthic alga with an ellipsoidal head, a stipe, and a simple holdfast – has been described from the Little Dal Group and from the Changlongshan Formation (*ca.* 800–900 Ma) in northern China. Individuals of *L. stipitata* sometimes occur in clusters (Figure 3C), like a bunch of inflated balloons tethered together. These carbonaceous compressions are probably multicellular eukaryotes, although *Chuaria* and *Tawuia* have been interpreted as colonial cyanobacteria by some palaeontologists.

Acritarchs, particularly those with more complex morphologies, became more diverse in the Early Neoproterozoic (Figure 3D). More than 20 acritarch species with complex morphologies have been described from Early Neoproterozoic successions. Except a few (e.g. *Tappania* and *Valeria lophostriata*), they are not known in the Mesoproterozoic. The Wynniatt Formation (*ca.* 800–900 Ma) of arctic Canada, the Svanberfjellet Formation (*ca.* 700–800 Ma) in Spitsbergen, and the Mirojedikha Formation in Siberia, for example, contain some of the best-preserved acritarchs in the Early Neoproterozoic. The Svanberfjellet

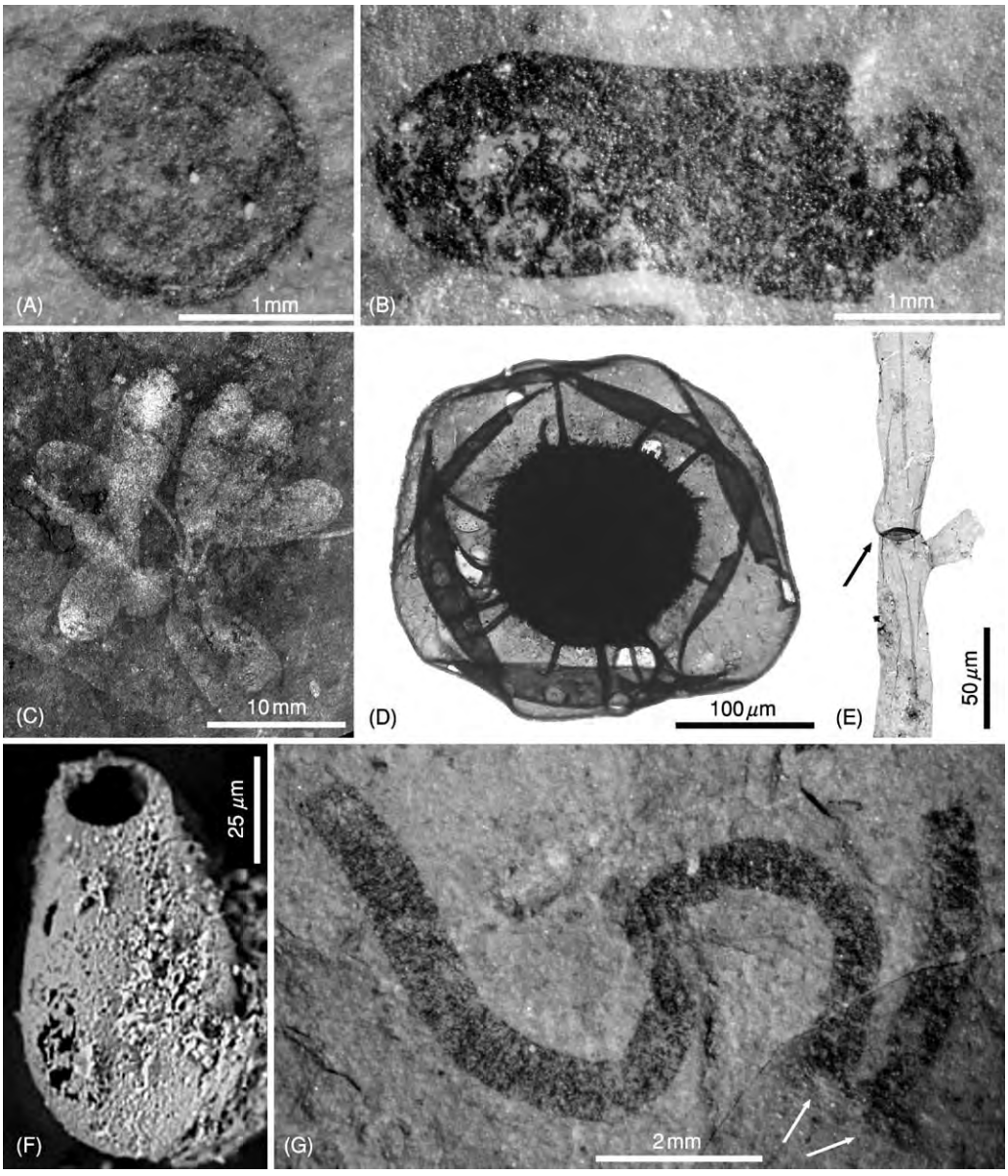


Figure 3 Early Neoproterozoic eukaryote fossils. (A) *Chuaria circularis* from the Huaibei Group, northern China. (B) *Tawuia dalensis* from the Huaibei Group, northern China. (C) *Longfengshania stipitata* from the Little Dal Group, north-western Canada. Reprinted with permission from Butterfield NJ, Knoll AH, and Swett K (1994) Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Fossils and Strata* 34: 1–84. (D) *Trachyhystrichosphaera polaris* from the Svanberfjellet Formation, Spitsbergen. Reproduced from Hofmann HJ (1985) The mid-Proterozoic Little Dal macrobiota, Mackenzie Mountains, north-west Canada. *Palaeontology* 28: 331–354 by permission of Taylor & Francis AS. (E) *Proterocladus major*, interpreted as a coenocytic green alga, from the Svanberfjellet Formation, Spitsbergen. Reproduced from Butterfield NJ, Knoll AH, and Swett K (1994) Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Fossils and Strata* 34: 1–84 by permission of Taylor & Francis AS. (F) *Cyclocyrillium simplex*, a vase-shaped microfossil interpreted as a testate amoeba, from the Chuar Group, Grand Canyon. Reproduced with permission from Porter SM and Knoll AH (2000) Testate amoebae in the Neoproterozoic era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. *Paleobiology* 26: 360–385. (G) *Protoarenicola baiguashanensis*, probably a benthic alga, from the Huaibei Group, northern China. Arrows point to holdfast structures. Photograph courtesy of Xunlai Yuan.

Formation has also yielded the earliest known green algal fossil–*Proterocladus* (Figure 3E).

Biologically controlled mineralization and the formation of skeletons is an evolutionary event that significantly enhances fossil preservation. All fossils discussed in the preceding sections are non-biomineralizing eukaryotes. The earliest known biomineralizing protists are probably from the Early Neoproterozoic. Silicified scale-shaped microfossils (less than 100 μm in diameter) from the Tindir Group (age poorly

constrained, but possibly 620–780 Ma) of north-western Canada can be compared to chrysophyte skeletons. Vase-shaped microfossils (Figure 3F) from the Chuar Group and other Early Neoproterozoic successions (e.g. the Visingsö Formation in Sweden and the Draken Conglomerate Formation in Spitsbergen) have been interpreted as testate amoebae whose mineralized tests are typically preserved in casts and moulds.

The Chuar testate amoeba fossils add another dimension to our consideration of the Neoproterozoic biosphere, and that is heterotrophy. Because testate amoebae are heterotrophic protists, the Chuar vase-shaped microfossils suggest that the Early Neoproterozoic biosphere was ecologically complex. Of course, heterotrophic eukaryotes must have evolved earlier. In fact, the earliest eukaryotes may be heterotrophic, given that eukaryotic autotrophy evolved through primary and secondary endosymbiotic events. However, evidence for heterotrophy in the fossil record is scarce. Vase-shaped microfossils and ciliate biomarkers from the Chuar Group are probably the earliest known evidence for heterotrophic eukaryotes.

The best-known heterotrophic eukaryotes are perhaps the animals. There have been many reports of animal fossils from Early Neoproterozoic and Mesoproterozoic successions, but their interpretation has been controversial. *Sinosabellidites huainanensis*, *Pararenicola huaiyuanensis*, and *Protoarenicola baiguashanensis* (Figure 3G), from the Huainan Group and Huaibei Group (*ca*. 740–900 Ma) of northern China, are some of the often-cited Early Neoproterozoic animal fossils. They are carbonaceous compressions of tubes of millimetric diameter and centimetric length with transverse annulations. The transverse annulations superficially resemble animal metameric segmentation. A few specimens bear poorly defined terminal structures that have been interpreted as proboscis-like structures. However, recent study has shown that these carbonaceous compressions are probably benthic tubular algae.

Middle Neoproterozoic

The Middle Neoproterozoic is characterized by multiple global glaciations, and is unofficially labelled as the Cryogenian Period by some Precambrian geologists. On a broad scale, acritarchs and other eukaryotes suffered significant losses of diversity in the Middle Neoproterozoic. The documented diversities of several Middle Neoproterozoic assemblages are extremely low, and such assemblages are typically dominated by *Sphaerocongregus variabilis* (or *Bavlinella faveolata*). This Cryogenian drop in eukaryote diversity may be a true evolutionary pattern that was related to the glaciation events. Despite the loss of diversity, the occurrence of red algae, green algae, photosynthetic stramenopiles, and testate amoebae in Mesoproterozoic and Early Neoproterozoic rocks suggests that some members of these groups must have survived the Middle Neoproterozoic glaciations.

Late Neoproterozoic

Eukaryote diversity rose sharply in the Late Neoproterozoic. Both acritarchs and multicellular algae reached unprecedented levels of complexity and diversity in the Late Neoproterozoic. Some of the multicellular algae are preserved in anatomical detail, allowing them to be placed within the red algae.

One of the most important landmarks in Late Neoproterozoic eukaryote evolution is the emergence of animals and animal biomineralization. Molecular-clock studies suggest that the deepest (protostomes–deuterostomes) divergence within the crown-group bilaterian animals probably occurred in the Mesoproterozoic or Early Neoproterozoic. But, as discussed above, there is no convincing palaeontological evidence to support these molecular-clock estimates. Some have suggested that perhaps the earliest animals were microscopic in size and would not be well preserved in the fossil record.

The Doushantuo Formation and Dengying Formation of South China provide several taphonomic windows onto the Late Neoproterozoic biosphere. Carbonaceous shales, cherts, and phosphorites of the Doushantuo Formation (*ca*. 600–550 Ma) preserve some of the most extraordinary eukaryote fossils in the Neoproterozoic. More than 20 taxa of macroscopic carbonaceous compressions have been reported from the Yangtze Gorges area and elsewhere in South China. Most of these compressions can be unambiguously interpreted as multicellular algae (Figure 4A). Some of them show clear evidence of holdfast anchoring, dichotomous branching, apical meristematic growth, and specialized reproductive structures. A few of these compressions (for example *Calyptrina striata* and *Sinospongia typica*; Figure 4B), however, have been interpreted as sponges or cnidarians, but such interpretations are not unique and an algal interpretation cannot be falsified conclusively. In any case, it is safe to conclude that none of the Doushantuo compressions can be interpreted as macroscopic bilaterians (bilaterally symmetrical animals).

Multicellular algae also occur in the Doushantuo cherts and phosphorites. Cellular features are preserved (Figures 4C–F), so anatomical detail can be deduced from these fossils. Many of these silicified or phosphatized algal fossils show pseudoparenchymatous thallus construction, apical meristematic

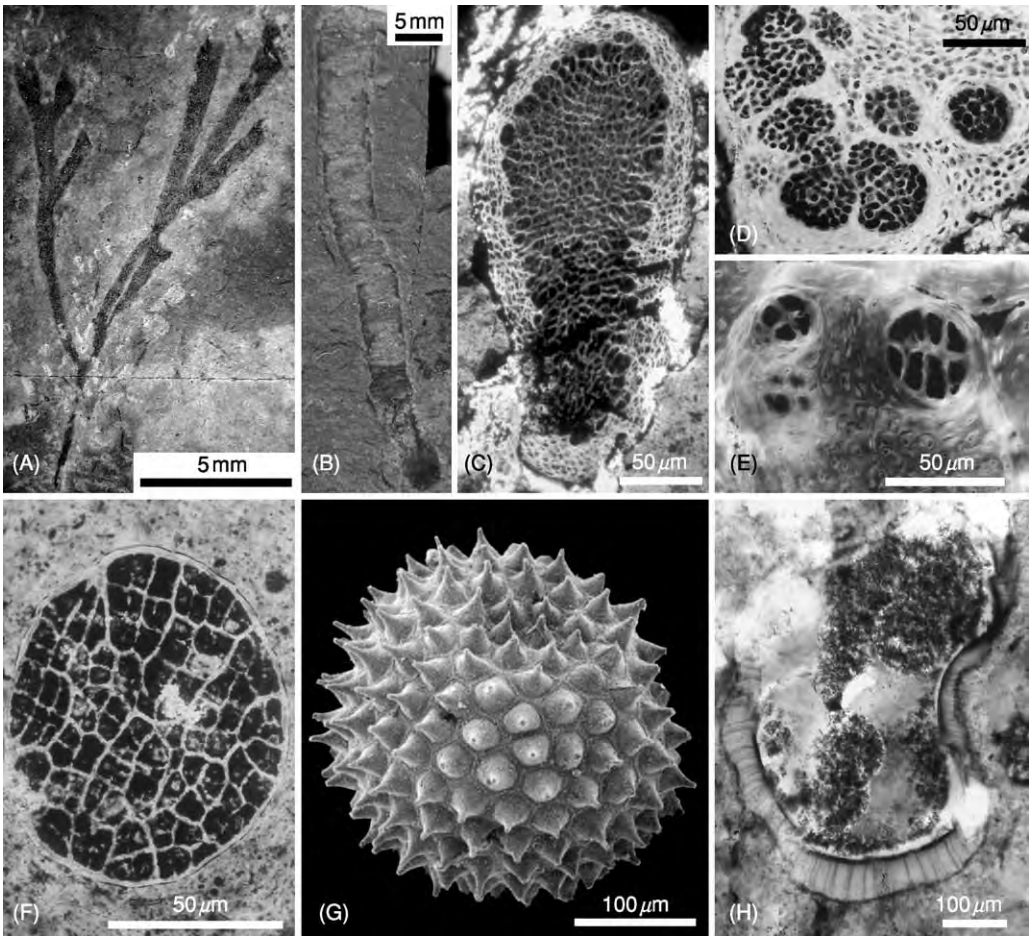


Figure 4 Late Neoproterozoic eukaryote fossils from the Doushantuo Formation, South China. (A) *Konglingiphyton erecta*, a dichotomously branching alga from the Doushantuo shales. (B) *Sinospongia typica*, possibly a benthic tubular alga from the Doushantuo shales. A, B, reproduced from Xiao S, Yuan X, Steiner M, and Knoll AH (2002) Macroscopic carbonaceous cpmpressions in a terminal Proterozoic shule: A systematic reassessment of the Miaohe biota, South China *Journal of Paleontology* 76: 345–374 with permission of The Paleontological Society. (C) A florideophyte red alga with medulla–cortex thallus differentiation from the Doushantuo phosphorites. (D, E) Florideophyte red algae with possible reproductive structures from the Doushantuo phosphorites: (D) larger and darker cells arranged in clusters are interpreted as possible carposporangia; (E) tetrads and octads embedded in algal thallus are interpreted as possible tetraspores and octospores. (F) *Wengania globosa*, a possible stem-group florideophyte red alga from the Doushantuo phosphorites. (G) *Meghystrichosphaeridium reticulatum*, an acanthomorphic acritarch from the Doushantuo phosphorites. (H) *Tianzhushania spinosa*, a large acanthomorph from the Doushantuo cherts. (A, C–F) are courtesy of Xunlai Yuan.

growth, thallus differentiation (Figure 4C), and specialized reproductive structures (Figures 4D and 4E). Some of them have been interpreted as stem-group florideophyte red algae or stem-group coralline algae, suggesting the presence of advanced red algae in Late Neoproterozoic oceans.

In addition, the Doushantuo cherts and phosphorites contain an assemblage of large (several hundred micrometres in diameter) acanthomorphs that are morphologically complex and taxonomically diverse (Figure 4G and 4H). This assemblage includes nearly 30 species, some of which are also known to occur in other Late Neoproterozoic successions, for example the Pertatataka Formation of central Australia, the Scotia Group of Svalbard, the lower Krol Group (Krol A) of Lesser Himalaya, and lower Vendian rocks in eastern Siberia. In the Yangtze Gorges area, elements of this acritarch assemblage (for example *Tianzhushania spinosa*) first appear just metres above Cryogenian glacial deposits, suggesting that the eukaryote recovery occurred shortly after the last Cryogenian glaciations.

The Doushantuo diversification of complex acritarchs appears to be ephemeral, however. Available evidence suggests that most of the Doushantuo acritarchs disappeared when Ediacaran animals

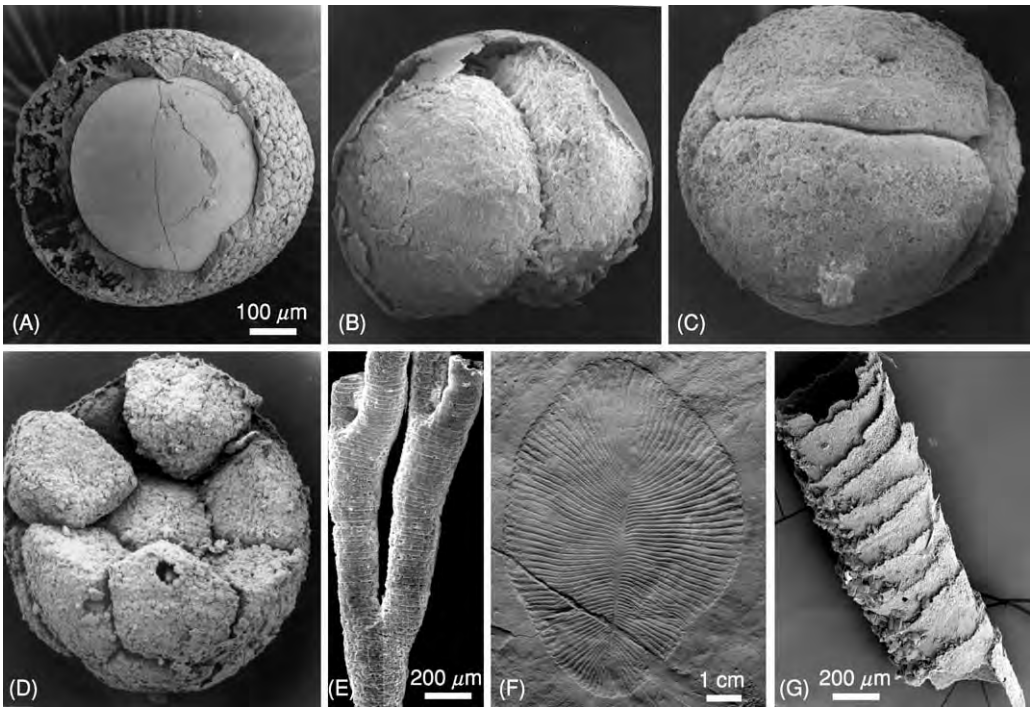


Figure 5 Late Neoproterozoic eukaryote fossils. (A–D) Animal eggs and blastulas at (A) one-, (B) two-, (C) four-, and (D) many-cell cleavage stages, Doushantuo phosphorites, South China. Scale bar in (A) applies to (A–D). Reproduced from Xiao S, Zhang Y, and Knoll AH (1998) Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* 391:553–558 with permission from Nature Publishing Group, and from Xiao S and Knoll AH (2000) Phosphatized animal embryos from the Neoproterozoic Doushantuo Formation at Weng'an, Guizhou, South China. *Journal of Paleontology* 74:767–788 with permission from The Paleontological Society. (E) *Sinocyclocyclicus guizhouensis*, interpreted as a stem-group cnidarian fossil, from the Doushantuo phosphorites, South China. (F) *Dickinsonia costata*, an Ediacaran fossil from the Flinders Ranges of South Australia. Photograph of a cast replicate preserved at the Botanical Museum, Harvard University. (G) *Cloudina riemkeae*, a biomineralized tubular animal fossil, from the Dengying Formation, South China. Photograph courtesy of Hong Hua.

began to diversify about 575 Ma ago. The disappearance is unlikely to be a preservational artefact: slightly younger phosphorites in the Dengying Formation contain no Doushantuo-type acritarchs, and basal Cambrian acritarchs are entirely different from those in the Doushantuo Formation. It is possible that the extinction of Doushantuo-type acritarchs is related to yet another small-scale glaciation in the Late Neoproterozoic.

Perhaps the most exciting fossils from the Doushantuo Formation are the submillimetric globular microfossils preserved in phosphorites at Weng'an in Guizhou Province. These globular microfossils have been interpreted as animal blastulae at successive cleavage stages, containing 2^n blastomeres within an envelope of roughly constant size (Figure 5A–D). The geometry of these microfossils is consistent with blastula cell division. It has not been determined, however, to which animal clade(s) these fossil embryos belong. Their chimeric combination of features that individually occur in crown-group sponges, cnidarians, and bilaterians suggests that these embryos may belong to stem groups at the animal, eumetazoan, or bilaterian levels. There are other microfossils in the Weng'an phosphorites that have been interpreted as adult sponges, adult cnidarians (Figure 5E), and putative bilaterian gastrulas. However, none of these are more than a few millimetres in size. Perhaps the earliest animals, at least bilaterian animals, were indeed microscopic.

Macroscopic bilaterian animals probably first evolved in the latest Neoproterozoic Ediacaran time. Among the best-known Ediacaran fossils, those in the Newfoundland assemblage (*ca.* 575–565 Ma) are probably the oldest, those in the White Sea and South Australia assemblages (*ca.* 555 Ma) are younger, and those in the Namibia assemblage (*ca.* 549–543 Ma) are the youngest. There do not appear to be any unambiguous macrobilaterians in the Newfoundland assemblage. The White Sea and South Australia assemblages, on the other hand, include body and trace fossils of macrobilaterians (e.g. *Kimberella quadrata*), as well as classic Ediacaran fossils such as *Dickinsonia* (Figure 5F) whose phylogenetic interpretations are still controversial. The Namibia assemblage and its equivalents (for

example the Dengying Formation in South China) include biomineralizing animals such as *Cloudina* (Figure 5G), *Namapoikia*, and *Namacalathus*. These Ediacaran assemblages, therefore, record a succession of evolutionary events leading to the rise of macrobilaterians and animal biomineralization. Most Ediacaran fossils, however, disappeared near the Precambrian–Cambrian boundary, closing the last chapter in Precambrian eukaryote evolution.

Acknowledgments

I would like to thank the National Science Foundation for funding my research on Proterozoic palaeontology. Nicholas J Butterfield, Hans J Hofmann, Hong Hua, Andrew H Knoll, Susannah M Porter, Bruce Runnegar, Leiming Yin, and Xunlai Yuan kindly provided the photographs used in this article.

See Also

Fossil Plants: Calcareous Algae. **Microfossils:** Acritarchs. **Origin of Life**. **Precambrian:** Prokaryote Fossils; Vendian and Ediacaran. **Sedimentary Rocks:** Chert. **Trace Fossils**.

Further Reading

Baldauf SL, Roger AJ, Wenk-Siefert I, and Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972–977.

Brocks JJ, Logan GA, Buick R, and Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* 285: 1033–1036.

Butterfield NJ, Knoll AH, and Swett K (1994) Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Fossils and Strata* 34: 1–84.

Fedonkin MA and Waggoner BM (1997) The late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism. *Nature* 388: 868–871.

Hofmann HJ (1985) The mid-Proterozoic Little Dal macrobiota, Mackenzie Mountains, north-west Canada. *Palaeontology* 28: 331–354.

Knoll AH (1996) Archean and Proterozoic paleontology. In: Jansonius J and McGregor DC (eds.) *Palynology: Principles and Applications*, pp. 51–80. Salt Lake City: American Association of Stratigraphic Palynologists Foundation, Publishers Press.

Narbonne GM (1998) The Ediacara biota: a terminal Neoproterozoic experiment in the evolution of life. *GSA Today* 8: 1–6.

Porter SM and Knoll AH (2000) Testate amoebae in the Neoproterozoic era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. *Paleobiology* 26: 360–385.

Schopf JW and Klein C (1992) *The Proterozoic Biosphere: A Multidisciplinary Study*. Cambridge: Cambridge University Press.

Vidal G and Moczydlowska-Vidal M (1997) Biodiversity, speciation, and extinction trends of Proterozoic and Cambrian phytoplankton. *Paleobiology* 23: 230–246.

Waggoner B (2003) The Ediacaran biotas in space and time. *Integrative and Comparative Biology* 43: 104–113.

Woese CR, Kandler O, and Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the USA* 87: 4576–4579.

Xiao S, Zhang Y, and Knoll AH (1998) Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* 391: 553–558.

Yuan X, Xiao S, Yin L, *et al.* (2002) *Doushantuo Fossils: Life on the Eve of Animal Radiation*. Hefei: China University of Science and Technology Press.

Zhang Y, Yin L, Xiao S, and Knoll AH (1998) *Permineralized Fossils from the Terminal Proterozoic Doushantuo Formation, South China*. Memoir 50. The Paleontological Society, Lawrence, Kansas, USA.

Prokaryote Fossils

M D Brasier, University of Oxford, Oxford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Conditions for the development and survival of life on the young planet Earth are thought to have been extremely harsh. Reconstruction of those conditions requires multidisciplinary and interdisciplinary research (combining astronomy, planetary science, microbial and molecular biology, genetics, geochemistry, petrology, palaeobiology, and geology). These disciplines are united in the emerging field of astrobiology (exobiology), which covers the phenomenon of life within the solar system and beyond. Astrobiology takes a particular interest in modern and ancient prokaryotic ecosystems on Earth and in the prebiotic–biotic transition and its definition.

Meteoritic and lunar research shows that accretion of planet Earth took place around 4550 Ma ago (*see* **Earth Structure and Origins**). This accretion was soon followed by melting of the planetary surface owing to the energy released through gravitational collapse. A cataclysmic collision with a Mars-sized planet is

thought to have produced the Moon within the first 100 Ma or so. The Moon played a major role in setting the equable conditions for life on Earth through its influence in stabilizing our planetary motion about the rotational axis and by providing stable tidal and seasonal oscillations. Earth's surface had cooled and the oceans had probably condensed by about 4000 Ma, as shown by *ca*. 4200 Ma old (hydrous melt) zircons in the oldest rocks of Canada and by putative sediments in the 3800 Ma Akilia and Isua Groups of Greenland. Even so, the study of craters on the Moon and Mars shows that their surfaces were bombarded by large asteroids between 4550 Ma and 3800 Ma. Similar bombardment is likely to have affected the surface of the Earth, although any evidence has been destroyed by the mobility of the planetary surface. These bombardments would have had a devastating effect on any early biosphere (see **Figure 1**; the so-called 'impact frustration of the origin of life'). For example, the impact of an asteroid more than 350 km across would release enough energy to boil off the oceans, while that of

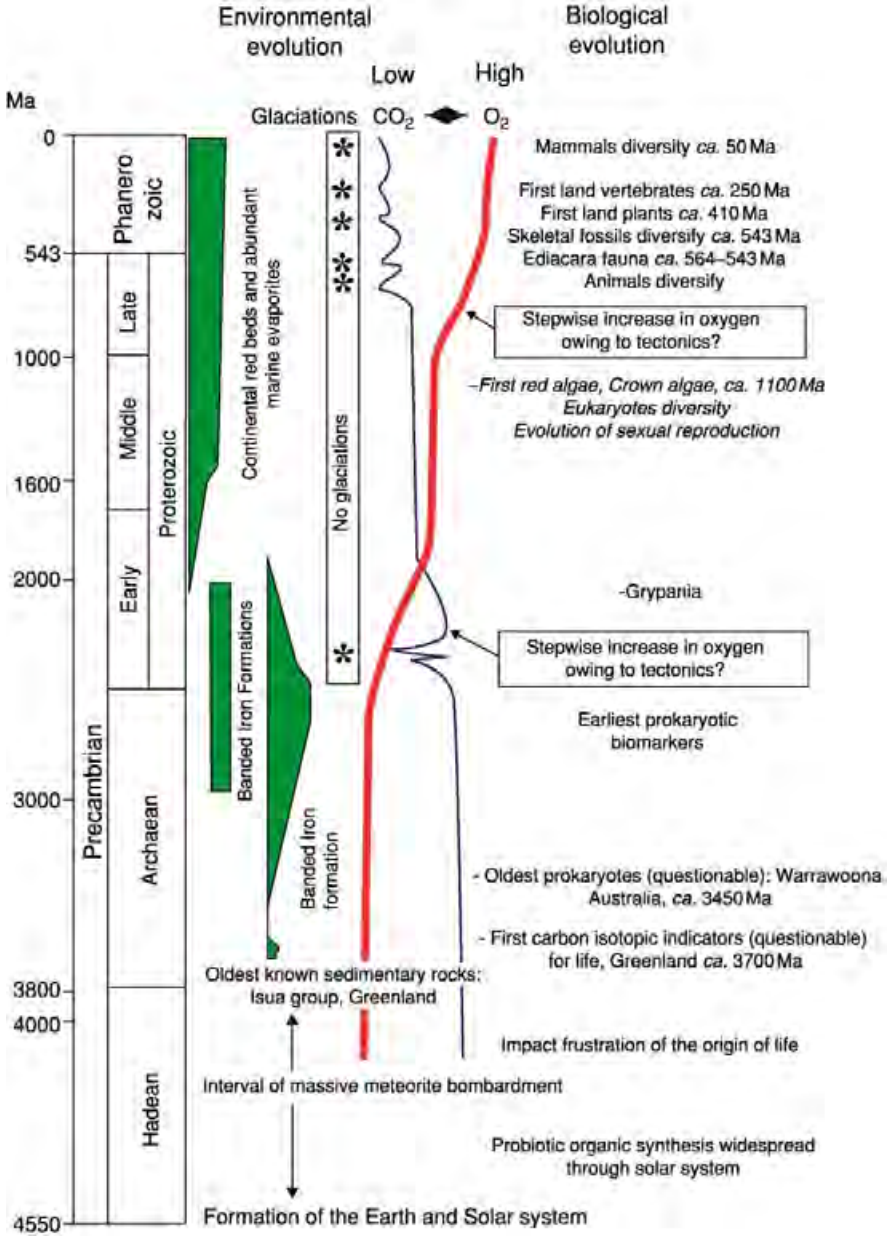


Figure 1 The main succession of events inferred for the evolution of the biosphere alongside geological evidence for changing levels of atmospheric oxygen and carbon dioxide during the Precambrian. The evidence for the first isotopic indicators (more than 3700 Ma) and the oldest prokaryotes (*ca.* 3450 Ma) is currently in dispute.

one about 150 km across could destroy the photic zone. Any life-forms synthesized prior to 3800 Ma could therefore have been repeatedly destroyed by episodic catastrophes. The only forms that could have survived such a holocaust would have been heat-tolerant hyperthermophile bacteria, which today are found living around volcanic and hydrothermal vents or deep in the Earth's crust. Such life-forms arguably originated from amino acid to RNA and DNA synthesis around hot alkaline hydrothermal vents. This is suggested not only by the antiquity of hyperthermophiles, as shown by molecular phylogeny (see below), but also by the ready availability of carbon sources (methane, carbon dioxide, carbon monoxide, aliphatics) and electron donors (metals, hydrogen sulphide, hydrogen) around such vents today, together with a plentiful supply of the phosphorus and transition metals that are needed to synthesize nucleic acids and enzymes. Against this, it has been argued that hydrothermal temperatures are too high to allow the necessary stability of complex organic molecules. But, on the early Earth, such hydrothermal vents were rich in metals (most of the world's metal resources come from the Archaean to Palaeoproterozoic crust) and clay minerals. These metals and clay minerals could arguably have allowed the stabilization of nucleic acids, even at relatively high temperatures. No other site on Earth could supply these essential building blocks in so many forms and so readily.

Asteroid bombardment of the Earth is thought to have declined progressively between about 3800 Ma and 3000 Ma. The hot, thick, and largely oceanic crust of this age can be envisaged as mantled by oceanic waters of moderate depth with a mean temperature of approximately 60°C. This means that aquatic habitats on Earth prior to about 3000 Ma may have been essentially hydrothermal in nature. It was not until about 3000 Ma that the wide shallow continental shelf seas, deep-sea troughs, and land areas began to develop. Crustal sources of free energy may therefore have begun to dwindle after about 3000 Ma, and the emerging biosphere may have been forced to switch to using solar energy, with one group perhaps converting from a heat-seeking catalyst to a light-seeking (photosynthetic) one.

Prokaryotes are considered to be the most primitive organisms on Earth and are assumed to have emerged early in the comparatively hot world of the Archaean. The cells of still-living forms are extremely small, generally less than 1 μm in diameter, and they lack a nucleus (hence the name prokaryote, which is derived from the Greek *pro* = before and *karyos* = nucleus). They may be single or colonial, the latter enclosed within a mucilaginous sheath called a capsule. Some living bacterial cells bear a whip-like thread (flagellum), and a few contain chlorophyll pigments for photosynthesis. Prokaryotes are important today in the formation of microbial sediments (*see* **Biosediments and Biofilms**), such as cyanobacterial mats and stromatolites, iron and manganese ores, carbonate concretions, and sulphide and sulphate minerals. They also yield important information about the early evolution of the cell and the histories of methanogenesis, photosynthesis, and biogeochemical cycles.

Molecular and Biochemical Evidence

Contrasts between the ribosomal RNA sequences of diverse bacteria, protists, fungi, plants, and animals indicate that life on Earth can be divided into three primary domains: the Archaea (or 'Archaebacteria'), which includes the methanogenic and sulphur bacteria; the Bacteria (or 'Eubacteria'), which includes cyanobacteria and other forms with photosynthetic pigments; and the Eucarya (or 'Eukaryota'), which includes all the protists, fungi, plants, and animals (Figure 2). It is notable that all the most deeply rooted branches of the tree of life are occupied by modern hyperthermophilic bacteria, which live at temperatures of 80–110°C and are seldom able to grow below 60°C. This has been taken to suggest that the

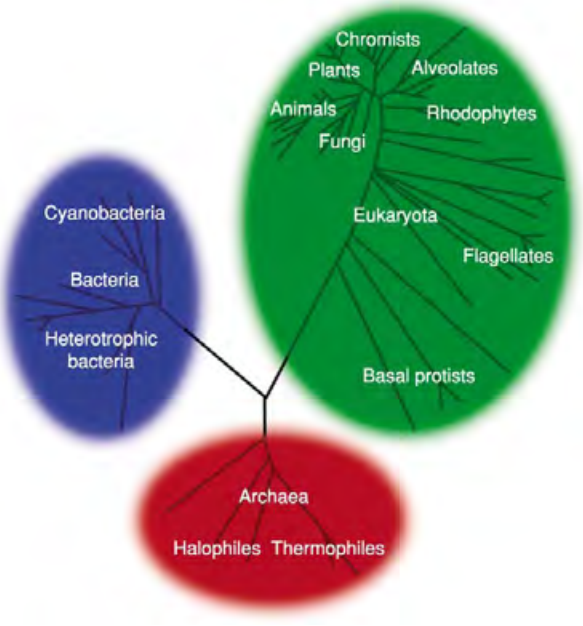


Figure 2 The three branches of the tree of life, in which all the deep-seated branches are hyperthermophilic bacteria (shown in bold). Time increases along the branches, but not necessarily in a linear fashion or at the same rate in each branch. Longer branches relate to faster evolution.

last common ancestor of all living organisms was a hyperthermophile, adapted to hot hydrothermal springs or to life deep in the Earth's crust. One explanation for this is that life originated in such conditions. Another explanation is that hyperthermophiles were pre-adapted to survive the catastrophic period of meteorite bombardment between 4500 Ma and 3800 Ma.

Biochemical evidence can be taken to suggest the following evolutionary sequence of autotrophic prokaryotes, each of which used carbon dioxide as their sole source of carbon.

1. Anaerobic chemolithotrophic prokaryotes, which mainly use hydrogen produced from inorganic reactions between rock and water as their main electron source.
2. Anaerobic anoxygenic prokaryotes such as green and purple sulphur bacteria, which use photosynthesis to reduce carbon dioxide to form organic matter, with hydrogen sulphide as the electron source, in the absence of oxygen.
3. Oxygenic cyanobacteria, which use photosynthesis to reduce carbon dioxide to form organic matter, with water as the electron source, releasing oxygen. These must have had an enormous impact on Earth surface processes and the biosphere, and considerable interest has been focused by astrobiologists upon their first appearance in the rock record. At the time of NASA's Viking missions to Mars in 1976, it was such photosynthetic autotrophy that scientists were hoping to find.

Heterotrophic prokaryotes do not synthesize organic matter. Like us, they use preformed organic matter as their source of carbon and can use a range of oxidants to break it down and release the energy bonds. Methanogenic Archaea are among the most primitive heterotrophs alive today, living in highly reducing sediments (such as peat bogs) and releasing methane gas. Sulphate-reducing bacteria use seawater sulphate (SO_4) ions in the absence of oxygen, but require a highly oxidized form of sulphur (SO_4), which may not have been widely available in the early ocean. Aerobic heterotrophic bacteria use freely available atmospheric oxygen and are unlikely to have radiated before the so-called Great Oxygenation Event, 2450–2200 Ma ago, when various indicators of the weakly reducing planetary surface (banded iron formations (*see* **Sedimentary Rocks:** Banded Iron Formations), detrital pyrite, uraninite, and siderite) begin to disappear from the rock record and red beds start to appear. This oxygenation event may relate in part to increasing rates of carbon burial in expanding cratonic basins and subduction zones and in part to the irreversible loss of hydrogen to space from the upper atmosphere. While oxygen producers and consumers could have existed prior to 2450 Ma, they were probably restricted to rather local oases of oxygenation. This inferred evolutionary sequence of methanogenic to sulphate-reducing to aerobic heterotrophic prokaryotes is likely, on the basis of evidence from living bacteria, to have been accompanied by an increasing yield of energy from the same amount of carbonaceous 'food'. Significantly, this evolutionary succession closely resembles the modern distribution of prokaryotic populations within marine muds, with methanogenic Archaea lying deep within the sediment pile, aerobic heterotrophs and photoautotrophs in the upper layers of the sediment, and sulphate reducers in between.

Evidence for the Earliest Biosphere

Biogeochemistry

The fossil evidence for life on Earth gets increasingly scarce as the age of rock units increases. This is because older rocks have suffered more exposure to erosion and have experienced a greater degree of alteration by metamorphism. Hence, the oldest rocks on Earth (approximately 3800–3700 Ma), from Isua and Akilia in Greenland, have been too heavily metamorphosed to yield morphological evidence of life. Possible traces of life must therefore be explored using biogeochemical techniques. Stable isotopes of carbon from Isua and Akilia, for example, are somewhat lighter than usually expected from an inert world (*ca.* −18‰ δ^{13} CPDB cf. Pee Dee Belemnite standard). This has been taken to imply that life was able to self-organize and survive the period of catastrophic meteorite impacts before about 3800 Ma (**Figure 1**). Such a view is now controversial for a variety of reasons. The sedimentary origin of the carbonaceous grains is questionable: fractionation of carbon compounds could also have resulted from abiogenic processes or even from carbonaceous meteoritic debris. The carbon may also be younger than claimed. While light carbon isotopes (*ca.* −40‰ to −25‰ δ^{13} PDB) are commonly encountered in rocks younger than about 3500 Ma, some of these hydrocarbon compounds may also have an abiogenic origin, and precise discriminators, such as C–H ratios, H isotopes, and aliphaticity, are needed to discount this possibility. Even so, most scientists assume that this 25–40‰ difference in carbon isotopes of carbonates and organic matter seen after about 3500 Ma provides key evidence that biological metabolic pathways (i.e. autotrophic fractionation of carbon isotopes) were in place by this time.

A trend towards highly negative carbon isotope ratios about 2800–2200 Ma ago has been explained as the result of a bloom in heterotrophic methanogenic and methanotrophic Archaea at that time, prior to the Great Oxidation Event. Sulphur isotopes can also be used to trace the history of prokaryotic metabolism. In this case, ^{32}S is preferentially taken up by sulphate-reducing bacteria, leaving the water column enriched in the heavier isotope, ^{34}S. Recent studies of the ratio between ^{32}S and ^{34}S in sedimentary pyrite and in supersaturated barite, gypsum, and anhydrite have been used to argue that bacterial sulphate reduction was not in place before about 2450 Ma. Prior to this, sulphur isotope fractionation seems to have been of the abiogenic mass-independent kind, more like that seen in the Martian atmosphere. This lack of bacterial sulphate reduction before about 2450 Ma may be explained by a lack of sufficient atmospheric oxygen to form the sulphate ions and/or by surface water temperatures that were too high in the Archaean to produce a measurable fractionation.

Organic geochemical 'biomarkers' can also be sniffed out in well-preserved rocks, at least as far back as 2700 Ma. Biomarkers called 2-methylhopanes have been reported from rocks of this age in Western Australia and have been taken to indicate the presence of cyanobacteria at that time. This line of uniformitarian reasoning assumes, of course, that such biomarkers were not present in any other prokaryotic group, living or extinct, which is rather difficult to refute. Unfortunately, most rocks older than 2700 Ma appear to be too 'cooked' to preserve complex organic molecules.

Stromatolites

Laminated domical structures known as 'stromatolites' (Figure 3A) have been described from carbonate rocks as old as 3450 Ma in the Pilbara Supergroup of Western Australia (Figure 3C) and in the coeval Swaziland Supergroup of South Africa. Although an origin from the accretion of prokaryotic and even cyanobacterial mats has often been inferred for these early Archaean examples, they do not contain microfossils and they show some features that render their biogenicity rather questionable: an association with, and continuation down into, hydrothermal dyke systems; close association with epigenetic crystal fans (e.g. after aragonite, barite, gypsum) and directly precipitated carbonates; intergradation with ripple-like forms having rotational symmetry (i.e. they look the same upside down); lack of multiple

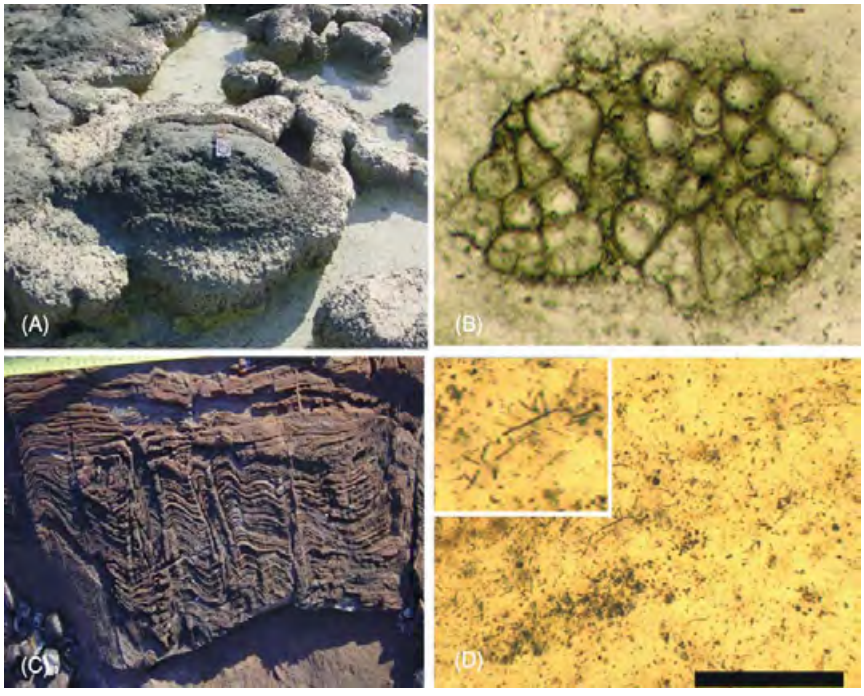


Figure 3 Stromatolites and Precambrian prokaryotes. (A) Living cyanobacterial–diatom stromatolites from the hypersaline Shark Bay in Western Australia; compass is *ca.* 6.5 cm wide. (B) The coccoid cyanobacterium *Eoentophysalis* preserved in cherts of the *ca.* 800 Ma old Boorthanna Chert of Australia; scale bar below = 100 μm. (C) A 30 cm wide exposure of 3400 Ma old stromatolites from the Strelley Pool Chert of Western Australia; it is unclear whether these are of microbial origin or are an abiogenic crystalline precipitate. (D) The 1900 Ma old filaments of cyanobacterial or iron bacterial origin preserved in the Gunflint Chert of Mink Mountain, Ontario, Canada; scale bar (main picture) = 40 μm; scale bar (inset) = 100 μm.

fractal dimensions; continuous laminae of barely varying thickness (typical of crystal growth); lack of wrinkle fabrics typical of biofilms; and lack of fenestrae or other signs of biological processing. Such simple stromatolites seem likely to have formed from largely physical processes such as direct chemical precipitation from seawater. It is only after about 3000 Ma that fractally complex stromatolites are found in carbonate platform settings, for which the presence of biofilms seems more plausible, if not always demonstrable. These tend to have wrinkle mat fabrics, discontinuous laminae, and laminar fenestrae (e.g. those from the 2700 Ma old Belingwe Group of Zimbabwe and the Fortescue Group of Western Australia). Microfossils associated with Archaean stromatolites remain questionable, however, until about 2530 Ma and are not really diverse until the 1900 Ma old Gunflint Chert of Canada (**Figure 3D**). Even here, the biogenicity of the associated stromatolites is open to question because they resemble abiogenic sinters.

Silicified Microbiotas

Precambrian oceans appear to have been supersaturated with silica (SiO_2) because it was not being removed from the water column by groups that evolved later, such as the diatoms, radiolaria, and sponges. In environments where early diagenetic silica was able to engulf prokaryotic populations, such as in peritidal hypersaline bacterial-mat settings, prokaryotic sheaths and even cell walls were sometimes (but still very rarely) well preserved in three dimensions. This prevalence of hypersaline settings may have produced an unfortunate bias towards prokaryotic assemblages in the early fossil record. In other words, the inferred dominance of prokaryotic microfossils within Precambrian cherts may be due to their restricted hypersaline setting rather than to evolutionary factors.

Silicified microbiotas are usually studied by means of standard (30 μm) to thick (300 μm) petrographic thin sections, at magnifications of up to about $\times$400. This thin-section technique is paramount because of the way in which it provides for contextual analysis, including three-dimensional morphology, mineralogy, rock fabric, and rock history. Other techniques, such as maceration (digestion in strong acids), etching of rock chips, scanning electron microscopy, and atomic force microscopy, are also used, but these do not provide the requisite information on context and are prone to the inclusion of structures that are later contaminants or 'artefacts' of the preparation process.

A classic example of a silicified Precambrian microbiota is the 1900 Ma old Gunflint Chert, which preserves about 12 taxa of prokaryotes, including forms that superficially resemble coccoid and filamentous cyanobacteria (**Figure 3D** and insert) but may be more closely allied to extant iron bacteria. The putative cyanobacteria *Eoentophysalis* (**Figure 3B**) and *Archaeoellipsoides* are thought to be present in the approximately 2100 Ma old cherts of West Africa and the 2000 Ma old Belcher Group cherts of Canada, respectively. The latter is claimed to preserve the specialist heterocyst cells used by cyanobacteria to help fix nitrogen in an otherwise oxidizing atmosphere. As one moves back into the Archaean, microfossils become both extremely rare and highly questionable, despite the great abundance of carbonaceous cherts and tufa-like carbonates. This may be explained partly by the inference that Archaean cherts were laid down in largely hydrothermal conditions that were often acidic and reducing, and partly by the scarcity of large and resistant cellular materials at that time. Bundles of silicified filaments and tiny calcified holes from stromatolites in the 2530 Ma old carbonates of the Transvaal Supergroup of South Africa may be the casts of coccoid and filamentous cyanobacteria but little of diagnostic significance is preserved. Intriguingly, such encrusted cyanobacterial filaments are rarely seen before about 1000 Ma, and endolithic microborings are not reported prior to 1500 Ma. A single microfossil-like structure from the 2700 Ma old Fortescue Group of Western Australia has been compared with a cyanobacterial filament but its biogenicity and context awaits full documentation.

Skeins of pyritic filaments found within carbonaceous cherts from 3200 Ma old black smokers of the Sulphur Springs Formation in Western Australia may be the remains of anaerobic hyperthermophile bacteria, though the indigenous and biogenic nature of these intriguing structures has yet to be demonstrated beyond question. The oldest cherts containing a supposed diverse prokaryotic microflora (**Figure 4A**) come from the 3450 Ma old Apex Cherts, which are intimately associated with ultrabasic and basaltic lava flows of the Apex Basalt in the Warrawoona Group of Western Australia. At least eleven different kinds of filamentous microfossil have been described from these rocks, some of which have been compared with cyanobacteria (**Figure 4A**). This has been taken to suggest that photosynthesis had begun to release oxygen into the atmosphere by 3450 Ma and that a substantial amount of evolution had taken place by this time. A critical re-examination of the context and fabric of these cherts suggests, however, that all these

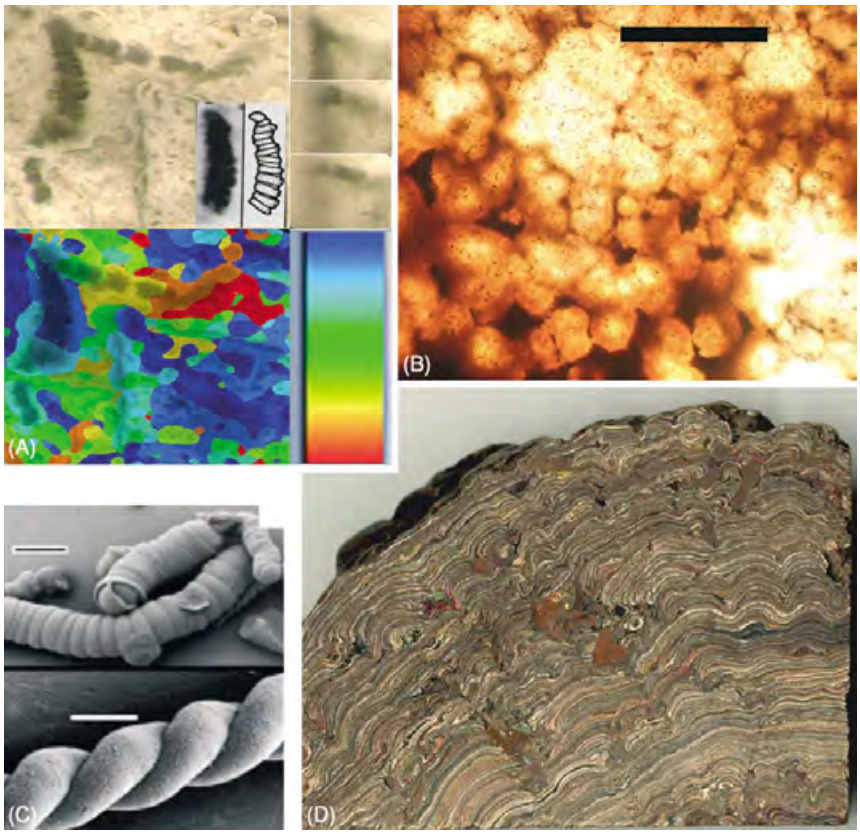


Figure 4 Pseudofossils. (A) Montage of the pseudofossil *Archaeoscillatoriopsis disciformis* holotype from the 3450 Ma old Apex Chert of Western Australia, showing (in black and white) its original interpretation and (in colour) its much more complex morphology when montaged (with topographic map) and its proximity to crystal growths. (B) Spherulites formed by the recrystallization of glassy hydrothermal silica with iron oxide impurities; recrystallization causes the formation of a spectrum of forms that range from abiogenic to microfossil-like pseudofossils; scale bar in B = 40 μm. (C) Complex carbonaceous artefacts formed in the laboratory; scale bars = 40 μm. (D) Complex stromatolitic growths can easily be formed artificially by over-spraying a surface with layers of paint; width of view = 12 cm.

structures occur deep within a cross-cutting hydrothermal vein. They comprise a continuous spectrum of arcuate to complexly branched carbonaceous reaction rims, formed around the edges of spherulitic silica botryoids, that appear to have formed during the recrystallization of hydrothermal silica glass to chalcedony and around rhombic crystal margins. Doubt is also cast on other 'microfossils' of about this age from the Warrawoona Group of Australia and from the Barberton mountains of South Africa. Their origin as mineralic pseudofossils or contaminants remains an open question.

Criteria for Biogenicity

The search for the signals left in rocks by very ancient prokaryotes requires a checklist of criteria to help establish biogenicity, not least because prokaryotic cells are so easily mimicked by physical artefacts and so readily introduced by contamination. A previous formula for the recognition of *bona fide* Archean microfossils has been that the putative microfossil structures must be unquestionably biogenic remnants that are indigenous to and syngenetic with the primary deposition of a sedimentary rock, which is itself of known stratigraphic and geographical source and of established Archaean age (i.e. greater than 2500 Ma old).

Until recently, the above criteria have been used to contrast the questionable 'microfossils' in Martian meteorite ALH 84001 with the 3450 Ma old supposedly *bona fide* 'microfossils' from the Apex Chert. It is now necessary to acknowledge, however, that complex structures of organic composition, like those in the Apex Chert, can be assembled very easily (e.g. Figure 4C), including abiogenic artefacts of

isotopically light carbon formed through late-stage remobilization of hydrocarbons with hydrothermal silica. Complexity *per se* is not a valid criterion for the recognition of life. Biological populations may be distinguished from abiogenic ones by the *lack* of a continuous spectrum of morphological variation, limited by the DNA-controlled genetic constraints. This narrow range can be contrasted with complex abiogenic artefacts, which will tend towards a 'symmetry-breaking cascade' of forms, ranging from a few biological look-alikes with rotational symmetry (**Figure 4C**) to highly information-rich forms (**Figures 4A** and **4B**) with poor symmetry.

Research into the earliest biosphere is now served by a marvellous range of new tools, including molecular (RNA and DNA) analysis of living-prokaryote phylogeny, biogeochemical analysis of ancient Earth rocks, and digital-image analysis of putative microfossils. These techniques should not be allowed to prosper, however, at the expense of more traditional techniques such as mapping and thin-section petrography. The latter are also vital for providing a context for early-life studies at a range of scales, from satellite images of greenstone belts, through field mapping, to microfabric mapping of thin sections. Future research will also require a much better understanding of the nature of complexity itself and of the multitudinous ways in which seemingly complex structures can self-organize. (**Figure 4** shows a range of microfossil-like and stromatolite-like objects of inferred non-biological origin.) This better understanding of complexity is needed, not only to understand the ways in which pseudofossils and pseudo-stromatolites can form, but also to comprehend the conditions in which life-forms and their building blocks were able to assemble and develop on the early Earth, and perhaps beyond.

See Also

Biosediments and Biofilms. **Earth Structure and Origins**. **Origin of Life**. **Precambrian:** Eukaryote Fossils. **Pseudofossils**. **Sedimentary Rocks:** Banded Iron Formations. **Solar System:** Mars.

Further Reading

Brasier MD, Green OR, Jephcoat AP, *et al.* (2002) Questioning the evidence for Earth's oldest fossils. *Nature* 416: 76–81.

Farquhar J, Bao HM, and Thiemens M (2000) Atmospheric influence of Earth's earliest sulfur cycle. *Science* 289: 756–768.

Furnes HR, Banerjee NR, Muehlenbachs K, *et al.* (2004) Early life recorded in Archean pillow lavas. *Science* 304: 578–581.

Garcia Ruez JM, Hyde ST, Carnerup AM, *et al.* (2003) Self-assembled silica-carbonate structures and detection of ancient microfossils. *Science* 302: 1194–1197.

Hofmann HJ, Grey K, Hickman AH, and Thorpe RI (1999) Origin of 3.45 Ga coniform stromatolites in Warrawoona Group, Western Australia. *Bulletin of the Geological Society of America* 111: 1256–1262.

Kazmierczak J and Altermann W (2002) Neoarchean biomineralization by benthic cyanobacteria. *Science* 298: 2351.

Kerr R (2004) New biomarker proposed for earliest life on earth. *Science* 304: 503.

Knoll AH (2003) *Life on a Young Planet*. Princeton, NJ: Princeton University Press.

Mojsis SJ, Arrenhius G, McKeegan KD, *et al.* (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* 384: 55–59.

Rasmussen B (2000) Filamentous microfossils in a 3,250-million-year-old volcanogenic massive sulphide. *Nature* 405: 676–679.

Rothman DH and Grotzinger JP (1996) An abiotic model for stromatolite morphogenesis. *Nature* 383: 423–425.

Schopf JW (1999) *The Cradle of Life*. New York: Princeton University Press.

Simpson S (2003) Questioning the oldest signs of life. *Scientific American* April 2003: 70–77.

Westall FM, de Wit MJ, Dann J, *et al.* (2001) Early Archean fossil bacteria and biofilms in hydrothermally influenced sediments from the Barberton greenstone belt, South Africa. *Precambrian Research* 106: 93–116.

Woese CR, Kandler O, and Wheelis ML (1990) Towards a natural system of organisms: proposals for the domains of Archaea, Bacteria and Eucarya. *Proceedings of the National Academy of Sciences USA* 87: 4576–4579.

Vendian and Ediacaran

M A S McMenamin, Mount Holyoke College, South Hadley, MA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The geological, palaeobiological, and earth systems (i.e., global geochemical) events associated with the Vendian period set the stage for all subsequent Earth history. Unprecedented change took place in the biosphere at this time, and the Vendian witnessed the appearance of the first Ediacarans, the first shelly fossils, and the earliest known animals. The Vendian saw the continuation of breakup of the earliest known giant supercontinent (Rodinia) and, at the end of the period, continuing into the Cambrian, the amalgamation of the supercontinent Gondwana. The Vendian also witnessed the termination of the worst series of glaciations known. The climate change was accompanied by dramatic fluctuations in the records of carbon and strontium isotopes. Linkages between supercontinent breakup and extreme climate change, and between biotic diversification and ecological change (e.g., as driven by metazoan disturbance of the marine substrate), are being actively explored by researchers following diverse avenues of investigation.

The Vendian period and system (called by some authors Lipalian or Ediacaran) spans the interval from 600–543 Ma. It was the final period of Proterozoic time (Neoproterozoic). Some accounts place the beginning of the Vendian as far back as 670–620 Ma, but this would push it into what should be the preceding period (the Sinian, 680–600 Ma). In any case, the Vendian represents the last interval of Precambrian time, and its end marks the most important division in the geological time-scale, the Proterozoic–Cambrian boundary. The importance of this boundary has led to escalating interest amongst researchers, and great advances have recently been made in understanding the Vendian. Many questions remain, however, and amongst them are some of the most contentious issues in contemporary earth science. The primary questions of interest are threefold. First, what caused global climate to warm after the worst glaciation on record? Second, what are the Ediacarans? Third, what events triggered the so-called Cambrian Explosion (*see* **Palaeozoic:** Cambrian) and the appearance of familiar animal types? A host of unsolved secondary questions follow. These include: What was the makeup of supercontinent Rodinia? What was the timing sequence and geometry of its breakup? What were the controls on eustatic sea-level change during the Vendian? Was there a mass extinction at the Proterozoic–Cambrian boundary? If so, what was its cause? What was happening to global geochemistry during this time? The implications of these questions are far reaching, and the discussion that follows is divided into three sections: Geological Events, Palaeobiological Events, and Earth System Events.

Geological Events

It can be said that the prelude to the Vendian world began at 1000 Ma with the amalgamation of the supercontinent Rodinia. The continental collisions that led to the formation of the supercontinent are generally called the Grenville Orogenic Event (*see* **Grenvillian Orogeny**). Named for 1000 Ma rocks that record an ancient episode of mountain building and continental collision in North America, Grenvillian rocks have now been recognized in places as distant as northwest India and Antarctica.

Composed of all the large continents of the Precambrian planet, Rodinia began to split apart beginning at about 750 Ma. Some geologists argue for a failed attempt at breakup at 850 Ma. Rodinia is thus a remarkably long-lived supercontinent, lasting some 250 million years. The rifting event was a drawn out affair that seems to have proceeded in two stages. Stage one began with the opening of the Pacific Ocean, as parts of what now constitute eastern Gondwana (Australia, Antarctica) split apart from what is now the west coast of North America. In stage two, fragments connected to the north-eastern and eastern parts of North America split off and began to collide with the continental blocks already set free by continental drift. The net result of all this continental motion, which was not complete until about the time of the Cambrian boundary, was the reorganization of the continental plate geometry of the planet with the sundering of Rodinia and the amalgamation of Gondwana. Some geologists believe that a short-lived global supercontinent (Pannotia) existed between the times of Rodinia and Gondwana, but evidence for this supercontinent is tenuous at best.

Block faulting and volcanicity associated with tectonic sundering characterize many Vendian stratigraphical sections. Deposition of the Vendian Tindir Group of east-central Alaska was influenced by block faulting in North American basement rocks. The Vendian La Ciénega Formation of northern

Mexico is punctuated by layers of porphyritic basalt erupted from fissures presumably related to tectonic extension.

Palaeomagnetic evidence suggests that the rates of continental motion near the Cambrian boundary were amongst the fastest ever measured. At one point near the boundary, the continents appear to have been moving synchronously, prompting what has been called inertial interchange theory, in which the Earth's crust is thought to have detached from the mantle for a relatively short time.

As the fragments of Rodinia went their separate ways, they set into motion a sequence of events that ultimately resulted in the drowning by seawater of all or most continental margins. As Rodinia split apart, the Rodinia's counterpart superocean, Mirovia, began to vanish beneath the subduction zones along the leading edges of the dispersing continental fragments. This led, overall, to an exchange of old, cold Mirovian seafloor crust for the relatively hot, less dense, and hence more buoyant ocean crust of the newly developing rift basins on the trailing edges of the continents opposite the subduction zones. Seawater was consequently displaced on a massive scale by buoyant, new mid-ocean ridges, and this led to the great marine transgression that began the Palaeozoic era.

The transgression began very slowly, as the early stage of breakup involved only a single (albeit long) continental margin, namely the split between Australia–Antarctica and North America. This gradual rise was then suddenly punctuated by a rapid fall in eustatic sea-level due to the onset of the first of several (Sturtian and Marinoan) Late Proterozoic glaciations. These were the worst glaciations known, leading many to believe that the Earth must have passed through what has been called 'White Earth' or 'Snowball Earth' conditions. The timing of these glaciations is somewhat uncertain, although two groups, an older (Sturtian, including the Rapitan, Chuos, and Stuartian glaciations) and a younger (Marinoan, including the Ice Brook, Ghaub, and Elatina glaciations), appear to each consist of roughly synchronous glacial events.

Compelling evidence exists during both glacial episodes for glaciers on land at sea-level at the equator, and some geologists suggest that the ice cap extended into the equatorial ocean as well. Regardless of the true extent of the ice cap, the glaciation was tremendously severe. The climatic aberration is made all the more curious by the character of the sedimentary layers deposited directly above the glacial deposits. Also, sedimentary (banded) iron formations reappeared during these ice ages after having virtually disappeared from the record for 1.5 Ga. These and other geochemical anomalies are interpreted to suggest almost incredible changes in the oxygen and carbon dioxide cycles of the planet – anoxia under the marine ice cap, leading to banded iron deposition; build-up of sufficient carbon dioxide in a life-depleted Earth to eventually trigger greenhouse conditions and melt the glaciers; and supergreenhouse conditions and supersaturation, leading to massive deposition of cap carbonates under unusual carbonate depositional conditions.

The cap carbonates consist primarily of abiogenically precipitated calcite and dolomite. Carbonate rocks of this nature are usually associated with deposition under very warm climatic conditions. Therefore, the juxtaposition of these sediments directly above deposits of the worst known glaciation is unusual in the extreme. The last Proterozoic cap carbonates were deposited at approximately 600 Ma and define the base of the Vendian system.

With the final melting of the ice, the transgression resumed its flooding of continental shelf areas. The transgression continued essentially unabated until well into the Cambrian. In many stratigraphical sections throughout the world, the base of the Cambrian is marked by a basal unconformity, although this is by no means universal, particularly in regions with a more or less complete Vendian section as well. The Vendian is thus defined in many places by a cap carbonate at its base (a pronounced aid to lithostratigraphical correlation when not associated with an unconformity), but an unconformity at its top (a decided hindrance to correlation efforts). Nevertheless, a combination of radiometric dates, lithostratigraphy, biostratigraphy, palaeomagnetic stratigraphy, and carbon and strontium isotope stratigraphy has rendered preliminary correlations possible between Vendian sections on different continents. Much work remains to be done to refine these correlations.

The Vendian is not characterized by major extraterrestrial impact events, but this may be an artefact of a less well-understood record of mass extinction during the period. The Acraman impact site (570 Ma) of South Australia, associated with a 160 km diameter crater, shock metamorphism, and shattercone development, is comparable in size to the 214 Ma, Manicouagan impact structure of Canada.

Palaeobiological Events

Although eukaryotic organisms are thought to have existed for more than 1000 million years before the beginning of the Vendian, they are rather inconspicuous until the Vendian begins. Thus, the Vendian marks the beginning of the Phanerozoic, the age of

visible life. The reason for the rapid expansion of eukaryotic life at this time is unclear, but some researchers have linked it to the climatic amelioration following the Proterozoic glaciations.

Biogenic stromatolites, known from the oldest rocks bearing microbial fossils, dramatically changed texture during the Vendian. The concentric lamination that characterized more ancient stromatolites gave way to a clotted, thrombolitic texture. This transition from stromatolite to thrombolite has been attributed to the burrowing and lamination-disturbing activities of early animals.

Biological inferences regarding the origin of metazoa, based on molecular clock data, indicate that metazoa appeared by at least 1000 Ma, but these inferences are beset by controversy. The date of origin of the animal kingdom is controversial as well, and the rock record before the Vendian does not provide many unambiguous clues to metazoan origins.

One thing is clear, however; by the Middle to Late Vendian, animals were present, as indicated by their trace fossils (*see* **Trace Fossils**). These trace or ichnofossils show evidence of peristaltic burrowing and displacement of sediment by burrowing activity, indicating that, by the end of the Vendian, animals with hydrostatic skeletons (coelomic spaces) were well established in the marine biosphere.

A difficulty with the study of Vendian animals is the fact that, apart from some phosphatized embryos that cannot be identified confidently to phylum, and some fossil sponges with preserved spicules, actual body fossils of these animals are rare. A comb-like structure from Russia, called *Redkinia,* may represent the flexible, filter-feeding mouthpart of a Vendian animal. However, apart from this, there are very few Vendian animal body fossils, particularly body fossils of the burrowing tracemakers.

The Cloudinidae consists of Vendian shelly fossils. The cone-in-cone tubular shells of genera such as *Cloudina, Sinotubulites* (**Figure 1**), and *Wyattia* are presumed to have been formed by worm-like animals. Cloudinids are occasionally associated with a bizarre, weakly calcified, stalked, goblet-shaped organism called *Namacalathus*. The biological affinities of cloudinids and namacalathids are not well understood, but they do seem quite unlike the more familiar skeletonized animals of the Early Cambrian. An algal affinity for the two has been suggested and cannot be ruled out. Large pores in the calyx of *Namacalathus* were probably filled with soft tissue in life; the pores might have served to admit light into the organism to sustain photosymbionts.

Ediacarans are the most puzzling part of the Vendian biota. These bizarre creatures grew to enormous sizes by Vendian standards, with some of the frondose forms reaching 2 m or more in length. The Ediacaran body is non-skeletal and seems to have been formed of a tough integument, in some cases partitioned into modules of similar shape. Unlike most other soft-bodied creatures, Ediacarans were capable of being preserved in sandstones. Ediacarans appear to be multicellular, but even this inference has been subjected to dispute. The concept of metacellularity (i.e., a body composed of uni- or polycellular partitions called metacells) has been applied to Ediacarans with some success.

Many palaeontologists have assigned these forms to conventional animal phyla, such as the Cnidaria,

Figure 1 *Sinotubulites cienegensis.* An early shelly fossil, this cloudinid is from the La Ciénega Formation of Sonora, Mexico. Holotype specimen. Length of largest tube, 12 mm.

Echinodermata, or Annelida, but the arguments supporting these assignments have not convinced everyone. Sceptics point out that not a single uniquely animalian trait has been identified on any of the thousands of Ediacaran fossils collected so far. A consensus for at least some of the fossils may be emerging, however, with recent new evidence showing trilobitoid arthropod features in *Spriggina,* and with the association of *Kimberella* with grazing traces.

In addition to the difficulties with systematic placement, there is the thorny problem of Ediacaran preservation. None of the Ediacaran fossils have skeletons, with the possible exception of the cloudinids and the namacalathids (whose relationship to the other Ediacarans is not known). Their cuticle also appears to have been relatively soft and flexible. Many of the Ediacaran fossils are preserved in relatively coarse sandy sediment, not ordinarily considered to be a good substrate for the preservation of non-skeletal organic remains. Thus, the preservation of Ediacarans must be explained either by unusual properties of the cuticle itself, or by anactualistic processes on the Vendian seafloor that could account for the preservation of soft tissue in sandstone. The latter possibility has evoked what is called the 'death mask' hypothesis, namely the idea that microbial mats in some way hardened the surface of dead Ediacarans, allowing them to be preserved as fossils. Unfortunately, the death mask theory is invoked by its supporters to explain all aspects of Ediacaran taphonomy – for example, the death mask mats supposedly formed beneath layers of storm sand and both above and below living Ediacarans, three highly uncertain propositions.

A key bedding plane surface from Newfoundland provides a new perspective on the death mask controversy (**Figure 2**). This bedding surface was smothered by a volcanic ash fall, entombing specimens of an unnamed spindle-shaped Ediacaran and a stalked form called *Charniodiscus*. The *Charniodiscus* was superimposed over the spindle form, and yet the morphology of the spindle creature shows clearly through the cuticle of the *Charniodiscus*. Thus, the death mask hypothesis is falsified, as there was no space for a post-mortem microbial mat to form between the *Charniodiscus* and the spindle. If the death mask hypothesis were correct, only the overlying *Charniodiscus* morphology should be preserved, but this is clearly not the case in this example.

Consequently, we are left with the idea that Ediacarans bore an unusually resilient cuticle. There is much going for this concept, as a link between Ediacarans and a problematic group known as conulariids (first suggested in 1987) has been dramatically confirmed by the discovery of the probable conulariid *Vendoconularia triradiata* in Vendian strata of the Ust'-Pinega Formation of the Onega River region, Russia. Conulariids appear to be a group of Ediacarans that survived well into the Palaeozoic. As with the Ediacarans themselves, most attempts at classifying

Figure 2 Cast of specimen of *Charniodiscus* superimposed over a spindle-shaped form from the Mistaken Point assemblage in Newfoundland. Fibreglass cast (Pratt Museum, Amherst College). Width of view, approximately 11 cm.

conulariids have tried to assign them to the Cnidaria, and have met with failure. Part of the problem seems to be that conulariids display, as do other Ediacarans, both triradial and fourfold radial symmetry. This translates in metacellular terms to the unipolar iteration of six, eight, twelve, or sixteen founding metacells. The phosphatic nature of the (also somewhat flexible) conulariid cuticle may provide an important clue to the nature of the Ediacaran cuticle.

The most difficult Ediacaran problem is that of body geometry. Attempts to classify Ediacarans as animals have been hindered by their unusual constitution, which can show combinations of triradial symmetry and glide symmetry (as in the genus *Pteridinium*; Figure 3); these have proven to be impossible to fit into a conventional metazoan body plan model. Some researchers have tried to break the impasse by arguing that Ediacarans are colonial communities of individual animals. This explanation is not adequate to the task either, as each Ediacaran seems to be a well-integrated individual holobiont rather than a loose collection of individuals in a colony. Furthermore, there is no evidence for the loss of morphological features in specialized individuals, as is the case in the highly integrated colonies of modern hydrozoans.

The concept of metacellularity may help to solve this morphological problem. Some Ediacarans are apparently modular partition creatures, and the individual units or partitions, referred to as metacells, are either uni- or polycellular pods that may be flattened, stretched, inflated, or repeated (to form an iterated chain of metacells) as required. Metacellular creatures of this style are known amongst the modern biota (e.g., characean pondweeds, such as *Nitella,* can have individual cells measuring up to 15 cm in length), but no other type of organism has explored the potential of metacellularity as thoroughly as have the Ediacarans. The concept of metacellularity seems to apply best to the Ediacarans of the Mistaken Point biota in Newfoundland, the morphologies of which (spindle forms, branch forms, pectinate forms) are strange even by Ediacaran standards.

The metacellularity concept does not solve the problem of whether Ediacarans are animals; for example, hexactinellid (glass) sponges can undergo a coenocytic or syncytial (plasmodial) stage that is quite uncharacteristic for animals. Metacellularity is unknown in living animals and, for this reason, a number of palaeontologists prefer to place at least some of the Ediacarans into their own Kingdom, the Vendobionta. These Ediacarans could conceivably bear a closer relationship to several protoctist groups (such as the xenophyophores) than they do to animals. A common ancestry with the protoctist group that gave rise to the animals is still plausible, though, and may even be likely. Even Ediacarans that would seem to have a more animalian character, such

Figure 3 *Pteridinium simplex*. Cast of specimen from the Nama Group, Namibia. The fossil is viewed from its underside; the undersides of two vanes are visible, and the basal part of the third vane is visible between them as a stitch-like chain of tube ends. Scale bar in centimetres.

Table 1 Ediacaran body forms, classified by metacellular grade

Original number of metacells	*Founding metacell(s) only, no iteration*	*Unipolar iteration of metacells*	*Bipolar iteration of metacells*
One	*Cyclomedusa, Evandavia, Parvancorina (?)*	*Charniodiscus, Charnia*	Spindle form of Newfoundland
Two	*Gehlingia*	*Dickinsonia, Phyllozoon*	*Windermeria, Ernietta*
Three	*Tribrachidium, Anfesta, Albumares*	*Swartpuntia*	*Pteridinium simplex*
Four	*Conomedusites*	*Spriggina (?), Marywadea (?)*	*Pteridinium carolinaensis*
Five	*Arkarua*	*Rangea (?)*	Unknown

Figure 4 *Evandavia aureola*. This hiemalorid Ediacaran occurs in the Clemente Formation, Sonora, Mexico. It is perhaps the oldest known (approximately 600 Ma) specimen of a complex life form. Scale bar in centimetres.

as *Yorgia waggoneri* from the White Sea region in Russia, show strange asymmetries in the 'head' region that challenge the animal interpretation.

Many Ediacarans may be placed in the metacellular grouping (Table 1). These forms manifest both a smooth central metacell surface and a surface formed of radial (and often bifurcating) tubes. *Evandavia aureola* is a non-iterated discoidal form from Sonora, Mexico, with a single metacell and radial tubes in its outer ring (Figure 4). Figures 5 and 6 show *Parvancorina minchami* from Australia, displaying a single trifold metacell with numerous branches coming from each branch of the initial metacell. *Parvancorina* serves as the subunit counterpart to *Tribrachidium* and *Gehlingia,* with the 'thumb structures' in the bilaterally symmetric *Gehlingia* and in the triradiate *Tribrachidium* homologous to the main medial branch in *Parvancorina* (Figure 7). Similar metacellular rearrangements may be observed in another Ediacaran clade, which includes the Newfoundland spindle-shaped form (Figure 8) and a 2 m long Newfoundland frond ('*Charnia*' *wardii*) shaped like an extremely elongate primary feather of a bird's wing. An interesting variation on the discoid body theme is a Chinese fossil from the Xingmincun Formation, southern Liaoning Province, consisting of a single, elongate metacell helically coiled into a flat disc. These specimens reach up to 4 cm in diameter.

Earth System Events

Eukaryotic phytoplankton are known from strata approaching 2 Ga in age, but, by the Vendian, these organisms began to resemble the modern phytoplankton at least in terms of diversity of tests, skeletons, and sheaths. Siliceous chrysophyte algae are known from Vendian strata of the Tindir Group, Alaska, and genera such as *Chilodictyon* and *Characodictyon* share a superficially diatom-like aspect.

Figure 5 *Parvancorina minchami*. A cast of a specimen from the Pound Supergroup of South Australia. Greatest dimension of specimen, 10 mm.

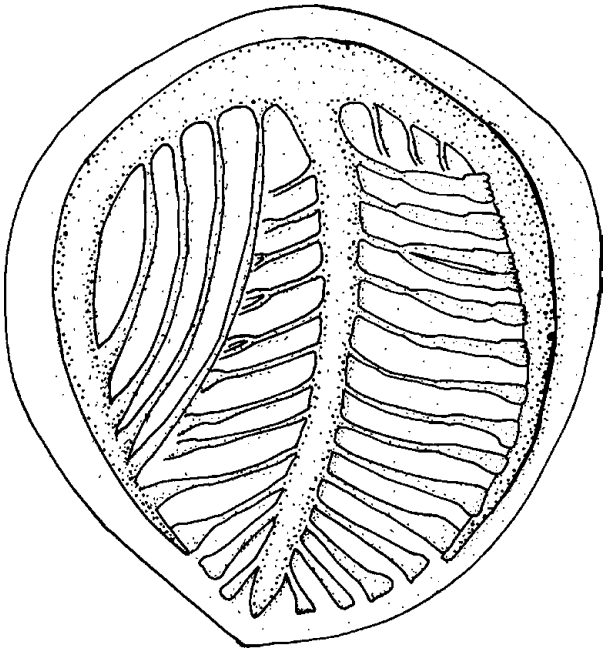


Figure 6 *Parvancorina minchami*. This Ediacaran poses difficulties for the arthropodan interpretation of this genus. Greatest dimension of specimen, approximately 10 mm.

In contrast, however, the Vendian seafloor had a distinctly primitive character. Microbial mats still blanketed the seafloor, as they had been doing since the Archaean, and stromatolites remained abundant even as burrowing began to alter their internal texture to the thrombolitic state. A wide variety of unusual sedimentary structures (with names such as *Arumberia*, *Kinneya*, and 'elephant-skin' texture) are known from fine clastic rocks of the Vendian. Lozenge-shaped structures (**Figure 9**) are often found on Vendian bedding surfaces; whether or not such structures might have use as a basis for interbasinal correlations is unknown. These primary sedimentary structures are puzzling, leading sedimentologists to invoke anactualist sedimentary processes to explain them. It seems quite reasonable to do so, as microbial mat carpeting of the seafloor would certainly influence the nature of marine sedimentation. Preston Cloud once called stromatolites 'organo-sedimentary structures', and the shelf, slope, and rise sediments of the Proterozoic may be thought of as a gigantic, connected, organo-sedimentary structure. Judging from the diversity of bedding plane texture types, never before or since have microbes had such a direct and intimate association with the basic processes of clastic sedimentation. Ediacarans evidently adhered to the surface of these mats or (in the case of forms with holdfasts and possibly in the case of *Pteridinium*) lived beneath them.

Metazoan burrowers apparently began their excavations beneath this mat surface, and evidence suggests that the earliest ichnofossil makers were submat burrowers. A good example is the Vendian trace fossil, *Vermiforma*, from the Carolina Slate Belt. This enigmatic and relatively large trace fossil, associated with other types of submat burrowers, consists of ten specimens that all follow the same rather tortuous track. As the traces are separated from one another by some distance, all underneath the mat, it is hard to

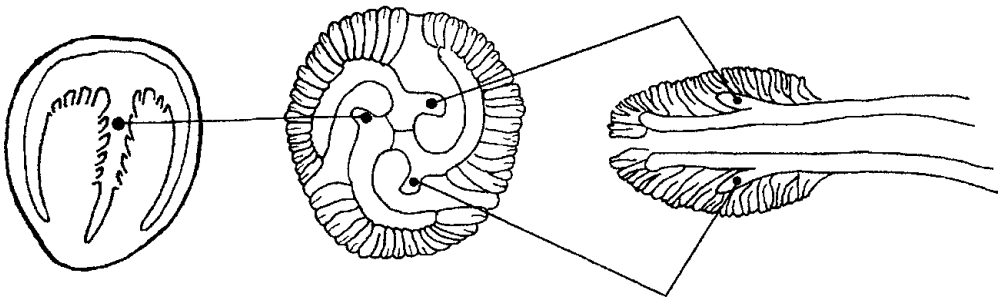


Figure 7 Ediacaran homology. Homologous structures ('thumb structures') linking the Ediacaran genera *Parvancorina* (left: length, 1 cm), *Gehlingia* (right: length, 8 cm), and *Tribrachidium* (centre: diameter, 1 cm). In spite of the vast differences in body symmetry type, the three genera are seen to be closely related.

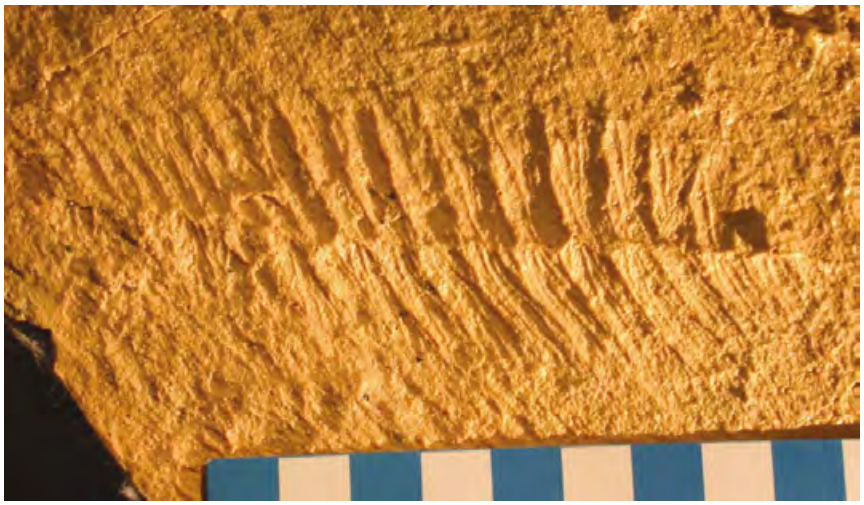


Figure 8 Spindle-shaped form from the Mistaken Point assemblage in Newfoundland. Fibreglass cast (Pratt Museum, Amherst College). Scale bar in centimetres.

imagine how and why the track paths were coordinated in shape. Arguments that *Vermiforma* is a pseudofossil are called into question by the presence of other types of trace fossils on the same bedding plane surface.

Throughout its duration the Vendian is marked by steadily increasing levels of bioturbation. The top of the Vendian system is currently defined by the appearance of the three-dimensional trace fossil *Trichophycus pedum*. The maker of this trace was able to excavate vertically as well as horizontally, and its activities (along with those of other vertical burrowers, such as the makers of *Skolithos*) tended to homogenize seafloor sediments. It would also presumably have punctured and shredded any microbial mats in the immediate vicinity.

It is thought that, towards the end of the period, the seafloor mat seal began to break down due to the intensity of metazoan burrowing activity. Such uncapping of the seafloor had dramatic consequences, both for the carbon budget of the planet (e.g., buried carbon was put immediately back into circulation) and for marine nutrient levels (e.g., sediment grains in suspension make wonderful substrates for nutritious bacterial growth). At about the same time, there was a tremendous flux of mineral nutrients to the oceans resulting from the fact that the Vendian saw a rare tectonic coincidence in Earth's history: the simultaneous occurrence of both divergent tectonics (final breakup of Rodinia) and convergent tectonics (formation of Gondwana, in an event known as the Pan-African Orogeny (*see* **Africa:** Pan-African Orogeny)), on a massive scale. As suggested by marine strontium isotopes of the Vendian, huge amounts of siliciclastic debris were shed from rift valley margins into the Vendian ocean. This would have been added to the volcaniclastic sediments derived from weathering of rift-associated basalts. The enhanced pool of igneous rock debris contributed mightily to an oceanic fertilization event that has been implicated in the emergence of skeletonized animals and the Cambrian explosion.

Figure 9 Lozenge-shaped structures from the Clemente Formation, Sonora, Mexico, found in association with *Evandavia aureola* and other Ediacarans. Such structures are known from Ediacaran-bearing strata on other continents. Scale bar in centimetres.

It appears that a massive palaeoecological reorganization of the marine biosphere took place at the end of the Vendian. With the exception of some very Late Vendian evidence for predatory activity in Sonora, Mexico, no evidence for large predators has been recognized in Vendian strata. In contrast, the Early Cambrian exhibits evidence for a great intensity of macropredation, involving forms such as the 2 m long predatory animal *Anomalocaris*. The Vendian biosphere has thus been referred to as a uniquely peaceful Garden of Ediacara, where large creatures, many with flattened bodies, partook of sunlight (via photosymbiosis), hydrogen sulphide (via chemosymbiosis), or osmotrophy (direct absorption of dissolved nutrients in seawater) on a mat-covered seafloor surface free of large predators. However, it was too good to last, and animals in the role of burrowers from below and predators from above triggered a rapid end to the Garden and the end of the Vendian.

The question of whether or not there was a mass extinction at the end of the Vendian is unresolved at present. A number of Ediacarans appear to have survived the Cambrian boundary, occurring with Cambrian fossils in Australia, Ireland, and elsewhere. Studies based on borehole data from the oil-producing strata in Oman have recently shown that cloudinids vanish from the record without a trace, and that this disappearance is not associated with any discernible lithological change. Thus, the disappearance of cloudinids from Oman cannot be attributed to environmental or facies change. Their loss has thus been interpreted as a major extinction event. Whether or not this localized loss of cloudinids represents a mass extinction, or an extinction event at all, is not known. Cloudinid-like fossils have been reported from the Cambrian Torneträsk Formation in northern Sweden.

Secular variations in secular isotopes in the Vendian are rather difficult to interpret, but, after a slight decrease from $-3‰$ $\delta^{13}C$ to $-5‰$ $\delta^{13}C$ after the Marinoan glaciation (Canadian and Namibian sections), the $\delta^{13}C$ values during the Vendian appear to undergo a fairly steady rise to a value of over $+3‰$ right before the Cambrian boundary (Dvortsy section, Aldan River, Siberia). The Vendian–Cambrian boundary itself is marked by a sudden $+3‰$ to $-1.4‰$ drop in $\delta^{13}C$ over a very short stratigraphical interval, as measured at the Dvortsy section.

This excursion has been linked to a variety of factors, including some sort of global environmental perturbation that also triggered the extinction of the cloudinids and many of the Ediacarans. Another way to look at it, though, is that the boundary excursion represents that moment in geological time when marine burrowing intensity crossed a threshold. At this critical point, the microbial mat seal on the seafloor may have been breached, resulting in previously immobile sediments (and their associated organic matter) becoming mobilized and injecting huge amounts of biogenic carbon into the water column. Owing to biogenic isotopic fractionation, this detrital organic matter was significantly depleted

in the heavy isotope of carbon and, as it went into circulation in marine water, caused the precipitous drop or boundary excursion in the $\delta^{13}C$ value.

It therefore seems reasonable to interpret the Vendian carbon isotope curve as a record of gradually increasing biotic productivity, with sequestering of much of the organic matter within and below the seafloor microbial mats, followed by a relatively sudden release of part of this organic matter deposited at the end of the Vendian as a result of increased burrowing intensity. Other factors, such as destabilization of gas hydrates in seafloor sediments, may also have been involved in these isotopic excursions.

A number of other phenomena can perhaps be traced to what has been called the Cambrian Substrate Revolution. The Vendian and Cambrian both saw an increase in the proportion of calcified filamentous microbes (such as *Girvanella*), which were perhaps less palatable to mat grazers than filamentous cyanobacteria and unprotected algae. With all the new sediment and organic matter in suspension, filter feeding probably became more possible throughout the water column, leading to the evolution of the first tiered filter feeders in the Cambrian. The only Vendian organisms that were likely to engage in suspension feeding were the cloudinids, which lived close to the sediment–water interface.

Finally, assuming that we are interpreting the secular carbon isotopic curve correctly, it is entirely possible that oxygen levels increased in the Vendian due to the sequestration of organic matter. Whether or not increasing oxygen levels influenced metazoan evolution is not known, although it seems fair to say that early burrowing animals would not have required high levels of oxygen. The earliest animal habitat appears to have been the submicrobial mat environment, where oxygen levels would probably have been rather low considering the relative abundance beneath the mats of hydrogen sulphide and other reduced compounds.

Glossary

abiogenically A term applied to rocks formed by processes not directly influenced by living organisms.

anactualistic processes Processes that occurred at one time in the Earth's past, but which are no longer operational today.

cloudinid A late Vendian calcareous shelly fossil consisting of closely nested, thin-walled tubes or cones. Thought to represent one of the earliest examples of a shelly animal fossil. Includes the genera *Cloudina* and *Sinotubulites*.

coelomic spaces The compartments that house the rigid, fluid-filled body cavity present in many animals. The coelom serves as a hydrostatic skeleton.

conulariid Any member of an enigmatic group of Vendian/Cambrian to Triassic shelly organisms. They formed conical, often pyramidal, tapering cones with transverse ribbing, composed of calcium phosphate.

Ediacaran Any member of a group of marine, megascopic fossils with a metacellular growth pattern. Found primarily in strata deposited before the Cambrian period. Assigned to extinct Kingdom Vendobionta.

frondose forms Ediacarans with a leaf, palm, or frond body form.

Garden of Ediacara A palaeoecological theory that holds that the marine ecosystems of the Vendian were largely free of megascopic predators and thus allowed organisms such as Ediacarans to survive unmolested using photosymbiotic, chemoautotrophic, and osmotrophic life styles.

holobiont A single integrated organism, as opposed to a colonial organism.

hydrostatic skeleton A fluid-filled internal organ or support structure within an animal's body that can be kept rigid or made limp by control of internal water pressure.

metacell A single or isolated modular unit of a metacellular organism; usually consists of a single enlarged cell.

metacellularity Term applied to organisms that are either multicellular (such as animals and plants) or consist of clusters or metacells (such as Ediacarans and certain types of aquatic algae).

Mirovia The Precambrian superocean that surrounded Rodinia.

molecular clock Any gene or gene sequence used by biologists in an attempt to determine the evolutionary time of divergence from a common ancestor between two or more groups of organisms belonging to different species.

osmotrophy A feeding strategy utilizing osmosis or direct absorption of nutrients.

peristaltic burrowing A burrowing strategy in metazoans that consists of rhythmic muscular contractions along the length of the body.

Rodinia A supercontinent consisting of all or nearly all of the continents. Consolidated one billion years ago (in an event referred to in North America as the Grenville Orogeny), this supercontinent broke up into smaller continents by the process of plate tectonics and continental drift before the Cambrian.

Sinian The Precambrian geological period immediately preceding the Vendian period.

Snowball Earth An extreme phase of glaciation in which glaciers reached tropical latitudes.

thrombolitic texture A texture in sedimentary rocks characterized by disrupted bedding lamination and a clotted fabric. Stromatolites develop this texture when influenced by burrowing metazoa.

Vendian The latest Precambrian geological period, immediately preceding the Cambrian period. Synonymous with Lipalian or Ediacaran.

Vendobionta The extinct kingdom to which Ediacarans are assigned.

See Also

Africa: Pan-African Orogeny. **Australia:** Proterozoic. **Biosediments and Biofilms**. **Grenvillian Orogeny**. **Palaeozoic:** Cambrian. **Precambrian:** Overview; Eukaryote Fossils. **Trace Fossils**.

Further Reading

Bottjer DJ, Hagadorn JW, and Dornbos SQ (2000) The Cambrian substrate revolution. *GSA Today* 10: 1–7.

Crimes TP (1999) Review of *Garden of Ediacara*. *Palaeogeography, Palaeoclimatology, Palaeoecology* 150: 357–358.

Crimes TP and Fedonkin MA (1996) Biotic changes in platform communities across the Precambrian–Phanerozoic boundary. *Rivista Italiana di Paleontologia e Stratigrafia* 102: 317–332.

Crimes TP, Insole A, and Williams BPJ (1995) A rigid-bodied Ediacaran biota from Upper Cambrian strata in Co. Wexford, Eire. *Geological Journal* 30: 89–109.

Donovan SK and Lewis DN (2001) The Ediacaran biota. *Geology Today* 17: 115–120.

Glaessner MF (1984) *The Dawn of Animal Life*. Cambridge: Cambridge University Press.

Hoffman PF and Schrag DP (2002) The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* 14: 129–155.

Ivantsov AYu (1999) A new dickinsoniid from the Upper Vendian of the White Sea Winter Coast (Russia, Arkhangelsk Region). *Paleontological Journal* 33: 211–221.

Ivantsov AYu and Fedonkin MA (2002) Conulariid-like fossil from the Vendian of Russia: A metazoan clade across the Proterozoic/Palaeozoic boundary. *Palaeontology* 45: 1219–1229.

McMenamin MAS (1987) The fate of the Ediacaran fauna, the nature of conulariids, and the basal Paleozoic predator revolution. *Geological Society of America Abstracts with Program* 19: 29.

McMenamin MAS (1990) Vendian. In: Briggs DEG and Crowther PR (eds.) *Palaeobiology: A Synthesis*, pp. 179–181. Oxford: Blackwell Scientific Publications.

McMenamin MAS (1996) Ediacaran biota from Sonora, Mexico. *Proceedings of the National Academy of Sciences (USA)* 93: 4990–4993.

McMenamin MAS (1998) *The Garden of Ediacara: Discovering the First Complex Life*. New York: Columbia University Press.

McMenamin MAS (2000a) The antiquity of life: From life's origin to the end of the Vendian Period. In: Margulis L, Matthews C, and Haselton A (eds.) *Environmental Evolution*, 2nd edn. pp. 158–169. Cambridge, MA: MIT Press.

McMenamin MAS (2000b) Out of the shadows. *Notes and Records of the Royal Society of London* 54: 407–408.

McMenamin MAS (2001a) The Garden of Ediacara and the appearance of complex life. In: Guerzoni S, Harding S, Lenton T, and Ricci Lucchi F (eds.) *Proceedings of the International School of Earth and Planetary Sciences, Siena, Italy, 2001*, pp. 61–68. Siena, Italy: Consiglio Nazionale delle Richerche, University of Siena.

McMenamin MAS (2001b) *Paleontology Sonora: Vendian and Cambrian*. South Hadley, MA: Meanma Press.

McMenamin MAS (2003) Origin and early evolution of predators: The ecotone model and early evidence for macropredation. In: Kelley P, Kowalewski M, and Hansen T (eds.) *Predator–Prey Interactions in the Fossil Record, Topics in Geobiology Series 20*, pp. 379–400. New York: Plenum Press/Kluwer.

McMenamin MAS (2003) *Spriggina* is a trilobitoid ecdysozoan. *Geological Society of America Abstracts* 35: 105.

McMenamin MAS and McMenamin DLS (1990) *The Emergence of Animals: the Cambrian Breakthrough*. New York: Columbia University Press.

McMenamin MAS and Weaver PG (1992) Proterozoic–Cambrian paleobiogeography of the Carolina Terrane. *Southeastern Geology* 41: 119–128.

Seilacher A (1997) *Fossil Art*. Drumheller, Alta: Royal Tyrrell Museum of Paleontology.

Seilacher A and Pfluger F (1997) From biomats to benthic agriculture: A biohistoric revolution. In: Krumbein WE, Paterson DM, and Stal LJ (eds.) *Biostabilization of Sediments*, pp. 97–105. Oldenburg, Germany: Bibliotheks und Informationssystem der Carl von Ossietzky Universitat Oldenburg (BIS) Verlag.

PSEUDOFOSSILS

D M Martill, University of Portsmouth, Portsmouth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

This entry discusses those objects that appear to be fossilized organic remains – often called pseudofossils – but which in reality are natural structures that merely resemble real fossils. Such structures occur commonly in sedimentary rocks and can easily deceive inexperienced, or even experienced, geologists and palaeontologists.

There are many types of pseudofossil, and they often occur associated with real fossils. They may be formed by both physicochemical and biochemical processes, some of which may be the same processes that are responsible for generating real fossils (*see* **Diagenesis, Overview**). Pseudofossils are highly variable in form, and the resemblance of a pseudofossil to a real fossil can be quite remarkable. To some degree an object's categorization as a pseudofossil is a reflection of the experience of the discoverer, and, as such, there is no strict definition of what is or is not a pseudofossil. Those with little or no experience of genuine fossils can easily believe that rocks and pebbles with striped or mottled patterning or unusual shapes might be genuine fossils; indeed, some pseudofossils are real fossils. Some calcareous algae can assume growth forms that resemble miniature versions of macro plant remains. A famous example of this is the so-called Landscape Marble from the Upper Triassic of the Bristol District of England. Here, intermixed arborescent and laminated growths of calcareous algae resemble scenes of forested hills (Figure 1).

Objects that are commonly confused with genuine fossils include both three-dimensional objects that represent some form of *in situ* mineral growth – called concretions or nodules – and largely two-dimensional entities occurring on rock surfaces that assume patterns resembling organic remains. In the latter category are the banded mineral growths called Liesegang bands (Figure 2A), dendritic mineral growths (Figure 2B), and various types of fracture surface (Figures 3 and 4).

Liesegang Banding

Liesegang banding (Figure 2A) is the result of the diffusion of soluble chemicals through micropores in a rock in differing concentrations and/or oxidation states and their precipitation to produce colourful (often orange, red, and brown) and variable banded patterns. Trace fossils (*see* **Trace Fossils**), such as filled burrows, can be prone to such diffusion and may resemble segmented worms and arthropods. In fact, such a structure would be both a genuine trace fossil (the burrow) and a pseudofossil (the imagined worm).

Liesegang banding occurs very commonly in chert nodules in the Cretaceous chalk formations of Europe. It also occurs commonly in many rocks where iron and manganese in a reduced state are weathering from the fresh rock to form oxides, such as goethite ($Fe_2O_3.H_2O$) and pyrolusite (MnO_2). Liesegang rings may penetrate the rock or may be restricted to surfaces such as bedding planes, joints, and microfracture planes.

Dendrites

Dendrites are arborescent patterns produced by the fractal growth of minerals as they precipitate from solutions that migrate through fractures in rocks. Commonly they occur as black traces of manganese dioxide (the mineral pyrolusite) or orange-brown precipitates of hydrated iron oxides such as goethite, although other mineral types may also occur. Such structures are commonly mistaken for the remains of soft foliage such as ferns and liverworts. Particularly beautiful examples occur in the Jurassic Solnhofen Limestone (Figure 2B) of Bavaria, Germany, and are often sold as ornaments.

Fracture Surfaces

The fractured surfaces of rocks may exhibit patterns that can be confused with organic remains. Very fine-grained and glassy rocks may split with conchoidal fracture – a series of concentric ripples emanating from the initial point of fracture (Figure 3A). Indeed, the very name suggests that the fracture resembles a shell. It is common for flint to fracture in this way, and such conchoidal fractures in flint may resemble the bivalve *Inoceramus* (Figure 3B), which is a genuine fossil that often occurs in Cretaceous flints.

Some fracture surfaces may exhibit feather-like traces, while others may show regular banding resulting from the intersection of cleavages or bedding planes with the fracture surfaces.

Figure 1 A real fossil resembling another. A cut surface through the Late Triassic 'Landscape Marble': both the laminated and the arborescent patterns were formed by calcium carbonate precipitation influenced by benthic photosynthetic microbes (probably blue-green bacteria). Such structures are known as stromatolites. In former times, small village scenes were painted onto these slabs to create picturesque landscapes.

Figure 2 Superficial pseudofossils on bedding planes of the Late Jurassic Solnhofen Limestone Formation from Bavaria, Germany: (A) Liesegang banding as a series of iron oxide (goethite) diffusion fronts, and (B) dendritic patterns in pyrolusite and goethite. Photographs by Robert Loveridge.

In an unusual example (Figure 4), a piece of highly compacted fossil wood has numerous parallel and orthogonal evenly spaced cracks that have divided the wood into a series of small cubes. Such a blocky fracture pattern is common in fossil woods and can be seen in most bituminous coals. In the specimen shown in Figure 4, each small coaly cube generated a conchoidal fracture when the rock was split to reveal the fossil wood (the split propagated through the fossil wood rather than around it). As a consequence, each cube has a near-spherical surface with concentric ripples. This circular structure resembles the leaf-scar pattern seen on many stems of fossil plants, but the structures are an artefact of splitting.

Cone-In-Cone Structures

Cone-in-cone structures are the result of an unusual growth of minerals in which fibrous crystals assume a cone-like growth form (Figure 5A). They are frequently mistaken for fossils, particularly fossil corals, because the outer surface appears as a series of concentric rings or raised discs (Figure 5B). Cone-in-cone structures are usually composed of fibrous calcite crystals, which nucleate on the surfaces of limestone bands, shell beds, or even large fossils and grow orthogonal to that surface into fine-grained strata, such as clays and shales.

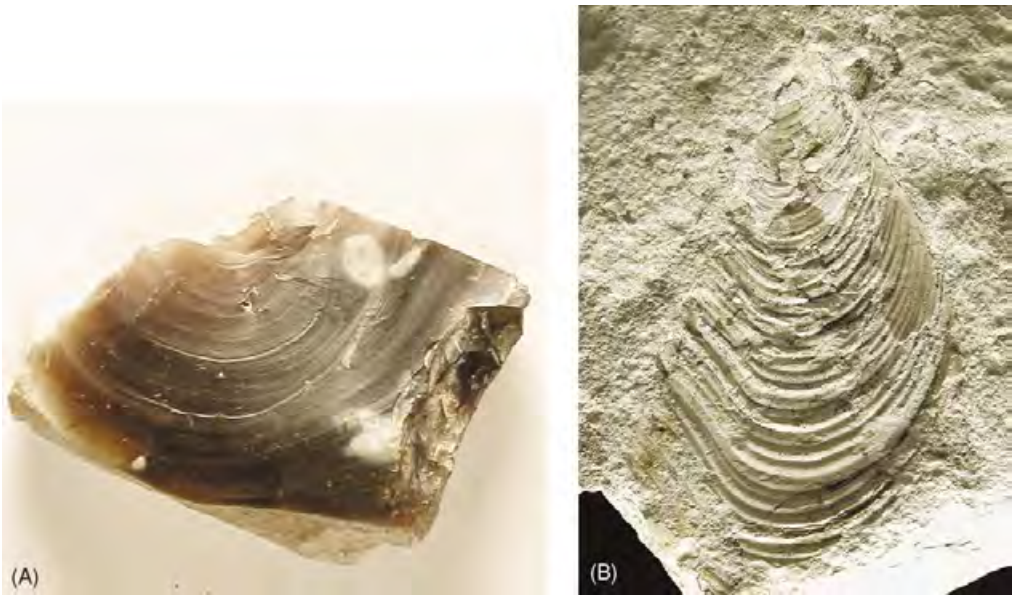


Figure 3 (A) Conchoidal fracture in flint. (B) The bivalve *Inoceramus*. *Inoceramus* occurs commonly in flint, and so conchoidal fractures may be easily confused with this bivalve. Photographs by Robert Loveridge.

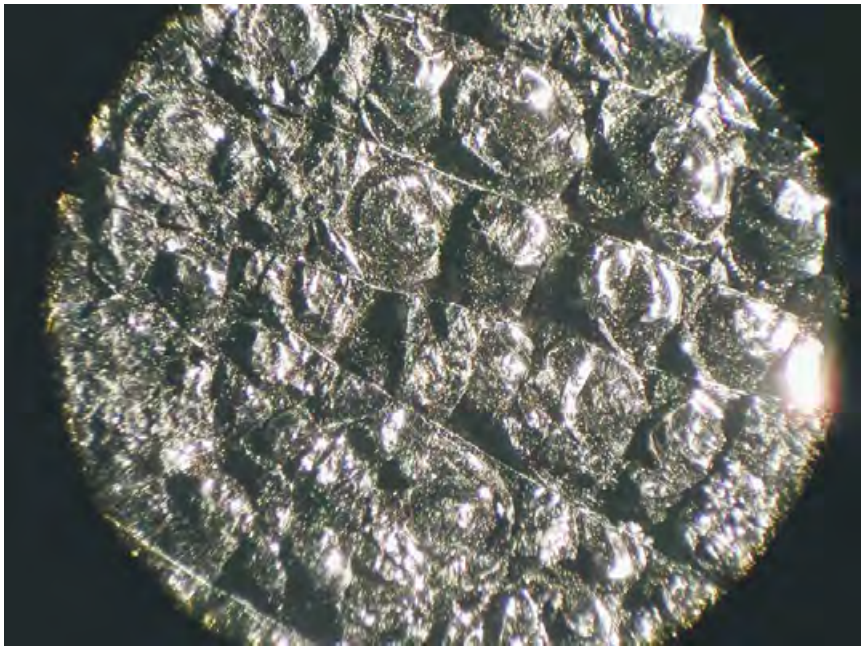


Figure 4 Conchoidal fractures in small cuboid segments of a fossil tree stem from the Lower Jurassic Posidonia Shale Formation of Dottenhausen, Germany. The conchoidal fractures resemble leaf scars but are a consequence of fracturing of the hard amorphous coal-like material. Photograph by Robert Loveridge.

Nodules and Concretions

Many of the common diagenetic minerals (those minerals that form in sediment and cement that sediment into rock) form nodular growths that can easily be mistaken for organic remains. Such mineral growths may assume bizarre shapes, influenced by the rate and direction of diffusion of solutes through the sediment and fluctuations in pH and eH. They may contain genuine fossils, and in some cases the external shape of the nodule assumes the shape of the enclosed fossil as it grows around it. The external appearance of nodules and concretions can be

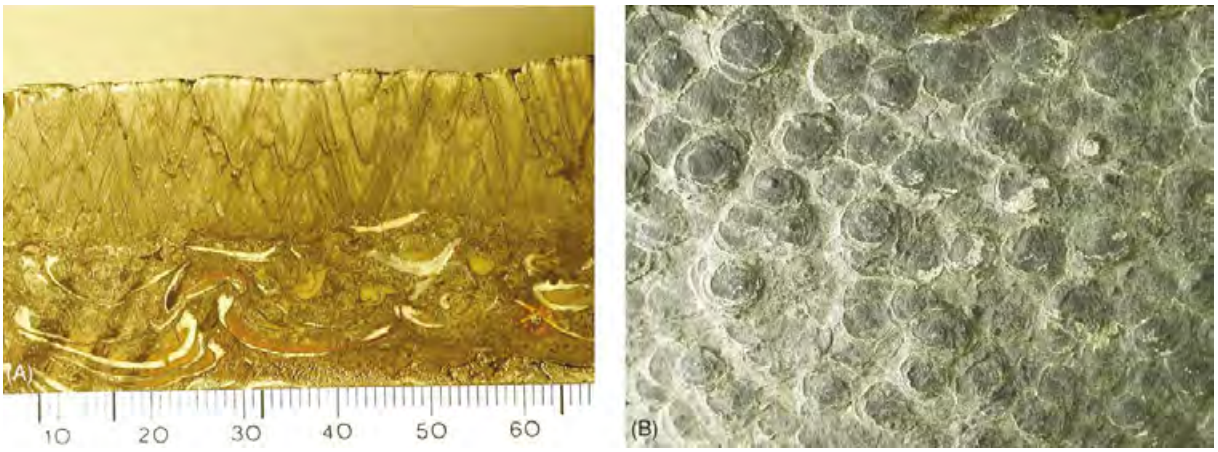


Figure 5 Cone-in-cone structures: (A) a cut section of a cone-in-cone that has grown on the surface of a shell bed, and (B) surface view – each cone appears as a circle or disc that can easily be mistaken for a fossil coral at first glance. Photographs by Robert Loveridge.

Figure 6 Nodules and concretions. (A) A spherical concretion of mudstone can easily be mistaken for a fossil egg. (B) The concretion in **Figure 6A** is split in half to reveal internal septarian cracking. Such cracks are often mistaken for fossils. Photographs by Robert Loveridge.

highly irregular, and they frequently assume shapes resembling isolated bones or even whole animals when there is no fossil enclosed. Highly spherical concretions (**Figure 6A**) are often mistaken for fossil eggs. Internally, concretions may have radial cracks or a honeycomb network of crystal-lined cracks (**Figure 6B**). Such concretions are called septarian concretions, and they are frequently mistaken for fossils.

Common mineral species that occur as concretions and nodules include the iron sulphide minerals pyrite and marcasite (nodules of the latter are often mistaken for meteorites), silica in the form of chert and flint, and calcite where it cements clay. Nodules sometimes grow in bands or strings, and, en-masse, they can resemble rows of vertebrae or ribs of large skeletons. A nodule of ironstone found within the rib cage of a dinosaur was mistakenly thought to be a fossil heart by a team of Canadian palaeontologists.

Very recently, experiments in which crystals of barite have been grown in gels under laboratory conditions have produced morphologies and sizes that are remarkably similar to structures found in Precambrian rocks that have previously been interpreted as fossil bacteria. The laboratory-grown crystals also resemble structures found in Martian-derived meteorites that were interpreted as evidence of extraterrestrial life. It looks as though these ancient and extraterrestrial microfossils might also prove to be pseudofossils.

See Also

Biosediments and Biofilms. **Diagenesis, Overview**. **Fossil Plants:** Calcareous Algae. **Sedimentary Rocks:** Chert. **Trace Fossils**.

Further Reading

Garcia-Ruiz JM, Hyde ST, Carnerup AM, *et al.* (2003) Self-assembled silica–carbonate structures and detection of ancient microfossils. *Science* 302: 1194–1197.

PYROCLASTICS

R J Brown, University of Bristol, Bristol, UK
E S Calder, Open University, Milton Keynes, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Pyroclasts are formed by the explosive fragmentation of magma (molten rock) during volcanic eruptions (*see* **Volcanoes**). They are carried away from the vent by buoyant eruption plumes, extensive umbrella clouds, or by destructive ground-hugging pyroclastic density currents. The largest explosive eruptions can produce pyroclastic deposits many 1000's km^3 in volume, which can be emplaced on a regional scale in a matter of hours. The study of the physical characteristics of pyroclasts and pyroclastic deposits can reveal much about the dynamic processes involved during explosive eruptions and in pyroclast dispersal and deposition. Recognizing, observing, and understanding pyroclastic deposits are vital first steps in assessing and mitigating volcanic hazard (*see* **Engineering Geology:** Natural and Anthropogenic Geohazards). This chapter summarizes the physical characteristics of the principal types of pyroclastic deposits and presents an introduction to their generation and emplacement mechanisms.

Generation of Pyroclastic Material

Explosive fragmentation of magma during volcanic eruptions can occur by two main mechanisms. The first involves the rapid exsolution of dissolved magmatic gases during rapid decompression events (magmatic eruptions), and the second results from the interaction of hot magma with external water sources (phreatomagmatic eruptions). Pyroclastic material can also be generated by rapid decompression and by autobrecciation processes during lava dome collapses.

Magma comprises three separate materials or phases: a viscous silicate melt (of varying composition), variable amounts of crystals (phenocrysts), and gas (volatiles) such as H_2O, CO_2, S, F, and Cl. There is a general positive correlation between the silica content of a magma and the degree of explosivity (Table 1). However, it is the quantity and behaviour of the gas phases that are critically important in determining the eruption style, because it is the rapid expansion of gas during decompression that drives explosive volcanic eruptions. Magma is stored at depth in magma chambers, under high temperatures and pressures. Magmatic eruptions are preceded by an increase in pressure and volume in the magma chamber. This is often attributed to the arrival of new magma into the chamber. The upper parts of many magma chambers are thought to contain a small volume fraction of gas bubbles (vesicles) due to supersaturation with volatiles, and seismic disturbance of these pre-existing bubbles can also lead to increases in magma chamber pressure (Figure 1). Crystallization, which enriches the melt in volatiles, can also act as a trigger. Once a critical point is reached, mechanical failure of the magma chamber roof occurs, allowing magma to rise, decompress, and exsolve gas in a runaway process (vesiculation) that can rapidly drive magma up the conduit at speeds of 200–400 m/s. Vesicle growth is controlled by the volatile content and by the physical properties of the magma (diffusivity rate, density, viscosity, and surface tension). The diffusivity rate is particularly important, and controls the rate at which gas bubbles escape from the magma: where escape is fast (in hot basic lavas), eruptions tend to be effusive or weakly explosive, but where escape is inhibited by high viscosities and low diffusivity rates (in intermediate and rhyolitic magmas), the exsolution of gas can explosively, and very violently, disrupt the magma. The expansion and coalescence of these bubbles forms a magma foam with radically different physical properties to that of the parent magma. During ascent, this rising vesiculated magma is fragmented into discrete particles and transforms into a gas-particle mixture, which accelerates up the conduit and is discharged into the atmosphere (Figure 1).

Phreatomagmatic fragmentation is driven by the volumetric expansion of external water after it has been rapidly heated by contact with magma. This mechanism is not restricted by magma type or vent type and it encompasses a spectrum of eruption styles

Table 1 Summary characteristics for the major types of explosive volcanic eruptions

Eruption type and examples	*C*	*H (km)*	*V (km^3)*	*D (km^2)*	*Common eruption processes*	*Pyroclast types*	*Duration*	*Deposits*
Magmatic								
Hawaiian Kilauea, Mauna Loa, Hawaii	Ba	$\ll$1	$\ll$1	<1	Sustained fire fountains; spatter fed lava flows; lava lakes; weak ash plumes	Scoria, ash, spatter, bombs, Pelee's hair and tears, reticulite	Days to years	Clastogenic lava; scoria cones; spatter cones; spatter ramparts; localized sheet-like ashfall deposits
Strombolian Stromboli, Italy	Ba	$\ll$1	$\ll$1	5	Intermittent discrete explosions; ballistic clasts; weak ash plumes	Ash, scoria, bombs	Persistent over centuries	Scoria cones; scoria fall deposits; ballistic bombs and blocks localized sheet-like ashfall deposits
Vulcanian Soufriere Hills (1995–present), Montserrat	Ba-An, An, Da	<5–20	<1	<500	Numerous discrete and violent explosions; sustained eruption plumes; block and ash flows; dome extrusion; ballistic clasts	Ash, pumice, scoria, bombs	Minutes to hours	Ash fall deposits; ballistic blocks and bombs; small volume block and ash flow deposits; pumice flows
Plinian Vesuvius (1631), Italy Mt Pinatubo (1991) Philippines Taupo (AD 181) New Zealand	Rh, Ph, Da, Tr,	<20–>35	<1–3000	500–50 000	Sustained moderately to very high eruption plumes; widespread pumice fallout; sustained PDCs[1]; caldera collapse; commonly zoned	Ash, pumice	Hours to days	Extensive pumice fall deposits; widespread ignimbrites; ballistic blocks; pyroclastic breccias
Pheatomagmatic								
Surtseyan Surtsey (1963) Iceland	Ba	<20	<1	<500	Closely-timed phreatomagmatic explosions; radially expanding PDCs[1]; ash plumes; peacock tail jets	Ash, pumice, scoria, glass	Days to months	Tuff cones; fall deposits; PDC deposits; ash fall deposits; ballistic clasts; accretionary lapilli
Taalian Taal (1965) Philippines	Ba	<20	<1	<500	Closely-timed phreatomagmatic explosions; radially expanding PDCs[1]; ash plumes	Ash, pumice, scoria, glass	Days to months	Tuff rings; PDC deposits; ash fall deposits; ballistic clasts; accretionary lapilli
Phreatoplinian Askja (1875) Iceland Oruanui (26.5 ka BP) New Zealand	Rh(Ba)	<40	>1–>1000	500 50 000	Sustained high to very high eruption plumes; widespread pumice and ash fallout; sustained PDCs[1]; caldera collapse	Ash, pumice	Hours to days	Extensive bedded ash fall deposits; widespread ignimbrites

Abbreviations: C composition (Ba basalt; An andesite; Da dacite; Ph phonolite; Tr trachyte; Rh rhyolite), H plume height; V volume, D dispersal, PDC pyroclastic density current.
[1]Pyroclastic density cements.

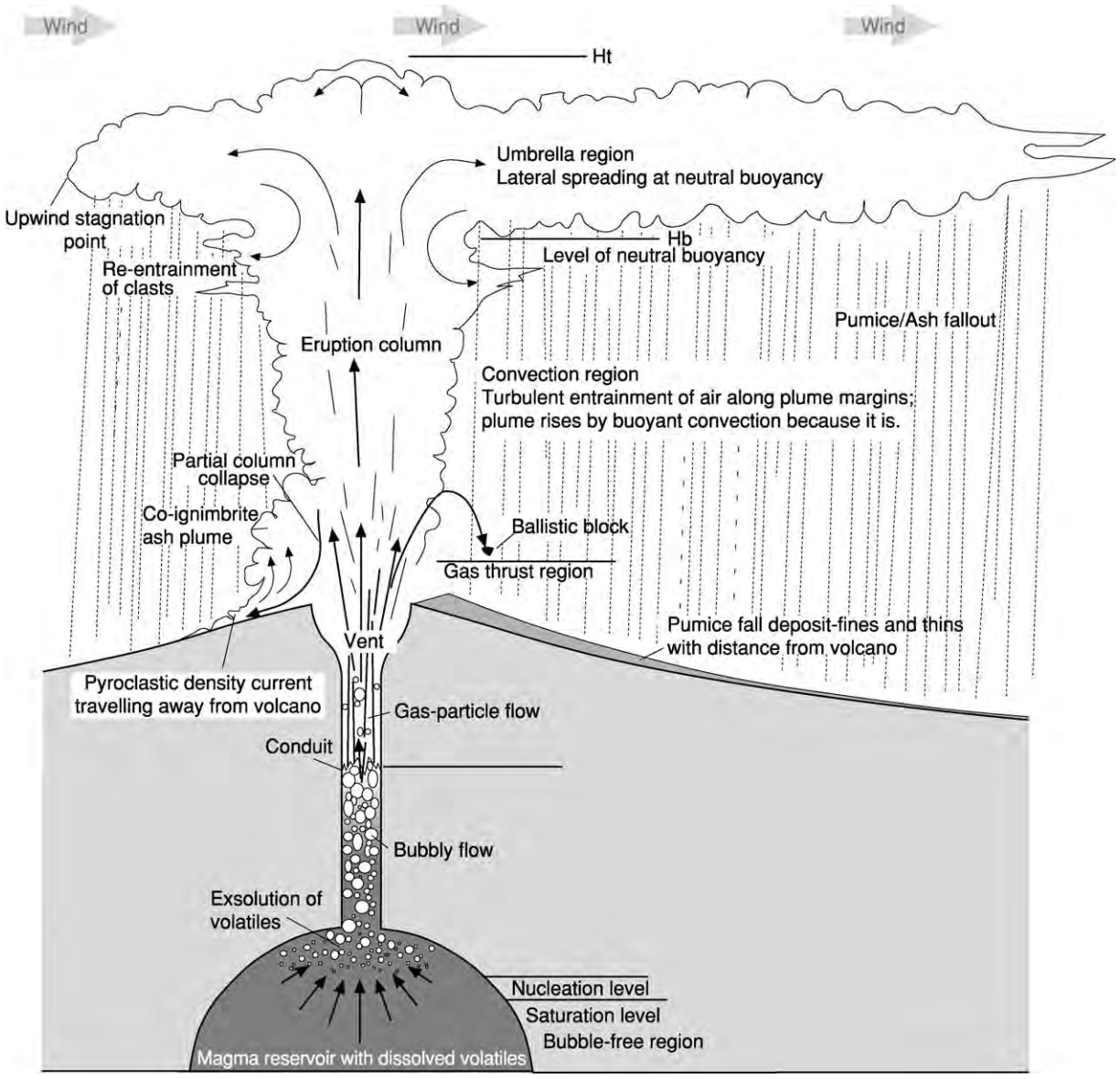


Figure 1 Cartoon illustrating the processes involved during an explosive eruption (in this case a Plinian eruption). Major processes in the magma chamber, conduit, and eruption plume are indicated.

from the shoreline fragmentation of lava (*see* **Lava**) entering the sea, to large volume phreatoplinian eruptions (Table 1). During a phreatomagmatic eruption, thermal energy is converted into mechanical energy (including seismic and acoustic energy) and kinetic energy (fragment motion). Fragmentation is controlled by the mixing ratio between water and magma, and by the magma viscosity, temperature/pressure, and the water/magma contact mode. As the eruption efficiency increases, more mechanical energy is released and fragmentation increases, resulting in a finer grain size and a greater distribution of pyroclasts.

The two fragmentation mechanisms are not mutually exclusive. Exsolving magmatic volatiles may contribute to expansion and fragmentation during phreatomagmatic eruptions, while phreatomagmatic phases are common during sustained magmatic Plinian and less extreme types of eruptions.

Eruption Plumes

Volcanic plumes are generated when explosive eruptions discharge a mixture of pyroclasts and gas into the atmosphere (Figure 2). These may either form buoyant convective plumes, which rise to great

Figure 2 Small Strombolian fire fountain feeding a weak ash plume (Etna, South Italy, 2003). (Photo courtesy of Tom Pfeiffer www.decadevolcano.com)

Figure 3 Pyroclastic density current generated during a Vulcanian eruption at Soufriere Hills volcano, Montserrat, West Indies. (Photo: E. Calder.)

heights in the atmosphere (<50 km) or they may collapse to form fountains (column collapse), which can generate flows of hot particles and gas that move laterally away from the volcano (pyroclastic density currents; Figure 3). Eruption plumes are termed 'maintained plumes' if the duration of discharge is sustained relative to the ascent time of the plume (e.g., Plinian eruptions) and 'instantaneous plumes' or 'thermals' if the discharge is short-lived relative to ascent time (e.g., Vulcanian explosions). The basic tripartite structure of these plumes is shown in Figure 1. In the gas thrust region the mixture of gas solids and liquid is expelled vertically from the vent as a jet with speeds up to a few hundred metres per second and a bulk density greater than that of the surrounding atmosphere. Air is mixed into the column, heated by the pyroclasts and expands, resulting in a decrease in the mixture density. If sufficient air is entrained and heated, the bulk density of the jet decreases below that of the surrounding atmosphere and the jet becomes buoyant, rising as a convective column. The plume continues to rise until it reaches a level of neutral buoyancy (Figure 1, 'Ht and Hb') at which it spreads laterally in an umbrella cloud. If however, sufficient air is not entrained, the jet decelerates until its velocity becomes zero and since the bulk density is greater than the surrounding atmosphere, the mixture begins to fall back to the Earth's surface.

Eruption Styles

Important measures of an explosive volcanic eruption are the magnitude (volume of erupted dense rock equivalent magma, in km^3), intensity (mass-flux), dispersal (areal extent of fall deposits, km^2), and driving mechanisms (magmatic or phreatomagmatic). These parameters are used to classify explosive eruptions; the characteristics of the major types are summarized in Table 1. Volcanic eruptions are extremely dynamic natural phenomena and many sustained eruptions undergo significant changes or fluctuations in eruption style through time, for example, changing from Surtseyan phreatomagmatism to Strombolian fire-fountaining. Likewise, the opening stages of Plinian eruptions can be phreatomagmatic or Vulcanian in style. Thus, the deposit of one eruptive sequence may comprise layers of contrasting pyroclasts that were formed by differing fragmentation mechanisms.

Pyroclast Types and Deposits

During explosive volcanic eruptions, magma can be transformed into pyroclasts (also called tephra) of widely varying physical properties, that range in size from submillimetric ash up to boulder size. Pyroclasts are classified according to their origin. Juvenile pyroclasts are derived from the erupting magma, and include pumice, scoria, volcanic glass, ash, crystals, and bombs (see Table 2 for descriptions). In magmatic eruptions, juvenile volcanic ash (vitric ash) comprises fine-grained cuspate, platy, or Y-shaped vesicle-wall shards. Vitric shards generated during phreatomagmatic eruptions can have blocky, fusiform, moss-like, platy, and spherical or drop-like shapes, reflecting the many different ways in which water and magma can explosively interact. Cognate pyroclasts are derived from the deposits of earlier eruptions at the same volcano, while accidental pyroclasts are those derived from the subvolcanic basement during the eruption or picked up by a pyroclastic density current during transport. Pyroclastic deposits typically comprise a mixture of juvenile, cognate, and accidental components. Due to the large differences in the physical properties of these components, and the sorting and segregation processes operating during pyroclast

Table 2 Characteristics of the main types of pyroclastic material

Pyroclast	*Description*	*Associated eruption styles*
Ash	Fine-grained (<2 mm diameter); whole or broken crystals, accidental lithic clasts or glass (vitric) shards; vitric shards can be cuspate, platy (broken vesicle walls), blocky (unvesiculated glass), or pumiceous	All
Pumice	Highly vesiculated volcanic glass; (density <1.0 g cm^3 – floats on water); fibrous or spherical/subspherical vesicles	Silicic explosive eruptions (particularly Plinian)
Scoria	Poorly vesiculated volcanic glass (density <1.0 g cm^3 – sinks in water; spheroidal vesicles; vesicularities 70–85%	All explosive eruptions; particularly Strombolian, Vulcanian
Bomb	Fluidal clot of magma that partially cools during flight; types include spindle, cow-pat, breadcrust, and cauliflower bombs	Hawaiian, Strombolian, Vulcanian
Spatter	Irregular fluidal clots of magma that can coalesce (weld) on impact	Hawaiian, Strombolian
Pelee's hair/ tears	Hair: long thin strands of glass Tears: irregular droplets with fluidal shapes (e.g., tear drop, ovoid and spheroid)	Hawaiian, Strombolian
Reticulite	Polygonal, lattice-like networks of very fine glass rods; very highly vesicular (95–99%)	Hawaiian
Accretionary lapilli	Comprise an ash core surrounded by one or more concentric laminae of usually finer-grained ash; also ash pellets, which lack concentric rims; Armoured or cored lapilli comprise a pumice or lithic lapilli core surrounded by a concentric rim of ash	All explosive eruptions that generate abundant fine ash; particularly phreatomagmatic (Surtseyan and Phreatoplinian); Magmatic eruptions during humid atmospheric conditions

Table 3 Particle size parameters for pyroclastic material

phi	*Mm*	*Class*
<3	<0.12	Very fine ash
3–1	0.12–0.5	Fine ash
1 to 0	0.5–1	Medium ash
0 to 1	1–2	Coarse ash
−1 to −3	2–8	Fine lapilli
−3 to −5	8–32	Medium lapilli
−5 to −6	32–64	Coarse lapilli
>−6	>64	Blocks and bombs

dispersal, the lateral and vertical distributions of these components may vary significantly within a deposit.

The grainsize terminology for pyroclastic material is given in Table 3. The term ash, when used to describe a pyroclastic deposit, refers to a unit comprised dominantly of ash-grade particles. When such a deposit is indurated or lithified, it is commonly called tuff. Deposits comprising ash and lapilli-grade particles are referred to as lapilli-tuffs, and deposits containing an abundance of blocks are called pyroclastic breccias. Pyroclastic successions can be best studied by utilizing a lithofacies-based approach. Lithofacies are commonly defined on a two-tier system of sedimentary structure and grainsize, although many classification schemes have been proposed and there is perhaps a strong case to be made for the standardization of pyroclastic lithofacies.

Pyroclastic Fall Deposits

Pyroclastic material carried up in eruption plumes eventually rains out from umbrella clouds to form pyroclastic fall deposits. These are classified according to a spectrum of increasing dispersal and fragmentation (Hawaiian, Strombolian to Plinian, Figure 4), but can also be distinguished on broad lithological grounds; scoria-fall deposits, pumice-fall deposits, or ash-fall deposits. Pyroclastic fall deposits are commonly well sorted and they maintain an even thickness over local topography (mantle bedding). Small eruptions (e.g., Hawaiian and Surtseyan) tend to produce weak, low-level plumes that form steep-sided, high aspect ratio fall deposits (cones), whereas Plinian eruptions produce high level (<50 km) columns that deposit extensive low aspect ratios (sheets).

An important concept in the formation of these deposits is that of terminal fall velocity, which depends on particle size and density and determines at what distance downwind from the source individual particles will begin to settle out. Large pyroclasts are often explosively ejected on ballistic trajectories (ballistic clasts) and are unaffected by the wind. The

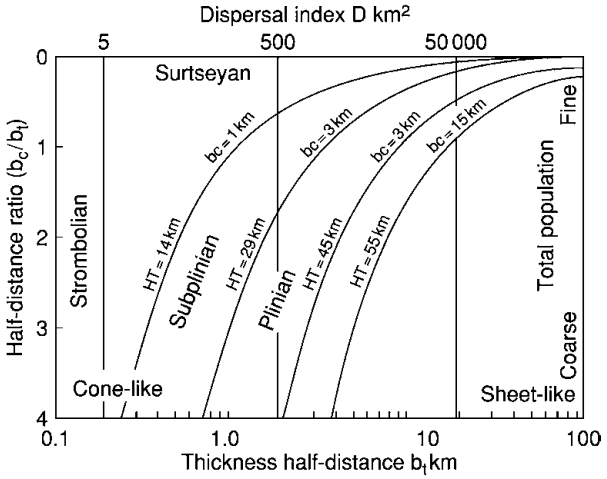


Figure 4 Classification scheme for fall deposits. The half-distance ratio represents the total grainsize population and the thickness half distance represents the dispersal. The figure is contoured for clast half-distance b_c and for total column height in km H_T. (From Pyle (1989).)

Figure 5 Parallel-bedded Plinian pumice fall deposits (Bandas del Sur Group, Tenerife, Canary Islands) overlying palaeosol (level with man's chest). Thin ash layers are visible in the middle of the deposit. White unit at top of photo is an ignimbrite. (Photo: M. Branney.)

thickness and grainsize of fall deposits decays approximately exponentially away from source and the rate of this thinning or fining is conventionally described by the 'half-distance', a measure of the distance over which the thickness or grainsize of the deposit halves. Quantification of fall deposits relies on the accurate mapping of the distribution and on detailed granulometric analysis. Isopach (contours of equal thickness) and isopleth (contours of equal grain-size) maps can be constructed and used to determine erupted volumes and investigate eruption dynamics. Isopachs and isopleths are commonly elongated downwind due to wind shear, although varying wind velocities and directions at different levels in the atmosphere, or changes in wind direction through time, can complicate dispersal patterns during sustained fallout.

Fall deposits are sensitive indicators of eruption dynamics. Barring interference from local atmospheric phenomena, quasi-steady eruption conditions should produce texturally homogenous, massive fall deposits. Such conditions are rare in nature and many fall deposits exhibit features indicative of eruption unsteadiness, for example, bedding (Figure 5), stratification, vertical grainsize variation (grading), and/or compositional variations (e.g., in abundance of accidental lithic lapilli). Such features are usually better developed in proximal deposits, becoming less well defined toward distal regions due to increased transport time. Table 4 summarizes some common features of fall deposits and their possible interpretations.

Pyroclastic Density Currents and Their Deposits

During an explosive volcanic eruption, pyroclasts can be carried laterally away from the vent as pyroclastic density currents (PDCs). These are ground hugging (i.e., they are denser than air) particulate gravity currents comprised of gas, ash, lapilli, and blocks. Pyroclastic density current is a general term given to a wide variety of phenomena generated by a number of very different mechanisms, for example, the collapse of an eruption plume, sustained boil-over fountain eruptions and directed lateral blasts (Figure 6). They can travel at high velocities (30–200 ms), can have high temperatures (>550°C), and can reach over 100 km from their source. PDCs tend to be controlled by the topography over which they flow, often being preferentially channelled down valleys but they can also surmount topographic obstacles more than 1.5 km high and travel many kilometres across open sea.

The nomenclature relating to the types of pyroclastic density currents and their respective deposits is notoriously complex. Pyroclastic density current deposits can be subdivided and classified in a number of different ways. Obvious distinctions can be made based on composition, deposit volumes, and dominant lithofacies, or on generic interpretations such as whether they resulted from either short-lived transient currents (e.g., during phreatomagmatic eruptions) or from sustained currents (during Plinian eruptions).

The term 'ignimbrite' is generally used to describe an ash and pumice-rich deposit generated during a Plinian-type eruption. Ignimbrites are some of the most impressive deposits on the planet: vast landscape burying sheets, emplaced over a few hours that can be several hundred metres thick and cover huge areas (>45 000 km^2). Ignimbrite range from small volume

Table 4 Major features of pyroclastic fall deposits and possible interpretations

Fall deposit characteristic		*Interpretation*
Geometry		
High aspect ratio cone		Fallout from a low (<2 km) eruption column; restricted dispersal
Low aspect ratio sheet		Fallout from a high eruption column (>2–>45 km); widespread dispersal
Circular isopachs/ isopleths		Fallout during no wind conditions; fallout from a powerful eruption column undeformed by wind
Elongate Isopachs/ isopleths		Strong distortion of eruption column by wind (elongation in downwind direction)
Changing isopach/ isopleth distribution with height		Variable wind directions at different heights; changing wind directions with time
Secondary thickening of fall deposit		Downwind aggregation of ash and premature fallout of large aggregates with greater settling velocities
Structure		
Grading	Inverse	Fallout from a growing eruption column (waxing eruption intensity); gravitational (downslope) remobilization of deposits (e.g., Strombolian deposits)
	Normal	Fallout from a waning eruption column (waning eruption intensity)
	Bedding	Fallout during unsteady eruption column dynamics (fluctuating eruption intensity)
Cross-stratification, lenses		Reworking of falling or deposited clasts due to surface winds, PDCs or gravitational remobilization (rolling down steep slopes)
Erosion surfaces (unconformities)		Removal of deposit due to mass wasting (landslides), PDCs, surface processes (alluvial/slope erosion)
Impact sags		Deformation by ballistic clasts
Compositional changes		
Chemical zoning of juvenile pyroclasts		Changes in the chemical composition of the erupted magma with time – eruption from a zoned magma chamber
Layers rich in accidental clasts		Explosive fragmentation of vent walls following blocking, collapse or increase in erosive power
Vertical changes in accidental clast composition		Vertical or lateral (from fissure vents) shift in locus of vent erosion
Layers rich in dense juvenile clasts (e.g., obsidian)		Explosive clearance of a cognate volcanic plug
Intercalated thin ash beds		Short-lived phreatomagmatic explosions which generate abundant fine ash; settling of fine ash from atmosphere during pauses in fallout activity
Pink ash coatings on pumice lapilli		Thermal alteration (oxidation) of pyroclasts
Accretionary lapilli; ash pellets; armoured lapilli		Moisture-driven aggregation of ash common during phreatomagmatic activity and during humid atmospheric conditions; electrostatic attraction of fine ash particles
Texture characteristics		
Poor sorting (presence of an ash matrix)		Partial phreatomagmatic activity (hybrid eruption); coeval deposition of fallout and fine ash from a dilute ash-rich PDC or a co-ignimbrite ash cloud
Rounded pumice clasts		Emplacement by lateral currents (PDC); reworking by alluvial processes
Welding		Proximal high temperature agglutination of hot (fluidal) pyroclasts; contact fusing of pyroclasts by an overlying welded ignimbrite

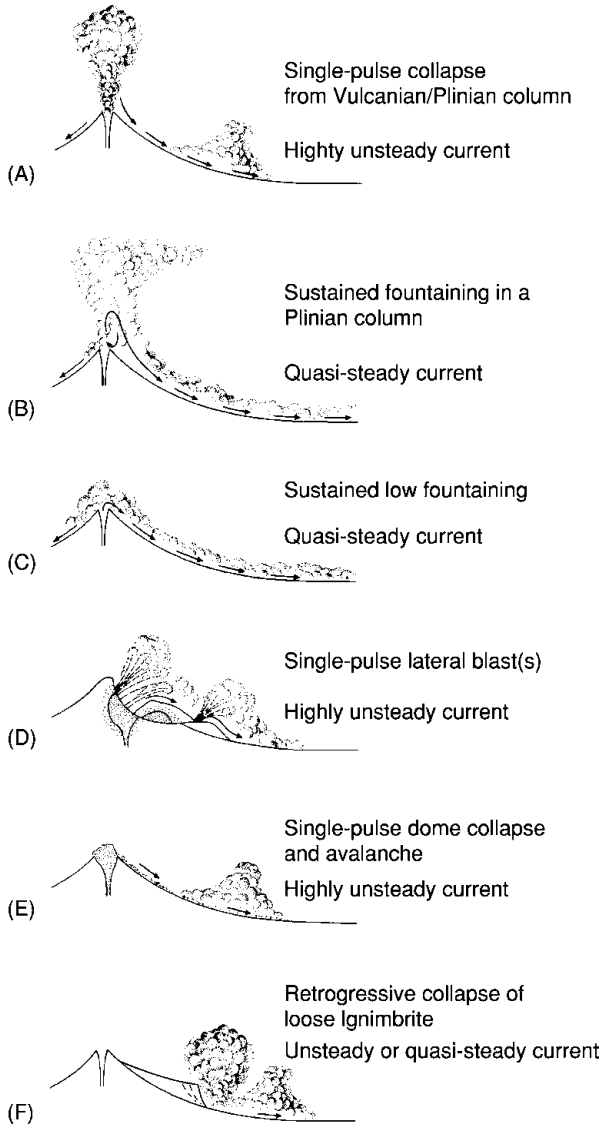


Figure 6 The origins of pyroclastic density currents. Most large volume ignimbrites are derived from B and C. Type D refers to transient pyroclastic density currents generated by lateral blasts or small scale phreatomagmatic activity (Surtseyan or Taalian eruptions). Type E commonly produces block and ash flows. (From Branney and Kokelaar (2002).)

ribbon-like deposits confined to valleys to extensive sheets distributed radially around the vent or across mountain ranges. Plinian eruptions are usually sustained for hours to days, and PDCs may be generated for all or part of that time, or intermittently. Ignimbrites exhibit a breath-taking diversity in their physical characteristics (e.g., geometry, bedding, sorting, grainsize, clast grading, geometry). Many are massive (Figure 7) and exhibit internal coarse-tail grading patterns the size grading shown by the coarsest particles in the deposit, but bedded, stratified, and cross-stratified lithofacies also occur. Detailed lithofacies

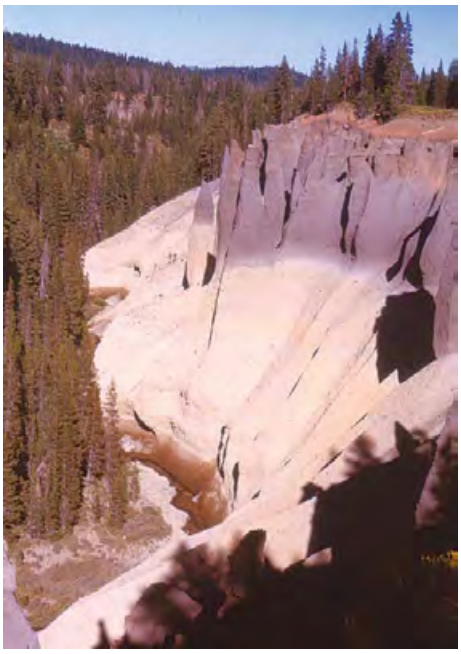


Figure 7 Compositionally zoned ignimbrite (Mount Mazama ignimbrite, Crater Lake, Oregon). (Photo: M. Branney.)

analyses, similar in approach to those applied to sedimentary deposits, have proved most successful in dealing with the complexity of ignimbrites and in elucidating important information about emplacement mechanisms. Moderate volume ignimbrites (1–10's km^3) often have complicated lateral, longitudinal, and vertical lithofacies transitions (ignimbrite architecture). Ignimbrites often show evidence for crystal enrichment relative to the juvenile pumice clasts, indicating significant loss of ash during transport (a process termed elutriation). This elutriated ash forms co-ignimbrite plumes, which rise vertically from moving flows and can later deposit extensive ash deposits.

Progressive evacuation of a chemically zoned magma chamber results in vertical chemical zoning of juvenile pyroclasts in an ignimbrite (Figure 7). Typically, juvenile pyroclasts become more basic with height (e.g., rhyolite to andesite; the inverse of vertical zoning in the magma chamber). Early phases of the eruption tap acidic magma (e.g., rhyolite) at the top of the chamber, while lower and more basic levels (e.g., andesite) are tapped during later phases. High temperature emplacement can lead to the agglutination of hot pyroclasts on deposition (welding). Pumice lapilli can flatten and deform to produce bedding-parallel *fiamme*, and columnar joints, similar to those developed in lavas, can develop. In extreme cases, the aggrading ignimbrite can start to flow as a dense viscous liquid and can end up looking superficially like lava.

Small volume ($<<1\,km^3$) well stratified pyroclastic density current deposits ('surge deposits') are typical products of monogenetic phreatomagmatic eruptions (e.g., Taal, Philippines) and directed lateral blasts (e.g., Mount St Helens, 1980). The pyroclastic density currents are typically transient single-surge types that travel out horizontally from source and rapidly wane. Each deposit can exhibit a wide range of sedimentary structures, and they characteristically show a proximal to distal lithofacies transition from massive coarse-grained lithofacies near source, stratified and cross-stratified deposits (with spectacular dune bedforms, Figure 8) further out, and planar-bedded deposits in more distal areas. Fine-grained ash fall deposits extend out from the distal limit of the sheet. Vertical lithofacies sequences often match the proximal–distal lithofacies transitions, reflecting deposition during strongly waning flow. Phreatomagmatic eruptions are characterized by numerous closely timed explosions, and during the course of the eruption, the deposits build up gently tapering cones with radii of several kilometres (called tuff rings or tuff cones). Due to the presence of external water, evidence for cool emplacement ($<100°C$) may be common in phreatomagmatic deposits, for example, abundant accretionary lapilli, vesicles in ash beds, rain-splash micro bedding, and the plastering of ash layers against objects.

Also included under the general term pyroclastic density currents, are hot avalanches of dense lava blocks and ash (block and ash flows) generated when portions of growing lava domes collapse due to gravitational instability. These events are short-lived and occur in discrete pulses. A classification has developed for these small volume pyroclastic flows depending on their generation mechanism. The term 'nuée ardent', favoured in some of the older literature, has also been applied in a rather general sense to most of these flow types. The generation of small pyroclastic flows by dome-collapse has been observed on numerous occasions (Montagne Pelée 1902; Fuego 1974; Mt Unzen 1990–1995; Merapi Volcano, and Soufrière Hills Volcano, Montserrat 1996–1998). These flows have been observed to move as far as a few kilometres from the vent at speeds of up to 60 ms. Due to the style of eruption, pyroclastic flows produced by gravitational collapse of lava domes occur repeatedly over significant time periods, typically, a few months to several years. This has enabled significant progress to be made in understanding the behaviour of these flow types in comparison to those of large volume ignimbrite eruptions. The deposits of these flows (block and ash flow deposits) vary from dense lava blocks and ash to dense semi-vesicular pumice and ash with only minor accidental components. Collapse volumes and local topography determine whether deposits are valley confined or sheet-like in nature. Small volume ($<0.5 \times 10^6\,m^3$) deposits tend to form well-defined lobate deposits with well-defined flow fronts and levees (Figure 9), while more extensive block and ash flow sheets tend to form deposits which thin laterally and have tapering margins.

Pyroclastic Density Current Transport and Deposition

The literature on pyroclastic density currents has been dominated by discussions on alternative concepts of transport and emplacement mechanisms. The physical nature of these flows and the mechanisms of sediment transport and deposition are still not fully understood. Conceptual models have been developed, ranging from dense concentrated suspensions dominated by particle interactions and fluidisation effects to very dilute

Figure 8 Cross-stratified pyroclastic density current deposit emplaced during a small volume phreatomagmatic eruption. Note spectacular dune bedforms interbedded with thicker massive deposits (Caldera del Rei tuff ring, Tenerife, Canary Islands). Current direction from right to left. (Photo: M. Branney.)

Figure 9 Well-developed lobes in block and ash flow deposits (Soufriere Hills Volcano, Montserrat). Small stripped trees for scale in left upper-middle of photo. (Photo: E. Calder.)

turbulent suspensions. In high particle concentration models, pyroclastic flows are considered as avalanche-like granular flows analogous to debris flows and rock avalanches. Variants of models involving high particle concentration granular flows emphasize non-Newtonian rheology, fluidisation effects, and hindered settling. Conceptual models have also been proposed that to varying degrees bridge the gap between the two rheological extremes. These latter models, which perhaps best describe the dynamics of PDCs generated during larger ignimbrite forming eruptions, propose separate low-concentration flows and higher-concentration depositional systems (flow-boundary processes). The historical development and evolving ideas surrounding this still somewhat controversial topic are covered in detail in several textbooks and major reviews (see **Further Reading** section at end of this article).

Pyroclastic density currents transport components of widely differing grainsizes and hydrodynamic properties (low density pumice, ash, and dense lithic clasts). During transport, these pyroclasts may be supported by a number of different mechanisms (fluid turbulence, saltation, rolling and sliding along an interface, and clast collisions) and adjacent clasts in a deposit may have different transport and particle support histories. Several field investigations support the idea that ignimbrites can be rapidly aggraded increment-by-increment from the bases of pyroclastic density currents (a process termed 'progressive aggradation', Figure 10A). Vertical chemical zoning of juvenile pyroclasts, gradual changes with height in an ignimbrite in grain fabric orientation, and lateral transitions from massive to stratified lithofacies, provide good evidence for incremental deposition. Deposition accompanies transport and the thickness of a deposit reflects the durations and rates of deposition, with basal parts deposited earlier than upper parts. Each particle carried in the current must pass through a lower flow boundary zone before it can be deposited (Figure 10).

The massive, poorly sorted nature of many pyroclastic density current deposits (Figure 7) indicates deposition from high particle concentration flows or flow boundary zones in which particles are supported by collisions (in granular flows, Figure 10D) and the upward flux of dusty gas produced as larger particles are hindered in settling, Figure 10E). Hindered settling results in partial fluidisation (sedimentation fluidisation) and greatly increases a current's mobility. Stratification and cross-stratification result from traction sedimentation involving fully dilute flow boundaries in which particle interactions have little role in particle support or segregation or in current rheology (Figure 10C). Small particles are supported by the turbulence in the surrounding dusty gas (air and suspended ash), whereas larger particles move by saltation along the base (as a bedload). Fully dilute conditions in the flow boundary zone result in more efficient segregation mechanisms and the deposition of relatively well-sorted stratified and cross-stratified deposits ('surge deposits', Figure 8). Small volume phreatomagmatic eruptions tend to produce currents initially of these characteristics, but which lose energy rapidly as they travel away from the vent. This results in higher rates of suspended load fallout, increasing particle concentrations and the deposition of more massive deposits with distance from source. Flow-boundary conditions in between the two end members outlined above (for example, in terms of particle concentration) will result in deposits that are gradational in character (diffuse stratified and bedded ignimbrites).

Inferences made from the deposits about the structure (e.g., particle concentration, density stratification) within the moving pyroclastic density currents have to be made with caution. However, the ability of some flows to travel large distances across water and scale mountains up to 1500 m high, >45 km from source, indicates that the greater thickness of these currents were dilute, low particle suspensions. Fluid turbulence is likely to be the dominant particle support mechanism in the upper parts of the currents. Additional evidence on the vertical structure of PDCs can be gleaned from ignimbrites that were emplaced over rugged topography. Parts deposited over high ground tend to be thinner and stratified or bedded, while contemporaneous deposits in depressions and valleys are more typically thick and massive (Figure 11). Such evidence suggests that conditions in the flow-boundary zone (e.g., in terms of current velocity, turbulence, and particle concentration) of the PDC varied significantly across topography and that aggradation rates were significantly higher in the valleys. What remains unclear for most pyroclastic density currents, however, is the partitioning of material between a higher concentration lower flow boundary zone and an overlying dilute and turbulent zone, noting that the vertical structure and the vertical density gradients are likely to vary considerably, both between and within PDCs, due to current non-uniformity and unsteadiness (see below).

Detailed lithofacies analyses of ignimbrite sheets provide important information on spatial variations (non-uniformity) and temporal variations (unsteadiness) in PDC flow-boundary conditions, induced by changes in eruption dynamics and by interactions between the flow and the substrate topography. Many large volume ignimbrite sheets were emplaced over minutes to hours during the sustained collapse or

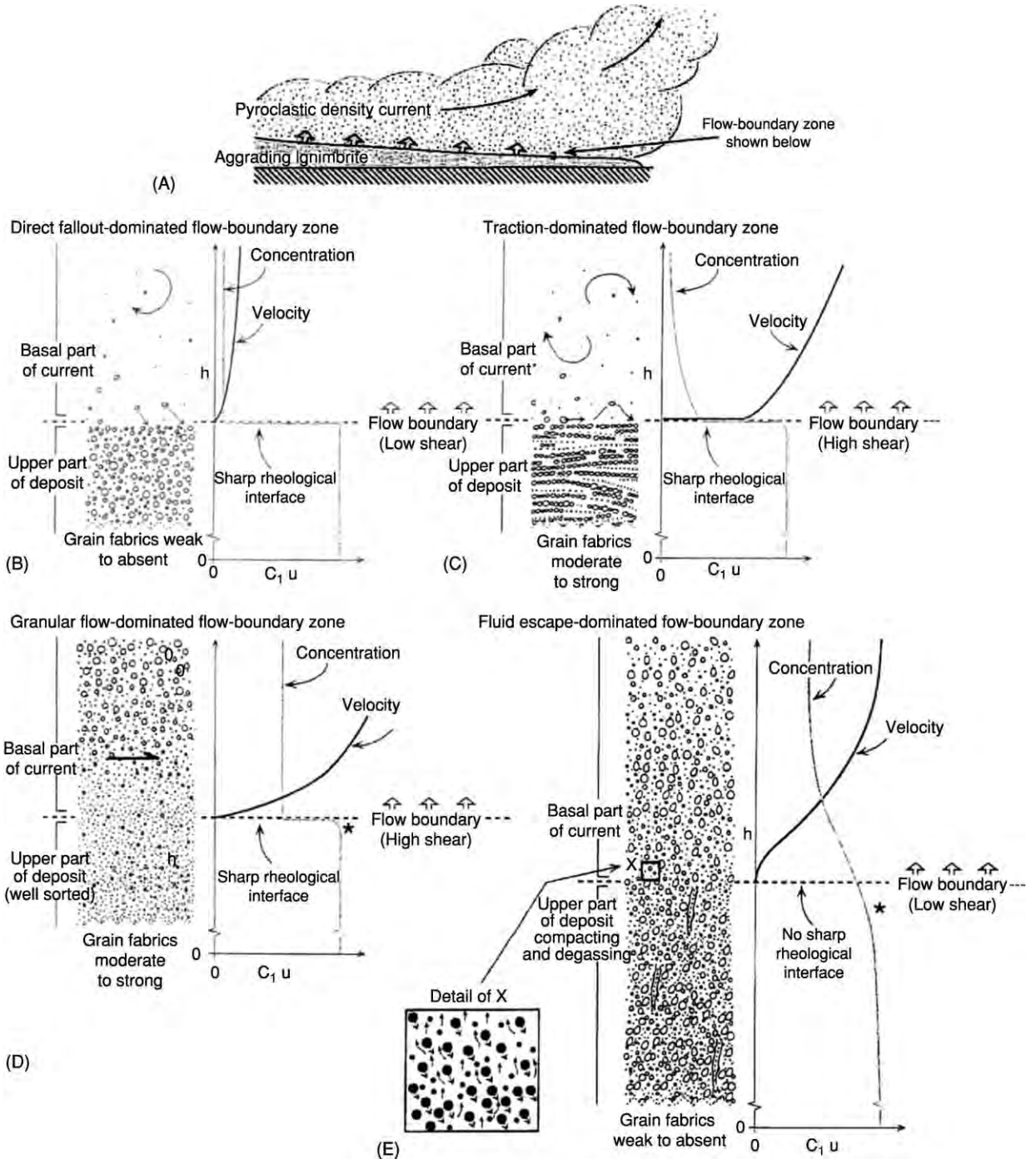


Figure 10 Four hypothetical types of pyroclastic density current flow-boundary zones during steady conditions. (A) Scale and location of flow-boundary zone emphasizing that depicted zones only include the basal part of the current and the uppermost part of the current. Most currents have flow boundaries intergradational between these types. (From Branney and Kokelaar (2002).)

fountaining of an eruption column. Sustained events are complex and can exhibit waxing and waning phases, pauses, and marked changes in eruption style. This can result in multiple flow-units and complex vertical lithofacies transitions and grading patterns. As the flows travel and deposit material, they substantially modify the substrate (e.g., by filling valleys, or by eroding loose material), and this, in turn, can affect sedimentation and flow boundary conditions up and downstream in subsequent parts of the current. Passage over uneven topography, changes in slope angle, and topographic blocking

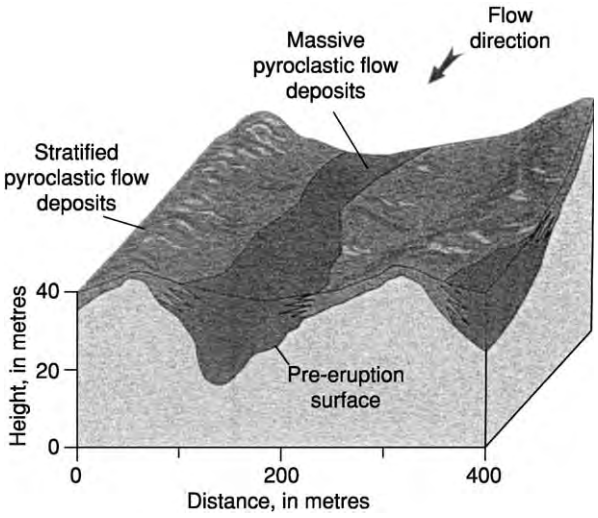


Figure 11 Lateral transitions from stratified ignimbrite (topographic highs) to massive ignimbrite (valleys) in the ignimbrite emplaced during the climactic 1991 eruption of Mount Pinatubo. (From Scott *et al.* in Newhall and Punongbayan (1996).)

(by mountain ranges), can all induce spatial changes in flow boundary conditions within a PDC and thus the deposit.

The dynamics of the small volume, end-member flows, such as those formed by fountain collapse during short-lived Vulcanian explosions and lava dome collapse are perhaps less controversial. Particle support mechanisms invoked for these small, relatively dense flows are commonly considered to be dominated by granular flow mechanisms (Figure 10D), where partial fluidisation may also play a role. The ash and gas components are considered important in friction reduction. Granular flow behaviour is controlled by a combination of the viscosity of the interstitial fluid and grain-to-grain interactions. In these dense small volume pyroclastic flows momentum is largely believed to be managed through grain-to-grain contacts. Low shear rate grain flows move as plug flows and become arrested en masse; high shear rate grain flows are smeared out during deposition as friction consumes momentum most strongly at the base while higher portions continue to move. Observations suggest separate mobile systems in which granular underflows or 'dense basal avalanches' occur beneath overriding turbulent ash clouds. The different properties of the two parts allow flow separation to occur, and the overlying ash cloud can move independently as a dilute pyroclastic density current.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards. **Igneous Processes**. **Lava**. **Volcanoes**.

Further Reading

Bonnadonna C, Ernst GGJ, and Sparks RSJ (1998) Thickness variations and volume estimates of tephra fall deposits: the importance of particle Reynolds number. *Journal of Volcanology and Geothermal Research* 81: 173–187.

Branney MJ and Kokelaar P (1992) A reappraisal of ignimbrite emplacement: progressive aggradation and changes from particulate to non-particulate flow during emplacement of high-grade ignimbrites. *Bulletin of Volcanology* 54: 504–520.

Branney MJ and Kokelaar P (2002) *Pyroclastic Density Currents and the Sedimentation of Ignimbrites*. London: Geological Society, London, Memoir 27.

Cas RAF and Wright JV (1987) *Volcanic Successions: Modern and Ancient*. London: Allen and Unwin.

Druitt TH (1998) Pyroclastic density currents. In: Gilbert JS and Sparks RSJ (eds.) *The Physics of Explosive Volcanic Eruptions*. Geological Society, London, Special Publications, 145, pp. 145–182.

Fisher RV and Schmincke H-U (1985) *Pyroclastic Rocks*. Berlin: Springer.

Francis P and Oppenheimer C (2004) *Volcanoes*. Oxford: Oxford University Press.

Freundt A and Rosi M (1998) *From Magma to Tephra: Modelling Physical Processes of Explosive Volcanic Eruptions*. Amsterdam: Elsevier.

Gilbert JS and Sparks RSJ (1998) *The Physics of Explosive Volcanic Eruptions*. London: Geological Society, London, Special Publications, 145.

Newhall CG and Punongbayan RS (1996) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*. Philippine Institute of Volcanology and Seismology and University of Washington Press.

Pyle DM (1989) The thickness, volume and grain size of tephra fall deposits. *Bulletin of Volcanology* 51: 1–15.

Sigurdsson H, Houghton BF, Rymer H, Stix J, and McNutt S (1999) *Encyclopedia of Volcanoes*. San Diego: Academic Press.

Sparks RSJ, Burnsik MI, Carey SN, *et al.* (1997) *Volcanic Plumes*. England, Chichester: Wiley.

QUARRYING

A W Hatheway, Rolla, MO and Big Arm, MT, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Quarries are resource extraction sites for removal of stone. The quarried stone is used as all manner of building materials, mainly as crushed-rock aggregate for concrete, road metal (crushed stone for road surfacing), asphaltic concrete pavement, and building construction stone. The stone is removed from 'pits' as cobbles, gravel, sand, and clay soil. Quarried stone is separated from outcrops along vertical working faces termed 'highwalls'. Highwalls are cut into a hillside of bedrock underlain at locations that have been selected to have little or no unconsolidated overburden (weathered rock, or 'soil'). Many quarries are developed as multisided rectangular excavations, and when the highwall exceeds a few tens of metres in height, the next level of extraction is 'benched' inward, leaving a stepped highwall, useful for control of inflowing groundwater. Where the host rock is not massive and may have some jointing dipping into the depression, bench heights are reduced to a general minimal height of 5–8 m. Stone removed from the quarry is hauled up a vehicle ramp for further processing at the company plant. Most stone is mechanically processed (broken and sized) to produce a size-graded product meeting engineering specifications of the buyer. When a quarry site is closed, environmental authorities typically require procedures to be followed as specified in a final closure plan; waste stone (spoil) is incorporated into the plan. Many urban-area quarries are closed out as disposal sites filled with environmentally inert construction and demolition wastes, mainly broken concrete and asphaltic concrete (Figure 1).

Creation of a geological quarry site model is essential for locating and proving quarry sites, in designing economical means for removing rock, and in devising environmentally sensitive ways for operation and closure. A number of geological factors and social/environmental considerations control the way in which quarries are located and operated:

- Stratigraphy and origin/mode of emplacement.
- Geomorphic setting.
- Postdepositional alteration of the host rock units and their rock fabric.
- Presence and orientation of discontinuities and potential layers of unacceptably altered weak rock.
- Nature and position of groundwater and management of its inflow, and ability to manage such water in terms of quarry operation and of mitigation of adverse impact on groundwater that is a neighbourhood resource.
- Ability to manage precipitation.
- Ability to manage rock extraction in such a way as to avoid instability of the highwalls, particularly in terms of quarry worker safety.
- Ability to manage waste rock (which typically amounts to about 5% of the extraction).
- Ability to operate in such a manner as to avoid danger to citizens and nearby residents, and to not degrade the environment.
- Development of a government-approved closure plan so that the closed quarry does not constitute a danger to the environment or to public health and safety.

Figure 1 Quarry operated for construction aggregate, road metal, and granular fill, in dolerite (diabase) and high-grade crystalline metamorphic country rock of Prince George's County, Maryland, USA; the vertical face is approximately 15 m high. The pit has reached the lateral limits of the deposit of usable rock and also has come to the maximum viable depth for maintenance of the pit, removal of groundwater seepage, and precipitation runoff. The site was being considered for conversion to a long-term secondary use compatible with environmental protection requirements. Photograph by the author.

Engineering Geology in Design and Operation of Quarries

Quarry development and operation has become a recognized field of practice requiring engineering geologists. As with many industries, quarry technology entered the machine age post-World War II, when excavation equipment became universally available to the family enterprises that owned most quarrying operations. Small operations are now rapidly disappearing and multinational companies control most of the stone that is mined today. In some instances, this equates to huge scales of magnitude (the 'superquarry'). Such are the economies of scale and the relatively low cost of bulk ocean transport that stone is imported over considerable distances, especially from quarries located close to coastal shipping.

Geological Factors Affecting Quarried Stone

Building stone is the traditional building material of mankind and still remains a critical construction material today, primarily as the aggregate, or bulking agent, for Portland cement and asphaltic concrete. Both cement and asphalt have been known to humans for several thousand years, yet their modern counterparts and their uses stem strictly from the industrial age.

Desirable Physical Properties of Quarried Stone

Quarried stone must have certain physical properties to serve as an effective building material. It must have strength, mainly compressional, to resist the load forces applied in engineered works; it must be durable and resistant to surface wear, especially when the use is for pavement; and it must be resistant to dynamic and chemical attacks, particularly from marine (saline) waters when the use is for port, harbor, and seacoast protective structures. Mined stone for building purposes must also be resistant to 'weathering' due to chemical attack by water and incident urban rainfall and must have consistency, be of uniform quality, and contain no deleterious minerals (e.g., pyrite, which causes staining of facing stone, and chert nodules, which cause eruptive pitting in concrete).

Desirable Mass Characteristics of Quarried Stone

Recovery cost for quarried stone is governed by several invariant cost factors inherent in the quarry site geology. The nature, pervasiveness, and orientation of discontinuities control the ease of extraction and the resultant stability of highwalls and benches. Stored tectonic stress may be a cost factor related to uneven responses to blasting patterns and explosive loading, due to release (relaxation) in strain-displacement relief ('pop-ups') of floor rock. The compressive strength of the quarry product, independent of geological structure, serves as a measure of the explosive force that will be needed (and thus the cost of explosives) to break the rock to requisite handling size.

Other Characteristics of Quarried Stone

Every stone quarry starts with an outcrop of durable stone or an excavated pit made in the overburden of surficial soils, often glacially deposited, so as to expose an artificial outcrop (working face) of stone. Once exposed, the stone must be subjected to efforts to dislodge it from the rock deposit. In most cases, the rock deposits selected for quarrying are cut by natural geological structural and depositional discontinuities such as bedding planes, foliation planes, and joints. Figure 2 depicts a rock face in which the supporting rock mass has three dominant joint 'sets', each representing a distinct array of joints, all of which will control the overall character and planar outline of a quarry developed in the rock mass.

Rock masses are virtually without strength in tension, and thus mechanical efforts to remove rock from a working face employ excavation machines to pull ('rip') discontinuity-bounded rock blocks from the quarry face (Table 1). Where the rock cannot be ripped, it is pulled from the working face by the judicial (planned) use of explosives set in geometrically arrayed and nominally vertical, drilled 'shot holes'. Discontinuities have value to the final product of the quarry only when the rock is destined to be recovered and sold as dimension stone, building stone, ornamental stone, or rip-rap. Rip-rap, used in shore protection works, is of great value when its quarry blocks are without further discontinuities, because the pieces can resist the constant pounding and surging action of ocean waves.

Many quarries also contain crystalline igneous and metamorphic rocks and the harder varieties of igneous and volcanic intrusive rocks. The many types of geological formations worldwide and their importance in quarrying operations are shown in Table 2.

Types of Rock Quarries

Quarries are established on the simple theme of locating and supplying a needed stone product at a competitive price in the free-enterprise system. There are some interesting relationships of quarry rock type and location worldwide (Table 2). Stone is intrinsically nearly valueless, and when obtained on a royalty

Figure 2 Rock highwall in rhyolite near Munsan, Republic of Korea, exhibiting three prominent joint sets of consequence to development of the site as a quarry. Such features should be taken into direct consideration in quarry dimensioning and in specification of the bearings of bounding highwalls; the horizontal width of the exposed joint faces is about 5 m, left to right. Photograph by the author.

basis nets the landowner literally pennies per tonne. Thus the primary quarry costs are essentially those related to extraction machinery/explosives costs, labour costs, and haul costs (for transportation from quarry to buyer).

Planning Considerations

Planning, regulatory involvement, and engineered design are required to develop a new stone quarry. The overall effort (Table 3) is aimed at minimizing both the cost of production and the operational and postoperational impacts on the local environment. Proposed quarries, and even expansion of existing quarries, must go through a typical sequence of feasibility studies, engineered design reviews, permit applications, operational environmental compliance reviews, and eventual closure and postclosure land-use reviews in accordance with conditions established prior to initiation of operations.

Engineering Geological Inputs

In addition to the geological tasks that are required to open a quarry and begin quarry operations, engineering geologists are necessary to the continued operation of quarrying. A stone quarry is a dynamic venture in which previously unidentified geological situations develop on a regular basis. There is an ongoing need to identify, classify, describe, record, and update all quarry site data in a three-dimensional geological model. Regular auditing of product and spoil utilizes digitized geological data and geostatistical analyses of the potential reserves (and thus the ongoing viability of the operation). These measurements and computations generally are compiled on 30-cm contour intervals at lateral scales of 1:1000. Geologists are the ideal staff members for the entire spectrum of detection, evaluation, interpretation, recording, and notification responsibilities necessary to proper management of the site. Geologists are also the most logical choice to serve as quarry managers, officers, and corporate directors.

Environmental Aspects

The most prevalent environmental issues in quarry operations are related to nuisances experienced by or perceived by a site's 'neighbours' (Table 4). Typical concerns are the noise emitted by the crushing, washing, and sizing operations and the noise and traffic caused by stockpile movement and loading and departure of haul vehicles. Though these concerns are not geological in origin, monitoring the disruptive operations often is best handled by geologists because of their skills in observation, location, mapping, and recording. An important aspect of environmental compliance is blast vibration; monitoring utilizes portable seismometers, in order to meet permit obligations and to provide defence against claims for structural damage allegedly caused by transitory ground motions from production blasts.

Careers in the Quarrying Industry

Technical and professional geological aspects of stone quarrying are fascinating and offer a distinct challenge to the geologist interested in the day-to-day practical application of their scientific knowledge. Quarries live and die on optimization of costs of producing buyer-acceptable products, and in doing

Table 1 Geological characteristics relevant to stone quarrying

Characteristic	*Importance*	*Geological considerations*
Discontinuities	Provide natural rock-break surfaces	Normally occur in predictable 'sets' of three-dimensional orientation in space; usually two sets represent each historical tectonic deformation phase of the region around the quarry
Uniaxial compressive strength	An engineering property that basically relates to the ability of explosives to break the rock deposit at the quarry face, in the absence of appropriate discontinuities	Individual explosive arrays, employed at 'shots', serve to lift, displace, and move the drilled and loaded face mass of rock outward, into the quarry, in a pile on the quarry floor, for subsequent working and removal
Structural domains	Rock masses with a geological history of tectonic, intrusive, or geochemical alteration have a characteristic three-dimensional variance in the orientation of discontinuities	Each 'domain' has definable geological contacts, within which the body of rock has predictable three-dimensional attitudes of its discontinuities; these control the whole activity of 'pulling' rock at the quarry face
Fault blocks	Quarries operated in tectonically damaged or seismically active terrain often show block displacement of masses of quarry rock, incorporating 10^5 or $10^6\,m^3$ of stone in the fault-bounded deposit selected for mining	Requires geological mapping and projection of geological structure, especially if the quality of fault-displaced portions of the quarry mass have different engineering properties or characteristics
Perched water	Water of all sorts is a hindrance to quarry operation	Will drain in hours and days and can generally be engineered so as not to reoccur; where encountered in the quarry face, must be channeled to a floor sump and pumped free of the quarry
Groundwater	Rock masses generally contain some groundwater and this will 'bleed' into the quarry over its operational life	Place the quarry at the highest elevation of the groundwater shed and thereby limit the movement of groundwater towards the quarry, which will otherwise perform as an undesirable master sump
Dykes, sills, and other igneous intrusives	Undesirable in all respects; these features follow pre-existing faults and shear zones and may foreshadow later discovery of fault-block displacement of the quarry deposit; intrusions generally are associated with some degree of lithological damage from ancient geochemical alteration	Engineering geological exploration and ongoing operational mapping should be guided so as to detect and avoid these conditions, which represent a high probability of the presence of undesirable rock properties and/or characteristics detrimental to the value of the stone produced at the quarry

so in compliance with environmental controls. There is little room for error or mistake, because stone must be handled effectively the absolute minimal number of times and supplied on demand to the client. The industry is highly competitive in terms of price and reliability of product and delivery. Solution of potential and actual geological problems associated with production and environmental compliance is a high-value return based on employment of competent professional geological staff in quarry operations. Undergraduate geologists giving consideration towards career employment in the stone industry may thus be duly rewarded. In fact, any geologist experiencing family demands towards a reduction in travel and time away from home may well want to exploit the opportunities in the stone industry. Technical expertise, when blended with managerial capacities, tends to be appreciated by the new-breed of large-industry quarry operators.

Underpinnings in geology required for success in the technical aspects of quarrying include the ability to make field geological maps and to operate at mapping scales of less than 1:1000; sound knowledge of igneous, metamorphic, sedimentary, and volcanic petrology; skill at petrographic analysis; a fundamental knowledge of hydrogeology; an interest in environmental protection and the supporting regulatory, permitting, and compliance issues; and an understanding of the national standards governing stone quality as a material of construction.

Seeking and Gaining Stone Industry Employment

There are two pathways towards professional-grade employment in the geological engineering field related to quarrying; both routes ideally require an informed and caring university faculty. A direct approach to the industry can often be effected through faculty

Table 2 Regional generalizations

Location	*Generalization*	*Geological considerations*
Singapore	Bukit Timah Granite and Gombak Norite quarried in the central third of the island for all hard-stone use	All pit and quarry space previously dedicated for follow-up use as potable water storage in accordance with the Master Plan of the Republic
San Francisco peninsula, San Francisco, California, USA	All bedrock essentially weathered and/or altered ultrabasic subduction rock	Market requires long-distance rail haul of suitable stone for aggregate
Los Angeles basin, Los Angeles, California, USA	Aggregate production mainly from coarse Pleistocene alluvial deposits; shortages made up from nearshore deposits of similar nature	Few quarries; mainly pits and these mainly converted to flood-control storage in recognition of the endemic cloudburst nature of rainfall in the region; some utilized as infiltration sites for treated wastewater effluent as 'water factories'
Missoula, Montana, USA	Coarse postglacial displaced outwash deposits of the Clark Fork of the Columbia River	Crystalline igneous and metamorphic rock
Most of Nebraska, USA	Coarse Plio-Pleistocene gravels of then-rising Rocky Mountains	Crystalline igneous and metamorphic rock contaminated with noticeable content of vertebrate faunal bones
Lower Mississippi River valley, USA	Huge open-pit quarries occupy the west bank of the Mississippi River, Missouri, below St. Louis; delivered by barge	Crushed Palaeozoic marine limestone of high carbonate content
Chicagoland: greater Chicago, Illinois, USA	Large, multisided quarries set below surrounding land surface	Palaeozoic dolomite; delivered by truck and rail
Greater Kansas City, Kansas, USA	Underground room-and-pillar limestone mines in dense, high-strength Mississippian-aged limestone	Modern mines engineered for long-term stability and conversion to high-value underground storage space
New York City/ Newark, New Jersey, USA	Triassic diabase (dolerite) quarries in New Jersey, with barge transport to New York City	High value for quarry end-use as disposal sites for inert construction and demolition waste

and professional society contacts; the secondary pathway is to take quality control employment with transportation or environmental agencies and become involved, respectively, in the review of specification quality of quarried stone products or in the environmental permitting and compliance oversight of producing quarries.

Literature Support for the Quarry Geologist

Quarry geologists should act quickly to assemble a reference library of relevant published materials (*see* **Building Stone**). Though much of the literature is relatively old and therefore available primarily from dealers of used and rare books, older works are easily searchable on the Internet. Worldwide, the current standard literature staples are the outstanding books on aggregates compiled by the Engineering Working Group of The Geological Society, London (*see* **Aggregates**). Subscriptions to trade journals are also essential, as are library trips to recover helpful geologically oriented articles from the historical literature archives. Though stone recovery methods have improved vastly since the end of Word War II, the underlying scientific principles are timeless.

Quarry Restoration

With the advent of environmental awareness in the 1970s, quarries in First World nations generally became subject to closure requirements as an integral part of environmental and/or land-use permit regulations. These requirements are well in place both in the European Union, particularly in the United Kingdom, and in Canada and the United States. In most cases, the requirements have in common the following parameters:

- Mandatory termination of extraction activities at pre-established bounds, both lateral and vertical.
- Filing of a closure plan and a postclosure plan with a regulatory agency, including appropriate engineering details of the concept by which the site is to be made safe for residents of the surrounding area and by which precipitation and groundwater inflow are to be managed in an environmentally acceptable fashion, to include safeguarding against dumping of waste at the closed facility.
- Placement of closure and postclosure funds in an escrow account, with amounts and payments tied to quantity of rock extracted.

Table 3 Typical sequence of events

Event	*Action agency*	*Geological considerations*
Market feasibility study	Proposed operator	Nature, value, and cost to produce the end product(s)
Feasibility of prospective quarry	Proposed operator and financial partners	Estimated unit cost of product(s); cost to develop plant, as amortized over estimated quantity of stone reserves; estimated cost to close the site at termination of productive life; nature and value of end-use of site
Development of topographic contour map	Consultant to proposed operator	Basis for planning of property geological exploration plan
Conceptual design of proposed quarry	Proposed operator, with assistance of consultants or corporate engineering staff	Location and bounds of the bedrock deposit; identification of potential geological constraints to operation and closure; management of groundwater and production spoil
Phase I geological exploration	Consultant to proposed owner or corporate engineering staff	Prove the existence of minimal reserve of suitable bedrock for production of stone product(s)
Tentative facility operational concept	Proposed operator, operator staff, or consultants to proposed operator	Identifies stages or phases of site development and product treatment; staging, groundwater, and surface water control features; and management of production spoil
Environmental impact report	By consultants to proposed quarry operator	Geological map of soil, bedrock, and groundwater conditions; hydrologic and hydrogeologic impact; control of dust, fugitive emissions, spoil, and runoff
Environmental impact statement	By state, county, or provincial public environmental agency	Identifies all physical and social environmental impacts and proposes appropriate mitigative measures; identifies and quantitifies risks to public health and welfare
Land-use zoning application	Proposed owner	Identifies geological constraints related to nature land use, as such impact provisions of land use specified in the General Plan of the governmental administrative body; shows that proposed quarry is an acceptable land use in consequence of facility operational concept
Permit application	By proposed operator to the public agency holding regulatory control	Mainly equivalent to that of the environmental impact report
Operation, compliance, expansion, and/or closure	Proposed owner, staff, and consultants	All above-noted geological considerations

Table 4 Major sources of environmental impact of quarrying

Impact	*Relevance*	*Geological considerations*
Alteration of groundwater regime	Quarries, as cuts and depressions, interrupt groundwater flow and create artificial flow gradients and discharge points/areas	Most of these 'interruptions' create, in fact, a 'shadow' of impact on the position and flow path of groundwater
Fugitive dust	Extraction of stone creates breakage and some comminution (size reduction), resulting in small particles subject to wind movement	Method of extraction, method of breakage, and method of size treatment for market product
Fugitive sediment	Results from dust caught up in the path of runoff	Reduction of amounts of dust and other small particles produced; methods of entrapment and sequestering from release to the environment
Loss of ground support to adjacent activities and structures	Slope instability could be induced here in the highwall, with the potential to ravel or migrate backward into adjacent land and buildings	Generally mitigated geologically by lateral buffer ground separating the highwall from adjacent activities and attention to the effects of adverse geological structure in the highwall rock (such as dips into the quarry)
Collateral damage from flying rock and ground motion of production blasting	Physical guidelines are well established for control[a]	Control related to empirical relationships of burden, ratio of explosive charge to mass of rock to be shot, and pattern and time delay of charges

[a]See guidelines of the United States Bureau of Mines, for example.

- An agreement on the postclosure land use for the property, including both the pit and highwall areas, the property roads, drainage, and the buffer zone surrounding the quarry.

In the twenty-first century, quarries have taken on a new postclosure value for several reasons, not the least of which includes the potential for new developable space following quarry backfilling with construction and demolition debris. Conversion to reservoir use for water supply storage, conversion to high-value residential use of the quarry rim when pit quarries are converted to recreational areas, and conversion to sanitary landfill use (if and where groundwater conditions and appropriate engineered design features can preclude environmental degradation) also are potentially valuable applications of former quarry sites. Conversion to some form of conservation use, often in agreement with local or regional public agencies, is another positive outcome for defunct stone mining operations.

See Also

Aggregates. **Building Stone**. **Engineering Geology:** Ground Water Monitoring at Solid Waste Landfills.

Further Reading

Department of the Environment, UK (1988) *Handbook on the Hydrogeology and Stability of Excavated Slopes in Quarries*. Geoffrey Walton Practice, Mining Geologists and Geotechnical Engineers. London: Her Majesty's Stationary Office.

Department of the Environment, UK (1988) *Technical Review of the Stability and Hydrogeology of Mineral Workings*. Geoffrey Walton Practice, Mining Geologists and Geotechnical Engineers. London: Her Majesty's Stationary Office.

Department of the Environment, UK (1992) *Landform Replication as a Technique for Reclamation of Limestone Quarries*. Limestone Research Group, Manchester Metropolitan University. London: Her Majesty's Stationary Office.

Department of the Environment, UK (1992) *Revised Planning Guidelines for Extraction of Aggregates* (replaces Master Planning Guidance Note 6). London: Her Majesty's Stationary Office.

Knill JL (ed.) (1978) *Industrial Geology:* Oxford, UK: Oxford University Press.

Latham J-P (ed.) (1998) *Advances in Aggregates and Armourstone Evaluation: Geological Society of London. Special Publication 13*, London: Geological Society.

Lent FA (1925) *Trade Names and Descriptions of Marbles, Limestones, Sandstones, Granites and Other Building Stones Quarried in the United States, Canada and Other Countries*, New York: Stone Publishing Co. (originally appearing in *Stone Magazine*, revised and corrected for the pamphlet).

Smith MR and Collis L (eds.) (2001) *Aggregates; Sand, Gravel and Crushed Rock Aggregates for Construction: Geological Society of London, Engineering Geology Special Publication No. 17*, 3rd edn. London: Geological Society.

United States Army (1967) *Pits and Quarries: Technical Manual TM 5-332*. Washington, DC: United States Government Printing Office.

REEFS

See **SEDIMENTARY ENVIRONMENTS: Reefs ('Build-Ups')**

REGIONAL METAMORPHISM

A Feenstra, GeoForschungsZentrum Potsdam, Potsdam, Germany
G Franz, Technische Universität Berlin, Berlin, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Regional metamorphic rocks form from other rocks (protoliths) by changes in mineralogy and texture in response to changing physical conditions (temperature, lithostatic pressure, and, in most cases, shear stress). These changes are essentially solid-state reactions, but very often a fluid phase is present, either participating in the reaction or as a reaction medium. Many regional metamorphic rocks have a chemical composition that is very similar to that of their sedimentary or igneous precursors, with the exception of removal or addition of volatiles (mainly H_2O and CO_2). This type of behaviour is termed 'isochemical metamorphism'. Metamorphism may also take place as a result of a change in chemical environment; this may occur by transport of elements between chemically contrasting rock types (e.g., formation of calc-silicate minerals at a quartzite–marble contact) or by circulation of fluids that dissolve some substances and precipitate others. This process of significant chemical change during metamorphism is known as 'allochemical metamorphism' or 'metasomatism', and rocks formed in this manner are metasomatic rocks. Metasomatism is, however, mostly of local significance, and the total volume of metasomatic rocks in regional metamorphic terranes is rather minor. The distinction between metasomatism and isochemical metamorphism is also a matter of scale. On the scale of individual grains, mass transport takes place during all phase transformations; on the scale of a thin section, it is probably the rule for regional metamorphism; on the scale of a hand (-sized) specimen, it can be observed frequently; and on a larger scale, it is the exception.

Regional metamorphism is nearly always accompanied by deformation and folding, which happens at converging plate boundaries and leads to orogenic belts. Therefore, regional metamorphic terranes follow the contacts between ancient lithospheric plates (suture zones), typically have an elongated shape that is parallel to the mountain belt, and cover areas of several 1000 to 10 000 km^2. Two basic types of regional (orogenic) metamorphism may be distinguished. The collision type relates to the collision of essentially continental plates and mainly produces metamorphic rocks at medium pressures and medium to high temperatures, often associated with partially molten (migmatites) and magmatic rocks. High- and ultra-high-pressure terranes ($P \approx 1.5$ GPa up to >3 GPa) from the deep roots of the collision zone can be preserved within the medium pressure (<1 GPa) belts. A variety of sediments from the continental margin and shelf (clay–shale, quartzofeldspathic rocks, limestone, marl), their metamorphosed equivalents, and acid magmatic rocks constitute the protolithic material of the collision type, whereas rock types of oceanic affinity are volumetrically unimportant. Typical examples include Mesozoic–Cenozoic metamorphic terranes in the Alps and Himalaya, Palaeozoic terranes in the Caledonides of Scotland and Scandinavia, as well as many Palaeozoic–Precambrian terranes in the shield areas of the continents.

The subduction type of regional metamorphism is associated with subduction of an oceanic plate underneath a continent and results in rocks formed at high pressure and relatively low temperatures (blueschist, eclogite). Basic igneous rocks (basalt, gabbro, volcanogenic rocks) together with clay-bearing clastic sediments are the dominant protolithic materials. Continental rocks may be included locally in the subduction-type terranes as well. Well-documented examples are found in the circum-Pacific region (Japan, California, New Caledonia, New Zealand), the Alps, and the eastern Mediterranean (Cycladic islands, Greece; south-western Turkey). In the case of a terminating oceanic plate, the oceanic subduction may be followed by a continent–continent collision, such that the two types may overlap each other in time.

Regional Metamorphic Processes

Large-scale geotectonic processes move rocks along particular paths in pressure–temperature–time (P–T–t) space. During movement along this path, the rocks re-equilibrate, and when chemical equilibrium is attained (when the reaction kinetics is fast enough), the mineral assemblage and the compositions of the individual minerals will react continuously by adapting to the new physical conditions. In addition to the mineralogy, the fabric of the rocks also changes, because regional metamorphism is a dynamic process typically accompanied by an anisotropic pressure field. Minerals with an elongated or platy habit will commonly align to the confining stress field. Mineral grains will also increase in size with increasing temperature, reducing the surface free energy (typically observed, for example, in metacarbonate rocks and quartzites). This process is termed 'static recrystallization'. Deformation in the anisotropic pressure field, however, reduces the grain size by various processes, depending on temperature and confining pressure; this is called 'dynamic recrystallization'.

Particularly at low-level metamorphic conditions, when deformation of the rock is non-penetrative, and in fluid-deficient rocks, equilibrium is often incomplete and minerals initially present in the protolith, and minerals formed during early metamorphic stages, may locally survive metastably. Such metastable persistence of minerals (e.g., as relict inclusions in newly formed minerals or as compositional zoning in minerals from core to rim) is fortunate for the petrologist in the way that it records the early geological history (the prograde path) of rocks. Petrological studies of metamorphic rocks traditionally aim to estimate the peak metamorphic conditions (pressure, temperature, and composition of coexisting fluid phase) that a rock has experienced, based on the assumption that the mineral assemblage essentially documents this peak equilibrium assemblage. Equilibrium among the minerals is commonly verified by textural evidence, i.e., that all minerals of the peak assemblage should be in mutual contact. Chemical evidence for equilibrium can be obtained from comparison of the composition of coexisting minerals in different bulk compositions; no crossing tie lines should be present and the tie lines (e.g., of garnet–biotite) should follow a systematic pattern. During the cooling and exhumation (the retrograde path), this peak metamorphic assemblage experiences only minor changes, because these processes are commonly associated with fluid-free conditions and minor deformation. A retrograde reaction, i.e., the formation of lower grade, mostly hydrous minerals from higher grade ones, thus typically occurs on a local scale. Examples include the replacement of garnet or biotite by chlorite or the replacement of plagioclase by calc-silicates such as (clino)zoisite, prehnite, or pumpellyite. Many occurrences of retrograde hydrous and carbonate phases are spatially connected with late (brittle) deformation, implying that deformation may facilitate local fluid introduction and thus formation of retrograde minerals.

In virtually all regional metamorphic terranes, mineral reactions are associated with deformation. Therefore, it is often possible to determine relative time relationships between growth of minerals and major deformation events characterized by development of a distinct foliation, folding, or shearing. Growth of a mineral can be specified as pre-, syn-, or post-tectonic with regard to specific deformation events. Relationships between mineral growth and deformation may be complex, however, because multiple or long-lasting deformation and mineral growth are the rule rather than the exception during orogenic cycles. Studies of regional metamorphic rocks conveniently separate the rocks' evolution into the three categories of prograde, peak, and retrograde stages using petrological and structural criteria. The researcher should be aware, however, that this is a simplified picture and the P–T–t paths of metamorphic rocks are in general quite complicated.

Regional Metamorphic Zones and Facies

It was recognized very early that metamorphic rocks can be classified according to different 'intensities' of metamorphism, or metamorphic grades. The concept of gradation is not applicable to the other families of sedimentary or magmatic rocks; a magmatic rock cannot be 'more magmatic' than another one, whereas a metamorphic rock can be metamorphosed more intensely. This concept also emphasizes that the processing of a sediment through diagenesis, then forming a metamorphic rock, is transitional, and it therefore is impossible to mark an unequivocal beginning of metamorphism (*see* **Diagenesis, Overview**).

The concept of metamorphic index minerals typical of metamorphic grade was first introduced by George Barrow in the south-east Highlands of Scotland in the late nineteenth and early twentieth centuries. Studying a sequence of metamorphosed clay-rich sediments, Barrow related the systematic change in mineralogy in these metapelitic rocks to changing metamorphic conditions. Barrow recognized chlorite, biotite, garnet, staurolite, kyanite, and sillimanite zones as representing increasing metamorphic grades

that are characterized by index minerals not present at lower grades. 'Mineral-appearance' isograds were mapped by drawing lines between locations where each index mineral first appeared. It later became common practice to replace these classic mineral isograds by 'reaction' isograds related to a particular metamorphic reaction and to use the equilibrium assemblage of that reaction as a zone boundary. The reaction isograd is defined by assemblages of reactant(s) on one side, of product(s) on the other side, and, if present, the equilibrium assemblage on the isograd line (surface).

It should be emphasized that in metamorphic petrology, an isograd is not exactly a line (surface) of constant temperature (and pressure). Various aspects of metamorphic reactions should be considered. (1) In cases in which minerals with variable chemical compositions (solid solutions) participate in the isograd reaction, the position of the isograd in pressure–temperature (P–T) space may depend on mineral compositions (and the bulk chemical composition of the rock). (2) The P–T position of reactions producing or consuming volatiles (H_2O, CO_2) may also depend strongly on the composition of the metamorphic fluid. The reaction kinetics may vary on a local or regional scale in metamorphic terranes, especially at the lower metamorphic grades. These aspects imply that mapped reaction isograds will only approximately reflect lines (surfaces) under identical peak temperature (and pressure) conditions.

Following the pioneering studies of Goldschmidt in southern Norway, in 1920, Pentii Eskola developed the concept of metamorphic facies. Studying mafic metamorphic rocks (mostly basaltic protoliths) in southern Finland, he recognized that "in any rock or metamorphic formation which has arrived at a chemical equilibrium through metamorphism at constant temperature and pressure conditions, the mineral composition is controlled only by the chemical composition". This concept has now been validated worldwide in innumerable regional field studies demonstrating that many rock compositions develop mineral assemblages typical of the physical conditions of formation. For a certain bulk composition, the mineral assemblages formed at these different metamorphic grades are then related by mineral reactions. Experimental determinations of the relative stabilities of metamorphic minerals and assemblages, followed by thermodynamic calculation of such phase equilibria, have allowed an approximate quantification of the pressure and temperature conditions of each metamorphic facies.

The metamorphic facies diagram shown in Figure 1 differs in some details from that proposed by Eskola, who named his metamorphic facies after

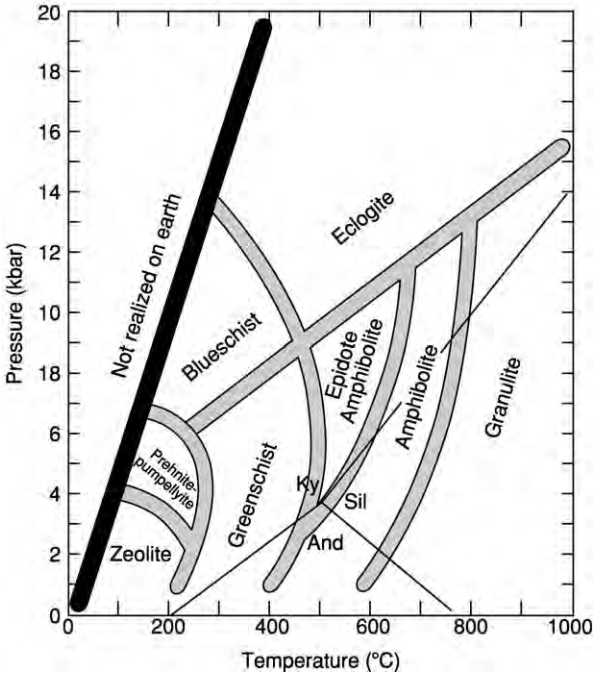


Figure 1 Pressure–temperature diagram, showing principal metamorphic facies; approximate boundaries are represented by thick grey lines. Thin black lines designate the stability fields of the three Al_2SiO_5 polymorphs (And, andalusite; Ky, kyanite; Sil, sillimanite). Reproduced with permission from Spear FS (1993) *Metamorphic Phase Equilibria and Pressure–Temperature–Time Paths*. Washington, DC: Mineralogical Society of America.

metamorphosed mafic rocks. Nowadays, differences in facies classification exist mainly in the transition from greenschist facies to amphibolite facies at higher pressure; some authors prefer, and others do not, to introduce an epidote–amphibolite facies (Figure 1). In addition, the classification of the facies below the greenschist facies is summarized by several authors as 'subgreenschist facies'. In any case, the borders between different facies are not sharp, and the classification should be used in a broad sense. It is schematic, but there is consensus on the temperature and pressure limits for each facies. The boundaries between the facies represent the P–T conditions at which key minerals or mineral assemblages appear or disappear.

Extensive petrological studies were carried out in metamorphic terranes around the world during the twentieth century. These studies revealed that distinct metamorphic facies series (high, medium, and low pressure and very low grade) correlating with specific geotectonic settings could broadly be distinguished.

Facies of High Pressure

High-pressure rocks of the blueschist and/or eclogite facies were metamorphosed at a low geothermal gradient, i.e., small increase in temperature with

depth (Figure 1). They are typically formed during subduction of a cold oceanic plate underneath a continent. Mafic protoliths best document such high-pressure conditions by their diagnostic mineral assemblages in blueschist (sodic amphibole bearing; the term 'blueschist' derives from the blue sodic amphibole, glaucophane) and eclogite (dominated by garnet + omphacitic clinopyroxene). Metapelitic rock compositions may develop diagnostic minerals or assemblages such as carpholite, phengite, jadeite + quartz, and talc + kyanite.

Facies of Medium Pressure

Most metamorphic terranes, especially in the internal parts of continents, are composed of greenschist, amphibolite, and granulite facies rocks, including migmatites at the high metamorphic grades. Such a facies series is related to a geothermal gradient of 20–35°C km^{-1}, typical of the Barrovian-type metamorphism as defined by Barrow in south-east Scotland. A well-documented example of a Barrovian-type metamorphic zoning is exposed on the island of Naxos, Greece (see later). As already discussed, medium-pressure regional metamorphic zones are best developed in Al-rich metapelitic rocks, which may contain either kyanite or sillimanite, or both Al-silicates. The formation of staurolite and the breakdown of chloritoid are generally considered to reflect the greenschist–amphibolite facies transition in pelitic rocks. In mafic rocks, the greenschist–amphibolite facies transition is designated by the appearance of hornblende and Ca-rich plagioclase and the disappearance of actinolite and Na-rich plagioclase. In metapelites, the amphibolite–granulite transition is considered by some workers to be indicated by the assemblage K-feldspar + Al-silicate forming at the expense of muscovite + quartz. Other workers consider the formation of orthopyroxene to be typical of granulite facies conditions, although this occurs at significantly higher temperatures (100–150°C higher) than does formation of K-feldspar + Al-silicate. In mafic rocks, the presence of orthopyroxene, which forms at the expense of hornblende and/or biotite, is considered indicative of granulite facies conditions. The role of water activity, however, complicates characterization of the amphibolite–granulite facies transition. If rocks are water saturated, they will, with a eutectic or eutectoid chemical composition near to granite, partially melt at P–T conditions in the muscovite + quartz stability field. When water activity is low, pervasive partial (dry) melting will occur at much higher temperatures and hydrous minerals (muscovite, biotite, and amphibole) may decompose to anhydrous ones by local dehydration melting.

Facies of Low Pressure

Low-pressure facies are in many cases typical for contact metamorphism (*see* **Thermal Metamorphism**), but especially in cordilleran-type mountain belts this contact metamorphism can be of regional extent due to long-lasting intrusive activity over a large area, producing a transition to regional metamorphism (Buchan-type in Scotland, Abukuma-type in Japan). It also may follow the medium-pressure facies. Metapelites are characterized by andalusite and cordierite.

Facies of (Very) Low Grade

Facies of very low grade are present in the transition zones of an orogenic mountain belt to the sedimentary foreland (e.g., Swiss Alps), but are often missing in regional metamorphic terranes, where the metamorphic rocks are overthrusted onto the foreland. They are typically best developed in regions with burial or hydrothermal metamorphism (e.g., New Zealand, Central Andes).

The Barrovian-Type Metamorphic Complex of Naxos

The island of Naxos is part of the Attic–Cycladic Metamorphic Complex that stretches from mainland Greece to south-western Turkey. Isoclinally folded sequences of metasediments (pelitic and psammitic schists and gneisses, quartzites, and calcitic and dolomitic marbles) and metavolcanics (amphibolites and mafic schists) are the dominant lithologies on Naxos (Figure 2). Ultramafic rocks occur in a few stratigraphic horizons and as isolated lenses. Metacarbonate units are widespread and contain in certain stratigraphic horizons diaspore and/or corundum-bearing lenses (up to ≈8 m thick; Figure 2), for which protoliths are bauxite formed in a karst soil. These metabauxite occurrences offer a unique opportunity to investigate the prograde metamorphism of a rock composition with excess Al.

The metamorphic pattern of Naxos was shaped during a late Alpine medium-pressure event that culminated 15–20 Ma; it consists of concentric zones of decreasing metamorphic grade, outward from a high-grade migmatitic core (Figure 2; Table 1). In the western part of Naxos, a ≈13-My-old granodiorite intruded the metamorphic complex and induced andalusite–sillimanite-type contact metamorphism in about a 1-km-wide zone. In this same region of Naxos, a post-intrusive upper unit, comprising Tertiary sediments and ophiolite-suite rocks, tectonically overlies the metamorphic complex and granodiorite. Petrological and geochronological studies indicate that relict

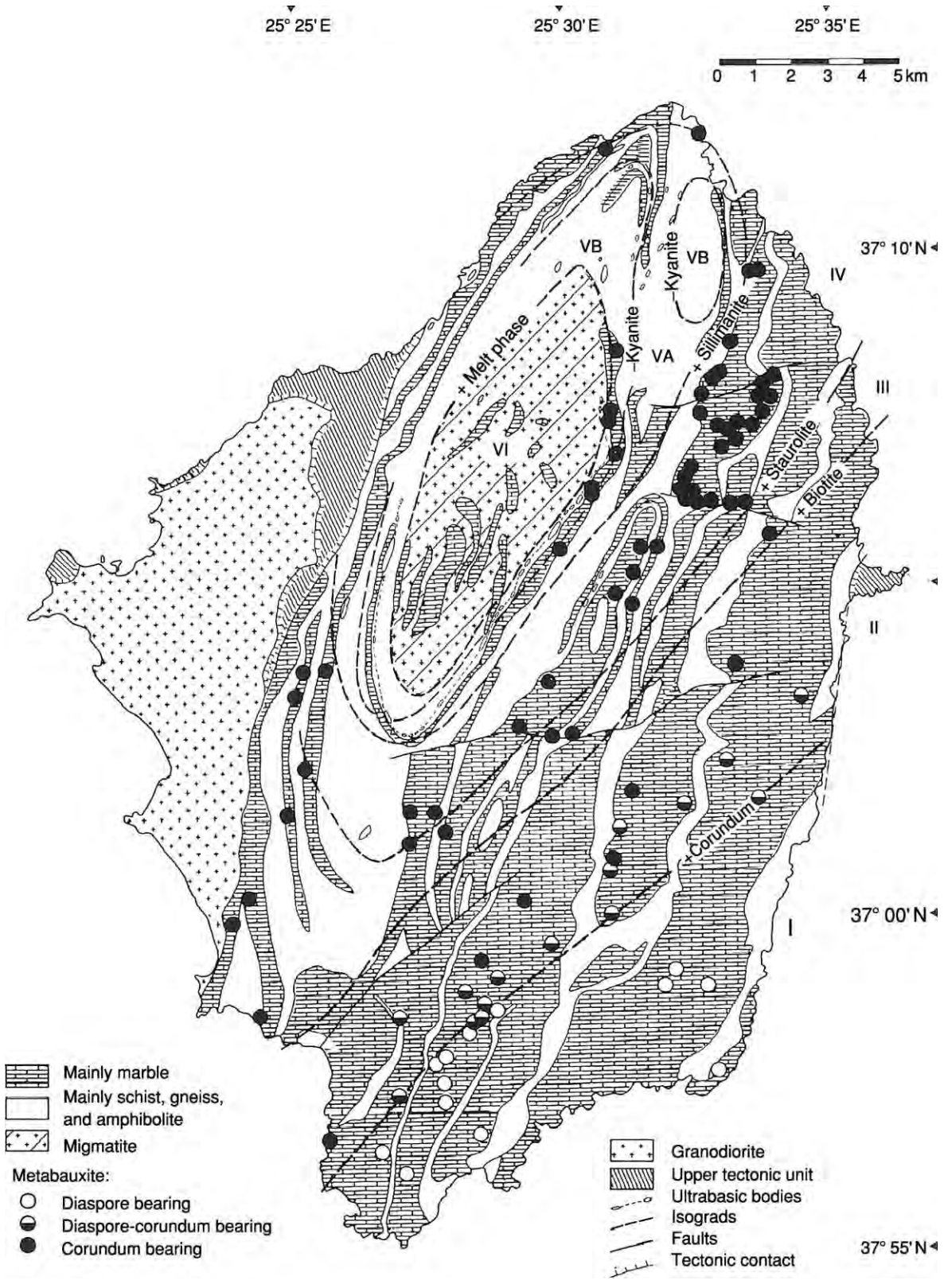


Figure 2 Geological–metamorphic map of Naxos (Cycladic Islands, Greece), showing main lithologies, isograds, and metamorphic zones (I–VI). The following isograds were mapped with increasing grade: corundum-in (metabauxites); biotite-in (metapelites); Fe-rich staurolite-in (metabauxites and metapelites); sillimanite-in (metapelites); kyanite-out (metapelites); melt phase-in (metapelites). (Adapted from Feenstra A (1985).)

Table 1 Mineral assemblages as a function of metamorphic grade in different rock types in the metamorphic complex of Naxos, Greece. (Data from Jansen and Shviling (1976) and Feenstra (1985))

Metamorphic zone	*Pelitic (+Qtz)*	*Mafic*	*Utramafic*	*Carbonate (+Cal ± Dol)*	*Bauxitic*
I	Ab + Ms + Chl ± Ctd ± Pg ± Ep	Ab + Ep + Act + Chl + Ms + Hem	Srp + Chl + Tlc	Ab + Chl + Ep/Zo + Ms + Hem + Qtz	Dsp + Cld + Hem + Rt ± Prl ± Ky ± Ms ± Pg ± Cal
+Corundum					
II	Ab + Ms + Chl ± Ctd ± Pg ± Ep	Ab + Ep + Act + Chl + Ms + Hem ± Bt	Srp + Chl + Tlc ± Phl	Ab + Chl + Ep/Zo + Phl + Ms + Tlc + Hem + Qtz	Crn + Cld + Hem + Rt ± Ky ± Ms ± Pg ± Mrg
+Biotite					
III	Pl + Bt + Ms ± Chl ± Ctd ± Grt ± Pg ± Ep	Pl + Ep + Act + Chl + Bt + Ms + Hem + Mgt	Chl + Tlc + Phl ± Srp	Pl + Chl + Ep/Zo + Phl + Ms + Tlc + Hem + Mgt ± Qtz	Crn + Cld + Hem + Mgt ± Ky ± Ms ± Pg ± Mrg
+Staurolite					
IV	Pl + Bt + Ms + St + Ky + Grt ± Chl ± Ep	Pl + Ep + Hbl + Act + Bt + Mgt ± Chl ± Ms	Chl + Tlc + Ol + Ath + Mgs + Phl	Pl + Ep/Zo + Phl + Tlc + Tr + Mgt ± Grs ± Qtz	Crn + St + Hem/Ilm + Mgt + Ky + Mrg ± An ± Ms ± Bt
+Sillimanite					
VA	Pl + Bt + Ms + St + Ky + Sil + Grt	Pl + Ep + Hbl + Bt + Di + Mgt ± Chl ± Ms	Chl + Tlc + Ol + En + Ath + Phl	Pl + Ep/Zo + Phl + Tr + Di + Mgt ± Grs ± Scp ± Ves	Crn + St + Ilm + Mgt + Ky + An ± Mrg ± Ms ± Bt
−Kyanite					
VB	Pl + Bt + Ms + Kfs + Sil + Grt ± St	Pl + Hbl + Bt + Di + Mgt + Grt ± Ms ± Ep	Ol + En + Spl + Ath + Phl + Di ± Tlc	Pl + Phl + Tr + Di + Mgt ± Grs ± Scp ± Ves ± Ep/Zo	Crn + Spl + Ilm + Mgt + An ± Ms ± Bt
+Melt phase					
VI	Pl + Bt + Kfs + Sil + Grt	Pl + Hbl + Bt + Di + Grt + Mgt	Not present	Pl + Phl + Tr + Di + Mgt ± Grs ± Scp ± Ves	Not present

Mineral abbreviations: Ab, albite; Act, actinolite; An, anorthite; Ath, anthophyllite; Bt, biotite; Cal, calcite; Chl, chlorite; Cld, chloritoid; Crn, corundum; Di, diopside; Dol, dolomite; Dsp, diaspore; En, enstatite; Ep, epidote; Grs, grossular; Grt, garnet; Hbl, hornblende; Hem, hematite; Ilm, ilmenite; Kfs, K-feldspar; Ky, kyanite; Mgs, magnesite; Mgt, magnetite; Mrg, margarite; Ms, muscovite; Ol, olivine; Pg, paragonite; Phl, phlogopite; Pl, plagioclase; Prl, pyrophyllite; Qtz, quartz; Rt, rutile; Scp, scapolite; Sil, sillimanite; Spl, spinel; Srp, serpentine; St, staurolite; Tlc, talc; Tr, tremolite; Ves, vesuvianite; Zo, zoisite.

mineral assemblages of an early Alpine high-pressure/low-temperature metamorphism (about 45 Ma) locally occur in south-eastern Naxos (zones I–II) and that part of the migmatite core (zone VI) is of pre-Alpine origin (from ≈300 Ma).

Temperatures during the Barrovian-type metamorphism are estimated to have ranged from ≈700°C in the migmatite dome to ≈400°C in south-eastern Naxos at pressures of 0.5–0.7 GPa. Six isograds mapped in metapelitic and metabauxitic rocks divide the metamorphic complex into seven Barrovian-type metamorphic zones (Figure 2; Table 1). The first three zones (I–III) belong to the greenschist facies: their subdivision is based on the first occurrence of corundum, by the reaction diaspore = corundum + H_2O (in metabauxites), and of biotite, by the reaction chlorite + muscovite = biotite + H_2O (in metapelites). In mafic rocks, biotite already occurs locally in zone II, whereas the first biotite in bauxitic rocks appears in zone IV.

The presence of staurolite, both in pelitic and bauxitic rocks, indicates the beginning of amphibolite facies conditions. Staurolite formation is related to complex breakdown reactions of chloritoid and approximately coincides with the disappearance of chloritoid. Zones IV, VA, and VB belong to the amphibolite facies; whether zone VI, the beginning of which is defined by partial melting of metapelites, is amphibolite or granulite facies is a matter of debate. The subdivision of the amphibolite facies zones of Naxos is based on the appearance of sillimanite and the disappearance of kyanite in metapelitic rocks. The Naxos rocks clearly illustrate that bulk composition controls the formation of particular minerals or assemblages. In common with other Barrovian terranes, the first kyanite in metapelitic rocks is found in staurolite-grade zone IV, whereas kyanite in bauxitic rocks is already formed in zone I by the reaction pyrophyllite + diaspore = kyanite + H_2O (Table 1).

The upper amphibolite or granulite facies zone VI contains areas where partial melting of SiO_2-rich rocks was intensive (melt phase has a granitic composition) and areas showing little evidence of melting. Variations in the degree of migmatization may relate to local variations in water activity and rock composition. The typical mineral assemblage in metapelitic composition is K-feldspar + sillimanite + biotite. Muscovite often occurs, but textural evidence suggests it is mostly of retrograde origin. Mafic rocks are devoid of orthopyroxene but the presence of sapphirine + spinel + corundum at a few localities in zone VI argues for granulite facies grade conditions.

See Also

Diagenesis, Overview. **Metamorphic Rocks:** Classification, Nomenclature and Formation; Facies and Zones. **Minerals:** Other Silicates. **Plate Tectonics**. **Tectonics:** Mountain Building and Orogeny. **Thermal Metamorphism**. **Ultra High Pressure Metamorphism**.

Further Reading

Barrow G (1893) On an intrusion of muscovite biotite gneiss in the S.E. Highlands of Scotland and its accompanying metamorphism. *Quarterly Journal of the Geological Society, London* 49: 330–358.

Barrow G (1912) On the geology of lower Deeside and the southern highland border. *Proceedings of the Geologist' Association* 23: 268–284.

Bucher K and Frey M (2001) *Petrogenesis of Metamorphic Rocks,* 7th edn. Berlin–Heidelberg–New York: Springer Verlag.

Eskola P (1920) The mineral facies of rocks. *Norsk Geologische Tidsskrift* 6: 143–194.

Feenstra A (1985) *Metamorphism of Bauxites on Naxos, Greece*. PhD thesis, Rijksuniversiteit Utrecht, Geologica Ultraiectina 39.

Jansen JBH and Schuiling RD (1976) Metamorphism on Naxos: Petrology and geothermal gradients. *American Journal of Science* 276: 1225–1253.

Spear FS (1993) *Metamorphic Phase Equilibria and Pressure–Temperature–Time Paths*. Washington DC: Mineralogical Society of America.

Winter JD (2001) *An Introduction to Igneous and Metamorphic Petrology*. Upper Saddle River, NJ: Prentice Hall.

Yardley BWD (1989) *An Introduction to Metamorphic Petrology*. Burnt Mill, Harlow, Essex: Longman Group UK Limited.

REMOTE SENSING

Contents

Active Sensors
GIS
Passive Sensors

Active Sensors

G Wadge, University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The amount of solar radiation reflected from rocks at longer wavelengths (>1 mm) is very small. To properly exploit this microwave part of the electromagnetic spectrum for remote sensing, artificial sources of radiation are used. These artificial, or 'active' techniques allow measurements to be made at night as well as during the day. The longer wavelength of microwave radiation compared to that in the optical part of the spectrum also makes clouds transparent, allowing microwave radiation to 'see' the ground surface in all weathers, day and night.

In radar (Radio Detection and Ranging) remote sensing, a series of timed pulses of microwave energy are transmitted through the air, strike targets (e.g., the ground) and are reflected back to the radar which receives the returned signals. There are parallels between radar and seismic exploration techniques. The principle of 'ranging' is common to all active sensors in remote sensing described here. Because the energy source is under the control of the instrument designer, the frequency and wavelength of the radiation is chosen with a very narrow range so that it can be recognized more easily on reflection and its arrival timed accurately. Knowing the two-way (to the ground and back to the sensor) travel time (t) and the speed of travel through the medium (c), the distance or range (R) to the ground is given by:

$$R = ct/2 \qquad [1]$$

After briefly reviewing the three allied 'active' techniques of radar, lidar, and sonar, attention is focused on radar and the basic principles behind its use and interpretation. There are three main application areas relevant to geology: structural and geomorphological mapping, exploitation of the ability to measure fine surface roughness, and the measurement of surface motions including those associated with earthquakes (*see* **Tectonics:** Earthquakes) and volcanoes (*see* **Volcanoes**).

Radars, Lidars and Sonars

Radars send pulses of microwave radiation (wavelength usually 3–30 cm) from antennas through the atmosphere to the ground, in our case, and receive the return signal. Lidars do the same but with optical wavelength light (800–1000 nm) from lasers. Lidars do not work through cloud, as radars do. Sonars are the oceanographic equivalent to radars that send pulses of energy (1–15 cm wavelength) through the water column to strike the seabed. Figure 1 shows the typical sensing arrangement of the three techniques. Radars have made the most impact on geology.

Lidars have been relatively little used in geological applications to date but the technology is newer. Usually flown from low flying aircraft, they are used as a simple track or swath altimeter with a small footprint. With their high vertical (~10 cm) and horizontal (~1–2 m) resolutions, detailed digital elevation models (DEMs) can be created. They will become increasingly useful to studies requiring detailed geometric knowledge of the terrain, such as process-based simulations.

Sonars have a much longer history of sea-floor mapping. The sensors are towed behind a ship at various levels above the seabed. Absolute positioning is a more involved process than for space-borne sensors. The range information received by the sonar is used to form bathymetric maps with accuracies of a few tens of metres in some cases. As with radar, the magnitude information is a function of the geometrical relationship of the sensor, and the seabed and the surface roughness. Particularly valuable results have been achieved at spreading ridges, accretionary prisms, and in mapping the apron of gravity collapse deposits around oceanic islands.

Radar Techniques

Altimetry

Radar altimeters are the simplest form of radar. They are usually simple ranging devices recording the distance from the radar to the ground, pointing vertically downwards from a satellite. Their footprint on the Earth's surface is large (kilometre scale) and as a result their profiling measurements are only easily interpretable on surfaces that are relatively flat at this length scale (e.g., oceans and ice sheets). Their importance to geology has been that they have provided a measure of the shape of the geoid in the ocean basins that reflects the dynamic processes of plate tectonics (*see* **Europe:** Mediterranean Tectonics): subduction and spreading ridges.

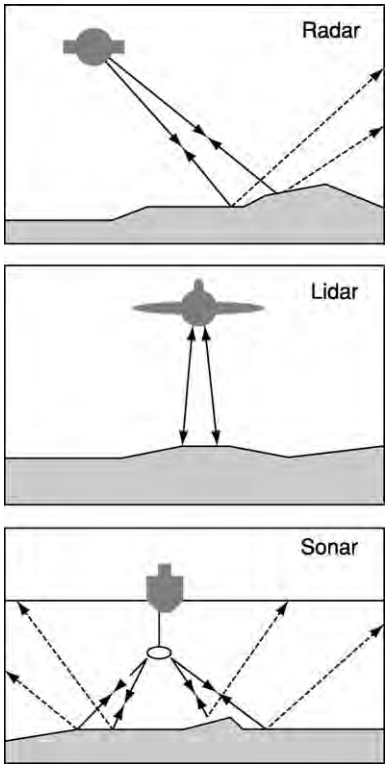


Figure 1 Schematic of the typical operating geometries of a spaceborne SAR, an airborne lidar and a multi-swath sonar.

Doppler Radar

In an active remote sensing device, the engineer can control the rate of pulsing of signals transmitted to the surface and the range of frequencies of radiation transmitted and received. This means that the measurements can remain 'coherent' from one measurement time to another, provided that the external conditions remain the same. Such a coherent radar signal has three measurable quantities: timing, magnitude, and phase. Any relative movement between the radar antenna and the reflecting surface from one pulse to another can be detected as a change in the frequency of return of these pulses – a Doppler shift. The size of this shift allows the relative motion to be calculated. Doppler radars of this type have been designed to measure the dynamics of volcanic ash columns. Lower frequencies of radar signal are more sensitive to larger fragments. Using multiple frequency bands in the same instrument, it is possible to measure the relative motions of different particle size fractions. Because many ash eruption columns are usually obscured by fine turbulent cloud, Doppler radars offer the only practicable way of seeing through this and measuring the internal dynamics of the column.

Imaging Radar

The majority of Earth science applications of radar employ imaging techniques. Table 1 summarizes the characteristics of the main space-borne radar imaging systems to date and those planned in the next decade. In addition there are a number of airborne systems and ground-based systems. Future development of more compact systems and the exploitation

Table 1 Selected orbital SAR systems

	Seasat	*ERS-1/-2*	*JERS-1*	*Radarsat*	*Envisat*	*Alos-Palsar*	*Terrasar-X*	*EVINSAR*[a]
Operation	1978 (3 months)	1991–2000 1995–	1992–1998	1995–	2002–	2004–	2007–	2009–
Country	USA	Europe	Japan	Canada	Europe	Japan	Germany	Europe
Frequency band	L	C	L	C	C	L	X	L
Repeat period (days)	17, 3	35 (1, 3, 176)	44	24	35	44	11	11
Incidence angle	23	23	26–41	20–60	20–50	8–60	20–55	30–46
Swath width (km)	100	100	85	10–500	100–485	70–350	10–100	250
Resolution (m)	25	28	18	10–100	25	10–100	1–16	45

[a]Proposed InSAR mission dedicated to ground motion measurement.

of the millimetre wave part of the spectrum will add considerably to these capabilities.

The motion of a plane or a satellite relative to the Earth's surface can be used to build up a 2D image of the radiation returning from its reflections, or backscatter. Such imaging radars are usually pointed to one side of the platform (Figure 1). This removes the ambiguity of signals returning from equal distances to the left and right of a downward pointing instrument. Most imaging radars are Synthetic Aperture Radars (SARs). In these instruments, the Doppler frequency information from radar returns received ahead of the moving radar (positive shift) relative to those behind (negative shift), are used to 'synthesize' electronically a larger antenna size. This is needed to improve the resolving power of the radar in the along-track or azimuth direction. The details of radar image formation and processing are complicated, but the resulting images essentially contain data on both the magnitude or amplitude and the phase (ϕ) of the returned signal reflected from the ground. Three main factors control these quantities in a radar image: the wavelength of the radar, the slope, orientation and roughness of the reflecting surface, and the dielectric properties of the surface.

The radar image amplitude is generally greater if the surface slope is close to perpendicular to the radar line-of-sight (LOS), which is often several tens of degrees from the vertical. It is also greater if the surface is 'rough' and if it is 'wet' (water content is the main determinant of dielectric behaviour). Within the instantaneous field of view of the radar there may be several individual surfaces or edges (e.g., rock facets) that return much stronger signals than intervening areas. Because the radiation is coherent, these strong signals may reinforce or destroy each other, depending on their relative positions. Depending on the arbitrary location of these strong reflectors on the surface then one radar pixel may have a high amplitude whilst its neighbour, comprising very similar surface material, has a low amplitude. This gives radar amplitude images their 'speckle' character. It also means that, in general, a meaningful measure cannot be retrieved of the surface properties from a single pixel of an amplitude image (unlike a passively acquired optical image). Instead a spatially averaged measure of backscatter amplitude is needed. A similar conclusion applies to the phase information measured. The cumulative sum of phase values from an individual pixel gives a cyclic angular measure ($0 < \phi < 2\pi$) that has no intrinsic significance. However, these values are not noise and the radar should retrieve exactly the same values on re-measurement. These phase data are used by taking temporal differences of phase in the technique of radar interferometry.

Radar Applications

Structural/Geomorphological Mapping

The ability to image large contiguous areas through cloud from airborne radars was first demonstrated in the 1960s. It enabled the first hundred kilometre scale reconnaissance structural mapping of cloud covered areas in the humid tropics. Slopes nearly perpendicular to the radar LOS appear bright in amplitude images and slopes away from this appear dark. This light and shade effect is the same as that seen in optical imagery. However, radar amplitude images suffer from distorting effects that complicate qualitative interpretation. Radar-facing slopes are foreshortened, whilst those facing away are extended. At very steep slopes in alpine-type terrane the foreshortening becomes 'layover' when the tops of peaks appear to overhang their valleys. Most of these distortions can be corrected quantitatively if a DEM is available. Suitably detailed surveys of several areas, particularly young fold and thrust belts in cloudy regions, have proven genuine exploration aids. The sensitivity of a radar amplitude image to a fault-controlled topography or any other geomorphological features is always a function of the angular relationship between the radar LOS and the feature on the surface. Features parallel to the LOS are hard to see. Airborne surveys can be designed with this in mind. The erosional morphology of different lithological terrains is also well expressed in radar amplitude imagery, permitting lithological discrimination (Figure 2). This is, of course, scale dependent, particularly where there is tree cover obscuring ground features less than about 100 m in size. With the increasing availability of high resolution DEMs for much of the globe then the value of imaging SAR at this reconnaissance scale is being superceded.

The most remarkable radar mapping survey yet undertaken was that performed by NASA's Magellan mission to Venus (*see* **Solar System:** Venus) from 1990 to 1992. Radar had to be used to penetrate the thick Venusian atmosphere. In addition to the huge scope of the global mapping undertaken, a lot of the geomorphological features seen had no obvious terrestrial analogues and hence stretched the interpretational skills of the Magellan science team.

Roughness Mapping

Radar backscatter is sensitive to the roughness of surfaces at length scales much less than the wavelength of the radar. For example, the C-band (wavelength = 5.8 cm) radar of the ERS/ENVISAT satellites are sensitive to surface roughness at the scale of millimetres. Thus, any changes in roughness in space or time that are relevant to geology can be utilized.

Figure 2 Radar amplitude image of part of north central Jamaica viewed by the SEASAT L-band radar in 1978. The radar illumination is from the left (west south-west) of the image at an incidence angle of about 20 degrees from vertical. Most of the scene is covered by ''dimpled'' pattern typical of tropical karst topography of Cenozoic limestones, except for the bottom left corner which is an inlier of Cretaceous volcanic and volcaniclastic rocks. In the northernmost central part of the image a straight normal fault scarp downthrows the limestones to the north. This is cut by one of two, subparallel curving fault scarps. The easternmost one shows the clear pattern of a scissor fault. To the south the fault shows a west-facing escarpment that is bright in the image and futher north this changes to an east-facing escarpment in shadow. The image measures about 36×24 km.

Satellite radar imagery can be used for offshore oil basin reconnaissance using this technique. Oil seeping from depth forms a film on the sea surface that dampens the backscatter roughness relative to surrounding areas. These areas of surface oil appear darker in amplitude images. Hydrocarbon-rich basins often have a variety of discrete oil seeps that help define the basin extent and character. These seeps can be detected by radar roughness mapping. Shorter wavelength radars (e.g., X-band, 3 cm) give a stronger roughness contrast than longer wavelength radars. The ambient sea-state also plays a role: contrast is enhanced by moderate wave formation and by viewing from perpendicular to the wave motion. In order to distinguish between natural seeps and ship pollution it is necessary to build up a time series of observations that establish the semi-permanent nature of seep locations.

Of course, most rock surfaces do not change their roughness over periods relevant to remote sensing. One environment that does, however, is arid playas. Here the evaporation of groundwater leads to the surface formation of evaporite minerals, particularly gypsum and halite. During the summer months, when evaporation is highest and rainfall negligible, the growth of these minerals, specifically halite, can produce a surface that becomes increasingly rough. The C-band radar of ERS is sensitive to this seasonal roughening of the surface of arid playas such that, using a time series of ERS radar amplitude images at 35-day intervals, it is possible to monitor quantitatively the evaporation rate from the backscatter values using models of roughening. Confidence in this approach only comes when the region is very flat and extensive, so that slope effects can be discounted.

Radar is very effective generally in deserts (*see* **Sedimentary Environments:** Deserts), because it can penetrate dry, unconsolidated materials. Where sheets of windblown metres-thick dry sand cover eroded surfaces in the Sahara, it is possible to image these buried surfaces and their subtle differences in roughness beneath their cover, even from space. This is best done with the long wavelength (24 cm) L-band radar which can penetrate several metres into the extremely dry sand by a process known as volume scattering. On reaching a buried rock surface, the radar signal it is reflected back through the sand. Shallow former river channels, eroded in wetter periods, form smoother reflectors relative to their surrounding eroded interfluves, revealing patterns totally hidden beneath their aeolian cover.

Ground Motion Measurement

The phase information from imaging radar can be used to measure the relative spatial motion of the ground surface between two separate images. This technique uses the principle of interferometric synthetic aperture radar (InSAR) and has made a major impact on the Earth sciences in the past ten years, since its practical value was first demonstrated using data from the ERS satellites. It relies on the ability to measure the change in phase at each pixel between two images. If the surface of part of the image moves towards or away from the radar LOS more than in another part of the image, then a phase gradient is measurable (Figure 3). Without independent control on motion, the technique known as differential InSAR only gives relative, within scene, motion-detection. The motion sensitivity is proportional to $4\pi/\lambda$, where λ is the wavelength. Typically for InSAR, using ERS data, LOS accuracies of about 1 cm can be achieved. After each 2π of phase change the phase wraps around and the interferogram image displays a fringe of interference. To retrieve a monotonic motion signal the

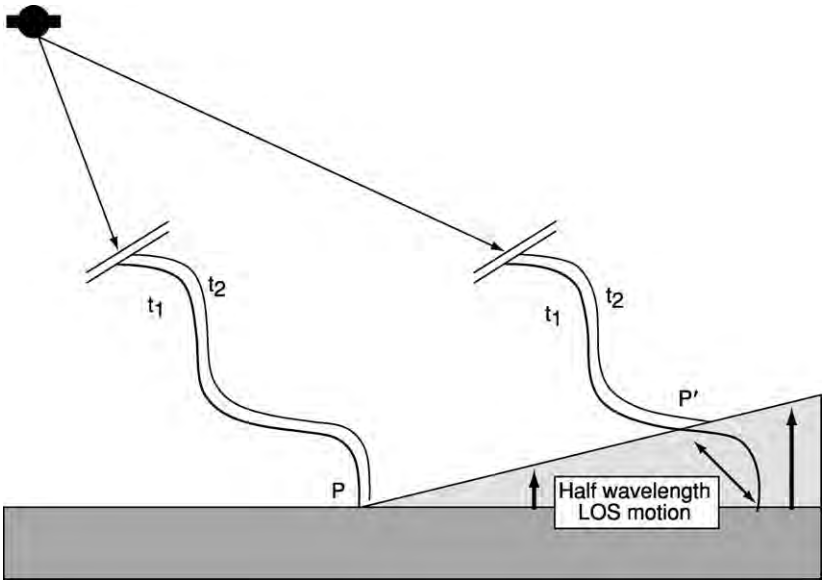


Figure 3 Schematic of the principle of differential InSAR. A radar images the earth's surface at time t_1 as represented by the thick cyclic lines reflecting from the surface at two places. After an interval the radar images the scene again, at time t_2, from the same orbit. During this time the ground surface has moved differentially as shown by the lighter shaded triangle such that there is now a difference in path length along the LOS of the radar of one half a wavelength of the radar between positions P and P′ on the ground. This corresponds to a gradient of 2π in phase, or one fringe in the interferogram made from the difference in phase of the two radar images.

phase fringing must be 'unwrapped' using automated algorithms.

InSAR has opened up a new form of space geodesy at the same time that GPS measurements of surface motion have become commonplace. The two techniques are largely complementary, InSAR has a dense spatial, but sparse temporal sampling, opposite to that of continuous GPS. InSAR also has the potential to work anywhere, without the need for ground instrumentation. However, there are a number of complicating factors that inhibit InSAR's practical usefulness at present. The technique requires the surface scatterers to remain unchanged between the two images of the interferogram. Vegetation change is a major inhibitor here at X- and C-band. Variable water vapour distribution in the troposphere produces unwanted path effects. Also, remarkably, none of the current space-borne radars were designed with InSAR in mind and so a lot of the radar parameters are not optimised for InSAR. Potentially, InSAR has a role to play in any Earth process that involves the motion of the Earth's surface at length scales of 100 m to 1000 km and at strain rates of 10^{-5}–$10^{-7}\,\mathrm{yr}^{-1}$. The Earth science fields that have made most use of it are the study of earthquakes and tectonics, volcanology, and the study of shallow subsidence processes.

Earthquakes and Tectonics

Seismology has dominated the study of earthquakes. In recent years, however, the ability to measure the surface ground motion field associated with fault rupture by InSAR has brought a new dimension to the study of shallow earthquakes. By inverting the co-seismic motion field of InSAR with numerical models of the fault plane motion in a 3D space, knowledge of the earthquake fault mechanism can be improved. Specifically, InSAR can greatly improve knowledge of exactly where the fault plane rupture was and estimates of its dip, whilst seismology is better at resolving the moment released. Immediately following a large earthquake the ground surface can move more slowly over the following months. This can be measured by InSAR to help constrain the physical processes potentially involved: afterslip on the fault plane, poroelastic pressure changes or viscoelastic relaxation in the lower crust/mantle. The other component of the earthquake cycle, the interseismic strain, can also be detected by InSAR, for example, on the San Andreas and North Anatolian Faults. This requires pushing the technique to its limits of sensitivity in measuring very slow strain buildup over several years across widths of about 100 km.

Not all Earth strain involves seismicity and sudden displacements. InSAR is starting to reveal a wealth of surface strain events in the continental plate boundary zones that whilst linked to fault systems, do not themselves display significant seismicity. Aseismic displacement events triggered by nearby seismic events has been measured and shallow folding deformation may be revealed in low competence

rocks. This raises the exciting prospect of being able to relate the current day surface deformation to neotectonic processes and geomorphological features of the past few hundred to thousands of years.

Volcanoes

Earthquakes and volcanoes can share a similar cycle of surface straining associated with the repeated charging of their systems by external forcing (plate motion and magma buoyancy) (*see* **Volcanoes**). In the volcanic case, dilational strain during charging of the magma reservoir is released suddenly, perhaps with shallower planar rupture deformation (dyking and surface fissuring). InSAR is having an impact on our knowledge of the volcano cycle as it has for earthquakes. In the volcano case, we also have the possibility of using the volume of erupted magma during an eruption to constrain the magma loss event in the reservoir. However, the complementarity of InSAR and seismicity in defining the source mechanism of earthquakes is not available in volcanoes. Because of this and the inherently more complex three-dimensional shape of pressured magma reservoirs, progress in refining their location and dynamics has been slower.

Many volcanoes are susceptible to gravitational instability and in particular to spreading under their own load if there is a suitable weak basal layer. This has been measured by InSAR. A complex mixture of spreading forces, local fault accommodation of these, regional tectonic movements, and pressurizing magma reservoirs mean that the INSAR-measured deformation at many volcanoes may reflect a much more complex system than hitherto realized (Figure 4).

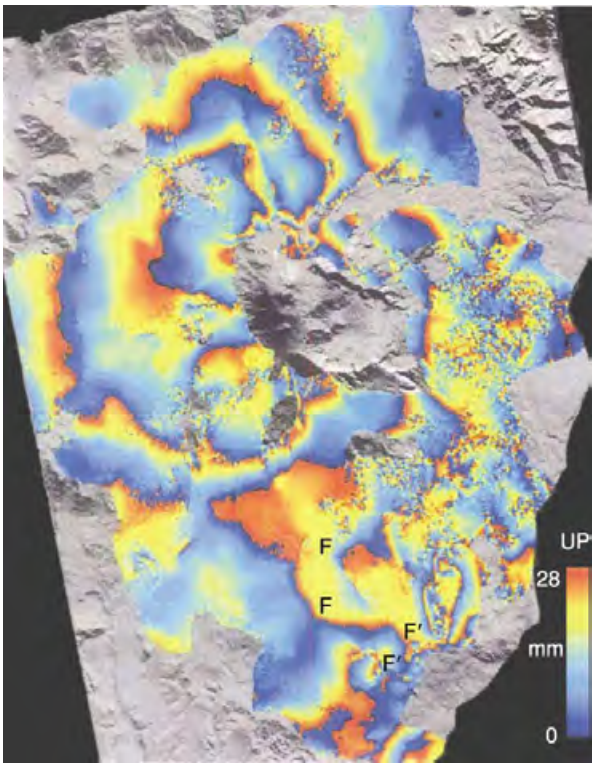


Figure 4 A differential radar interferogram of the Etna volcano in Sicily created using two ERS SAR images from 1995 and 2000. Each colour cycle from blue to red represents one fringe of relative motion of 28 mm. The parts of the scene where the phase data are too noisy are replaced by a shaded relief image of the topography. The displacement pattern of the ground surface shows apparent LOS uplift of >10 cm of the summit region of the volcano relative to the flanks. Also indicated by the F-F′ symbols is evidence of relative movement along two faults on the southern flanks. The image is about 40 km wide. The ERS data were supplied by the European Space Agency.

Subsidence

There are a number of natural and anthropogenic or human-induced mechanisms that can produce shallow and hence local subsidence of the Earth's surface: for example, soluble rock dissolution, landslides, mining, groundwater/oil extraction, and tunnelling. InSAR has been shown to be valuable in monitoring all of these. Because of their small spatial extent, it is easier to calibrate the absolute motion.

Local subsidence events are of considerable concern in urban environments and a variant of the InSAR technique has been developed to take particular advantage of this. Buildings and roofs of a variety of shapes and orientations serendipitously provide a source of very strong radar scatterers. In the 'permanent scatterer' technique of InSAR, the phase change information from these scatterers alone is used in a time series of images (typically at least 20–30). By making some assumptions about the nature of the atmospheric effects and with good knowledge of the topography, these sources of error can be removed and a line of sight (LOS) motion history of the set of scatterers can be measured to accuracies of a few millimetres.

See Also

Engineering Geology: Natural and Anthropogenic Geohazards. **Europe:** Mediterranean Tectonics. **Remote Sensing:** Passive Sensors. **Sedimentary Environments:** Deserts. **Solar System:** Venus. **Tectonics:** Earthquakes. **Volcanoes**.

Further Reading

AGU (1992) *Magellan at Venus*. reprinted from Journal of Geophysical Research 97 (E8, E10), Washington, DC: American Geophysical Union.

Amelung F, Jonsson S, Zebker H, and Segall P (2000) Widespread uplift and 'trapdoor' faulting on Galapagos volcanoes observed with radar interferometry. *Nature* 407: 993–996.

Blondel P and Murton BJ (1998) *Handbook of Seafloor Sonar Imagery.* Chichester: J Wiley and Sons.

Elachi C, Roth LE, and Schaber GG (1984) Spaceborne radar subsurface imaging in hyperarid regions. *IEEE Transactions Geoscience and Remote Sensing* GE-22: 383–388.

Ferretti A, Prati C, and Rocca F (2001) Permanent scatterers in SAR interferometry. *IEEE Transactions Geoscience and Remote Sensing* 39: 8–20.

Hanssen RF (2001) *Radar interferometry. Data Interpretation and Error Analysis.* Dordrecht: Kluwer Academic Publishers.

Henderson FM and Lewis AJ (1998) *Principles and Applications of Imaging Radar.* Manual of Remote Sensing, 3rd edn. Vol. 2. New York: J Wiley and Sons.

Koopmans BN (1983) Side-looking radar, a tool for geological surveys. *Remote Sensing Review* 1: 19–69.

Massonnet D, Rossi M, Carmona C, *et al.* (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364: 138–142.

Massonnet D and Feigl KL (1998) Radar interferometry and its application to changes in the earth's surface. *Reviews of Geophysics* 36: 441–500.

Trevett JW (1986) *Imaging Radar for Resources Surveys.* New York: Chapman and Hall.

Ulaby FT, Moore RK, and Fung AK (1986) *Microwave Remote Sensing*, vols. I, II and III. Norwood MA: Artech House.

Wadge G and Archer DJ (2003) Evaporation of groundwater from arid playas measured by C-band SAR. *IEEE Transactions Geoscience and Remote Sensing* 41: 1641–1650.

Wehr A and Lohr U (1999) Airborne laser scanning – an introduction and overview. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 68–82.

GIS

P J Mason, HME Partnership, Romford, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Geological phenomena vary in space and time and this basic trait makes geology a particularly suitable application field for Geographical Information Systems (GIS). A GIS is a computer-based tool for the mapping and analysis of things that exist on, and events that happen on Earth. It is, therefore, a sophisticated but general-purpose tool and one which can be applied in many ways to many problems, and especially so in geosciences.

There have been a number of technological advances, including the development of GIS, which have transformed the work of the geoscientist over the last 15–20 years. Until the 1980s, GIS was used and developed mainly by the geographical community and generally in government organizations and research institutions. The wider geological community has been, until relatively recently, a little sceptical about its operational success in solving complex geological problems.

The first GIS systems were pretty cumbersome hardware and software suites, with high-end functionality that required programming skills to operate. These have evolved into highly functional desktop systems, which anyone can learn to use, and which are part of a growing, multibillion dollar, worldwide industry. GIS functionality developed very quickly, with the advanced statistical and 3D analytical modules, which had been considered requirements only of the advanced GIS community, becoming the selling points for desktop GIS. Multivariate and geo-statistical methods laid the foundations of the geological application of GIS. Specialized exploration software, in the late 1970s and 1980s, had many features present in current GIS suites, such as those of Zycor, Z-Map Plus, Strat Works, GeoData Works, Earth Cube, and StratSoft's Statistica. The costs of these commercial software packages were generally prohibitive and the wider geological community continued to rely on the more readily and cheaply available geo-statistical packages, many of which were available in the public domain.

The technological developments in GIS were accompanied by, and to a large extent enabled by, huge leaps in computer power and development. Large mainframe machines gave way very rapidly to desktop, laptop, and now palmtop processing. Software evolved in parallel with these changes, and GIS mapping can now be done in your hand, in the field, enabling great savings in time and money. The growth of the World Wide Web (WWW) and wireless communications has enabled the generation of, and access to, accurate and timely information, in addition to fast and reliable transport of information from office to field and vice versa.

The end of GPS selective availability in 2000 and the reduced cost and increased availability of commercial satellite imagery, also make 'remote' digital mapping cheaper, more accurate and more accessible. One milestone in this part of the 'geospatial revolution' was the successful launch of the Earth observation satellite, Landsat 7 in 1997, whose data were

commercially available at unprecedented low prices and, for the first time ever, were browsable on the Internet only a short time after acquisition. The launch of ASTER on TERRA AM-1, in 1999, has broadened geological mapping capabilities and reduced costs even further. Sadly, Landsat 7 is no longer fully functional but ASTER is and this series of satellites is likely to evolve further.

This article not intended to provide an exhaustive technical explanation of what GIS is, how it functions or of the technical details of various software suites. Rather, it aims to give an overview of the state of the art in 'geo-GIS', i.e., those aspects of GIS which are specifically relevant to geoscientists and their activities. The Further Reading section at the end of this article includes a number of standard texts on geographical information systems and science, and on supporting subject matter, in addition to some rather more specific publications of interest to the geoscientist working with GIS.

The Geological Information System

Most geologists would agree that a geological map could be considered a database in its own right, as the information contained on the map and in the legend is, by definition, spatially referenced and attributed. To realize the full potential of that database, the map must be properly converted, and it is at this point that the power and tools of GIS come into play. They enable far more to be gained from the original map than that which first meets the eye. Fundamental to the conversion are the mapped spatial relationships, which are interpreted by the GIS through topological rules. Successful digital conversion requires the understanding of the geological and implied topological relationships described by the map.

Components of a geological map are stored as records, or objects, in a spatially referenced GIS database, which can be displayed spatially, as a map, or in graphical or tabular form (as attribute tables). Standard querying enables selection and extraction and/or the linking of related attribute tables.

Attributes, i.e., any kind of descriptive information, can be attached to a record or object in the database. Geological attributes, attached to a map object can be likened to the association of multi-element geochemical data to a single sample locality. GIS allows this attribute information to be viewed and edited in tabular form, where each object represents a record, and each column represents a field containing a different attribute. The objects can then be reclassified, according to the various attributes, and displayed in map form, revealing patterns and relationships not always detectable in tabular form.

Spatial Data, Models and, Structures

There are two basic methods of representing information in a GIS – raster and vector. Geological maps and data can be displayed using either method, but if a layer is to act as a database, it must be properly converted into or captured in vector format.

i. The raster or 'grid'
   The raster displays spatial information as a series of contiguous cells (or pixels) of uniform size, and each raster grid represents the variance within one attribute. This method of storage and display of information is appropriate for data that has high spatial variability, such as geophysical survey data or satellite images. As the raster grid (or lattice) is regular, spatial relationships are implicit and therefore, explicit storage of spatial (and topological) relationships is not necessary, as it is for the vector data model.
ii. Spatial or vector data comprises any or all of the following basic components:
   - Lines or arcs (e.g., representing contacts and boundaries)
   - Areas or polygons (e.g., representing closed areas and regions defined by surrounding lines)
   - Symbols, points or nodes (e.g., such as a sample point, borehole, or a strike and dip measurement at a specific location).

To correctly display a map in vector form, certain information about the geometric and spatial relationships of the map objects is required and this is where topological rules become necessary. For example, the GIS needs to be able to determine whether a fault is connected to another fault, or where one outcrop borders another. Figure 1 illustrates the appearance of a simple digital geological map in vector and raster form, and the effect of varying raster resolution, on map detail.

There are advantages and disadvantages to both methods, but the fundamental requirement in both cases is that linear features (e.g., boundaries, contacts, and faults) are digitally captured only once. These linear features are then used to construct the areas or polygon objects (e.g., lithological outcrops) that comprise the map. The result is that all adjacent or connected objects have topological connectivity and contiguity, so that spatial relationships between them are understood by the GIS. Topologically correct data capture and database construction should be done from the outset.

The Open GIS Consortium, Inc. (OGC) leads in the development of open, interoperable, geo-processing computing standards and specifications. OGC is involved with government, private industry, and

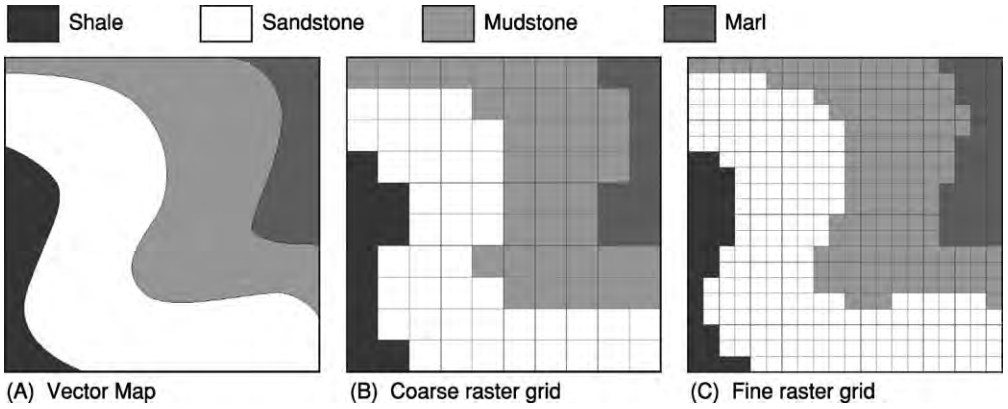


Figure 1 Digital representation of a simple digital geological map; (A) vector or spatial data map; (B) coarse raster grid; and (C) fine raster grid (illustrating the effect of spatial resolution or grid cell size on the detail stored in the representation).

academia, and its adopted GIS standards and specifications provide a valuable resource; they are available at the OGC's website (http://www.opengis.org/about/).

Database Design, Quality, Precision

The design and structure of the GIS database being developed should be governed by the application, and by the intended use of the database. The quality, accuracy, precision, and relevance of the captured data are all significant considerations. It has been said many times that the GIS analysis is only as 'good' as the information that is input.

Georeferencing

To make GIS work, there must be techniques for assigning 'coordinates' of time, location, and attribute in a way that is commonly understood. Basic tabulation, and calendar and time systems are not a problem, so the recording of location is the most vital reference information stored in GIS. Some georeferences are based on names, others on measurements, which are sometimes referred to as 'metric' georeferences. They include latitude and longitude and various other kinds of projected coordinate system, and provide infinitely fine resolution and enable accurate measuring capabilities.

Georeference, geolocation, georegistration, and geocoding are all common terms for the recording of an object's position and shape on a map, in the form of a map projection and datum. The need for map projections is fundamental – the world is a three-dimensional object, roughly oblately spherical in shape, which needs to be represented on a two-dimensional piece of paper or screen. We also need to measure positions, distances, and heights, all of which would be arbitrary and relative, if not made with reference to some known reference datum or ellipsoid. Many coordinate systems, projections, datums, spheroids, and ellipsoids have been developed over the years for various regions of the world and for the whole world. Integrating multisource, multiform data into the GIS presents the fundamental need to ensure coordinate reference system conformality. This can be time consuming but most GIS software contains tools for transforming datasets from one system to another. There are also some very comprehensive reference texts, and now Internet resources, to help deal with this issue.

Visualization

GIS visualization tools (two- and three-dimensional) within GIS enable and rely on the ability to share and integrate multisource data, models, and interpretations. The simplest form of visualization involves the display of two-dimensional images with conventional cartographic symbols. The geological map can be recreated using conventional geological symbology, which can easily be incorporated within the GIS, through the use of special fonts. Thus structural information can be presented to give the appearance of a published geological map (Figure 2).

The ability of GIS to bridge the gap between two- and three-dimensional display and analysis is especially useful in geosciences, because depth and 3D structure are fundamental considerations. The ability to integrate surface and sub-surface data is therefore an essential function of any digital mapping technology in geoscience.

GIS enables exploration and interrogation in several dimensions for improved and enhanced understanding of structures and relationships. It is possible to conduct virtual geological field visits, which can assist in the logistics and improve understanding prior to setting foot in the field, thus saving valuable time. Most GIS software provide such visualization tools

Figure 2 Data capture and conversion from a simple field sketch map (left) to the visualization of a digital geological map (right) within GIS.

for the manipulation of surfaces, images, and maps in pseudo three-dimensional space, and the mathematical derivation of other products, such as slope angle, aspect or azimuth, curvature, line-of-sight, watersheds, and catchments. Examples of software providing excellent 3D manipulation and viewing capabilities are ERDAS Imagine's Virtual GIS, ER Mapper, ArcGIS 3D Analyst, and Geomatica Fly.

Visualization in pseudo-3D requires a valuable commodity – a digital representation of a surface, whether it is topography or other feature data, such as an interpreted seismic horizon or geophysical survey data. The attribute information in this 'surface' layer provides the third dimension or 'z' value for the display.

There are many ways to obtain and derive digital surface topographic data or DEMs. Many datasets can now be browsed and obtained on the Internet from organizations such as the United States Geological Survey and the National Imagery and Mapping Agency (NIMA). In February 2000, the Shuttle Radar Topographic Mapping Mission (SRTM) carried out radar interferometric mapping of the Earth between +60 and −60 degrees latitude, in fewer than 11 days. SRTM is an international project, headed by NIMA and the National Aeronautics and Space Administration (NASA), with the objective of generating global DEMs. DEM data, gridded at 90 m are now available for much of the world, and at 30 m for the USA only, and are downloadable, free of charge, from ftp://edcsgs9.cr.usgs/pub/data/srtm. Sophisticated software exists (such as PCI Geomatica, Socket Set, ENVI ASTER module, Virtuozo) to generate DEMs from stereo airborne and space-borne photography and imagery. One such source is Terra-1 ASTER, whose stereo capability means that high-quality DEMs can be readily generated at very low costs and at spatial resolutions suitable for 1:50 000 scale mapping (http://asterweb. jpl.nasa.gov/).

Geological Application Fields

The application areas to which GIS makes a significant contribution are as varied as they are numerous. In geosciences there are some key areas where the necessity for and contribution of GIS are clear and significant. They range from the basic requirement of producing maps through the solving of complex geological challenges, to advanced spatial analysis.

Field Mapping

The most basic of activities for the geoscientist, mapping, has been changed fundamentally through the use of GIS. Though the hammer may remain ubiquitous, weatherworn field slips and notes are being replaced. Field mapping essentials now include laptop and palmtop computers (**Figure 3**), connected to GPS hand sets (and mobile telephones) so that positional data can be fed directly and dynamically into the GIS, to record current positions for waypoints or sample localities. Alternatively, tracks can be recorded as you walk along, so that you can re-trace your steps at any time, and boundaries and contacts can be walked out to provide direct mapping to the GIS database. Even sketches and notes can be immediately input through palmtop graphical interfaces and handwriting transcribers.

These processes are commonly facilitated by Earth observation image data (from both satellite and airborne platforms), which are easily integrated in the

Figure 3 GIS in the field: (A) Staff and students record their observations on a digital map, in real-time, using a GPS handset, GIS and laptop, during fieldwork in Spain; (B) Palm-top computer, displaying a geo-referenced Landsat 7 satellite image being used as a base map, within GIS, for field geological mapping.

GIS. They form ideal basemaps and provide hard evidence of Earth surface features and changes.

Natural Resources

A necessary and important area of growth for GIS is the promotion and employment of GIS in developing countries for the management of all natural resources. One of the most important global challenges, which will affect both developed and developing worlds, is water resource management. GIS can be used for almost all applications related to water resource management, such as groundwater targeting, resource estimation, and groundwater recharge estimation. Groundwater risk assessment can also be carried out; GIS enables mapping and analysis of contaminant plumes (caused by oil, chemical, and radioactive pollutants, for example), ground water vulnerability assessment, and environmental impact evaluation. The United States Geological Survey provides useful web based information on water resources and hydrological data (http://water.usgs.gov/maps.html).

Exploration and Production

The identification of potential new exploration targets is vital for exploration mapping and management and the use of GIS in mineral and petroleum exploration is now widespread. One of GIS's most useful contributions in this context is the capacity for integration of disparate digital datasets into a single database. All available data, geological, geochemical, and geophysical, can be compiled, analysed, and interpreted, to produce exploration potential maps and images (such as in **Figure 4**).

In quantifying prospectivity, through spatial analysis, the development of an analytical model is as important as the quality of the data; identifying significant factors, evaluating their significance, and establishing relevant weights for the layers used in the model are all fundamental considerations. The resultant prospectivity map reflects the depth and accuracy of the input information, the relevance of the model, and the geoscientists experience and local knowledge.

Environment and Geohazards

Environmental Quality

All aspects of environmental monitoring and management are particularly suited to the application of GIS. Using digital terrain models, geochemical and hydrological data, pathways for pollutants can be assessed and the probability of contamination calculated for areas where no sampling has been performed. The Interdepartmental Committee on the Redevelopment of Contaminated Land (ICRCL) publishes guidelines and trigger concentration values, based on appropriate usage in, for example, the UK, on the Internet (http://www.contaminatedland.co.uk/std-guid/icrcl-l.htm). Such values provide an ideal starting point for establishing 'decision rules' within the GIS model. Analytical tools, found in any GIS, enable generation of basic mapping, site selection, and advanced spatial analysis for environmental impact assessments, environmental audits and risk assessment, in addition to soil and groundwater investigation, remediation, and quality control.

Geohazards

The term geohazard is used in reference to those potentially dangerous phenomena that are caused by natural, geological processes, such as landslides, ground subsidence earthquakes, and volcanic eruptions. GIS has an important role to play in understanding and quantifying the spatial variability of the phenomena, in predicting its future behaviour,

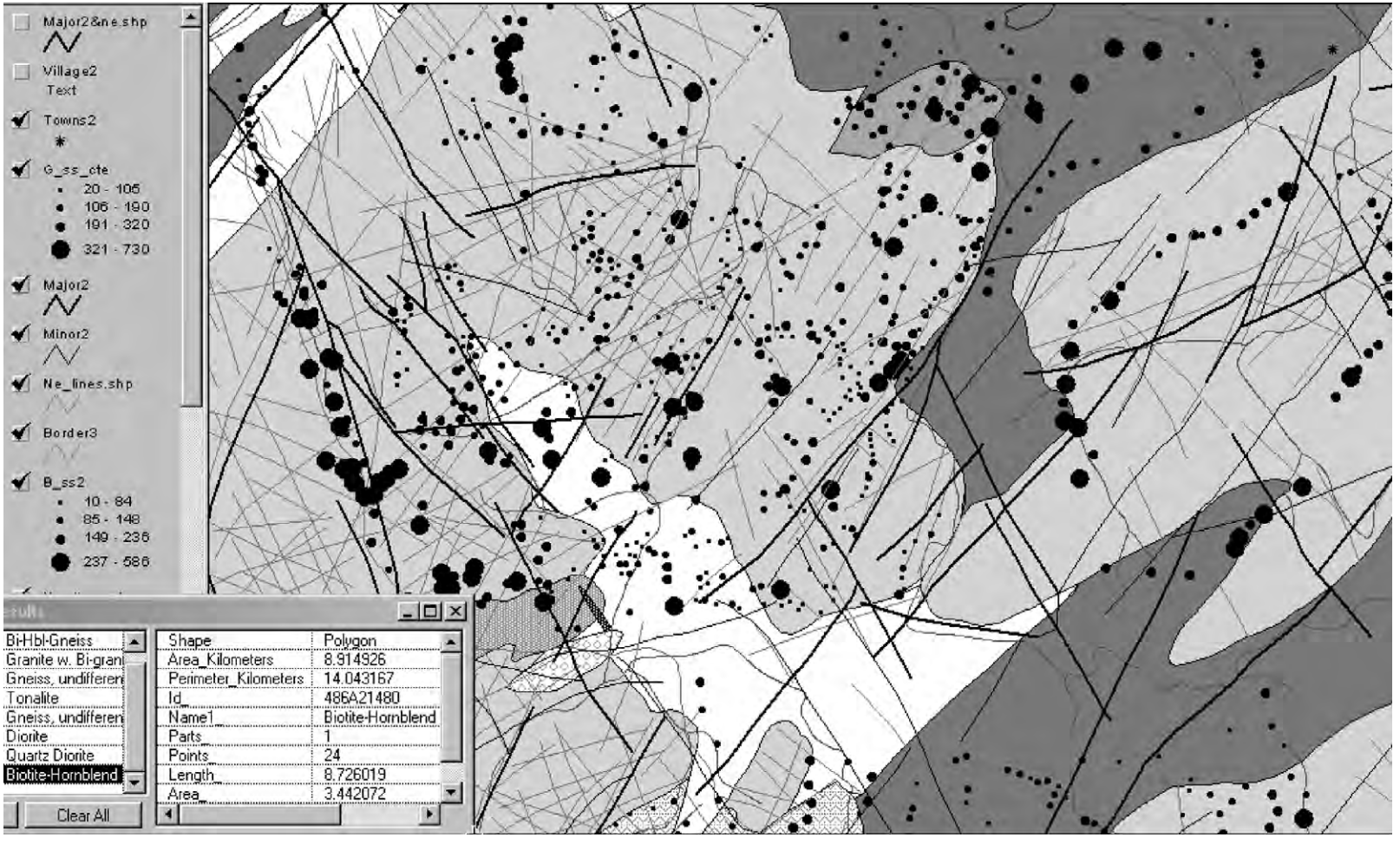


Figure 4 Display of vector polygon (lithological), line (structural features) and point (geochemical sample) data in an exploration GIS database.

and frequently in helping to identify appropriate remediatory activities.

The Integrated Global Observing Strategy (IGOS) is a strategic partnership linking research, and long-term monitoring and management with data users and suppliers, for global environmental observations of the atmosphere, oceans, and land. Its themes include all the main geohazards (earthquakes, volcanoes, subsidence, and landslides) and one of its key objectives is "to develop and integrate Earth observation data, *in-situ* spatial data organisation, spatio-temporal data modelling and analysis, using GIS, for a global, integrated geohazard monitoring system". Information about IGOS, its contributing partners and its goals, can be found on its website at http://www.igospartners.org.

Landslides Landslide hazard assessment involves the classification of the land surface into areas according to the degree of potential hazard posed by mass movements; an example of a landslide hazard assessment map is shown in Figure 5. Many methods and techniques have been proposed and used to quantify causative factors and GIS is commonly used to produce maps representing the probability of occurrences, on the basis of occurrences in the past. This 'direct' method of hazard assessment consists of geomorphological mapping of past and present landslides, and identification of factors leading to instability. GIS is then used for classification, or zonation, to reveal sites where future failures are most probable. The alternative 'indirect' method includes two different approaches, namely the didactic or statistical (data driven) and heuristic (knowledge driven) techniques. The didactic approach involves construction of an abstract model and proposal of hypotheses, followed by experiments and data collection, testing of the hypotheses and finally the construction of landslide instability rules. In the heuristic approach, landslide-influencing factors, such as slope angle, lithology, landform, and land-use, are ranked and weighted according to their expected significance in causing slope instability, and the success of the result is tested against known cases before rules are constructed (*see* **Sedimentary Processes:** Landslides).

Vulcanology GIS contributes many tools to manage field data (such as steam, water, and soil geochemistry), enabling patterns of output and geochemical changes to be identified effectively, and presented graphically. Another important role for GIS is in the modelling and prediction of the behaviour of ejected materials and mass movements after eruptions. Using a Digital Elevation Model (DEM) images, and maps, the existing topography can be analysed, enabling paths, discharge rates, and velocities of potential

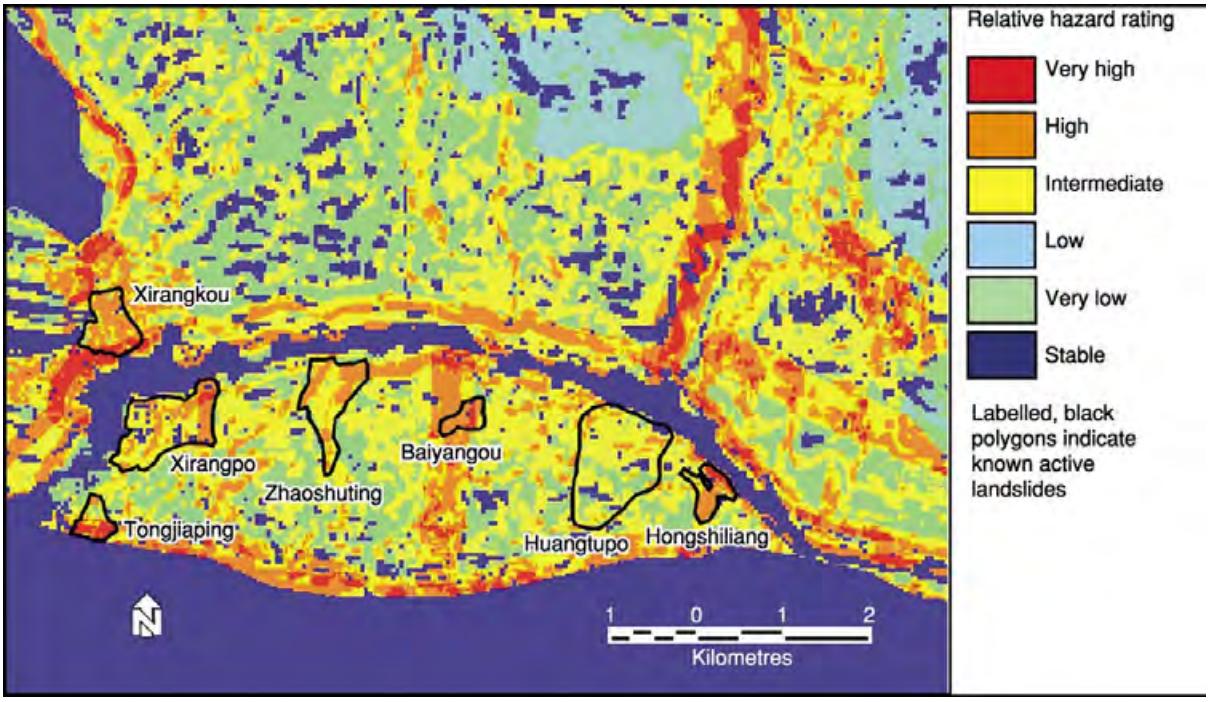


Figure 5 Map showing relative landslide hazard levels as derived from a multi-criteria elimination and characterization model, of an area of the Three Gorges region in China. The locations of the landslides, which are known to be active in this area, are indicated by black polygons.

lahars and mudflows to be predicted. Pre- and post-eruption terrain modelling also allows the monitoring and detection of landform changes, and the preparation of susceptibility maps (*see* **Volcanoes**).

Earthquakes Seismic hazards caused by ground motion and failure are characterized by their spectral response, and by peak ground acceleration and velocity. The spatial distribution and patterns of ground motion can be determined using deterministic and probabilistic analysis within GIS. For an event of a certain magnitude, soil conditions and attenuation relationships can be used to predict the amount and distribution of ground shaking. Another significant role for GIS is the modelling and simulation of motion intensity and damage under various scenarios (*see* **Tectonics:** Earthquakes).

An example of the operational use of GIS in this field is the HAZUS programme (HAZards in the United States), under the control of the Federal Emergency Management Agency (FEMA), which has been developed to estimate losses caused by earthquakes, hurricane and wind erosion, and flooding. HAZUS takes the form of a software 'plugin' that exploits functionality within ArcView or MapInfo GIS packages. A new, HAZUS-MH (Multi-Hazard) version is due for release and more information can be obtained from http://www.hazus.org and http://www.fema.gov/hazus.

Tools for Spatial Analysis

Management, integration, manipulation, and visualization of high volumes of data require sophisticated tools for effective storage, interrogation, and analysis, to generate suitability maps based on a hypothesis. A GIS contains a variety of simple and complex tools for analysis of single data layers, pairs of layers, or multiple layers, enabling the extraction of a single composite result.

Individual Layers

Spatial analysis of individual maps and layers involves two-dimensional processing and geo-statistical methods, such as reclassification and thresholding, neighbourhood functions using spatial filters, distance, and buffer calculations, 2D spatial transformations and, importantly, gridding or interpolation.

Geo-statistical methods, involving the application of probabilistic methods to geographically related phenomena, can be used to highlight spatial correlation within a data layer. This idea is based on the assumption that points located close to one another, should also be close in value. Existing data are then used to interpolate into areas where no data exists.

Multiple Layers

Simple tools and processes can also be used effectively to manipulate a multisource, multilayer dataset, to produce a single, composite, and meaningful result. The term 'Map Algebra' is commonly used in reference to mathematical calculations carried out between two or more GIS layers, according to some expression, to produce a new layer. For example, to derive a map showing the depletion of an aquifer over a set period of time, layers representing groundwater levels as measured at set times are required. One layer is simply subtracted from the other to produce a layer representing the changes between the two dates (Figure 6). GIS enables the modification and classification of each layer and the combination of the modified layers, mathematically, to produce the final result. Through the combination of several, simple operations in this way, either collectively or in sequence, complex problems can be addressed using GIS.

Multicriteria Evaluation and Uncertainty

Multisource datasets are rarely complete and, in Earth sciences, are seldom defined by rigid, Boolean-type, observable boundaries. Uncertainties, arising from data and decision rules, are common. GIS provides a number of specialised, multicriteria, analytical tools to deal with these uncertainties, which draw on the principles of, for example, Bayesian probability and fuzzy logic.

Multicriteria evaluation is primarily concerned with the combination of multiple datasets to produce a single index of evaluation. In the simplest cases, such as those involving Boolean constraints, this combination can be achieved using union (logical OR) or intersection (logical AND) rules. For factors that display continuous variability, some form of ranking and weighting is required. As the criteria are measured and recorded on different scales, they also require standardization so that all factors are quantified on the same scale. It is also convenient if they are all positively correlated with 'suitability'. Deciding on ranking and weighting is the most difficult aspect, and it commonly requires discussion, field verification, and iterative modification, for which a common technique is the comparison matrix (illustrated in Table 1).

Quantifying uncertainty Ground truth data rarely provide pure examples of the classes they represent, and boundaries between one type of natural phenomenon and another are rarely crisp or clearly defined; there is usually a continuum between them. Modifications in the choice of function, and the replacement of 'hard' for 'soft' decision rules, are often

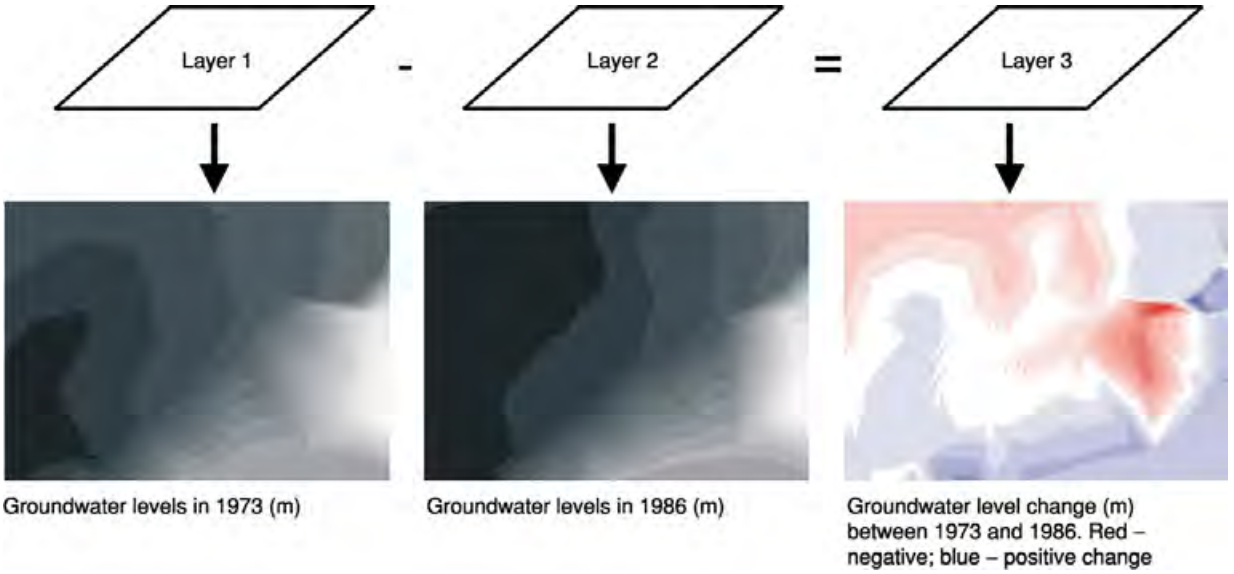


Figure 6 A simple map-algebraic example, in which spatial changes in groundwater levels, over a fixed time period (between 1973 and 1986) are derived by layer subtraction.

Table 1 Example of a pairwise comparison matrix, used to assess relative factor significance and to calculate factor weights (for landslide hazard assessment in this case)

	Slope	*Aspect*	*FS*	*DEM*	*Distance (from roads)*	*Distance (from drainage)*	*Soil moisture*	*Iron-oxides*	*Clay content*
Slope	1								
Aspect	1	1							
Factor of safety (FS)	1/3	1/3	1						
DEM	1/6	1/6	1/5	1					
Distance (from roads)	1/4	1/4	1/3	2	1				
Distance (from drainage)	1/7	1/7	1/4	1	1/5	1			
Soil moisture	1/5	1/5	1/3	2	1/2	2	1		
Iron-oxides	1/5	1/5	1/3	1	1/2	1	1/2	1	
Clay content	1/7	1/7	1/5	1	1/2	1	1	1	1
Factor weights	0.278	0.278	0.151	0.037	0.086	0.034	0.057	0.041	0.039

necessary within the analysis. Bayesian probability theory, Dempster–Shafer theory, and fuzzy set membership are commonly used to incorporate uncertainty in GIS analysis in geological applications, for example, landslide hazard assessment and mineral prospectivity mapping.

Considerable research has been carried out into the extraction of mineral prospectivity maps from multisource datasets. In many cases, identification of the very lowest and very highest prospectivity has not been in dispute but uncertainty arises in definition of the regions of intermediate prospectivity, which then require further analysis and interpretation. Suitability or prospectivity must be treated as a continuous phenomenon representing a measure of confidence in the outcome.

A fuzzy set describes a continuous membership function where 0 represents a non-member and a value of 1 represents a member, and the values between them represent the increasing possibility of membership. The fuzzy set membership function can be most readily appreciated with reference to a simple linear function (Figure 7). The fuzzy function, may in fact be linear, sigmoidal or 'J' shaped, and monotonic or symmetric. The threshold values, which define it, will depend on the phenomenon and desired outcome

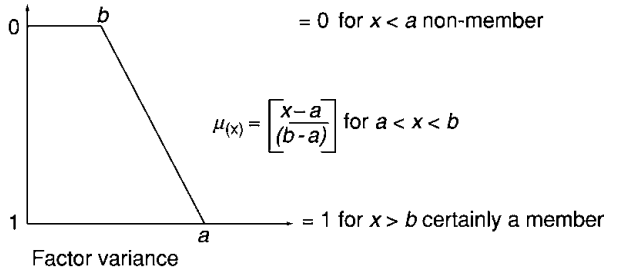


Figure 7 A simple, linear fuzzy set membership function, which can be used to standardize factors and to allow for a measure of uncertainty.

of the operation, i.e., the threshold values applied to each membership function reflect their influence on the result. The resultant fuzzy set layers can be combined in a number of ways, for example, using Boolean logic and fuzzy algebra or 'set theory'.

One novel technique, known as the Vectoral Fuzzy Logic method, produces two maps: one of calculated prospectivity and the other of confidence (represented by the similarity between the two inputs). Using the vectoral fuzzy logic method, null data and incomplete knowledge can be incorporated into the multicriterion analysis.

One of the most widely used statistical, multicriteria techniques is the Weights of Evidence Method. Here the quantitative spatial relationships between datasets representing significant criteria and known occurrences are analysed using Bayesian probability analysis. Predictor maps are used as input evidence and the end-products are maps representing the probability of occurrence of a phenomenon and of the uncertainty of those probability estimates.

The 'weighted linear factors in combination' method is also common, but can, in some instances, be considered too 'liberal' in its handling of the data in the system. Its factor aggregation method can be likened to a parallel connection system, which allows all input criteria to survive to the end. It is also possible that the relationships between the geological factors are not linear, in which case, a more complex model would be required. In many cases, this parallel system may be appropriate but in others a more strict system is required, which enables certain factors or combinations of factors to be eliminated completely from the system. A model based on the simple calculation of the geometric mean of scaled input factors (as opposed to arithmetic mean) can be thought of as a sequential connection system in which the occurrence of a zero rating, at a particular location, terminates the system and eliminates that location from the analysis. The method used to generate the landslide hazard map in Figure 5, incorporates this concept. Ideally, the method should incorporate both of these types of data treatment. Commonly the terms factor and 'constraint' are used to distinguish between information which is measured on a relative scale indicating a variable degree of likelihood, and that which is Boolean, respectively.

One difficulty with many of the approaches employed is the assumption that all factors contribute positively towards the suitability; consideration may also need to be given to evidence that appears contradictory. Dempster–Shafer theory provides a method of establishing suitability (or unsuitability) as well as a measure of plausibility for each hypothesis. Bayesian probability allows any new evidence to be incorporated, so updating the probability of suitability each time, but data are required for every possible occurrence, i.e., the absence of evidence is assumed to support the alternative hypothesis (unsuitability). The Dempster–Shafer method allows for 'Ignorance', i.e., the lack of evidence does not necessarily support the alternative hypothesis. A measure of plausibility (and of belief) is incorporated in the outcome. The theory recognizes a total of six concepts–Basic Probability Assignment, Ignorance, Belief, Disbelief, Plausibility, and a Belief Interval. Dempster–Shafer theory, therefore, provides a very useful basis for establishing the value of information and the design of further site investigation, including where new data would be most effective in further reducing the uncertainty.

Data Sharing and the Internet

The alliance between GIS and the Internet is both a natural and an extremely powerful one. Internet GIS databases allow access to vast resources of information, not only geological but also for public interest and attention. Database architectures and data formats, specially designed to be served through intranets and the Internet, enable huge volumes of information to be effectively served to computers anywhere.

Over the last few years, data compression software and specially designed internet servers have been developed, for example, Earth Resources Mapping's wavelet transform based data compression format (.ecw) and *Image Web Server*, which enables vast raster image datasets to be served and shared with incredible speed. Examples of the internet map data serving, through the use of intelligent databases and systems, include the NIMA Geospatial Engine (http://geoengine.nima.mil/), US based GIS Data Depot (http://data.geocomm.com/) and the British Geological Survey's Geoscience Data Index

Table 2 Commercially available GIS software suites, their manufacturers and application fields

Manfacturer	*Professional*	*Palmtop mapping*	*Internet mapping*	*Free viewer*
ESRI	ArcInfo, ArcGIS, & ArcView	ArcPAD	ArcIMS	ArcExplorer
Leica Geosystems	ERDAS Imagine	-	ESRI	Mapsheets Express
MapInfo	MapInfo	MapXtend	MapXtreme	ProViewer
Autodesk	AutoCAD/World	OnSite	MapGuide	AutoCad LT
Intergraph	GeoMedia & GeoMedia Pro	IntelliWhere	GeoMedia Web Map	GeoMedia Viewer

(http://www.bgs.ac.uk/geoindex/index.htm). These allow the user to search, view, and in the case of NIMA, actually download free any globally available geological, image, and map data (*see* **Geological Field Mapping**).

Software

The GIS software suites now commercially available provide a dazzling choice of functionality to suit all applications and budgets. The following table (Table 2) provides a brief overview of the main software products and their variants, as designed for different markets and applications.

Conclusions

GIS provides a very convenient framework for map-analysis and modelling functions which can be applied in many areas of the geosciences. It is expected that there will be continuing demand for GIS functionality to become faster, cheaper, and more complex, with continuing open standards. There will also be demand for the worlds national authorities to release more GIS data, free of charge, to the public. The vision of OGC is "a world in which everyone benefits from geographic information and services made available across any network, application or platform". Geoscience represents an expanding application area for GIS, both for research and commercial development.

A minor, but general, failing of standard GIS software functionality is its inability to properly deal with geological structures in three dimensions. There are, however, other software suites on the market, which are specially designed for this purpose and so the focus of GIS's contribution in geology lies in data integration, management, and spatial analysis.

There are more automated techniques for solving complex, spatial, multicriteria problems, such as neural networks and data mining, but these are in some respects at a disadvantage, compared with GIS, as they require very large and homogeneous datasets for training the systems. Data mining involves the development and use of effective tools for direct access and extraction of appropriate datasets and patterns from very large volume databases and represents a logical step in the evolution of effective data supply and analysis. Both data mining and neural networks have some overlap with GIS in functionality and objectives, and are anticipated to have some impact on GIS in the coming years. Commonly though, the geoscientist prefers a more intuitive, graphical, and less formal approach, so that even in complex analysis, interpretive skills are still important.

GIS owes its power and success to its flexibility and potential for cross-disciplinary application. GIS (and other related technologies) enable the analysis of vast and expanding datasets, allowing a better understanding of terrestrial and anthropogenic processes, the development of economic sustainability, and the improvement of environmental quality.

See Also

Geological Field Mapping. **Remote Sensing:** Active Sensors; Passive Sensors. **Sedimentary Processes:** Landslides. **Tectonics:** Earthquakes. **Volcanoes**.

Further Reading

Bonham-Carter GF (1994) *Geographic Information Systems for Geoscientists*. Pergamon.

Bonham-Carter GF (1996) *Geographic Information Systems for Geoscientists: Modelling with GIS*, 2nd edn. Pergamon.

Burrough PA (1986) *Principles of Geographical Information Systems for land resources assessment*. Oxford: Clarendon Press.

Eastman JR, Jin W, Kyem PAK, and Toledano J (1995) Raster Procedures for Multi-Criteria/Multi-Objective Decisions. *Photogrammetric Engineering and Remote Sensing* 61(5): 539–547.

Illife JC (2000) *Datums and projections for remote sensing, GIS, and surveying*. Caithness, Scotland: Whittle Publishing.

Isaaks EH and Srivastava RM (1989) *An Introduction to Applied Geostatistics*. New York, Oxford: Oxford Univ. Press.

Knox-Robinson CM (2000) Vectoral fuzzy logic: a novel technique for enhanced mineral prospectivity mapping, with reference to the gold mineralisation potential of the Kalgoorlie terrane, Western Australia. *Australian Journal of Earth Sciences* 47(5): 929–941.

Liu JG, Mason PJ, Clerici N, Chen S, and Davis AM (2003) Landslide Hazard Assessment in the Three Gorges Area of the Yangtze River using ASTER Imagery: Zigui-Badong. *Geomorphology* 61: 171–187.

Longley PA, Goodchild MR, Maguire DJ, and Rhind DW (eds.) (1999) *Geographical Information Systems – Principles, techniques, applications and management*. New York: John Wiley.

Mason PJ and Rosenbaum MS (2002) Geohazard mapping for predicting landslides: the Langhe Hills in Piemonte, NW Italy. *Quarterly Journal of Engineering Geology & Hydrology* 35: 317–326.

Mather PM (ed.) (1994) *Geographical Information Handling – research and application*. Wiley and Sons.

Zadeh LA (1965) Fuzzy sets, Information and Control. 8: 338–353.

Passive Sensors

J G Liu, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

What is a Passive Sensor?

A sensor detects and quantitatively records the electromagnetic radiation (EMR) from an object as 'sensing'. For an object to be sensed, it must either reflect or transmit the radiation impinging on it from an illumination source or radiate its own energy. If a sensor provides its own illumination source it is an active sensor. Synthetic Aperture Radar (SAR) is a typical active sensor system as it sends microwave radiation pulses to illuminate the target area and receives the returned signals to produce an image. If the sensor depends on an independent illumination source, such as the Sun, or the radiation from the target itself, such as the Earth's thermal emission, it is a passive sensor. The most commonly used sensors for remote sensing systems are passive sensors, which are the topic of this article.

A passive sensor images the Earth by recording either the reflected solar radiation or the emitted radiation from the Earth. An ordinary camera is a typical passive sensor system. A daytime outdoor colour photograph taken by a camera is a record of reflected solar radiation in the visible spectral range from the scene that is illuminated by the Sun. A camera can become an active sensor when it is used with a flash light in darkness. In this case, the camera provides its own light source to illuminate the object and then takes a picture.

The Sun is the primary illumination source for the Earth. For Earth observation remote sensing, most passive sensor systems operate under solar illumination during daytime, ranging from aerial photography to satellite multispectral scanners. These sensors pick up the reflected solar energy from the land surface to produce panchromatic and multispectral images. The features in these images are mainly influenced by two types of information: spatial pattern (e.g., topography) and spectral signatures. Ignoring minor factors, we can present such an image as

$$M_r(\lambda) = \rho(\lambda)E(\lambda) \quad [1]$$

where $M_r(\lambda)$ is the reflected solar radiation of spectral length λ by the land surface, or an image of spectral band λ, $E(\lambda)$ is irradiance, the incident solar radiation energy upon the land surface, while $\rho(\lambda)$ is the reflectance of land surface at wavelength λ.

If solar illumination is constant, $E(\lambda)$ is effectively the topography, as it governs the geometry of a surface in relation to illumination. In general, considering solar radiation as a parallel radiation source illuminating the Earth, with constant incident radiant flux density M_s, then the solar irradiance upon the land surface, E, varies with the angle between the land surface and the incident solar radiation, γ. When the land surface is perpendicular to the incident solar radiation M_s, E is at its maximum and equal to M_s. If the solar radiation has an incident angle θ_1 (from nadir) and azimuth angle ϕ_1, then the irradiance upon a land surface with slope angle θ_2 and aspect direction ϕ_2 is determined by:

$$\begin{aligned} E &= M_s \sin\gamma \\ &= M_s[\sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2) + \cos\theta_1 \cos\theta_2] \end{aligned} \quad [2]$$

The spectral reflectance, $\rho(\lambda)$, is a physical property, quantitatively describing the reflectivity of the exposed materials on the land surface at wavelength λ. The selective absorption and reflection of a material result in variation of spectral reflectanc across a sensor's spectral range, giving a unique signature for this substance. It is, therefore, possible to work out the land cover types or mineral compositions of the land surface based on spectral signatures using multispectral image data. Reflective spectral remote sensing is among the most effective technologies to study the Earth surface composition and has become an essential tool for geological mapping.

A thermal sensor is also a passive sensor but the energy that a thermal sensor detects is radiation emitted from the Earth's surface, rather than reflection. Thermal sensing does not need an illumination source as the target itself is the illumination source. The Earth's surface can be approximated as a black body of 30 K and using Wien's law, we can calculate that the radiation peak of the Earth is at about 10 μm. In this spectral range, the radiation can be sensed and measured by temperature, rather than brightness, as measured for the visible spectral range; it is, therefore, called thermal sensing. Different materials on the land surface have different thermal radiation properties; thermal sensors are, therefore, useful tools for geological and environmental studies.

According to Wien's law, the higher the temperature of an object, the shorter the wavelength of its radiation peak. Thus, passive sensors designed for shorter wavelengths, though usually used to detect reflected solar radiation from the land surface, can also be used for detecting emitted energy from targets as long as the temperature of these targets is high enough. For instance, Landsat Thematic Mapper (TM) shortwave infrared band 5 (1.55–1.75 μm) and band 7 (2.08–2.35 μm) are far more effective to map fires on a land surface than the thermal band.

The land surface is also an effective radiator at microwave range, though the microwave emission is significantly weaker than thermal emission. As another type of passive sensor, radiometers are designed to image the emitted radiation from the Earth's surface at this spectral range.

For geological studies, the most effective passive sensor systems are of medium to high spatial resolutions with image pixel size ranging from a few to a few hundred metres and are often called Earth observation systems. The following sections use the most common Earth observation systems as examples for further discussion of passive sensors.

Passive Sensor Imaging Technology

With a few exceptions, most passive sensor systems are essentially an optical camera system. As shown in Figure 1, the incoming energy to the sensor system goes through an optical lens and is focused onto the rear focal plane of the lens where the energy is recorded by a sensor device of radiation sensitive media, such as film or CCD (Charged Coupled Device).

Aerial photography, using a large format camera, was the earliest operational remote sensing technology for topographic survey as well as for geological investigation. The space-borne remote sensing for Earth observation began with Landsat-1 with a MSS (Multi-spectral Scanner) and RBV (Return Beam Vidicon) on board, launched on 23 July 1972. This and subsequent landsat instruments collect, record, and transmit electronically to ground stations, images of the Earth. The images are then distributed to users in a digital format as digital imagery data. The concept and technology of digital imagery gave birth to satellite remote sensing. An Earth observation satellite images the Earth's surface continuously from its orbit and sends the images back electronically to the receiving stations on the ground.

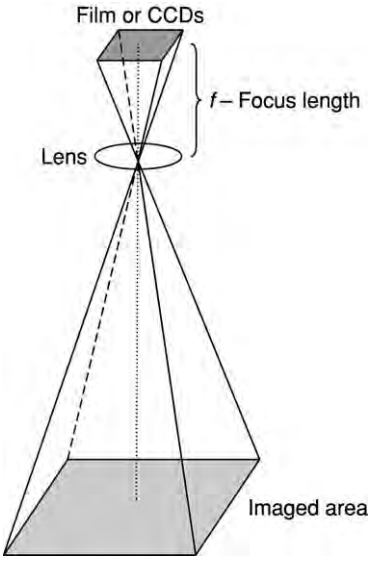


Figure 1 Basic structure of an optical sensor system.

The development of sensor technology is mainly focused on improving the spatial and spectral resolution. For a given sensor system, its spatial resolution is dictated by the minimal energy level of electromagnetic radiation (EMR) that can make a signal distinguishable from the electronic background noise, i.e., dark current, of the instrument. This minimal energy of EMR is proportional to the product of radiation intensity over a spectral range, IFOV (Instant Field Of View), and dwelling time.

The IFOV (Figure 2) is decided by spatial sample density of an optical sensor system and determines the pixel size of the image. For a given dwelling time (equivalent to exposure time) and spectral range, the larger the IFOV, the more energy will be received by the sensor, but the lower the spatial resolution. To improve the spatial resolution, IFOV needs to be reduced while maintaining the same energy level; either dwelling time or spectral range, or both have to be increased. When a sensor, which has its dwelling time fixed by the sensor design and platform orbit parameters, receives the reflected solar radiation from the Earth, it may record the energy in a broad spectral range as a single image. This generates a

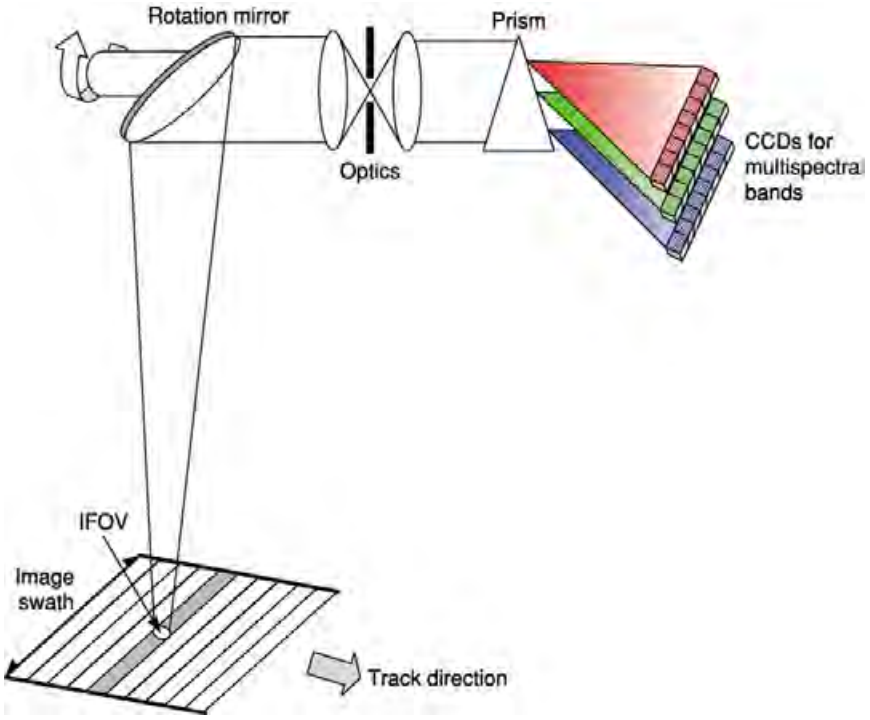


Figure 2 A schematic diagram of an across-track mechanical multispectral scanner.

panchromatic image at a relatively high resolution. It may also split the light into several spectral bands and record them separately into several images of narrower spectral range: the multispectral images. In this case, the energy that reaches the CCDs of each narrow spectral range is significantly weaker than in the panchromatic mode. To achieve the same energy level, the only solution is to increase the size of IFOV, or reduce spatial resolution. This is the reason that nearly all the optical imaging systems achieve higher resolution in a panchromatic band than in multispectral bands (Tables 1 and 2). For instance, SPOT 1–3 HRV (High Resolution Visible) Panchromatic bands have 10 m resolution while its XS (multispectral) bands have 20 m resolution.

There is another way to improve the spatial resolution for both panchromatic and multispectral imagery; by increasing the dwelling (exposure) time. This has been an important consideration in sensor design, although the capacity for increasing the dwelling time is limited; for both airborne and space-borne remote sensing, the image is taken from a moving object and so a long exposure time will blur the image.

Across-Track Mechanical Scanner

The early design of space-borne passive sensor systems was constrained by the technology of the key sensor unit: the CCD. In many systems, a mechanical scanner has been an ideal solution to achieve multispectral imaging at relatively high resolution. Such scanners are based on a simple mechanical device using a few CCDs. Figure 2 is a schematic diagram of the principle of an across-track, mechanical multi-spectral scanner. The optical part of a scanner is essentially a camera but the image is formulated pixel by pixel, scanned across track by a rotating mirror, and line by line, as the platform (aircraft or satellite) flies along track over an area. The range of a scan line is called the swath. The light reflected from the land surface reaches the rotating mirror, which rotates at a designed speed and thus views different positions on the ground across a swath during a rotation scan cycle. The rotating mirror diverts the incident light through the scanner optics and then the light is dispersed into several spectral beams by a spectral splitting device (a prism or interference filters). The multispectral beams are then received by a group of CCDs which sample the light at regular time intervals. When the scanner finishes scanning one swath, the satellite or aircraft has moved forward along its track to the position for the next scan. One scan swath can be a single image line or several image lines depending on the sensor design and the synchronisation between flying speed, swath width, altitude of the satellite or aircraft, and required image resolution. In this way, a scanner can image a large area using a limited number of CCDs. A scanner, though mechanically complicated, relies less on CCD technology.

Table 1 Comparison of spectral bands of TM/ETM+, ASTER, and SPOT

	Sensor systems								
	Terra-1 ASTER			*Landsat 3–7 TM/ETM+*			*SPOT 1–3 HRV, SPOT 4 HRVI, SPOT 5 HRG*		
Spectral region	*Band*	*Spectral range (μm)*	*Spatial resolution (m)*	*Band*	*Spectral range (μm)*	*Spatial resolution (m)*	*Band*	*Spectral range (μm)*	*Spatial resolution (m)*
VNIR				1	0.45–0.53	30			
	1	0.52–0.60	15	2	0.52–0.60		1	0.50–0.59	20
	2	0.63–0.69		3	0.63–0.69		2	0.61–0.68	
	3N	0.78–0.86		4	0.76–0.90		3	0.79–0.89	
	3B	0.78–0.86		ETM+ Pan	0.52–0.90	15	Pan	SPOT 1–3: 0.51–0.73	10
								SPOT 4: 0.61–0.68	10
								SPOT 5: 0.48–0.71	2.5–5
SWIR	4	1.60–1.70	30	5	1.55–1.75	30	4	1.58–1.75	20
	5	2.145–2.185		7	2.08–2.35		Band 4 is only in SPOT 4 HRVI (High Resolution Visible Infrared) and SPOT 5 HRG (High Resolution Geometric)		
	6	2.185–2.225							
	7	2.235–2.285							
	8	2.295–2.365							
	9	2.360–2.430							
TIR	10	8.125–8.475	90	6	10.4–12.5	TM			
	11	8.475–8.825				120			
	12	8.925–9.275							
	13	10.25–10.95				ETM+ 60			
	14	10.95–11.65							

Table 2 Some satellite high resolution broadband sensor systems

		Spatial resolution (m)			*Spectral range (μm)*		
Satellite	*Launch time and Status*	*Pan*	*MS*	*Hyp*	*Pan*	*MS*	*Hyp*
Ikonos 2	24 Sept. 1999 In operation	1	4		0.45–0.90	0.45–0.53 0.52–0.61 0.64–0.72 0.77–0.88	
QuickBird	18 Oct. 2001 In operation	1	4		0.45–0.90	0.45–0.52 0.52–0.60 0.63–0.69 0.76–0.89	
Orbview 3	26 June 2003 In operation	1	4		0.45–0.90	0.45–0.52 0.52–0.60 0.625–0.695 0.76–0.90	
Orbview 4	Failed at launch in Sept. 2001	1	4	8	0.45–0.90	0.45–0.52 0.52–0.60 0.625–0.695 0.76–0.90	0.45–2.50 200 bands

The MSS with 4 spectral bands on board Landsat 1–3 is a classical example of a mechanical scanner. It is a one-way scanner that scans in one direction of mirror rotation only, and with an empty return run. Such a design makes compensation of the Earth's rotation easier, as the Earth rotates a fixed distance along the swath direction for each scanning cycle. However, the inactive return would waste valuable time for imaging, resulting in a shorter dwelling time in the active runs and thus reduce spatial resolution.

The Thematic Mapper (TM) on board of Landsat 4–7 is a significantly improved scanner with six reflective spectral bands and one thermal band (Table 1). It is a two-way scanner that scans in both directions. For the same width of swath, the two-way scan allows the mirror to rotate slower than a one-way scan, thus increasing the dwelling time of CCDs at each sample position. This configuration improves both spatial and spectral resolutions. The geometric correction, to compensate for Earth rotation effects, is more complicated for TM than for MSS, however, because one scan direction is for, and the other, against the Earth's rotation, respectively.

Along-Track Push-Broom Scanner

With the rapid development of CCD technology, a more advanced scanner, the so-called 'push-broom scanner', has become dominant in passive sensor design since the successful launch of SPOT-1 in 1986. As shown in Figure 3, the key difference between a push-broom scanner and a mechanical scanner is that a push-broom scanner does not have a mechanical mirror for pixel-by-pixel scanning across along the swath direction. In a push-broom scanner, a line array panel of CCDs, covering the whole imaging swath, is mounted at the rear of the spectral dispersal device. Push-broom scanner images an area line by line across the whole track when the sensor platform (a satellite or an aircraft) flies along track, like pushing a broom forward to sweep a floor.

Since one image swath is generated simultaneously, the dwelling time for each CCD representing an image pixel can be as long as a whole swath scanning time for a mechanical scanner. With significantly increased dwelling time, a push-broom scanner achieves much higher resolution. Based on advanced CCD technology, the push-broom scanner is simpler than the mechanical scanner in structure, and the geometric correction is less complicated. Without a mechanical part, the system is robust and reliable. The number of CCDs in a line array in swath direction decides the size and the resolution of the imagery data generated. For instance, the SPOT HRV has 6000 CCDs per line for its panchromatic band while 3000 CCDs per line for its multispectral bands. The system, therefore, produces panchromatic images with 6000 pixels per line at 10 m resolution and multispectral images with 3000 pixels per line at 20 m resolution.

The push-broom design also allows greater flexibility to manoeuvre the sensor to point to off-nadir positions. As shown in the Figure 3, a mirror is fixed in front of the lens like a mechanical scanner. This mirror is not for scanning; it has several fixed angle positions. At different angles, the mirror directs the sensor system to view and image different scenes off-nadir along the track.

Digital Cameras

In the last few years, digital cameras have started to compete in the market of traditional film photog-

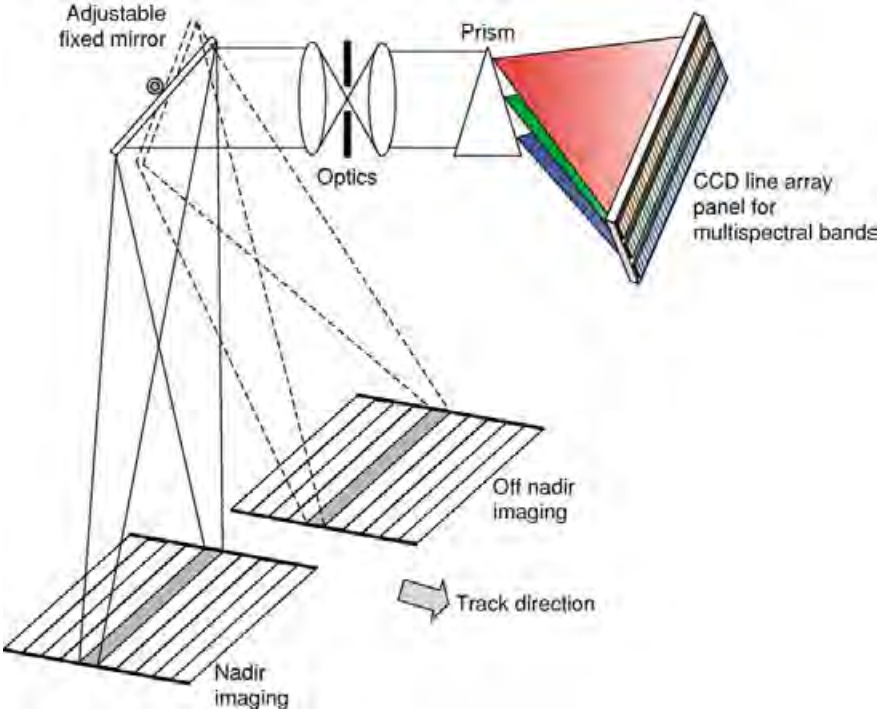


Figure 3 A schematic diagram of an along-track push-broom multispectral scanner.

raphy. A digital camera is built on a full 2D (two dimensional) CCD panel. The only difference between a digital camera and a film camera is that it records an image through a 2D CCD panel to a memory chip instead of using a film. The new generation of passive sensors will be largely based on digital camera technology that takes an image by an instantaneous frame rather than scanning line by line. The consequence is that the constraints on platform flight parameters can be relaxed, image resolution (spatial and spectral) can be improved, and image geometric correction processing can be further streamlined.

Broadband Reflective Multispectral Imagery

Broadband reflective multispectral sensors and thermal sensors are addressed separately being based on different physics. In practice, however, these two groups of passive sensors are often mounted in the same instrument as different bands. This is true for many sensor systems, such as TM, Enhanced Thematic Mapper Plus (ETM+), and ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) (Table 1), used for geological applications.

Broadband reflective multispectral image data are the most widely used for geological studies and regarded as an effective operational tool for mapping tectonic structures and lithology, for mineral exploration, for logistic planning, and for field survey navigation. Extending the scope of applications from Earth observation to planetary study, this group of sensors are often the major tools for data collection from other planets. A typical example is the exploration of Mars.

The American Landsat satellite family TM, ETM+, and French SPOT (Système Pour 1'Observation de la Terre) satellite family high resolution visible (HRV) are the most commonly used Earth observation systems, providing broadband multispectral and panchromatic imagery data of global coverage. As shown in Table 1, these types of sensor systems operate in: the visible spectral range with bands equivalent to three primary colours; blue (380–440 nm), green (440–600 nm), and red (600–750 nm); the near infrared (NIR) range (750–1100 nm), and the short wave infrared (SWIR) range (1550–2400 nm). The number of bands and the spectral width in VNIR (visible near infrared) and SWIR spectral ranges depend on the atmospheric windows and sensor design. For instance, the spectral width of SWIR bands needs to be much wider than visible bands if the same spatial resolution is to be achieved, as is the case for TM bands 5 and 7, because the solar radiation in the SWIR spectral region is significantly weaker than that in the visible spectral range.

In general, the 'broad' band means that the spectral range is significantly wider than a few nanometres, except in the case of hyperspectral sensor systems described below. Broadband reflective multispectral sensor systems are a successful compromise between spatial resolution and spectral resolution. With relatively broad spectral bands, such a sensor system offers reasonable spatial resolution with high SNR (Signal Noise Ratio) and meanwhile, operating in a wide spectral range from VNIR to SWIR, such a system can provide images of multispectral bands, enabling identification of major ground objects and discrimination of various land cover types. With the dramatic improvement of sensor technology, from mechanical scanners to push-broom scanners, and to CCD digital cameras, the spatial resolution of broadband multispectral imagery is improving all the time. For Sun synchronous near polar orbit satellites, the spatial resolution of this type of sensors has been improved from 80 m (Landsat MSS) in the 1970s to a few metres, on current systems, as shown by the examples in Table 2.

The VNIR spectral range is used by nearly all broadband reflective multispectral sensor systems. This spectral range is within the solar radiation peak and thus allows the generation of high resolution and high SNR images. It also covers diagnostic features of major ground objects such as the few examples below:

- *Vegetation*: minor reflection peak in green, absorption in red, and then significant reflection peak in NIR, often called the 'red edge'.
- *Water*: strong diffusion and penetration in blue and green, and nearly complete absorption in NIR.
- *Iron oxide (red soils, gossans, etc.)*: absorption in blue and high reflectance in red.

Many satellite sensor systems did not use the blue band, in order to avoid strong Rayleigh scattering effects occurring in the atmosphere that can make an image 'hazy'. A popular configuration is to offer three broad spectral bands in green, red, and NIR, such as the case of SPOT and the most recent commercial high spatial resolution space-borne sensors (Tables 1 and 2). In a computer graphic system, we can display the three band images as a colour composite with NIR displayed in red, red in green, and green in blue. Such a colour composite image is called a standard false colour composite. This image is the most effective for mapping healthy vegetation.

The SWIR spectral range is regarded as the most effective for lithological mapping and mineral exploration, because most rock types have high reflectance in 1.55–1.75 μm and clay minerals (often products of

alteration) related to mineralisation, have various absorption features in the spectral range 2.0–2.4 μm. These two SWIR spectral ranges, corresponding to Landsat TM band 5 and 7, are preferred by geologists. SWIR sensor systems are technically more difficult and complicated because the SWIR detectors have to operate at low temperatures, which therefore require a cooling system (a liquid nitrogen coolant or a cryocooler) to maintain the detectors at about 80 K.

With six broad reflective spectral bands, Landsat TM provided the best spectral resolution among broadband sensor systems for many years. The six broad reflective spectral bands are very effective for discrimination of various ground objects but they are not adequate to achieve specific identification in particular rock types and major mineral assemblies relating to mineral deposits. This requires a sensor system with much higher spectral resolution at a few nanometre bandwidth to resolve their subtle spectral signatures. This demand has led to the development of the hyperspectral system.

ASTER (a push-broom scanner for VNIR and SWIR bands), on board the Terra-1 satellite, is a representative of a transitional sensor system between broadband multispectral and hyper-spectral narrowband sensing. It is an integrated system of three scanners: a VNIR push-broom scanner with three broad spectral bands; a SWIR push-broom scanner with six narrow spectral bands; and a TIR across-track mechanical scanner with five thermal bands (Table 1). The system combines good spatial resolution in the VNIR bands and high spectral resolution in SWIR and thermal bands and was specifically designed for geological applications. The three 15 m resolution VNIR bands are adequate for distinguishing broad categories of land cover such as vegetation, water, red soils, urban areas, superficial deposits, and rock outcrops, while the six narrow SWIR bands of 30 m resolution provide potential for mapping major mineral assemblies of rock types and alterations. Another unique advantage of ASTER is that it has along track stereo capability. The VNIR scanner has a backward viewing telescope to take NIR images beside its nadir telescope for the three VNIR bands. Thus nadir and backward viewing NIR images are taken simultaneously, forming along track stereo image pairs. The along track stereo image pairs enable generation of DEM (Digital Elevation Model) data.

Thermal Infrared (TIR) Sensors

Various minerals, rock types, and other ground objects have different thermal properties, such as thermal inertia, thermal emission, and thermal absorption. For instance, quartz has distinctive high emissivity around 9 μm and strong absorption in 10–11 μm; dolomite can be distinguished from limestone by its significant lower emissivity in 8–11 μm.

There are several airborne thermal IR sensor systems for example, the Thermal Infrared Multispectral Scanner (TIMS) with 6 bands in the 8.2–12.2 μm spectral region developed in 1982. Of the satellite systems, the Landsat Thematic Mapper system has a broad thermal band, TM band 6, at a wavelength range of 10.4–12.5 μm. So far the only space-borne multispectral thermal system is ASTER on board of Terra-1 satellite, which has 5 thermal bands as shown in Table 1.

In general, a broadband TIR sensor operating in the 8–14 μm spectral range images the radiation temperature of the land surface while the narrower band of the multispectral thermal imagery data detect the thermal spectral signatures of materials on the land surface. It is important to note that daytime thermal images are fundamentally different from the night time thermal images. Daytime thermal images are dominated by topographic features, governed by the geometry between slopes and solar radiation described previously, while night time thermal images are solely determined by emission from the Earth surface. They therefore detect the thermal properties of ground materials more efficiently.

For both systems, TM and ASTER, the spatial resolutions of their thermal bands are significantly lower than their reflective multispectral bands, as shown in Table 1. One reason is that the interaction between thermal energy (or heat) and the atmosphere is more complicated than the case of VNIR and SWIR energy. Heat can be transmitted in air, not only by radiation, but also by air circulation. Both solar radiation to the Earth in TIR spectral range and direct thermal emission from the Earth are very weak, compared with the energy intensity of the Earth reflected solar radiation in the VNIR and SWIR spectral ranges.

Most thermal sensors are of cross-track mechanical scanner type, as shown in Figure 4. The major difference of a thermal scanner from a reflective multispectral scanner is that it needs a cooler system to maintain the TIR detector at very low temperature for maximum sensitivity. For instance, the thermal sensor of Landsat Thematic Mapper is surrounded by liquid nitrogen at 77 K stored in an insulated vessel. In the ASTER system, a cryocooler is used to maintain the detectors for TIR bands at 80 K. A blackbody plate is used as an on-board calibration reference that is viewed before and after each scan cycle, providing estimation of instrument drift. This is essential to maintain the accuracy and consistency of a TIR instrument. The temperature sensitivity of a modern TIR sensor system can be as high as 0.1 K.

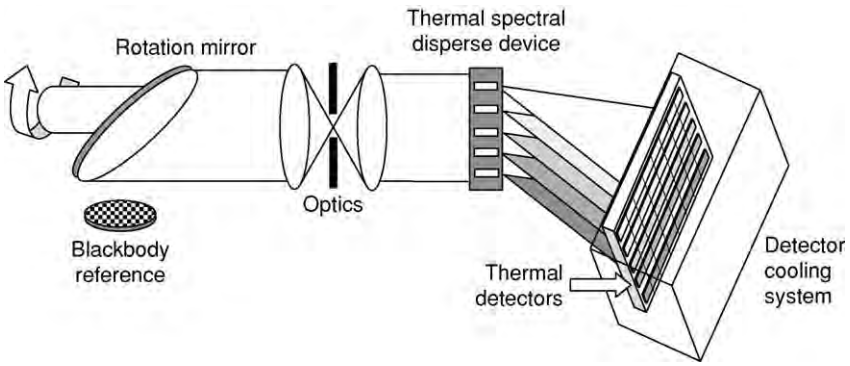


Figure 4 A schematic diagram of a thermal scanner.

To fully represent the sensitivity, many thermal IR sensors use 10–12 bit quantisation to record data, for example, ASTER multispectral thermal band images are 12 bit integer data.

Hyper-Spectral Sensors (Imaging Spectrometers)

The development of passive sensor technology is always aimed at higher spatial and spectral resolutions. Hyperspectral sensor systems represent a revolutionary development increasing passive sensor spectral resolution to as high as a few nanometres. This means that they are capable of generating nearly continuous spectral profiles for land surface materials. A hyperspectral sensor system is an imaging spectrometer. It combines the spatial imaging capacity of an imaging system with the spectral analytical capabilities of a spectrometer. Such a sensor system may have up to several hundreds narrow spectral bands with spectral resolution on the order of 10 nm or narrower. Imaging spectrometers produce a complete spectrum for every pixel in the image. The resulting dataset allows identification of materials rather than mere discrimination as with broadband sensor systems. The data processing methodology and strategy are, therefore, different from that employed for broadband images in many aspects. It is more important to analyse the spectral signature for each pixel, rather than to perform general image enhancement to improve visualization, though the latter is still essential for data overview.

One of the earliest and the most representative hyper-spectral systems is JPL's Advanced Visible Infrared Image Spectrometer (AVIRIS) (Table 3). Figure 5 shows the general principle of hyperspectral systems. The incoming EMR from the land surface goes through the sensor optics and is then split into hundreds (e.g., 224 for AVIRIS) of very narrow spectral beams by a complicated spectral dispersion device (e.g., interference filters) and finally the spectral beams are detected by arrays of CCDs corresponding to, say 224, spectral bands. A hyperspectral system can be either an across-track mechanical scanner with a small number of detectors for each band, or an along-track push-broom scanner with a panel of hundreds of line arrays of CCDs. Hyperspectral sensors are so far operating in either VNIR only or VNIR and SWIR spectral ranges (Table 3), although a similar evolution is taking place in TIR systems.

There is only one experimental, space-borne hyperspectral sensor, Hyperion. Several satellites with a hyperspectral system on board have failed, for various reasons, before the sensor system even started to operate (Table 2). Nevertheless, with the success of several airborne hyperspectral systems (Table 3), it is only a matter of time and money before hyperspectral satellite system become operational.

Table 3 Some airborne Hyper-spectral sensors

Instrument	*Spectral range (nm)*	*Band width (nm)*	*No. bands*
AVIRIS	400–2400	9.6	224
AIS	1200–2400	9.6	128
SISEX	400–2500	10–20	128
HIRIS	400–2500	10–20	128
MIVIS	430–1270	20–54	102

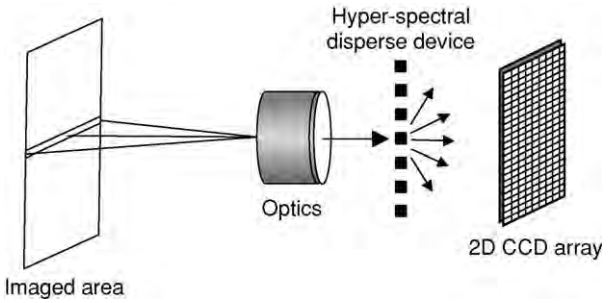


Figure 5 Principle of an imaging spectrometer (hyper-spectral system).

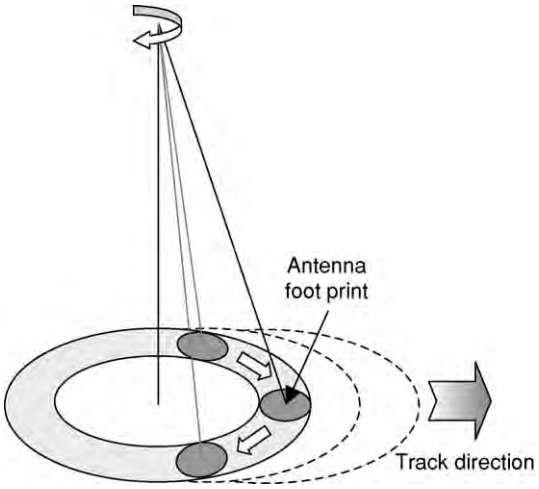


Figure 6 Conical scanning mechanism.

Passive Microwave Sensors

Thermal radiation from natural surfaces, such as the land surface, extends from its peak in the thermal infrared region into the microwave region. An Earth observation microwave imagining radiometer operates in this spectral region to receive the microwave radiation from the Earth. As a passive sensor system, it is important to understand that a microwave radiometer is fundamentally different from a radar sensor, that is a ranging system. The only resemblance between the two is that they both operate in the microwave spectral range. A passive microwave sensor system, a microwave imaging radiometer, works more like a thermal sensor system. It collects emitted energy radiated from the Earth in the microwave spectral range that provides useful information relating to surface temperature, topography, and material dielectric properties. This type of sensor has been used for global temperature mapping, polar ice mapping, and regional soil moisture monitoring.

A space-borne microwave imaging radiometer is often a multichannel scanner, such as the SMMR (Scanning Multichannel Microwave Radiometer), which was on board of Seasat and Nimbus in 1978 and the Microwave Imager on board the Tropical Rainfall Measuring Mission (TRMM) satellite in 1997. It is composed of an antenna together with its scanning mechanism, a receiver, and a data handling system. The received emitted microwave signals are highly related to the observation angle and the path length in the atmosphere. Ensuring that these scanning parameters are constant, thus they can significantly increase the accuracy of the derivation of surface parameters from the microwave brightness temperature. A conical scan configuration is, therefore, popular for passive microwave scanners. As shown in Figure 6, the antenna observation direction is offset at a fixed angle from a nadir rotating scan around the vertical (nadir) axis and thus sweeps the surface of a cone. If the scan is configured for a full 360°, double coverage fore and aft of the spacecraft is obtained. With the moving forward of a satellite along its orbit, a belt of land surface is imaged. Obviously, in a conical scanning geometry, the observation angle and distance to any scanned position are constants.

Space-borne passive microwave scanners are usually of low spatial resolution from several kilometres to several tens of kilometres, because of weak signals in the microwave spectral range.

See Also

Geological Field Mapping. **Remote Sensing:** Active Sensors.

Further Reading

Drury SA (2001) *Image Interpretation in Geology,* 3rd edn. Chelternam: Blackwell Science.

Elachi C (1987) *Introduction to the Physics and Techniques of Remote Sensing*. New York: John Wiley & Sons.

Lilesand TM and Kiefer RW (2000) *Remote Sensing and Image Processing,* 4th edn. New York: John Wiley & Sons.

Colwell RN (ed.) (1983) *Manual of remote sensing,* Vol.1, 2nd edn. Theory, instruments and techniques. Falls Church, Va: American Society of Photogrammetry.

Robinson N (1966) *Solar Radiation*. Amsterdam, London and New York: Elsevier.

Sabins FF (1996) *Remote Sensing: Principles and Interpretation,* 3rd edn. Basingstoke: WH Freeman.

ASTER user Handbook v.2, on line at http://asterweb.jpl.nasa.gov/.

Landsat-7 handbook, on line at http://landsat.gsfc.nasa.gov/.

Remote sensing learning sources of Canada Centre for Remote Sensing, on line at http://www.ccrs.nrcan.gc.ca/ccrs/learn/learn_e.html

The SPOT Payload, on line at http://www.spotimage.fr/automne_modules_files/standard/public/p229_file-LINKEDFILE_Spot_payload.pdf.

RIFT VALLEYS

See **AFRICA: Rift Valley**

ROCK MECHANICS

JP Harrison, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Although rock has been used as an engineering material since time immemorial, rock mechanics as a specialist engineering discipline has existed only since about the middle of the twentieth century. The particular engineering endeavours in which rock mechanics is used run from traditional applications such as mining and tunnelling, through more modern applications such as large underground caverns for various civilian uses (e.g., fuel storage and transport infrastructure), to deep wellbores for petroleum extraction, and on to novel issues such as underground disposal of nuclear waste. There is also a growing realization that the principles of rock mechanics as used by engineers are applicable to problems in geological science, such as fault movement, seismicity, and plate tectonics. So, techniques that have their origin in engineering are more and more being applied to problems such as the stability of natural slopes, fracture development in extensional basins, and fluid flow through deforming rock.

Rock mechanics arose, and continues to develop, because, as far as engineers are concerned, rock is an awkward material. By this is meant that it is not like the usual well-specified and controlled engineering materials of steel, concrete, or polymers. Its natural basis means that it is likely to be variable in its character and to contain features that are weak; both of these properties present particular difficulties to an engineer. Thus, rock mechanics as an engineering discipline is concerned with the development of an understanding of rock as an engineering material, such that safe and economic design approaches can be identified. Generally, the geology of a particular rock is of secondary importance in rock mechanics: it is the behaviour of the rock, rather than its geological characteristics, that is of principal interest to the engineer. Notwithstanding this, an understanding of the geological setting of a rock is useful in rock mechanics, because a range of geological processes (e.g., the development of schistosity, or the form of a weathering profile) can impact greatly on the behaviour of rock as an engineering material.

It is now recognized that a key issue in understanding and applying rock mechanics is that, on the scale of most engineering projects, rock is fractured. As a result, rock often cannot be considered as a continuum, and a rational approach for dealing with the resulting discontinuum is to consider rock mechanics as a subject that comprises a number of clearly defined topics. To help understand the subject, these topics can be placed into three major categories: (1) the basic components of *in situ* stress, and the physical characteristics and behaviour of both fractures and intact rock; (2) rock masses as an amalgamation of these components; and (3) the interaction between rock engineering activities and rock masses.

The Basic Components: *In Situ* Stress, Fractures, and Intact Rock

The Earth's crust is almost everywhere subjected to a non-zero state of *in situ* stress. This state of stress is disturbed when any engineering operation takes place, and the perturbed stresses can cause a range of unwanted effects, such as stress-related failure of the rock. As a result, assessment of the *in situ* state of stress is a critical component of rock mechanics. Over the years, various techniques have been developed that allow engineers to determine, with varying degrees of completeness, the state of stress in a rock mass. Two of the most commonly used techniques are hydraulic fracturing and various so-called overcoring methods. In hydraulic fracturing, a short length of a borehole is filled with fluid and is pressurized. A fracture forms in the rock of the borehole wall when the fluid pressure reaches a critical value, and the magnitude of this pressure is used to calculate the *in situ* stress (Figure 1). Overcoring makes use of sophisticated strain measuring devices that are fixed with epoxy cement into a small-diameter borehole (Figure 2). A larger diameter borehole is then drilled coaxially with the first, and the device recovered. Continuous recording of the strains induced within device, and hence within the core of rock, then allows the *in situ* stress to be calculated.

Existing fractures are particularly important to engineers because they are usually much weaker than is the intact rock, and are hydraulically very conductive (i.e., readily permit fluid flow). Assessing the geometry, mechanical, and hydraulic characteristics of the fractures at a project site are thus key engineering activities. Simplified models are usually used to describe such characteristics. For example, fractures may be considered to be planar and of infinite extent, and to occur in sets in which the fractures are mutually parallel. The complete rock mass may then be considered to comprise a number of such sets. For

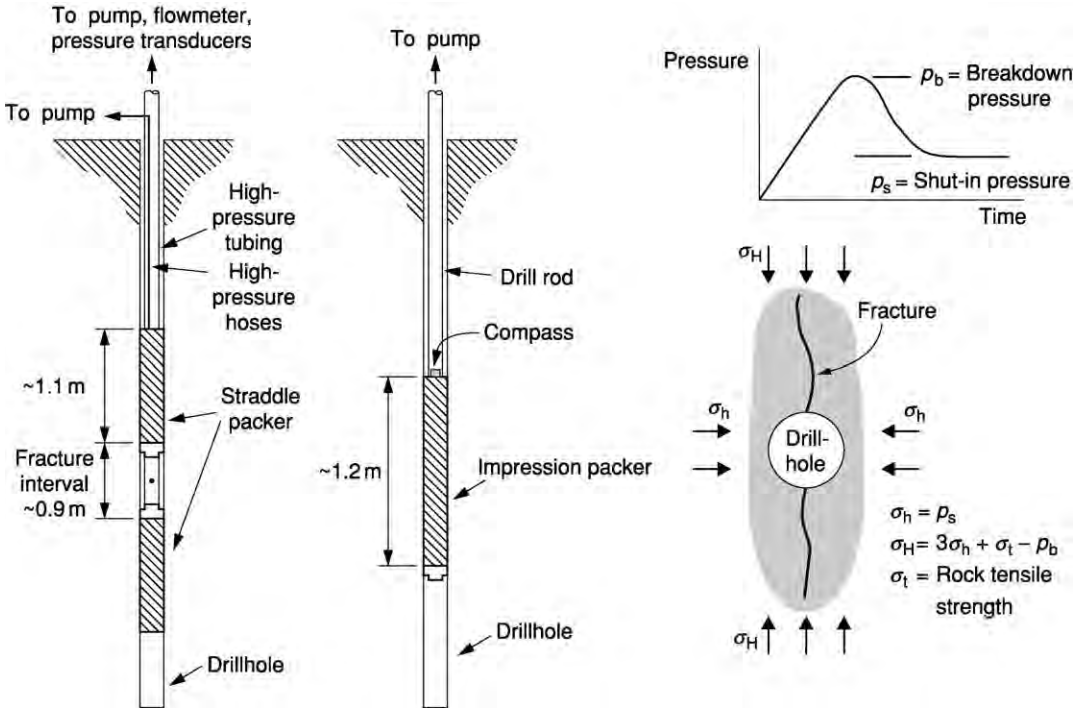


Figure 1 Hydraulic fracturing for estimation of *in situ* stress. Two longitudinal sections of the boreholes show the downhole equipment; the cross-section shows the form of the induced fracture and the equations governing its development.

Figure 2 An overcoring device for stress estimation, sectioned after recovery from the test borehole. The control wires enter at the left; the part of the cell carrying the strain gauges, and which is cemented into the borehole, is at the right.

mechanical characteristics, a simple zero-cohesion Coulomb model is often used for frictional strength, with the shear and normal stiffnesses of a fracture being modelled as linear or non-linear springs. Hydraulic models range from those whereby each fracture is simply represented by a pair of smooth parallel plates, through to sophisticated statistical models in which the complex variation of fracture aperture is reproduced. Although many of the models of fracture geometry and strength are simple, they nevertheless allow valuable insight into the behaviour of fractured rock masses. For example, superposition of the frequencies of a number of fracture sets allows assessment of the variation of frequency with orientation through the rock mass (Figure 3), with the result that the direction in which a borehole, for example, would intersect most fractures can be seen. This has ramifications for optimal borehole orientation in endeavours such as water, petroleum, and geothermal energy extraction.

Most rock mechanics investigations involve determination of various mechanical properties of rock, particularly strength and elastic moduli. Laboratory testing of intact rock primarily makes use of cylindrical specimens, loading them in a state of triaxial compression in specialized servo-controlled machines that allow strain, rather than stress, to be the independent variable (Figure 4). Using strain as the independent variable allows the post-peak behaviour of a rock to be determined; this is critical for accurately establishing the engineering behaviour of structures, such as support pillars in underground mines, that undergo strain-controlled loading. Mechanical testing of fractures is usually undertaken using a shear box, in which a specimen of rock containing a through-going fracture is mounted and subjected to both normal and shear loads (Figure 5).

The strength of intact rock is often modelled using a linear Coulomb criterion, but because the relation

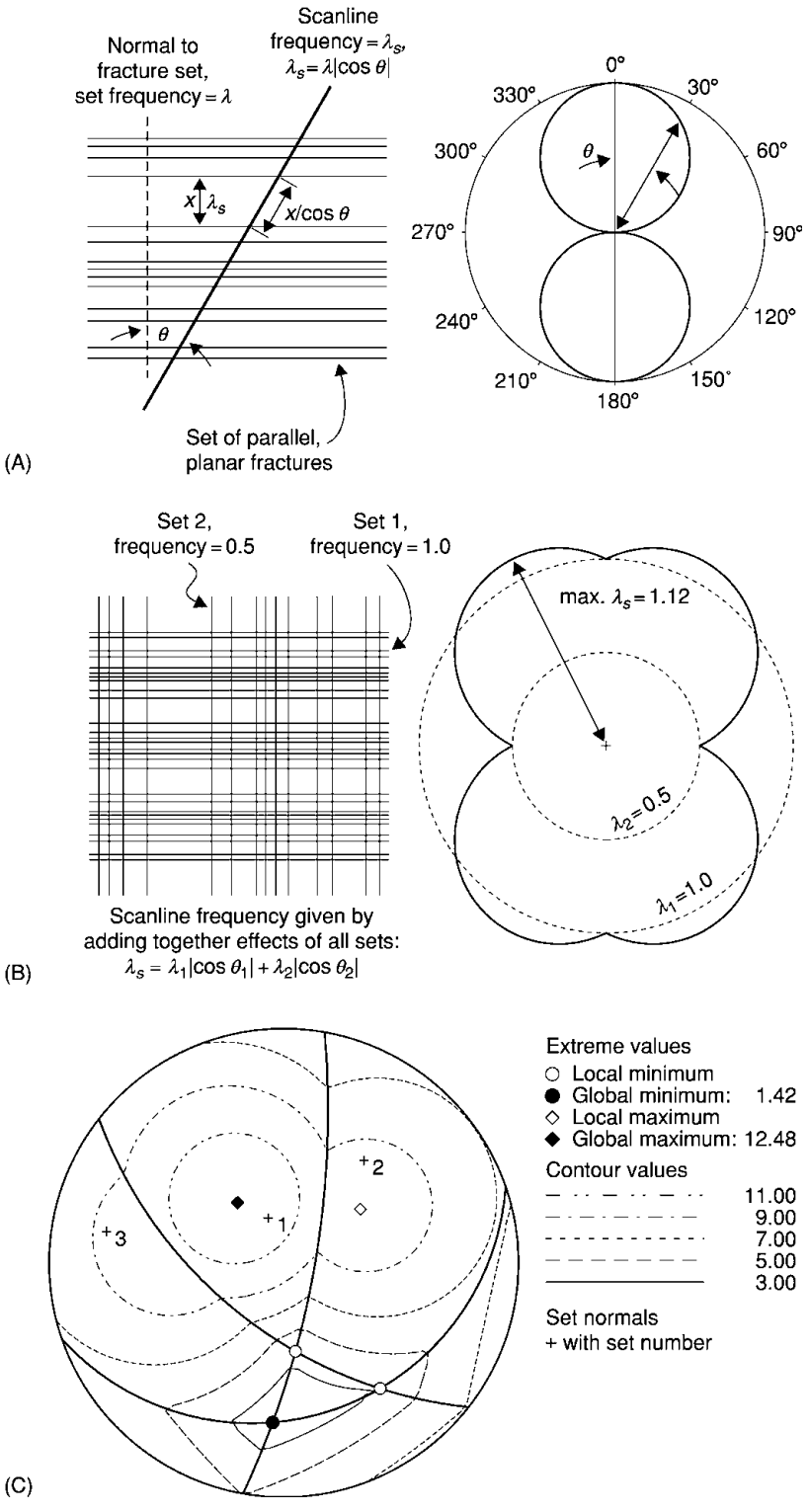


Figure 3 Predicting fracture frequency within a rock mass. (A) Frequency variation due to one set. (B) Frequency variation due to two orthogonal sets. (C) Frequency variation in three dimensions due to three sets. The fracture set orientation (dip direction/dip angle) and frequency values are as follows: set 1, 161/23 and 7.72 m^{-1}; set 2, 218/58 and 3.07 m^{-1}; set 3, 100/75 and 5.34 m^{-1}.

between strength and confining stress is known to be non-linear, increasing use is made of empirical non-linear strength criteria. One of the most popular of these is the Hoek-Brown criterion (Figure 6). Although the shear strength of fractures is similarly non-linear, for ease of calculation a simple linear Coulomb criterion is usually used for modelling.

The Engineering Material: Rock Masses

Rock masses can be usefully considered as being a composite material that comprises blocks of intact rock surrounded by fractures. Because both of these elements have their own mechanical characteristics, an estimate of the strength of the composite rock mass can be obtained by superposition of the component behaviours (Figure 7). This single-plane-of-weakness theory is widely used both for understanding the general strength of a rock mass and for determining the forces that tunnel support is required to supply. Although this approach can be used to examine the effect of more than one set of fractures, the usual approach for predicting the strength of fractured rock masses is through the use of non-linear empirical criteria. Over the years, many such criteria have been developed, but the one that appears to have become most widely used is the Hoek-Brown criterion. In its simplest form, this is a straightforward extension of the criterion used for intact rock, but with the addition of a parameter that describes the degree of fracturing of the rock mass (Figure 8).

It is often appropriate to consider intact rock as being impermeable, with the result that any fluid flow through a rock mass occurs along the fractures. A full, three-dimensional analysis of flow through a rock mass is not trivial, but a two-dimensional analysis can be undertaken relatively simply using a computer spreadsheet to solve simplified equations of flow. Figure 9 shows the results of an analysis undertaken for a two-dimensional network, in which the fractures have been modelled as pairs of parallel smooth plates, all with the same aperture. Rather than simply assessing the strength of a rock mass, an alternative approach is to assess its general engineering behaviour. Although this is a more nebulous concept compared to strength, through the use of a few, easily determined properties of a rock mass, a number of rock mass classification schemes have been developed that allow calculation of a single parameter that somehow uniquely quantifies the mass. Two of the most popular schemes in widespread engineering use are the so-called Q and rock mass rating (RMR)

Figure 4 A servo-controlled testing machine. To the left is the stack of control electronics. In front of the loading frame (to the right) is an assembly comprising a triaxial test cell, axial strain transducers, and load cell.

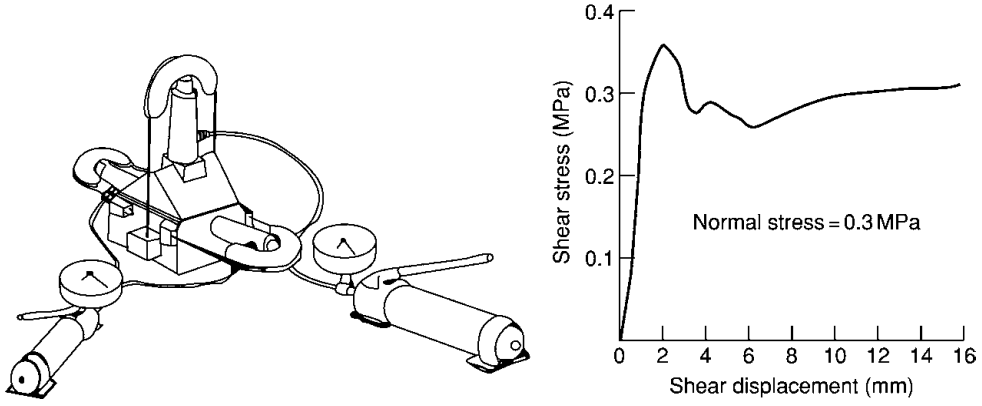


Figure 5 Schematic of a shear box for strength testing of fractures and a typical stress/shear displacement curve. The shear box is loaded by means of the hydraulic hand pumps (one for normal, the other for shear load).

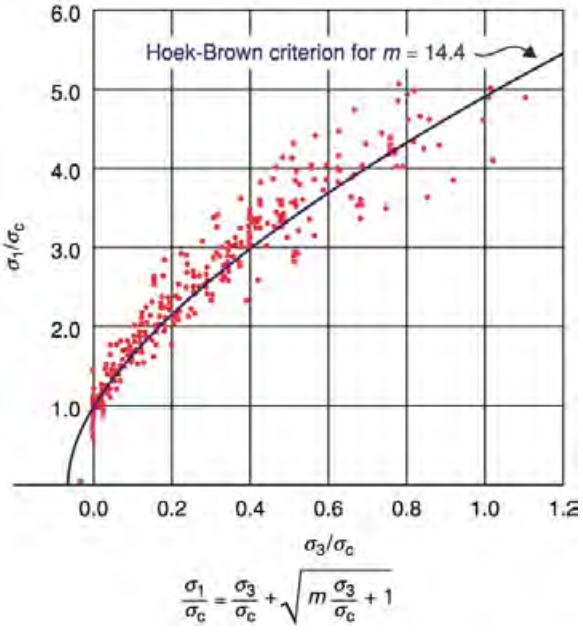


Figure 6 The equation and plot of the Hoek-Brown criterion for intact rock, where σ_c is unconfined compressive strength of the particular rock under consideration.

systems (Figure 10); for both of these, extensive tables are readily available that indicate what parameter values should be used for particular conditions. Over the years, correlations have been developed between these rock mass classification values and various engineering aspects such as ease of mechanical excavation, blasting requirements, and tunnel support requirements. As a result, and due to their inherent simplicity, the use of rock mass classification schemes is widespread.

Basing an appreciation of rock masses on the concept of a composite material comprising intact rock and fractures as two separate materials leads automatically to an awareness of the variability of rock masses with respect to location and orientation. Techniques for the treatment and understanding of variability are still in their infancy, but two acronyms that are useful in helping to concentrate on the issues are CHILE and DIANE. The first of these represents 'continuous, homogeneous, isotropic, and linearly elastic' behaviour; the second, which is somewhat more contrived, represents 'discontinuous, inhomogeneous, anisotropic, and not elastic' behaviour. CHILE rock is a hypothetical material, but its existence is relied on extensively in rock mechanics. For example, the presence of a CHILE material may be

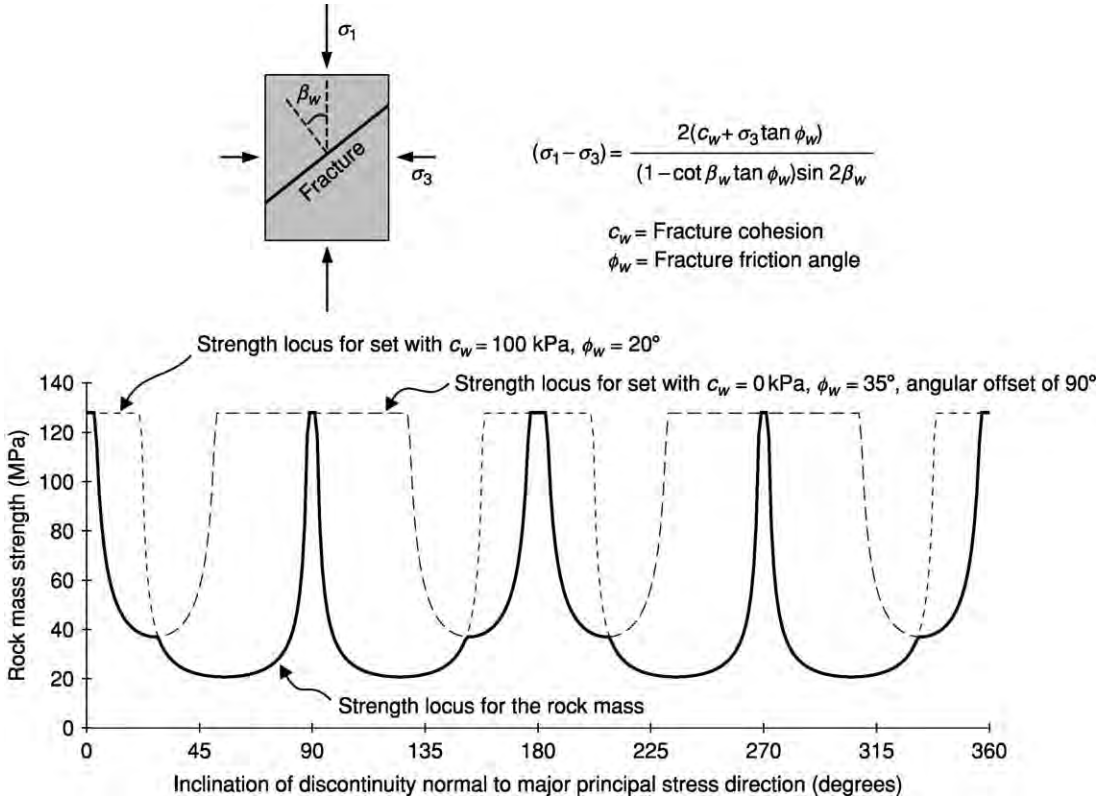


Figure 7 The single-plane-of-weakness theory and its extension to the strength of a fractured rock mass.

The Hoek–Brown criterion for fractured rock masses $\frac{\sigma_1}{\sigma_c}=\frac{\sigma_3}{\sigma_c}+\sqrt{m\frac{\sigma_3}{\sigma_c}+s}$	Values of s	Dolomite, limestone, marble	Mudstone, siltstone, shale	Sandstone, quartzite	Andesite, dolerite, rhyolite	Gabbro, gneiss, granite
		Values of m				
Intact rock samples Laboratory-sized specimens free from fractures	1.0	7	10	15	17	25
Very good quality rock mass Tightly locking undisturbed rock with unweathered fractures, spacing 1 to 3 m	0.082	2.40	3.43	5.14	5.82	8.56
Good quality rock mass Fresh to slightly weathered rock, slightly disturbed fractures, spacing 1 to 3 m	0.00293	0.575	0.821	1.231	1.395	2.052
Fair quality rock mass Several sets of moderately weathered fractures, spacing 0.3 to 1 m	0.00009	0.128	0.183	0.275	0.311	0.458
Poor quality rock mass Numerous weathered fractures, spacing 30 to 500 mm and with some gouge	0.000003	0.029	0.041	0.061	0.069	0.102
Very poor quality rock mass Numerous heavily weathered fractures, spacing less than 50 mm, gouge filled	0.0000001	0.007	0.010	0.015	0.017	0.025

Figure 8 The Hoek-Brown criterion for fractured rock masses, with typical values of the strength constants m and s for various qualities of different rock types.

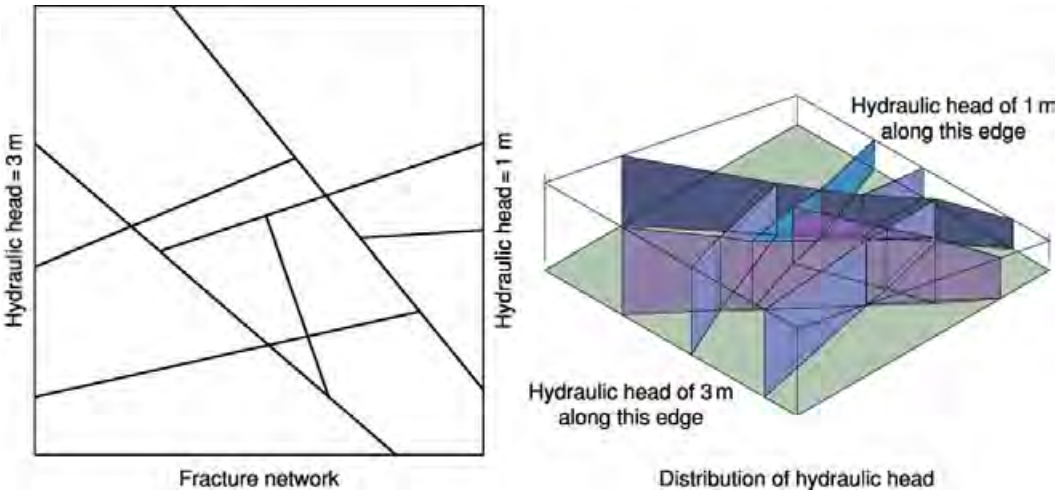


Figure 9 Distribution of hydraulic head in a two-dimensional fracture network.

invoked in the determination of *in situ* stress, characterization of intact rock, and analysis of stresses induced around underground excavations. The DIANE nature of real rock masses means that the effects of fractures, variability with respect to location and orientation, and the fact that stress changes will usually lead to irreversible deformations in the material must always be considered. A further important consequence of the DIANE nature of rock masses is the scale effect. In general, because of the presence of fractures, it is found that large volumes of rock and small specimens of intact rock have different properties. This means that *in situ* rock mass strength and stiffness will be much lower than that determined in the laboratory. Similarly, the large-scale permeability of a fractured rock mass is very likely to be much greater than that of a specimen of intact rock. The scale effect is thought to apply to other properties, such as *in situ* stress, but in these cases the consequence of it is less predictable. For example, under

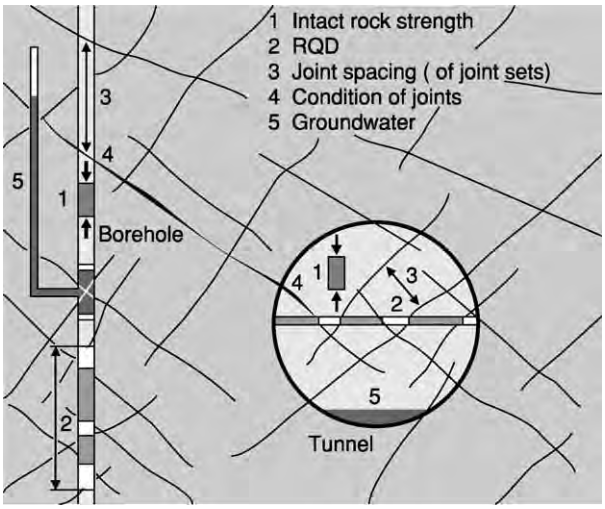


Figure 10 Schematic of typical rock mass properties included in a rock mass classification scheme, and how they may be assessed in both boreholes and tunnels. RQD, Rock quality designation.

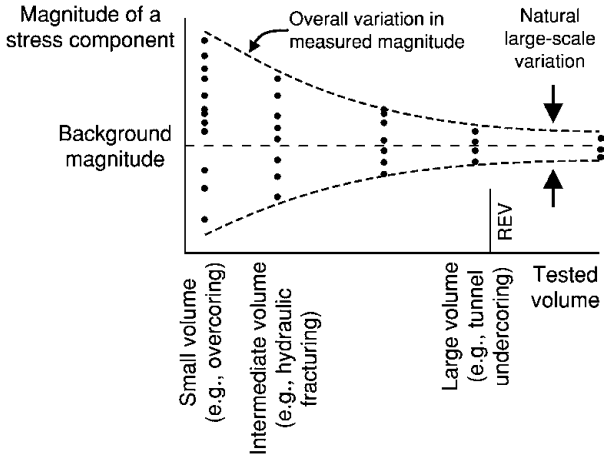


Figure 11 Illustration of scale effect and the representative elemental volume (REV) for *in situ* stress.

different circumstances, the measured state of stress in a small volume of rock could be less, the same as, or greater than the actual state of stress in a large volume. In all cases with the scale effect, the concept of a representative elemental volume (REV) is useful. As Figure 11 shows, there may be substantial variation in measured values or properties at small scales, but as the REV is approached the variation reduces until, at the REV and above, the variation is sufficiently small that it can be disregarded. Figure 11 also shows that the distribution of values at any particular scale may differ from those at another scale. This implies that it is usually not possible to use measurements of properties at one scale to estimate properties at another scale, and hence exemplifies the ideal requirement of performing rock mechanics tests on a volume of rock comparable to that of the final engineering structure. Unfortunately, this is seldom possible due to the constraints of cost and time.

Rock Engineering Topics

The development of rock mechanics concepts has in large part been driven by the requirement to solve problems in rock engineering. For example, although the rock engineering associated with mining and quarrying operations has a long history, the desire to exploit reserves at greater depths and hence in more onerous geomechanical environments has required a more rigorous approach than custom and practice can deliver. Endeavours such as the drilling of wellbores thousands of metres deep for petroleum extraction, or the requirement to ensure the long-term stability of steep rock slopes many tens of metres high adjacent to highways, has required additional developments in rock mechanics. Development efforts continue, driven by problems without precedent, such as those concerning the ultra-long-term stability of engineering designs for radioactive waste repositories. Regardless of the specific engineering application and its environment, the fundamental effects that an excavation has on a rock mass must always be borne in mind. These are summarized in Figure 12.

Fundamental techniques used in rock engineering are applicable to a range of engineering topics. These

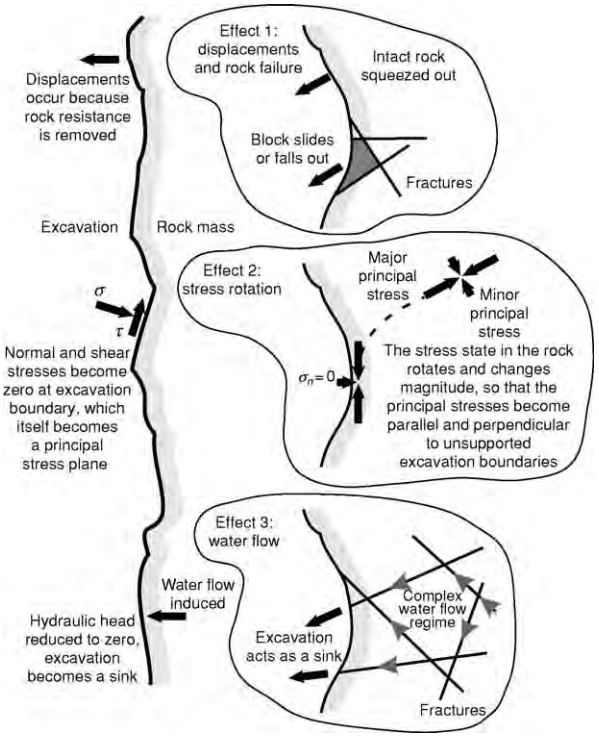


Figure 12 The fundamental effects of an excavation on a rock mass.

techniques can conveniently be classified in terms of those applicable to rock that is obviously discontinuous, and those that are strictly valid only for continuous rock. This effectively separates the applications into those for near-surface structures (such as slopes, foundations, and tunnels) and those for structures at depth (such as mine excavations and petroleum wellbores).

Engineering in Fractured Rock

Near-surface engineering is often controlled entirely by the strength and geometry of fractures. It is common, and conservative, to assume that the strength of fractures is purely frictional, i.e., the fractures have zero cohesion. This has an added advantage in that, in the absence of any external forces such as those generated by water pressure, an analysis of static equilibrium reduces to one that is governed by the orientation of the fractures. These orientations may conveniently be determined using standard hemispherical projection techniques. Of course, the blocks of rock that exist in a rock mass may possess a great many shapes and sizes. In order to develop a tractable problem, only tetrahedral blocks of the maximum size that can fall or slide into an excavation are analysed. Such tetrahedral blocks are generally formed from either two fracture surfaces and two excavation surfaces or from three fracture surfaces and one excavation surface. Polyhedral blocks (i.e., those possessing five or more surfaces) are usually ignored, both for analytical convenience and because such a block can be cut from, and hence will be smaller and less critical than, a tetrahedral block.

An example of how the hemispherical projection is used in rock block stability analysis is given by the case of a rock slope containing a wedge of rock that is delineated by two fracture surfaces, as shown in Figure 13. This figure also shows how the stability of this wedge is a function of the friction angle of the fractures and three angles that relate to the geometry of the fractures. These latter three angles can be measured directly and quickly from the hemispherical projection. Further simplification of this analysis is possible by ignoring the stabilizing effect of the wedge factor, i.e., the quotient $\sin\beta/\sin 1/2\xi$. This leads to kinematic analyses, in which only the feasibility of instability, given the interaction between the fracture and excavation geometry, is considered.

Kinematic analyses test for whether the geometry of the fractures and the excavation surface are capable of defining blocks that can be released into

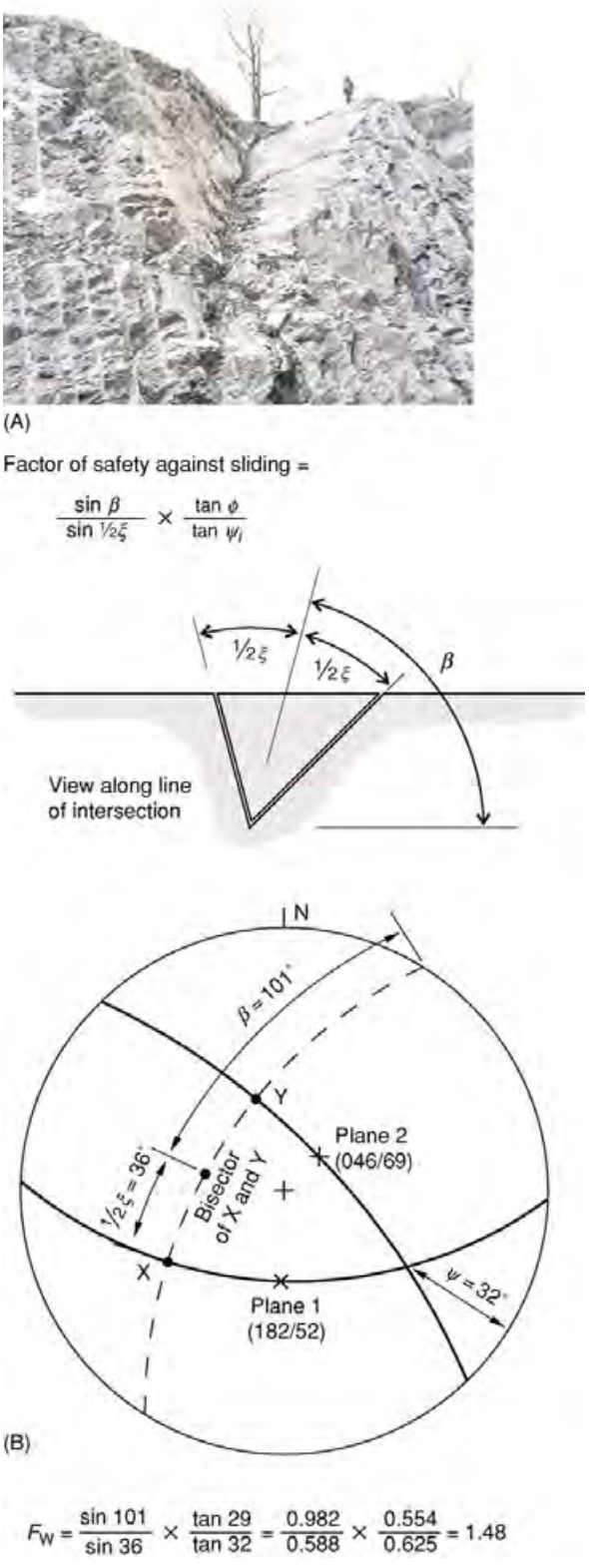


Figure 13 Illustration and analysis of wedge instability in a rock face. (A) Basic equation for factor of safety against sliding, where β and ξ are shown in the sketch, ψ_i is the plunge of intersection and ϕ is the friction angle of the fractures. (B) The factor of safety against wedge sliding, F_W, is 1.48 for the case of two fractures of orientation 046/69 and 182/52 and a friction angle of 29°. Dashed line = plane whose normal is the intersection of planes 1 and 2.

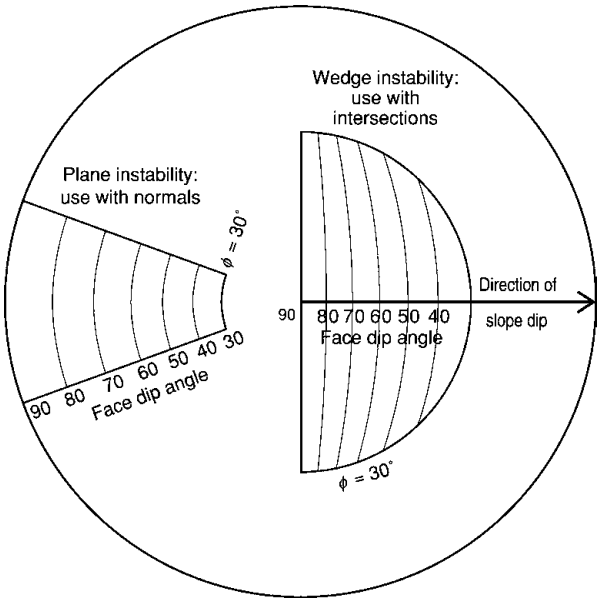


Figure 14 Example of kinematic assessment overlay for rock slope instability due to plane and wedge sliding.

the excavation, and therefore may be unstable. Again, for no external forces and fractures possessing only frictional strength, the kinematic criteria relate to orientations only, with the result that, once more, the hemispherical projection can be used rapidly to assess stability. Figure 14 shows an overlay that can be used for assessing instability due to plane and wedge sliding in a rock mass for which the friction angle is 30°. To use this overlay, the fracture data (both poles and great circles) are plotted on a sheet of tracing paper and placed on a hemispherical projection net, with a second sheet of tracing paper on which the assessment overlay is drawn placed over the top. The overlay is then rotated to identify those slope orientations that give rise to either poles or intersections falling within the unstable zones. Because the overlay also shows various slope angles, it can quickly be used to select a slope angle that avoids the kinematic conditions leading to instability.

A similar analysis can be undertaken for underground excavations. Figure 15 shows three hemispherical projections illustrating the unstable kinematic conditions that apply to a tetrahedral block formed by three fractures and the horizontal roof to a tunnel or cavern. These blocks are identified on the projection as areas bounded by three great circles. If the vertical direction (i.e., the centre of the projection) lies within the projection of the block, this indicates that the block is able to fall. In the case of sliding instability, the direction of sliding is the steepest of either a line of maximum dip or the intersection of two planes. For blocks that are identified as being

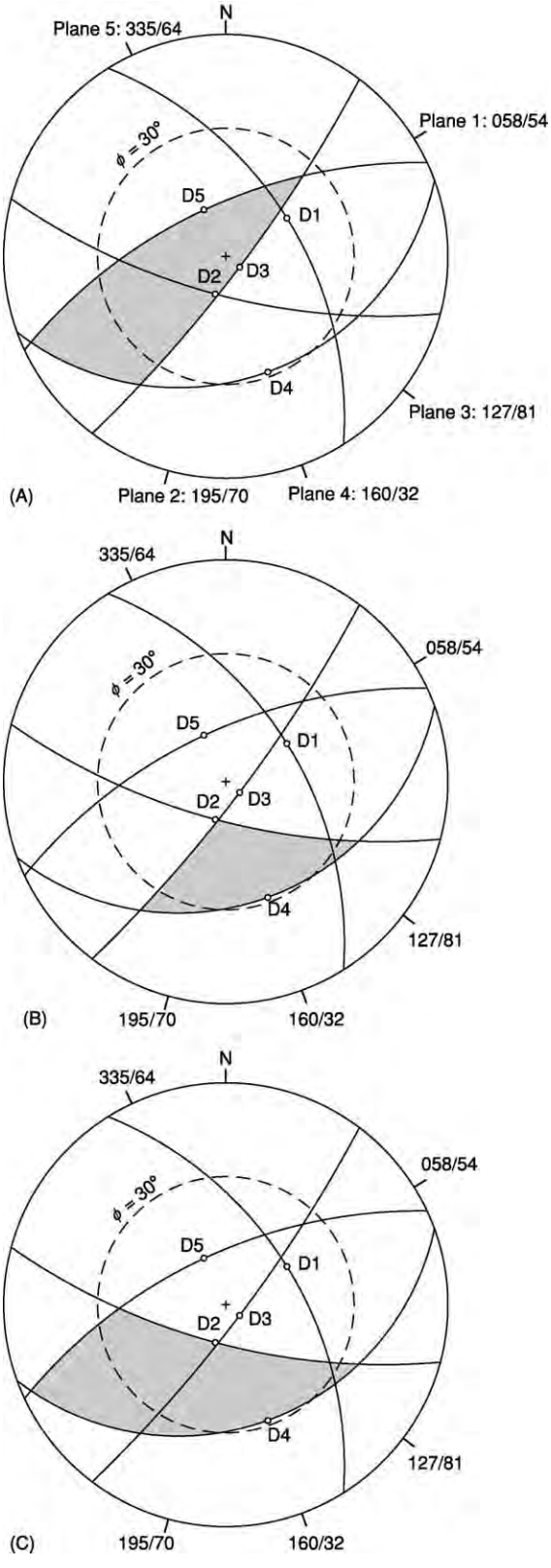


Figure 15 Identification of falling, sliding, and stable blocks in the roof of an underground excavation. (A) Block 345 falls vertically. (B) Block 234 slides on the intersection of planes 2 and 3. (C) Block 245 slides on the line of maximum dip of plane 2.

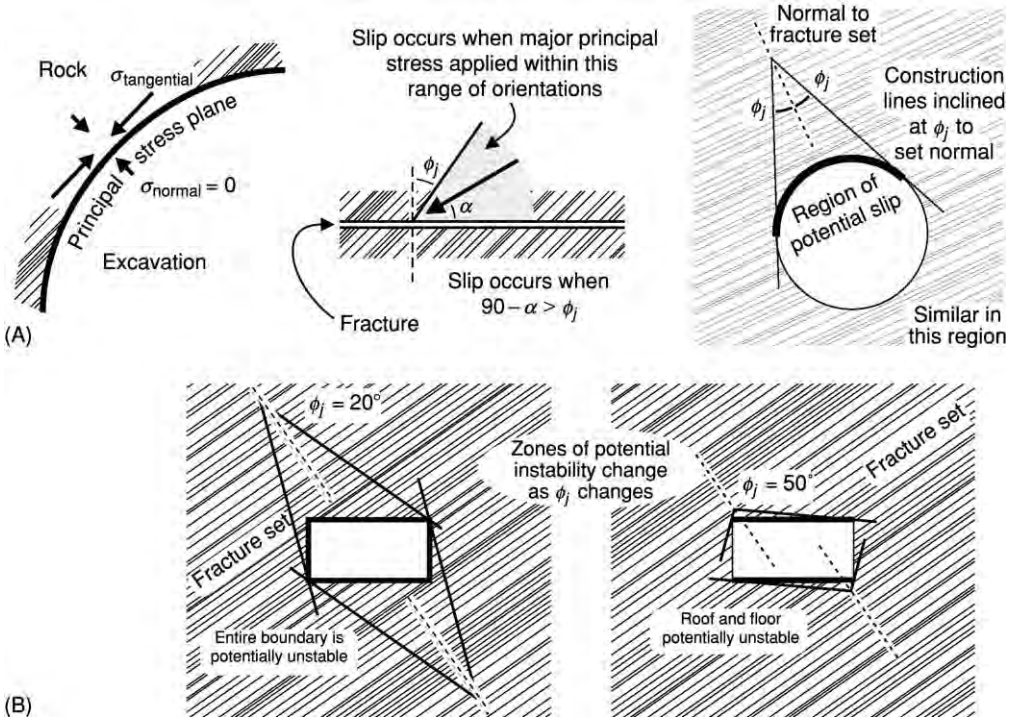


Figure 16 The ϕ_j theory for underground excavations. (A) The fundamental stress conditions leading to slip, and how these are identified in the ϕ_j theory. (B) Application to underground excavations in stratified materials with different frictional properties.

unstable, a further rigorous equilibrium analysis can be undertaken in order to determine their size and hence support requirements. These analyses can be time consuming and prone to error if undertaken by hand, and so are usually performed by computer programs.

Another form of kinematic analysis that is widely used is the so-called ϕ_j theory. This uses the principle that, if shearing along a fracture is to be avoided, the direction of action of the major principal stress acting on the fracture must lie within an angle ϕ_j, the friction angle, of the normal to the fracture. Now, close to an underground excavation the major principal stress is tangential to the boundary. This means that a geometric analysis can be used to identify those zones on the boundary where the orientation of the major principal stress is such as to induce shearing. This is illustrated in Figure 16, for the case of a rock mass that contains one dominant set of fractures, and hence can be assumed to be transversely isotropic.

Techniques for the analysis of foundations on fractured rock are not as straightforward as for slopes or underground excavations. However, with most engineering structures it is possible to use foundations that are sufficiently large as to impose stresses that are small enough to prevent instability from occurring. If it can be assumed that the rock beneath the foundation is continuous, then the induced stresses may be determined using analytical solutions developed from the theory of elasticity. This approach is widely used in soil mechanics, wherein the assumption of a continuum is often valid, but it is seldom used in rock mechanics for the reason that most near-surface rock masses are fractured. However, the concept of 'bulbs of pressure', or contours of major principal stress, from these analyses is useful in rock mechanics. As the ϕ_j theory shows, as the direction of action of the major principal stress is constrained by the presence of a fracture, the shape of the resulting stress contours is controlled by the fracture geometry. Figure 17 shows a simplified solution for the radial stress that is applicable to the case of rocks that contain one dominant set of fractures, and so can be assumed to be transversely isotropic. The contours of major principal stress are influenced strongly by the fracture geometry, showing that foundations on rock can induce high stresses in zones far removed from the foundation.

Engineering in Continuous Rock

For underground excavations in continuous rock, or in locations where the *in situ* stresses are sufficiently high as to allow a fractured rock to act as if it were continuous, it is the stress induced around the excavation that leads to instability. Simple analytical solutions for the induced stresses are available for only a restricted set of excavation geometries, namely, circular and elliptical openings that are in a state of plane strain. This last constraint requires the excavations to be long, of constant cross-section, and with

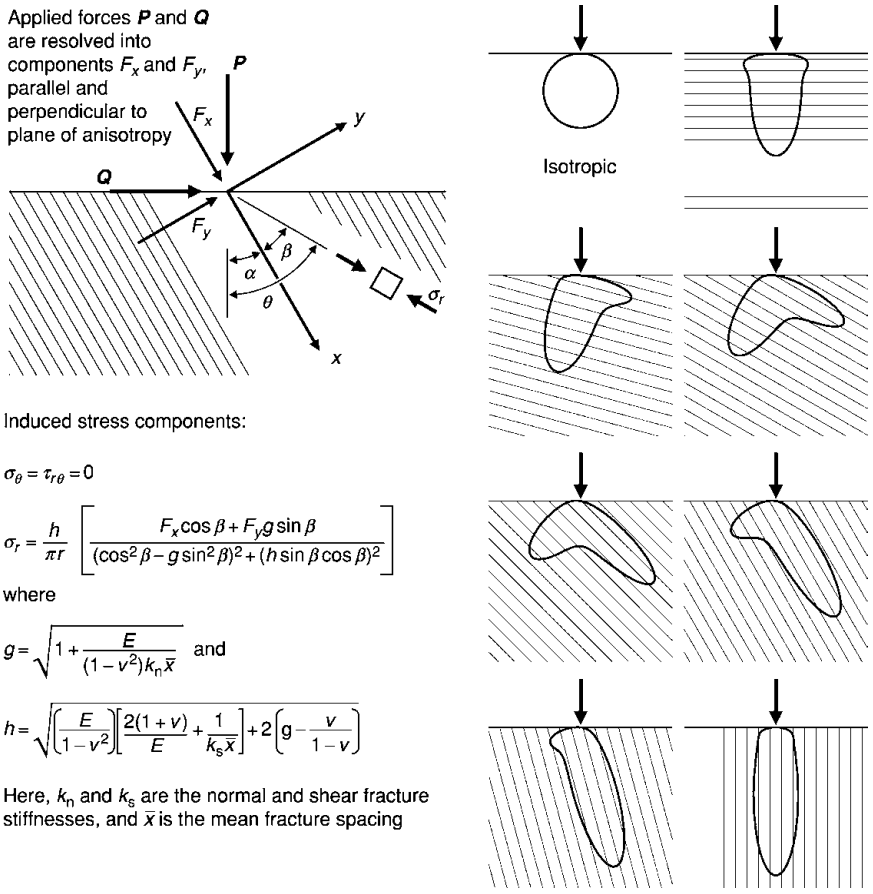


Figure 17 Simplified relations for contours of constant radial stress for a foundation in stratified material resembling a transversely isotropic rock.

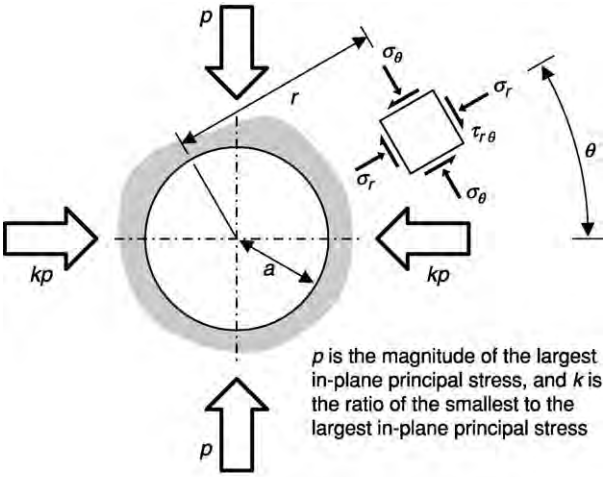
their axial direction parallel to one of the *in situ* principal stresses. Despite these severe restrictions on geometry, the fact that engineers are commonly required rapidly to assess the stability of underground excavations means that these solutions are used extensively.

Figure 18 shows a cross-section through a circular opening and presents the well-known Kirsch equations for the induced-stress components. These equations are used for many applications involving the analysis of stress around tunnels, wellbores, and mine shafts. As an example, Figure 18 also shows simplified relations for the specific case of the maximum and minimum values of tangential stress induced at the boundary of an excavation, and how these components vary as the *in situ* stress ratio varies. The figure shows how the maximum and minimum stresses are induced under uniaxial conditions, represented by $k = 0$, and that for isotropic conditions, represented by $k = 1$, the induced stress is constant around the boundary.

For the cases of stress induced around an elliptical excavation, a particularly elegant solution is available for the tangential stress induced at the boundary. This is shown in Figure 19. Also, for the particular case of an elliptical opening with a horizontal major axis, the extreme boundary stresses, which are induced in the side walls, roof, and floor, are also given. These final two equations can be equated, and doing so shows that the tangential stress is constant around the boundary in the case of an elliptical opening that has an aspect ratio equal to the *in situ* stress ratio. This elegant result is a valuable design guide for engineers wishing to minimize the range of stresses induced around the boundary of an excavation.

Numerical Analysis

With all of the fundamental analyses that are routinely used in rock mechanics and rock engineering, various simplifying assumptions are required in order to make the problem tractable. These simplified procedures are perfectly acceptable for the majority of applications, but for applications requiring an accurate understanding of the behaviour of a structure, more complex analyses are necessary. These are performed using computer programs that implement sophisticated numerical techniques. The techniques most often used in rock mechanics are finite element, boundary element, and finite difference methods for continua, and various discrete element methods



$$\sigma_r = \frac{p}{2}\left[(1+k)\left(1-\frac{a^2}{r^2}\right)-(1-k)\left(1-4\frac{a^2}{r^2}+3\frac{a^4}{r^4}\right)\cos 2\theta\right]$$

$$\sigma_\theta = \frac{p}{2}\left[(1+k)\left(1+\frac{a^2}{r^2}\right)+(1-k)\left(1+3\frac{a^4}{r^4}\right)\cos 2\theta\right]$$

$$\tau_{r\theta} = \frac{p}{2}\left[(1-k)\left(1+2\frac{a^2}{r^2}-3\frac{a^4}{r^4}\right)\sin 2\theta\right]$$

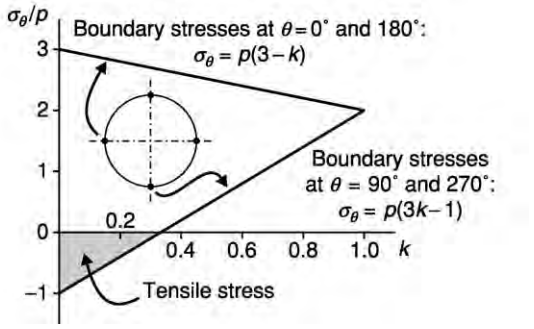
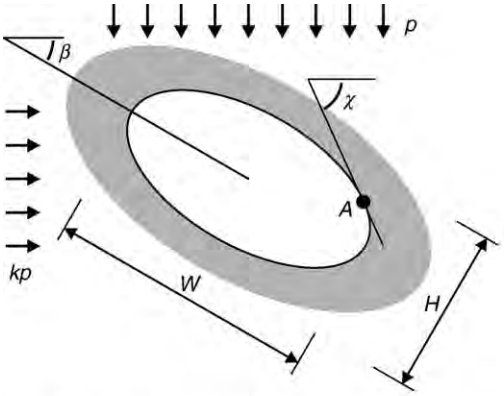


Figure 18 The Kirsch solution for stresses around a circular opening in isotropic continuous rock, under conditions of plane strain.



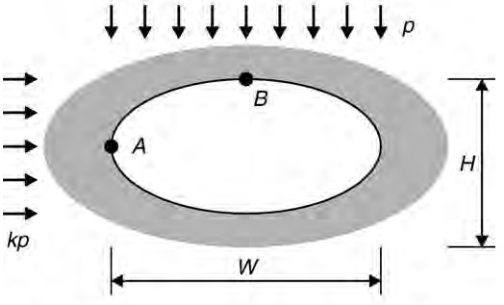
Boundary stress at point A:

$$\sigma_\theta = \frac{p}{2q}\{(1+k)[(1+q^2)+(1-q^2)\cos 2(\chi-\beta)]$$

$$-(1-k)[(1+q^2)\cos 2\chi+(1-q^2)\cos 2\beta]\}$$

where $q = W/H$,

p is the magnitude of the largest in-plane principal stress, and k is the ratio of the smallest to the largest in-plane principal stress



Extreme boundary stresses:

$$\sigma_A = p(1-k+2q)$$

$$\sigma_B = p\left(k-1+\frac{2k}{q}\right)$$

Figure 19 Simplified relations for the tangential stress at the boundary of an elliptical opening in isotropic continuous rock, under conditions of plane strain.

for discontinua. The background theory to these methods is complex, and their use is not straightforward. However, they are able to produce models and simulations of great accuracy and detail, and as researchers continue to implement new developments, the results obtained using them become ever more realistic. There is no doubt that the use of such procedures will become more widespread in the future.

See Also

Geological Engineering. **Geotechnical Engineering**. **Quarrying**. **Soil Mechanics**.

Further Reading

Amadei B and Stephansson S (1996) *Rock Stress and Its Measurement*. London: Chapman & Hall.

Harrison JP and Hudson JA (2000) *Engineering Rock Mechanics: Illustrative Worked Examples*. Oxford: Pergamon.

Hoek E and Bray JW (1981) *Rock Slope Engineering*, 3rd edn. London: Institution of Mining and Metallurgy.

Hoek E and Brown ET (1980) *Underground Excavations in Rock*. London: Institution of Mining and Metallurgy.

Hudson JA (ed.) (1993) *Comprehensive Rock Engineering*. Oxford: Pergamon Press.

Hudson JA and Harrison JP (1997) *Engineering Rock Mechanics: An Introduction to the Principles*. Oxford: Pergamon.

Jaeger JC and Cook NGW (1979) *Fundamentals of Rock Mechanics*. London: Chapman & Hall.

ROCKS AND THEIR CLASSIFICATION

R C Selley, Imperial College London, London, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction: What on Earth is a Rock?

To a geologist the term 'rock' is applied to all the solid materials of the Earth, whether they are hard enough to hit with a hammer or soft unconsolidated gravel, sand, or mud. The nomenclature and classification of rocks is a prerequisite to understanding them. One of several definitions of science is that it is organized knowledge. It is the organization and classification of knowledge that provides insight into processes and adds predictive value. The Periodic Table of the Elements illustrates these points admirably. Classification is an integral part of geology in all its forms. It extends from palaeontology, where the classification of fossils according to the binomial nomenclature of Linnaeus is a prerequisite to demonstrating evolution and understanding palaeoecology, to mineralogy, where minerals are grouped into like families according to their chemistry and crystallography.

The object of this chapter is to review the fundamental classification of rocks as a foundation for discussing igneous, sedimentary, and metamorphic rock nomenclature and classification (*see* **Sedimentary Rocks:** Mineralogy and Classification, **Metamorphic Rocks:** Classification, Nomenclature and Formation).

Rocks are traditionally grouped into three main classes, according to their genesis. These will now be defined and described. All generalizations are dangerous, however, including this one. Thus, a detailed examination of the classification (*see* **Igneous Processes**) of rocks soon reveals discrepancies and inconsistencies (*see* **Metamorphic Rocks:** Classification, Nomenclature and Formation). It may, therefore, be helpful to examine these subsequently.

A Classification of Rocks

Rocks are of three genetic types: igneous (formed from cooling magma), sedimentary (formed by the breakdown of pre-existing rock), and metamorphic (formed by the action of heat and pressure on pre-existing rock) (Figure 1 and Table 1).

Igneous Rocks

Magma is molten material from deep beneath the surface of the Earth. Through a variety of processes discussed elsewhere in this encyclopaedia, magma may move up through the crust towards the surface. It may cool very slowly beneath the surface to form coarsely crystalline rocks. These include silica-rich rocks, such as granite (*see* **Igneous Rocks:** Granite), and ferromagnesian-rich rocks, such as gabbro. If the magma reaches the surface, it erupts in volcanoes (*see* **Volcanoes**), and the melt cools quickly to form microcrystalline or even glassy rocks. Depending on their chemistry and texture, volcanic rocks may be classified as silica-rich rhyolite or ferromagnesian-rich basalt.

Thus, owing to their mode of formation, igneous rocks tend to be structureless when viewed at outcrop or in a hand specimen. Under the microscope they are seen to consist of a random arrangement of interlocking crystals of a range of different minerals, which formed when the melt crystallized. Porosity is generally absent.

Sedimentary Rocks

Sedimentary rocks are formed from the detritus of pre-existing rocks, which may be igneous, metamorphic, or sedimentary. The way in which rock is

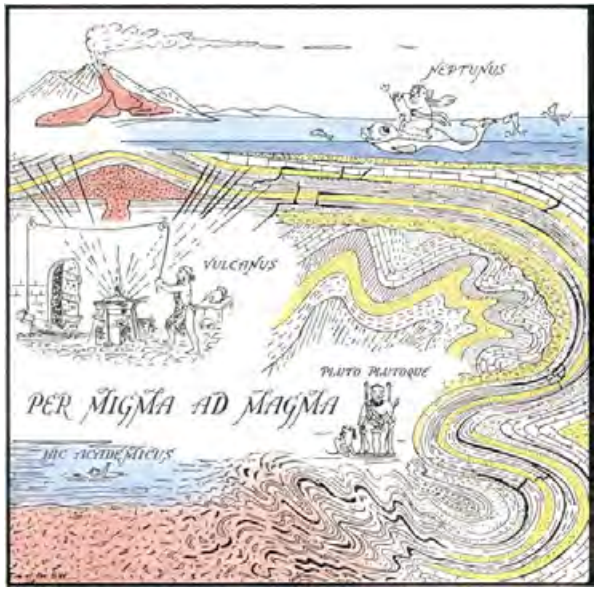


Figure 1 The Rocks Display'd. An illustration of the modes of formation of sedimentary, igneous, and metamorphic rocks, accompanied by their classical deities Neptune, Vulcan, and Pluto (together with his synonymous dog). Reproduced by permission of The Geologists' Association from *Proceedings of The Geologists' Association*, Wilson G, in Read HH, Meditations on Granite 2, volume 55, pp. 4–93. Fig. 1. © 1944 The Geologists' Association.

Table 1 The three main classes of rock, their origins, and examples

Class of rock	*Origin*	*Examples*
Sedimentary	Deposition of particles of pre-existing rocks	Gravel, sand, and mudstone
	Chemical replacement and precipitation	Limestone dolomite and salt
Metamorphic	Heated by igneous intrusions	Hornfels
	Heated and subjected to high pressure by deep burial	Slate, schist, and gneiss
Igneous	Volcanic eruptions	Basalt, andesite, and rhyolite
	Deep cooling of magma	Granite, diorite, and gabbro

weathered, eroded, transported, and deposited is discussed in detail elsewhere (*see* **Weathering, Sedimentary Processes:** Erosional Sedimentary Structures, **Sedimentary Environments:** Depositional Systems and Facies). Sediments possess a wide range of particle sizes, ranging from boulders to clay, and of chemical compositions, including silica, lime, and ferromagnesian volcanic detritus. These parameters of particle size and composition are used to classify sedimentary rocks (*see* **Sedimentary Rocks:** Mineralogy and Classification). Sedimentary rocks commonly exhibit two properties that may be used to differentiate them from igneous and metamorphic rocks. Where they crop out at the surface of the Earth, sedimentary rocks generally show stratification. The strata indicate successive episodes of deposition. Layering is usually absent from igneous rocks, as discussed above, but is found in some metamorphic rocks, as discussed below. When examined under the microscope, sedimentary rocks are generally seen to consist of particles. Void space (porosity) is commonly present between the constituent grains. If there are interconnected pores, these give the rock permeability. Permeability allows fluids to migrate through the rock and enables rock and soil to drain. Fossils are found only in sedimentary rocks, some of which are, indeed, made up of nothing else.

Sedimentary rocks include sands, including terrigenous silica-rich sandstone, and carbonates, termed limestone. Sediments composed of rounded pebbles are termed conglomerates, while those composed of angular clasts (fragments) are termed breccias.

Metamorphic Rocks

Metamorphic rocks are so termed from the Greek *meta-*, meaning altered, and *morphos-*, meaning shape. The word is applied to rocks, of whatever previous origin, that have been changed as a result of being exposed to high temperature, high pressure, or both. Two types of metamorphism are recognized, regional metamorphism, and thermal or contact metamorphism. When magma moves up through the crust it creates a metamorphic aureole in the adjacent cover rocks. This is termed contact or thermal metamorphism. By contrast, rocks that are subjected to high temperatures and pressures during deep burial are said to have undergone regional metamorphism.

Common thermally metamorphosed rocks include hornfels and quartzite. Regionally metamorphosed rocks include schist and gneiss.

Thermally metamorphosed rocks are often crystalline and devoid of porosity. They may be massive or layered, according to the character of the original rock. Regionally metamorphosed rocks are also commonly crystalline and devoid of porosity. They are, however, often banded. The banding seldom resembles the regular stratification of sediments, but often takes the form of curvaceous foliation and cleavage (*see* **Regional Metamorphism**).

Rock Classification: A Simple Summary

Igneous and metamorphic rocks are both normally composed of a mosaic of interlocking crystals that formed at high temperatures and pressures; they thus usually lack porosity and permeability. In the petroleum industry igneous and metamorphic rocks are generally grouped together under the term 'basement', indicating that it is pointless to continue drilling deeper into such tight rocks. Igneous rocks are generally massive and structureless, while metamorphic rocks are often banded.

Sediments, on the other hand, are deposited with space between the constituent particles. Thus, most sedimentary rocks, by their very nature, tend to be porous and permeable. Sediments are deposited episodically and thus characteristically exhibit stratification. They are also the only one of the three main classes of rock to contain fossils. Igneous, metamorphic, and sedimentary rocks may therefore be differentiated from one another using macroscopic structure and microscopic texture as shown in Figure 2.

Complications and Anomalies

When the tripartite classification of rocks was introduced at the beginning of this article it was pointed out that, like all classifications, it has complications

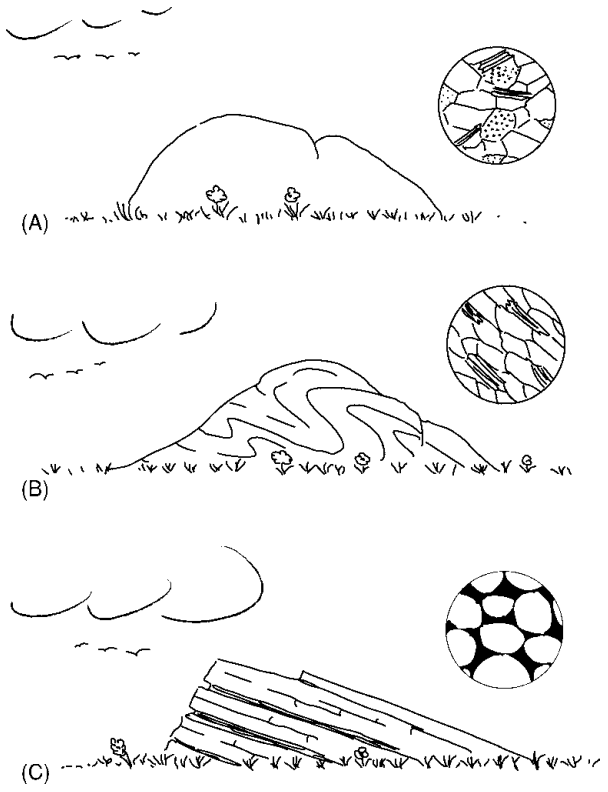


Figure 2 A rough guide to how igneous, metamorphic, and sedimentary rocks may be differentiated from one another by macroscopic structure (left) and microscopic texture (right). As discussed in the text, however, there are exceptions to these simple guidelines. (A) Igneous rocks are generally structureless in the field, with a random mosaic of crystals when viewed in thin section under the microscope. (B) Metamorphic rocks are generally banded (often curvaceous) in the field, with a linear arrangement of crystals when viewed in thin section under the microscope. (C) Sedimentary rocks are generally stratified in the field and exhibit rounded grains with intergranular pores when viewed in thin section under the microscope.

and anomalies. These will now be examined for igneous, metamorphic, and sedimentary rocks in turn.

Igneous Rock Anomalies

When describing igneous rocks it was stated that they form from cooling magma and are generally composed of a mosaic of interlocking randomly arranged crystals. Thus, in hand specimen, and at outcrop, igneous rocks present a massive appearance. They generally lack the foliation of metamorphic rocks and the stratification of sediments. Now for the exceptions, which occur in both volcanic lavas and major intrusions.

Lava is extruded from a volcano episodically, on time-scales ranging from thousands of years to minutes. Major sequences of lava flows give rise to a very typical landform, with plateaus of lava bounded by cliffs in which successive lava flows can be discerned, separated from one another by weathered soil horizons that formed between individual flow events. Superficially this gives rise to landscapes analogous to those created by stratified sedimentary rocks. Plateau basalts are widespread around the world, from the Deccan Traps of India to the Tertiary volcanic province of the Scottish Hebridean islands. When examined closely, lavas commonly exhibit flow-banding on a variety of scales. This includes the 'ropy' structure of some basalt flows, the so-called 'pahoehoe' lavas of Hawaii, and a diverse range of structures that are morphologically similar to the post-depositional structures of sedimentary rocks (*see* **Sedimentary Processes:** Post-Depositional Sedimentary Structures). For example the Devonian rhyolites of the Glen Coe cauldron subsidence in Scotland exhibit excellent examples of slumps, slides, load-casts, and flame structures.

It was also noted earlier that igneous rocks generally lack porosity. Some gassy lavas, however, consolidate with contained bubbles of gas, giving rise to what are termed vesicular lavas. Vesicular lavas thus exhibit porosity, but, unless the bubbles are abundant and interconnected, there is no significant permeability.

Deep-seated igneous intrusions also sometimes show features indicative of sedimentation. These are most commonly found in ultrabasic ferromagnesian-rich intrusives, such as the gabbroic Skaergaard complex of Greenland and the Bushveld intrusion of South Africa. These intrusions show a gravity segregation of minerals, with the densest, such as chromite and iron ores, at the base, overlain by ferromagnesian minerals, and then feldspars above. It is clear that the minerals have crystallized out of a melt, and have settled under gravity. There is more. This gravity separation may occur not only on an intrusion-wide scale, but also on a scale of metres or centimetres, resulting in size-graded 'beds' analogous to those deposited by sediment gravity flows. The evidence suggests that convection currents within the magma chamber permit successive episodes of deposition from density flows. Inclined bedding analogous to that produced by traction currents in running water is also known (*see* **Unidirectional Aqueous Flow**).

Sedimentary Rock Anomalies

Sedimentary rocks are defined as those that are composed of fragments of pre-existing rocks, of a wide range of particle sizes, that are deposited under ice or water or in the atmosphere. They are thus commonly stratified and contain fossils, and usually exhibit porosity between the constituent sediment grains. This definition is adequate for gravels, sands, and muds and for the conglomerates, sandstones, and shales that are their lithified equivalents. There is,

however, a large group of sedimentary rocks where these characters are largely absent. These are the chemical sediments. The chemical sediments are a diverse group that includes limestones, dolomites, ironstones, evaporites, chert, hydrocarbons (peat, lignite, and coal), and phosphate (*see* **Sedimentary Rocks:** Chert; Dolomites; Evaporites; Ironstones; Limestones; Phosphates). Limestones are frequently composed of clasts and exhibit sedimentary structures analogous to those of sandstones. The other chemical rocks, however, owe their peculiar properties very largely to post-depositional chemical changes during diagenesis (*see* **Diagenesis, Overview**). In the extreme case of the so-called evaporites, it is now recognized that they should be more properly termed 'replacementites' since their original chemical composition has been completely changed, though primary sedimentary features may be discerned as ghostly relics. Indeed it could be argued that the evaporites/replacementites are really metamorphic rocks.

Stratification is sometimes absent in sedimentary rocks. Massive structureless sediments include 'loess' (wind-blown silt) and grain-flow and mass-flow deposits such as debrites and olistostromes.

Metamorphic Rock Anomalies

The previous paragraph on anomalous sedimentary rocks provides a useful introduction to problems of metamorphic-rock classification. Just as the boundary between chemical sediments and metamorphic rocks is gradational, so too is the boundary between detrital sediments and metamorphic rocks. During burial gravel lithifies to form conglomerate, sand forms sandstone, and mud forms shale. With increasing burial temperature and pressure, porosity is destroyed, and texture and mineralogy are modified. There is therefore a gradation between highly cemented sedimentary rocks and low-grade metamorphic rocks. Similar gradations occur when superficial igneous rocks undergo deep burial, with concomitant mineralogical and textural changes as they metamorphose. Historically nowhere has this problem been more acute than in the controversial theories about the origin of granite. Granites were believed to have formed from the slow cooling of magma that had intruded shallower rocks. It was pointed out, however, that, when considering the vast granite masses at the eroded roots of mountains, there was a space problem. This was resolved when it was realized that wholesale replacement of pre-existing rock had taken place. Intrusive granites were found only at shallow depths, having moved upwards from deep parent batholiths.

Rocks and their Classification: Conclusion

Rocks are of three genetic types: igneous, formed from cooling magma; sedimentary, formed mainly by the breakdown of pre-existing rock; and metamorphic, formed by the action of heat and pressure on pre-existing rock. Igneous and metamorphic rocks are both normally composed of a mosaic of interlocking crystals that formed at high temperatures and pressures; they thus usually lack porosity and permeability. In the petroleum industry igneous and metamorphic rocks are generally grouped together under the term 'basement', indicating that it is pointless to continue drilling deeper into such tight rocks. Igneous rocks are generally massive and structureless, while metamorphic rocks are often banded.

Sediments, on the other hand, are deposited with space between the constituent particles. Thus, sedimentary rocks, by their very nature, tend to be porous and permeable. Sediments are deposited episodically, and thus characteristically exhibit stratification. They are also the only one of the three main classes of rock to contain fossils.

All classifications have their illogicalities and inconsistencies. Thus sedimentary rocks grade with increasing burial into metamorphic rocks, and metamorphic rocks ultimately melt to form magma, which cools to form igneous rocks, which are weathered, eroded, transported, and redeposited as sediments.

See Also

Diagenesis, Overview. **Igneous Processes**. **Igneous Rocks:** Granite. **Metamorphic Rocks:** Classification, Nomenclature and Formation. **Regional Metamorphism**. **Sedimentary Environments:** Depositional Systems and Facies. **Sedimentary Processes:** Erosional Sedimentary Structures; Post-Depositional Sedimentary Structures. **Sedimentary Rocks:** Mineralogy and Classification; Chert; Dolomites; Evaporites; Ironstones; Limestones; Phosphates. **Unidirectional Aqueous Flow**. **Volcanoes**. **Weathering**.

Further Reading

Norton WH (2002) *The Elements of Geology*. Indypublish.com, Cybercity.

Peters KE (1997) *No Stone Unturned*. San Francisco: WH Freeman.

Rothery D, Turner D, and Wilson RCL (2001) *Geology*. Milton Keynes: Open University Worldwide.

RUSSIA

A S Yakubchuk, The Natural History Museum, London, UK
A M Nikishin, Lomonosov Moscow State University, Moscow, Russia

Copyright 2005, Natural History Museum. All Rights Reserved.

Introduction

The Russian Federation is the world's largest country, spread across 10 time zones from west to east and from the Arctic to Subtropical climates from north to south. Her territory is mostly plains in the lowlands of Eastern Europe and western Siberia and highlands in eastern Siberia. The country's mountain ranges are the Urals, Verkhoyansk, Koryak, Sikhote-Alin, Altai, Sayan, and Caucasus.

Geological exploration of this vast area started in the sixteenth century, when Russia began its advance to the east. However, it became systematic only in the eighteenth century when Russia invited European scientists and many Russian people received education in Europe, many in mining centres. This stimulated development of base metal and silver deposits in the Urals, Altai, and Transbaikalia provinces. Gold and platinum were discovered only in the late eighteenth to early nineteenth century in the Urals. This boosted exploration, and many prospectors moved to Siberia and the Russian Far East, where they found many rich gold placers before the Golden Rushes of the nineteenth century in the West.

These discoveries required systematisation, and the Russian government established the Geological Committee in 1882. After the 1917 Revolutions, these works expanded across the Soviet Union, and its first geological map at a scale of 1:5 000 000 was compiled by 1937. In the 1940–1970s, regional tectonic maps were compiled and many world-class deposits of Ni, Au, diamonds, and oil were discovered. During this period, Russian geology evolved in isolation, but under the significant influence of the German geological school. As a result, the fixistic approach dominated in geology and many local names were proposed for orogenic events, such as Baikalides, Salairides, etc. As a result, Russia has its own system of Proterozoic subdivisions, such as Riphean (1600–680 Ma), that incorporates the Middle and Upper Proterozoic in the International Scale, and Vendian (650–540 Ma), roughly corresponding to the Eocambrian. Phanerozoic subdivisions generally correspond to the International Scale.

After the introduction of plate tectonics, Russian geologists remained mostly non-mobilistic in their thinking. Its regional geology was revised in line with plate tectonic theory only in the 1980s, which was also a period of active study of the Russian shelves and participation of many geologists in international projects. During these years, Russian geological knowledge advanced significantly, with superdeep drilling in the Kola Peninsula, Urals, Central Asia, and western Siberia.

After the breakup of the Soviet Union in 1991, there has been a significant reduction of geological work. However, interesting compilations were produced, as well as detailed petrological studies. The country is now an area of international exploration activity.

General Geology

Russian territory consists of Precambrian cratons and Neoproterozoic to Cenozoic orogenic collages, overlapped by the Mesozoic-Cenozoic sedimentary basins (Figure 1). The East European and Siberian cratons are the oldest structures in continental Russia. On the Arctic shelf, west of the Gakkel Ridge in the Arctic Ocean, there are the Svalbard and Kara blocks. The Hyperborean Craton is prominent in the Arctic shelf east of the Gakkel Ridge and north of Chukotka. The cratons are bordered by the Neoproterozoic orogens of the Baikalides and Timanides, which probably continue on to the Arctic shelf. Between these cratons is the Palaeozoic orogenic collage of the Altaids, mostly hidden under the West Siberian Mesozoic-Cenozoic sedimentary basin. The Altaids extend to Transbaikalia (Figure 2). To the south of the Siberian Craton is the Mongolide orogenic collage. Between the Siberian and Hyperborean cratons and along the Far East coast are the Mesozoic-Cenozoic Circum-Pacific orogens. The Alpine orogenic belt borders the East European Craton in the south.

East European Craton

The East European Craton consists of shields, basement rises, and sedimentary basins. The Baltic Shield outcrops in the Kola Peninsula and Karelia. In the south, the crystalline basement occurs in the so-called Voronezh Rise. There are several sedimentary basins – Baltic, Moscow, Mezen, Volga-Ural, Pachelma, Ulianovsk-Saratov, Peri-Caspian – and the Peri-Urals foredeep.

The craton formed prior to 1.6 Ga (Figure 3). In the Baltic Shield, there are the Karelia, Central

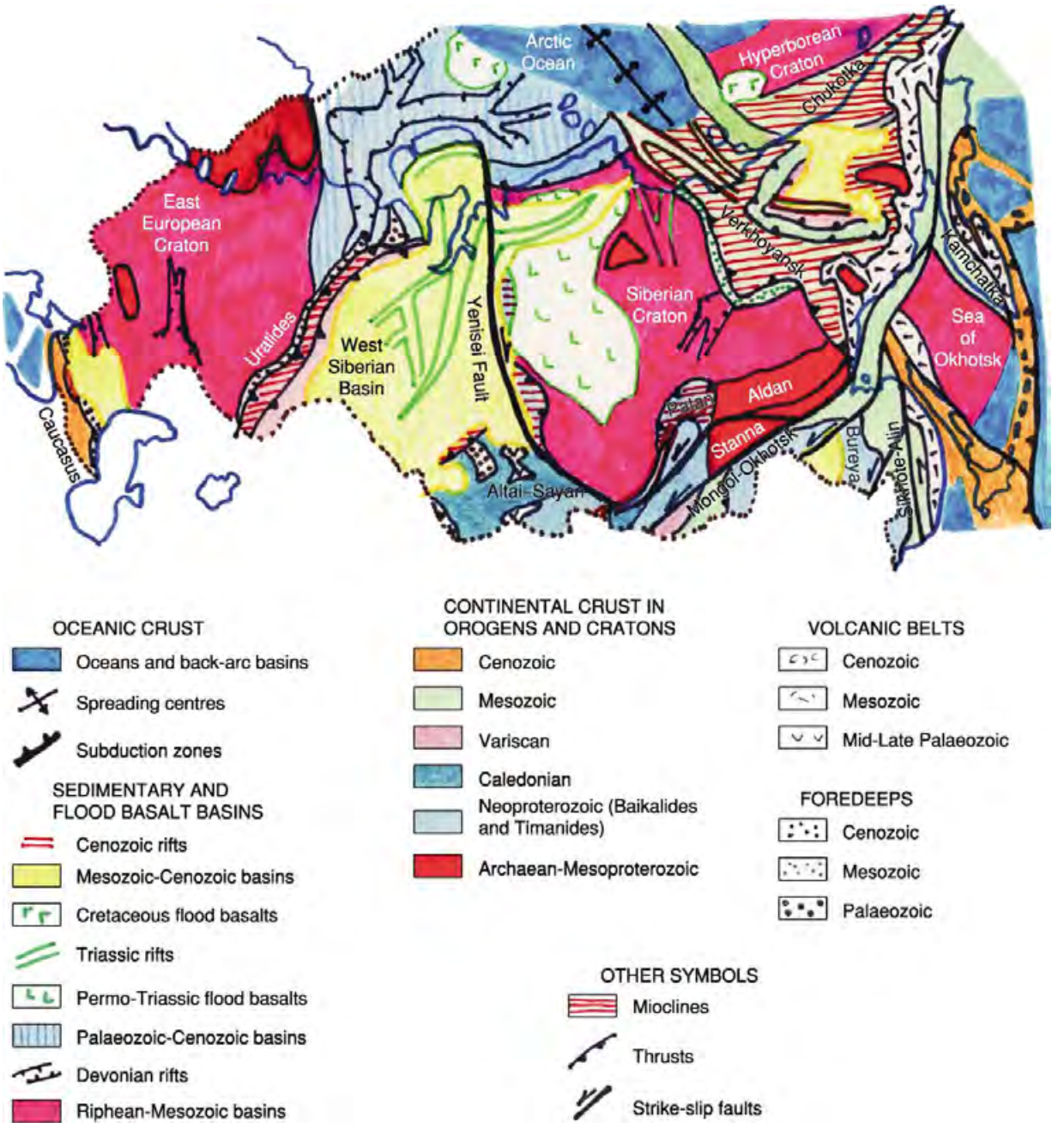


Figure 1 Outline Tectonic map of Russia and adjacent shelves (newly compiled by Nikishin and Yakubchuk). Note the presence of vast Mesozoic-Cenozoic sedimentary basins superimposed onto the Palaeozoic Collage between the East European and Siberian cratons. Devonian rifts were superimposed onto cratons. Red cratons: deep blue, Ocean basins: light blue, Palaeozoic tectonic areas: horizontal red shading, Mesozoic to Cenozoic areas: yellow, Mesozoic-Cenozoic sedimentary basins.

Kola, and Murmansk Archaean terranes. The Karelia Terrane is a typical 3.0–2.7 Ga granite-greenstone province. The Central Kola Terrane is a classic block with amphibolite metamorphism and it consists of various volcanic and sedimentary rocks with numerous granitoids and gneisses. The Murmansk Terrane consists of Late Archaean granitoid batholiths. Between the Karelia and Kola terranes is the linear White Sea Belt, in which the Archaean crust was reworked in the Early Proterozoic, producing 1.9 Ga granulites. The Archaean nuclei of the Baltic Shield are overlain by Early Proterozoic sedimentary basins and volcanics. Similar terranes exist in the unexposed part of the craton.

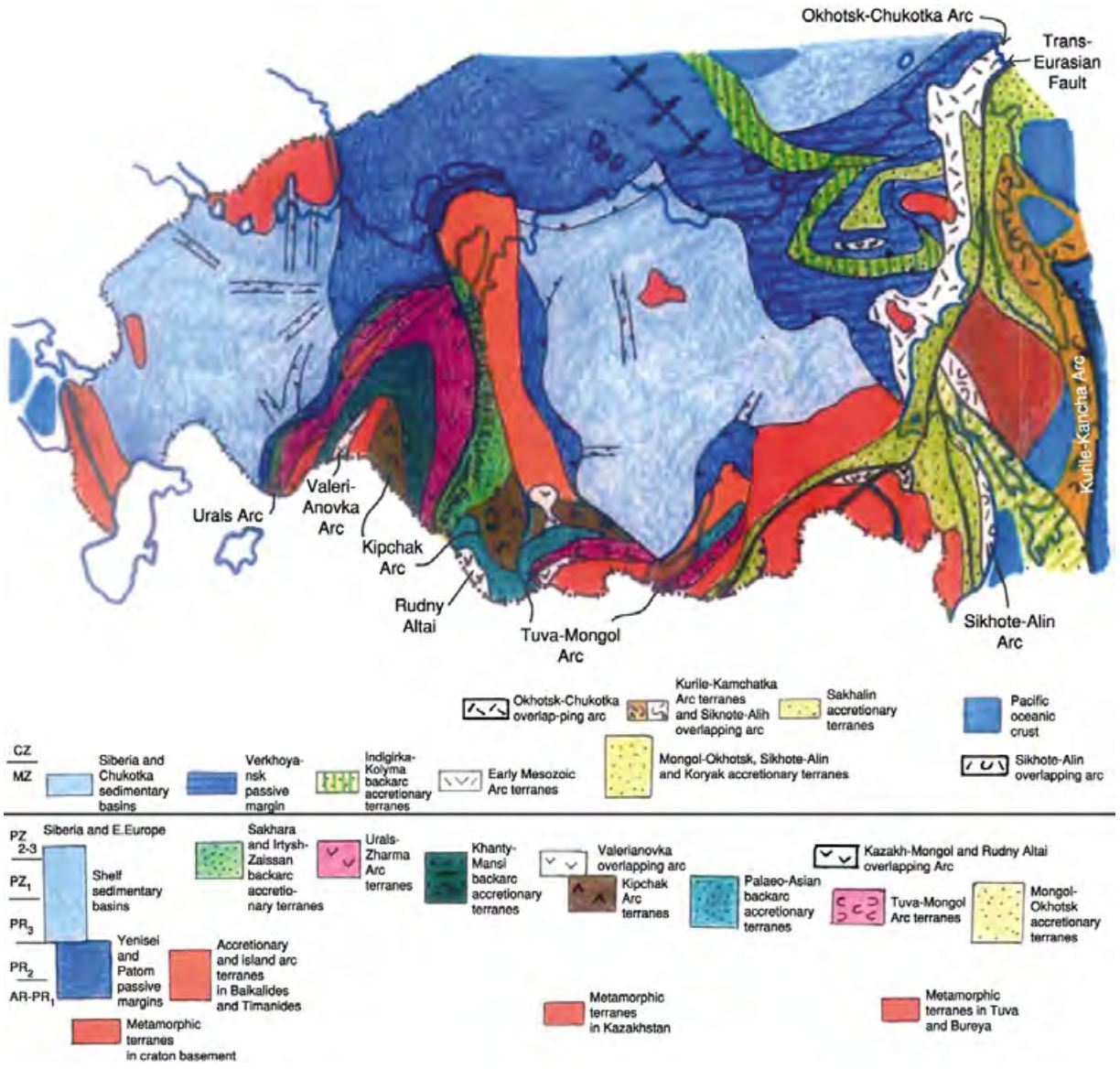


Figure 2 Tectonic map of Russia and adjacent shelves without Mesozoic-Cenozoic sedimentary basins (newly compiled by Yakubchuk and Nikishin). Note the oroclinal structure in the Palaeozoic and Mesozoic orogenic collages. Riphean rifts are shown on the major cratons. Late Palaeozoic to Early Mesozoic Trans-Eurasian strike-slip faults separate the Altaid Collage into two domains.

The sedimentary basins of the East European Craton (Figure 4) consist of several stratigraphic units: (i) Riphean (1.6–0.65 Ga); (ii) Vendian–Early Cambrian; (iii) Middle Cambrian–Early Devonian; (iv) Middle Devonian–Permian; (v) Triassic; (vi) Jurassic–Lower Cretaceous; (vii) Upper Cretaceous–Eocene; (viii) Oligocene–Neogene; and (ix) Quaternary.

Riphean palaeorifts (or aulacogens) are filled with basalts, continental and marine clastic rocks, and carbonates. The Riphean stratigraphy is poorly studied. Three epochs of rifting at 1.6–1.3 Ga, 1.3–1.0 Ga, and 1.0–0.65 Ga have been identified, although some dates are approximate. Palaeorifts mainly occur on the flanks of the craton, but Central Russian palaeorifts crosscut the entire craton.

Vendian-Early Cambrian rocks form the sedimentary cover of the entire craton. The Lower Vendian (~650–600 Ma) consists of tillites, representing the Lapland glaciation. On the south-western edge of the craton are Lower Vendian Volynian plateau basalts (in Ukraine), which were formed prior to the opening of the Tornquist Ocean. Upper Vendian to Early Cambrian rocks are mostly shallow-water sandstone and clay. Upper Vendian rocks have many occurrences of the Ediacaran fauna (*see* **Precambrian:**

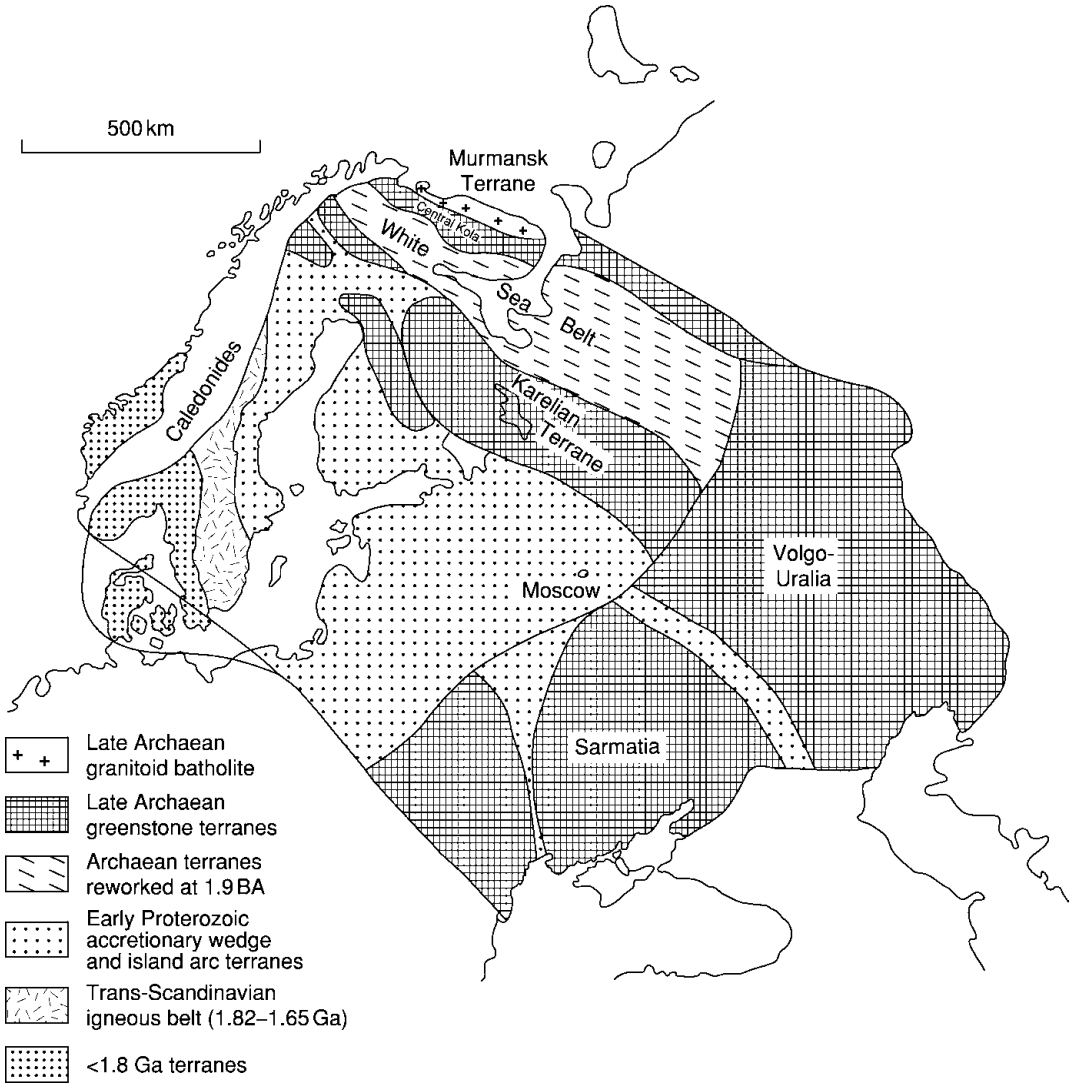


Figure 3 Crustal provinces of the East European Craton (compiled using data by Gorbatschev and Bogdanova).

Vendian and Ediacaran). Accumulation of the sedimentary basins was controlled by the Iapetus and Tornquist Oceans to the west (modern coordinates) and the foreland basin of the Timanides to the east (*see* **Europe:** Timanides of Northern Russia). The craton was elevated at the end of the Early Cambrian synchronously with the final orogenesis in the Timanides.

Middle Cambrian to Early Devonian rocks form a wide basin that might have fully covered even the Baltic Shield. The principal basins are Baltic, Moscow-Mezen-Pechora, Peri-Urals, and Peri-Caspian, but they formed a united system of epicontinental seas. In the Ordovician, to the east of the craton, was the Urals Ocean or oceanic back-arc basin, and the poorly studied Scythian Ocean was to the south. Marine sedimentation was controlled by climate: for the Middle Cambrian to Middle Ordovician strata, clays, sandstones, and carbonates are typical; in the Late Ordovician to Silurian, carbonate rocks are dominant; in Early Devonian times, red clastic rocks began to dominate against a background of arid climate and regression, culminating in total sea regression.

The Middle Devonian to Permian cover was formed in the epicontinental marine basin. The Middle Devonian rocks are typical clastic and carbonate sediments. At the Givetian-Frasnian transition was a magmatic event and rifting, with a system of rifts (Figure 4), such as Pripyat-Dnieper, Donets, Peri-Caspian, Barents Sea, Kola, Timan-Pechora, and Vyatka. During that time, diamondiferous kimberlites formed in the Arkhangelsk region, and the world's largest alkaline ultramafic massifs were emplaced in the Kola Peninsula. Middle to Late

Figure 4 Sedimentary basins of the East European Craton (compiled by Nikishin and Fokin).

Devonian rifting and magmatism can be explained by the influence of mantle plumes, in combination with regional continental extension and possible back-arc processes.

The Middle Devonian to Permian rocks are mostly carbonates. An epicontinental marine basin changed its depth many times due to sea-level fluctuations. Carboniferous strata include typical buried river valleys and limnic coal. The Peri-Caspian Basin was a deep-water basin framed by reefs. In the Donets Basin, there was a large delta system that accumulated 15–20 km of sediments with numerous coal seams. In the Peri-Urals Basin, the Domanik Horizon, enriched in organic material, was formed during the Frasnian, and it is a regional oil-bearing stratum. In Kungurian time, the Peri-Caspian Basin was almost completely separated from the ocean by the Uralide Orogen, and in it was accumulated up to 5 km of evaporites. In the Late Carboniferous, the main Urals collision began, and the Peri-Urals foredeep formed in front of the Uralide Orogen in the Permian. From the mid-Permian, there was a major regression, and mostly shallow-water marine, limnic, and continental sediments were accumulated in the Late Permian. By the end of the Permian, most of the craton had experienced regression, but the giant Pechora-Barents Sea sedimentary basin survived in the north, accumulating 3–5 km of sediments.

At the Permian-Triassic transition, the East European Craton was affected by the Siberian Superplume (Figure 4). The main Triassic sedimentary basins formed on the flanks of the craton are the Peri-Caspian, Pechora-Barents Sea, and Baltic-Polish-German. On the craton, thin continental sediments accumulated in the Moscow-Mezen and Dnieper basins. At the Triassic-Jurassic transition, the

Caucasus-Black Sea-Turkish Orogen formed to the south of the craton; to the north was the Novaya Zemlya orogen. Almost complete regression took place synchronously with these deformations and it continued until Aalenian time.

In Middle Jurassic to Early Cretaceous times, the main sedimentary basins were in the south and north of the craton. In the south were the Peri-Caspian, Ulianovsk-Saratov, Dnieper, and Peri-Black Sea Basins of the Northern Peri-Tethys area. In the north was the Pechora-Barents Sea Basin of the Boreal area. The Boreal and Peri-Tethys basins were periodically linked to form a single basin. The Peri-Tethys area has typical shallow-water marine carbonate and clastic rocks, whereas the Boreal Basin has marine and continental clay and sandy rocks. In the internal part of the craton are shallow-water marine and continental clastic sediments with multiple hiatuses. Organic-rich sediments are typical in the Late Jurassic.

In the mid-Cretaceous, the Peri-Tethyan and Boreal basins were separated. Carbonate rocks dominated in the Late Cretaceous to Eocene in the Peri-Tethys basins, whereas sandy-clay sediments dominated in the Boreal Basin. In the Late Cretaceous-Paleocene there was an intracratonic inversion of palaeorifts.

Oligocene-Neogene shallow-water sediments accumulated in the south of the craton against a background of maximum orogeny in the Caucasus and a colder climate.

In the Quaternary period, there were five major glacial and interglacial epochs. The main glaciers were on the Baltic Shield. To the south, they expanded as far as the Ukrainian Shield, producing moraine and fluvioglacial accumulations.

Siberian Craton

Most of the Siberian Craton is covered by its sedimentary and flood basalt basins. Its basement outcrops are in the Anabar Shield in the north, the Aldan-Stanovoy Shield in the south, and in the Sayan-Yenisei area in the south-west. East of the Anabar Shield is the small Olenek Rise. All the basement rocks are older than 1.6 Ga, but the southern part of the Aldan-Stanovoy Shield was magmatically reworked in the Palaeozoic and Mesozoic. The Sayan-Yenisei area was deformed in the Neoproterozoic.

The Siberian Craton is the largest craton on Earth, with a pre-1.6 Ga crust. Its structure is dominated by Archaean terranes. Some of them were reworked during Early Proterozoic times, and juvenile Early Proterozoic crust occupies a relatively small area (Figure 5). It amalgamated as a single craton by the end of the Early Proterozoic.

There are three principal sedimentary basins (Figure 6): Vilyui in the east, Tunguska in the north-west, and Taseevo-Angara-Lena (or Irkutsk Amphitheatre) in the south-west. The thickness of the sedimentary cover is up to 5–8 km and it consists of several rock complexes: (i) Riphean (1.6–0.65 Ga); (ii) Vendian–Early Devonian; (iii) Middle Devonian–Early Carboniferous; (iv) Carboniferous–Permian; (v) Triassic; (vi) Jurassic–Cretaceous; (vii) Tertiary; and (viii) Quaternary.

Riphean palaeorifts (aulacogens) are widely distributed and are filled with carbonates, continental clastic rocks, and basalts. Their stratigraphy is poorly studied, but there are at least three epochs of rifting distinguished at 1.6–1.3 Ga, 1.3–1.0 Ga, and 1.0–0.65 Ga. Most of the palaeorifts are perpendicular to the craton margins: their palaeotectonic setting is not clear, but some data indicate that 0.8 Ga rifts in the east of the craton could be related to the opening of the ocean. In addition to the palaeorifts, wider Riphean cover overlaps northern parts of the craton.

Vendian to Early Devonian rocks cover almost the entire craton. The Vendian, known in Siberia as the Yudomian, consists of clastic rocks at its base which were overlain by carbonates. During the Early Cambrian, carbonate sedimentation dominated in the Yudoma-Olenek shelf in the east, and a carbonate-evaporite shallow-water basin existed in the Tunguska-Irkutsk Basin. These two basins were separated by carbonate reefs with archaeocyathid-algae buildups. From the Middle Cambrian to the Late Ordovician, carbonates dominated in all basins. Silurian rocks are carbonates, clastic sediments, and evaporites. The maximum regression occurred in the Late Silurian-Early Devonian and coincides with orogeny in Transbaikalia, where the Patom Highlands were thrust onto the Siberian Craton, causing thin-skinned deformations of its cover north of Lake Baikal.

In the Middle Devonian to Early Carboniferous, there was rifting and magmatism. The main rifts formed in the east (Figure 6). Vilyui was the largest system, consisting of the Ygyata, Kempendyai, and other rifts. Rifting started with flood basalt magmatism and the intrusion of basalt dyke swarms. Later, there was the main phase that continued until Tournaisian times. In the Early Carboniferous, the Verkhoyansk branch of the rift system was transformed into the passive margin of the oceanic back-arc basin. As a result, the continental rifts became extinct, but the passive continental margin continued to evolve until Jurassic times. Middle-Late Devonian rifts possibly formed in the Tunguska Basin and

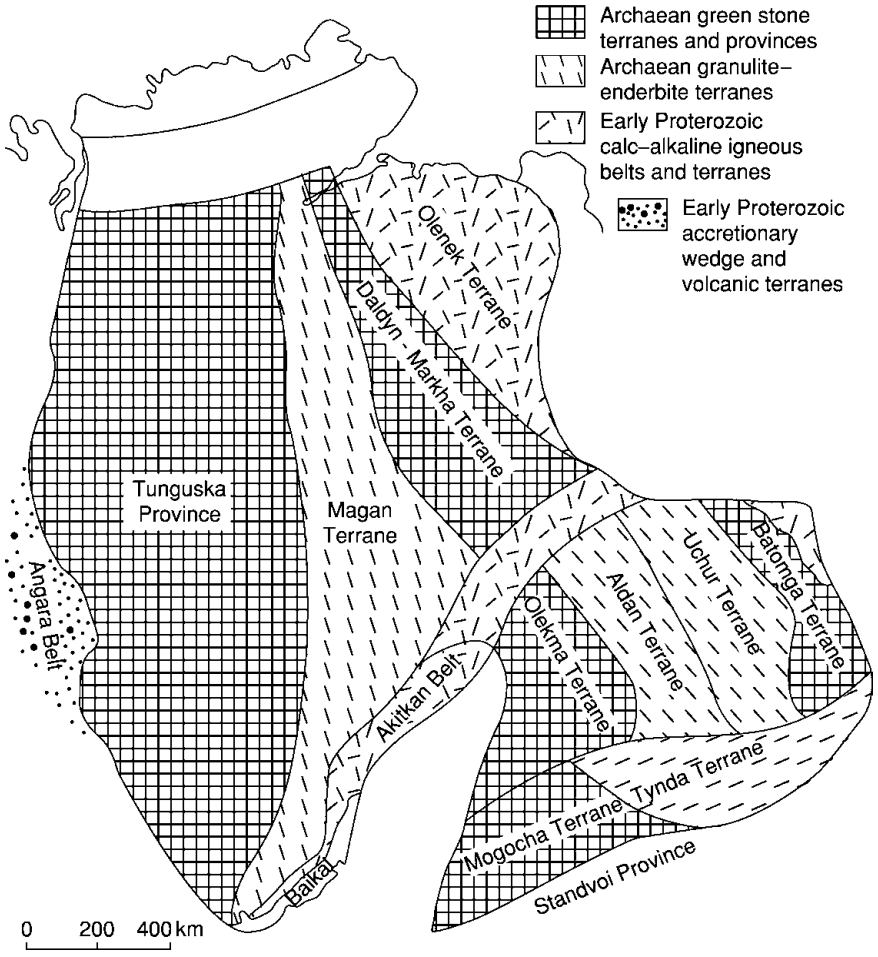


Figure 5 Crustal provinces and terranes of the Siberian Craton (simplified and modified after Rozen).

along the northern flank of the craton. In addition, numerous diamondiferous kimberlites were emplaced into Archaean nuclei, reflecting the impact of Middle-Late Devonian plumes, in a similar way to the East European Craton.

Carboniferous and Permian rocks occur in the Vilyui and Tunguska basins. In the Vilyui Basin, Middle Devonian rocks form thick grey clastic shallow-water and limnic-alluvial formations. This setting persisted until the end of Jurassic time, when numerous coal beds accumulated in the Vilyui Basin. At the end of the Jurassic and in the Early Cretaceous, orogenesis occurred in the Verkhoyansk Orogen, and a foredeep basin formed in its front. In the Carboniferous to Permian, the Tunguska Basin was filled with clastic continental and shallow-water marine sediments with numerous coal beds.

The latest Permian and Triassic was dominated by mantle plume tectonics. Large-scale flood basalt magmatism (often termed the Siberian Traps) took place in the north of the Siberian Craton and southern Taimyr (a probable northern continuation of the craton). The total thickness of basalts in the Tunguska Basin reached 2–4 km. This was the largest intracontinental Phanerozoic magmatic event (Figure 6), and it affected an area extending from the northern Urals to Verkhoyansk and from Taimyr to Central Asia, revealing it as a superplume event.

Jurassic rocks occur in the Vilyui and Irkutsk Basins. They also form a chain of foredeep basins between the Aldan and Stanovoy blocks. Typical rocks are continental coal-bearing sediments. The Stanovoy Block was an Andean-type margin in Jurassic times. Cretaceous clastic rocks are known in the Verkhoyansk foredeep.

During Tertiary times, the Siberian Craton was relatively elevated, possibly in response to the India-Asia Himalayan collision. The altitude of its plateaux was 1–1.5 km. Tertiary sediments occur locally and consist mostly of continental clastic rocks. The Popigai meteorite crater of Eocene age (80 km in diameter) occurs north of the Anabar Shield.

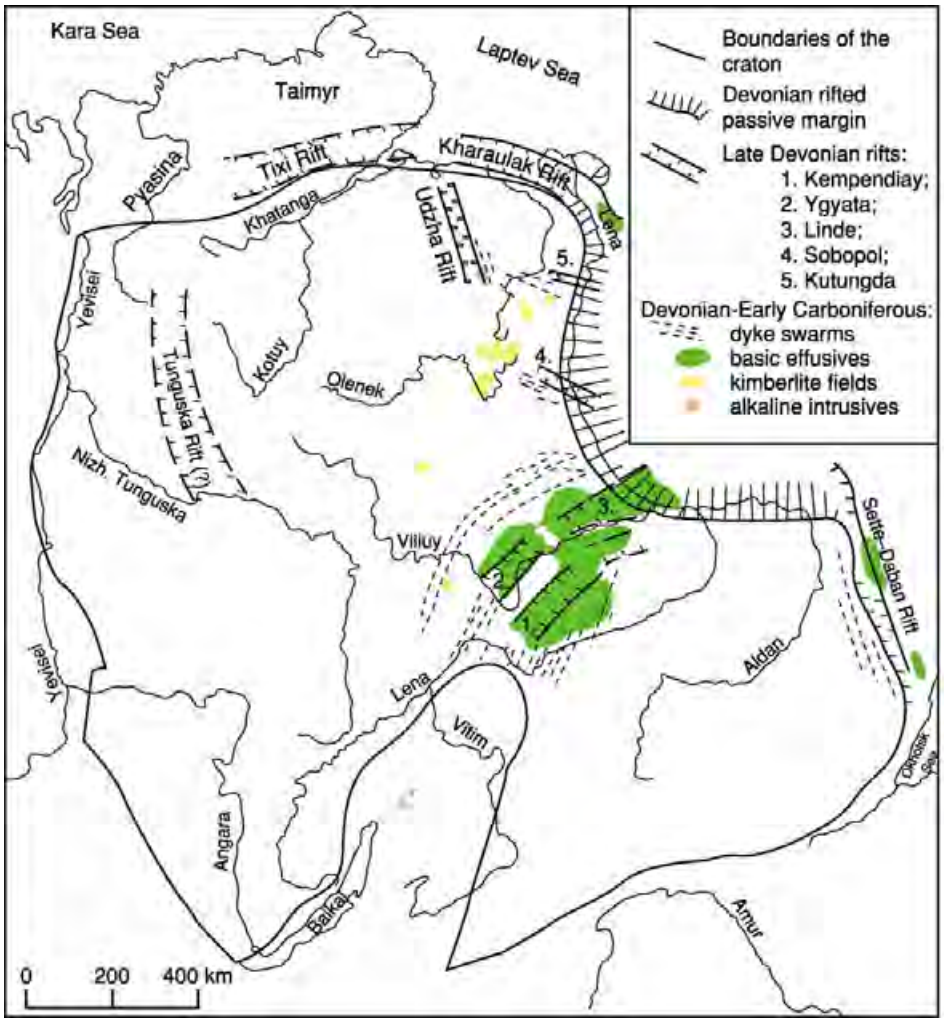


Figure 6 Sedimentary basins of the Siberian Craton (compiled by Nikishin and Fokin).

In Quaternary times, the craton was glaciated many times, but thick glaciers formed only in its northern part and the rest of the area was affected by thick permafrost that survives until the present day.

Neoproterozoic Orogens

Shatsky proposed the term Baikalides for the Neoproterozoic orogens on the flanks of the major cratons of Russia. Their type locality is in the Patom Highlands in northern Transbaikalia, but they were also distinguished in the Yenisei ridge, northern Taimyr, and Timan. Their slivers are recognized inside the Palaeozoic to Cenozoic orogens.

Until the 1980s, the interpretation of Baikalide age was mostly based on stratigraphic data. Recently, the Neoproterozoic age of the type locality in northern Transbaikalia was questioned. Some researchers suggested that the Baikalides are Early or Middle Proterozoic structures, whereas others proposed a Caledonian age. However, the deformed sequences in the Patom Highlands are unconformably overlapped by less deformed Vendian to Cambrian sedimentary rocks, indicating Neoproterozoic deformations. In addition, isotopic dating of syn- and postmetamorphic granitoids in the Yenisei ridge, Taimyr, and Timan also demonstrate the validity of Neoproterozoic deformations.

Patom Highlands This is an oroclinal embayment in the south of the Siberian Craton, extending from Lake Baikal towards the Patom Highlands (Figure 1). In the 1970–1980s, the Baikalides were interpreted as an Early Proterozoic orogen. Subsequent regional mapping re-confirmed that the orogen is Neoproterozoic. Most of this area is occupied by the Barguzin granitoid batholith. In the 1960s, it was considered as Proterozoic. In the 1970–1980s, the first K-Ar data showing its Early Palaeozoic age were obtained. However, application of more advanced techniques

in the 1990s revealed a Carboniferous age. In addition, isotopic dating of some regionally metamorphosed rocks in the Patom Highlands also revealed a Carboniferous age. This raised speculations that the Baikalides might be a Caledonian Orogen.

Geologically, the Patom Highlands consists of two major parts. In the north is a Baikal-Patom passive margin of Middle to Upper Riphean carbonate-terrigenous rocks deformed into linear folds. These rocks are unconformably overlain by Vendian-Cambrian molasse. Together they were thrust onto the Siberian Craton in the Middle Devonian. To the south is the Barguzin-Vitim Terrane which is separated from the passive margin by a suture with Riphean ophiolites. The Barguzin-Vitim Terrane is exposed in small fragments, consisting of metamorphosed Riphean accretionary wedge and island arc rocks.

Yenisei Ridge The Yenisei Orogen in the west of the Siberian Craton (Figure 1) is considered as an equivalent and continuation of the Baikalides. Deformed passive margin rocks occupy the bulk of the ridge. It has been demonstrated that in the west there is a suture with Late Riphean ophiolites and an island arc terrane. The folded structure is intruded by Neoproterozoic granites. Passive margin rocks were thrust onto the Siberian Craton, possibly in mid-Palaeozoic times.

Northern Taimyr, Arctic Shelf and Timanides The east–west trending Taimyr Orogen is an imbricated structure produced during Mesozoic thrusting to the south (Figure 7). It has been suggested that this orogen was a result of collision between the Kara Plate to the north and the Siberian Craton to the south. Northernmost Taimyr is a deformed Neoproterozoic to Early Palaeozoic passive margin of the Kara Plate, consisting of flysch and carbonate sequences which were regionally metamorphosed. They were earlier considered as Precambrian rocks: however, recent study has showed that metamorphism is Late Palaeozoic in age and might have taken place during the emplacement of 300–264 Ma collisional and post-collisional granitoids. The Main Taimyr Thrust separates northern and central Taimyr. The latter hosts two terranes (Mamont-Shrenk and Faddey) with Neoproterozoic island arc rocks and ophiolites (740–820 Ma) which were obducted 650 Ma ago onto Neoproterozoic carbonates, which are also thrust on top of Vendian to Early Carboniferous terrigenous-carbonate rocks of central Taimyr, part of the Siberian Craton. Another ophiolite belt inside northern Taimyr might be a root zone where the actual suture might be located. If so, northern Taimyr should be separated into two terranes, the northern one representing the Kara Plate and the southern one a deformed passive margin of the Siberian Craton.

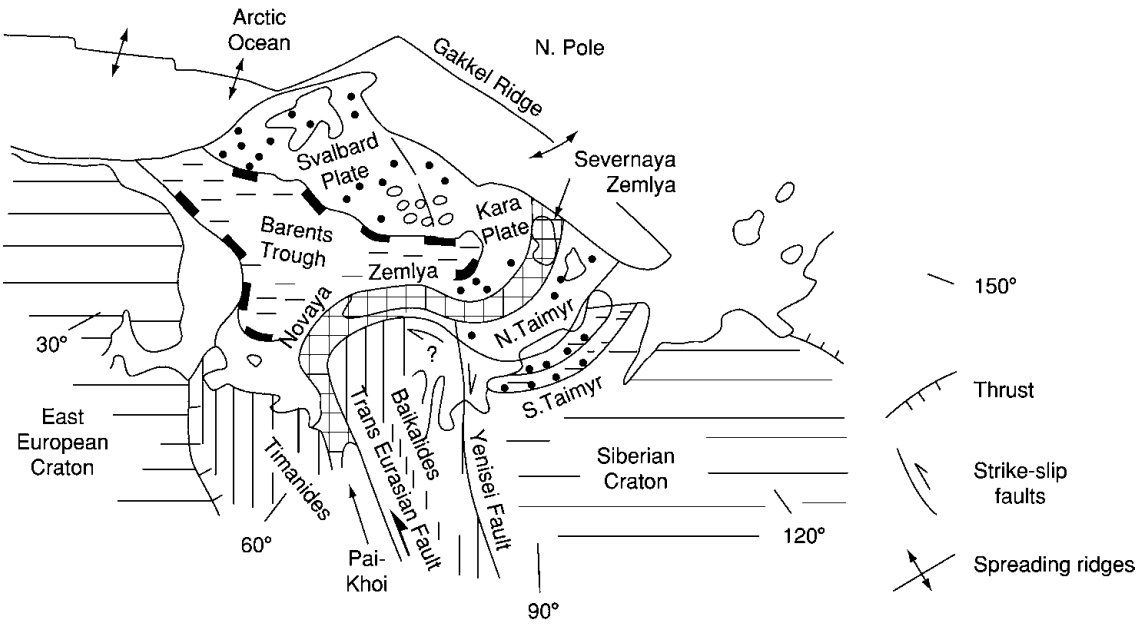


Figure 7 Tectonics of the Arctic shelf and adjacent regions (compiled by Yakubchuk, using data by Bogatskiy *et al.* and Bogdanov *et al.*). The Yenisei Fault separates the Arctic shelf into the Kara and Svalbard plates. Severnaya Zemlya is probably a continuation of Novaya Zemlya. The Barents extensional trough bounds the Svalbard Plate in the south. A major strike-slip fault might separate the west-east-trending structures of the Arctic shelf and Siberian Craton.

Central Taimyr was thrust southward to form the southern Taimyr thrust zone which consists of an imbricated Middle Carboniferous to Triassic sequence. Its stratigraphy is generally similar to the Siberian Craton in the Tunguska Basin.

The position of Taimyr relative to other structures has been controversial. It has been proposed that it might represent a continuation of Novaya Zemlya. However in the 1980–1990s, the Arctic shelf was actively explored for oil. New data were obtained, and it was suggested that the northern Taimyr is an exposed part of the Kara Plate, which occupies the adjacent Arctic shelf. To the east, the Taimyr structures are truncated by Arctic oceanic floor. Its western continuation is far from clear, but magnetic data suggest correlation of Pai-Khoi and Novaya Zemlya with Severnaya Zemlya and not with Taimyr (Figure 7).

The Severnaya Zemlya islands consist of 6–7 km of Ordovician to Devonian deformed shallow-water sedimentary rocks with evaporites and felsic alkaline volcanics (456 Ma). They conformably overlie Cambrian rocks and their fossils are similar to those of the Novaya Zemlya archipelago. Folded Ordovician to Devonian rocks are unconformably overlain by Carboniferous-Permian redbed sequences. The timing of deformation is assumed to be Late Triassic, as in Taimyr, but no structural relationship has been recorded.

Novaya Zemlya and Pai-Khoi consist of Riphean to Lower Cambrian metamorphic rocks unconformably overlain by deformed Ordovician-Devonian terrigenous-carbonate rocks. These are also unconformably overlain by Carboniferous-Triassic continental rocks deformed prior to Jurassic times.

Pai-Khoi, Novaya Zemlya, and Severnaya Zemlya reveal similar lithologies and similar unconformities. The two regions are dextrally offset relative to each other along the Yenisei strike-slip fault that separates the Arctic shelf into the Kara and Svalbard plates and bounds the western margin of the Siberian Craton. Tectonically, Severnaya Zemlya and Novaya Zemlya might represent an inverted rift structure.

West of Pai-Khoi and Novaya Zemlya are the Timanides, where Puchkov reported relic Late Vendian fold belts (600–540 Ma). The Timanides are poorly exposed, and an understanding of their structure is based on geophysical and drilling data. There are two major terranes: in the south-west is the Izhma-Pechora passive margin with Riphean shales, felsic volcanics, and minor ultramafic intrusions; in the north-east is the Bolshaya Zemlya Terrane with Riphean intermediate volcanics and granitoids. The passive margin of the terrane can be traced along the northern coast of the Kola Peninsula, whereas the island arc terrane is truncated by the East Barents Trough and Pai-Khoi structures.

The Timanides are overlain by Devonian to Jurassic rocks of the Pechora sedimentary basin. The whole orogen and its sedimentary basin were thrust onto the East European Craton prior to Jurassic times. Numerous listric compressional faults in the Pechora basin reflect the Urals collision.

In the middle of the Arctic shelf is the 1500 km long East Barents depression, extending parallel to the proposed Pai-Khoi–Novaya Zemlya–Severnaya Zemlya inverted rift. Aplonov and his colleagues proposed that this depression might have a Palaeozoic oceanic crust over which lies up to 15–17 km of Middle Palaeozoic to Cretaceous sediments. In the south, the East Barents depression continues as continental rifts in the base of the Pechora Basin. In the north-east, it is truncated by the Arctic oceanic floor.

Altaid Collage

The area between the East European and Siberian cratons and south of the Arctic shelf is occupied by the Altaid Orogenic Collage (Figure 2). The Russian portion of the Altaid Collage is concealed under the Mesozoic-Cenozoic sedimentary basin of western Siberia, being exposed in isolated orogens in the Urals, Altai-Sayan, and Transbaikalia. Other exposures are in adjacent Central Asia (*see* **Asia:** Central; **China and Mongolia**), Mongolia, and China.

The Altaid Collage is separated into two domains along the giant Trans-Eurasian strike-slip fault, which strikes from Novaya Zemlya towards western Mongolia and then to the Mongol-Okhotsk suture, south of the Siberian Craton. To the west is the Kazakhstan-Khingan Domain and to the north-east is the Altai-Mongol Domain. Both domains have similar Terranes which were sinistrally offset during the Late Palaeozoic to Triassic by 1000 km. The internal structure of the domains is oroclinal, and their axes are also sinistrally offset by 1000 km.

Altai-Mongol Domain

The Altai-Mongol Domain consists of the 'Caledonides' and 'Variscides' (these terms are used here in different ways from their West European counterparts). The Caledonides, which occur in Altai-Sayan and Transbaikalia, represent a portion of the Altaid Collage attached to the Siberian Craton and Neoproterozoic orogens (Figure 8). In the south-west, it is truncated by the Trans-Eurasian Fault. The 'Variscides' occur in the Tom-Kolyvan Terrane in Siberia and the Khangai-Khentei Superterrane in Transbaikalia.

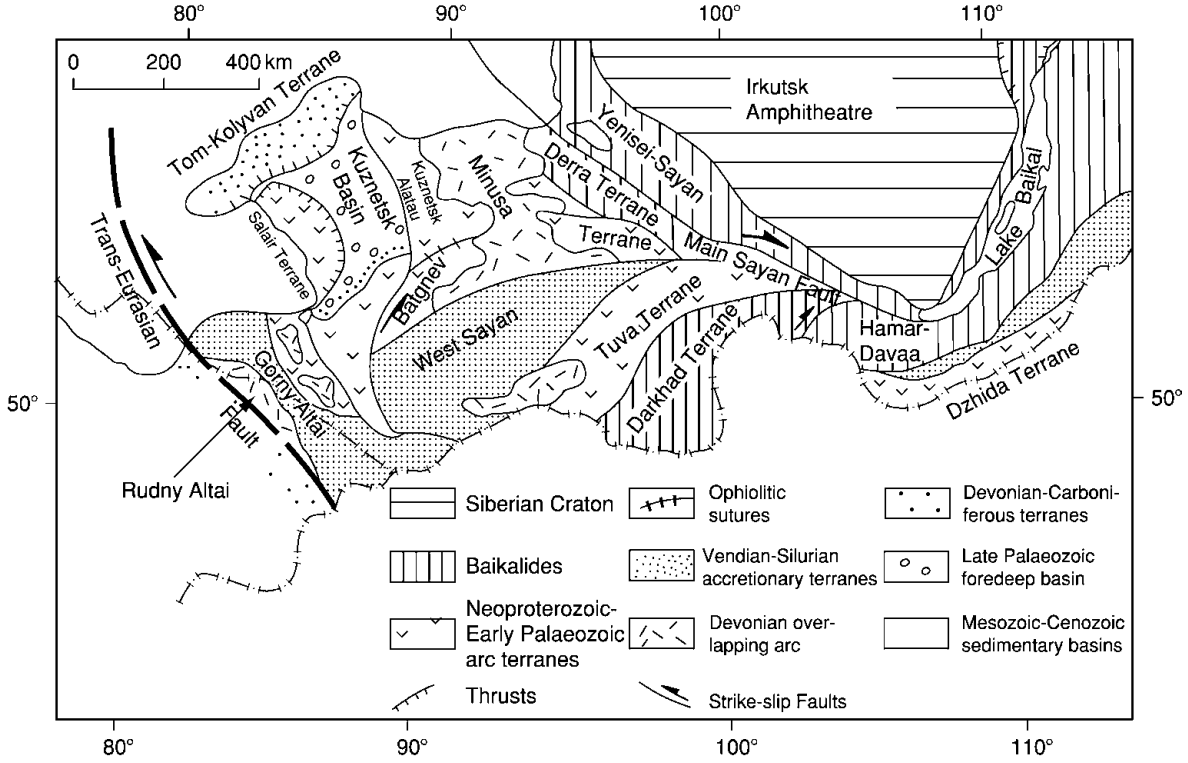


Figure 8 Tectonics of the Altai-Sayan area (compiled by Yakubchuk). The mosaic pattern reflects significant strike-slip faulting in the area. The Main Sayan fault is a southern continuation of the Yenisei fault in the north that separates the Baikalides and Siberian Craton (**Figure 7**). Offset equivalents of the West Sayan and Tuva terranes are proposed in southern Transbaikalia.

The 'Caledonides' incorporate Neoproterozoic metamorphic terranes and Vendian to Early Palaeozoic magmatic arc and turbidite terranes. In Russia, there are terranes of two Neoproterozoic to Early Palaeozoic arcs – Kipchak and Tuva-Mongol, overlapped by the Middle Palaeozoic Kazakh-Mongol and Rudny Altai and Late Palaeozoic to Early Mesozoic Orkhon-Selenga-Khanka arcs.

The Kipchak Arc includes the Salair and Batenev immature arc terranes separated by sutures with Vendian-Early Cambrian ophiolites from the adjacent Neoproterozoic terranes of Kuznetsk Alatau and Derba. These structures are stitched by Cambrian to Silurian granitoids, but they are offset by the Kuznetsk-Teletskoye sinistral strike-slip fault for as much as 200 km. In the north, the Variscan Tom-Kolyvan Terrane is thrust southward onto the 'Caledonides'. It consists of flysch that might represent an accretionary wedge of an arc, presently hidden under the West Siberian sedimentary basin.

To the south and west, the 'Caledonian' island arc terranes are bound by the accretionary and forearc complexes of Gorny Altai and West Sayan. They consist of Vendian to Silurian flysch with slivers of cherts, basalts, and Vendian–Early Cambrian ophiolites. In the east, they have a T-shaped junction with the Main Sayan Fault, a continuation of the above-mentioned Yenisei Fault that strikes for more than 3000 km. It truncates the accretionary terranes of West Sayan, and their dextrally offset equivalents can be found 600 km to the south-east in Transbaikalia.

South of the accretionary terranes are the 'Caledonian' terranes of Central Mongolia. These are Precambrian metamorphic and Vendian-Cambrian immature arc terranes, known as the Tuva-Mongol Arc. The arc starts near the Stanovoy ridge of the Siberian Craton and continues to the west. It then turns to the south and east, forming the Mongol and then the Khingan oroclines. The 'Caledonian' terranes are stitched by Cambrian-Ordovician and Silurian granites and Devonian calc-alkaline and alkaline volcanicsin Minusa, Rudny Altai, and other locations.

In the core of the Mongol Orocline are Neoproterozoic to Early Mesozoic accretionary wedges of the Khangai-Khentei Superterrane. In Russia the superterrane is exposed in southern Transbaikalia, from where its analogues can be traced to the Circum-Pacific orogens via the Mongol-Okhotsk Suture. In

Mongolia, these terranes are overlain by the Carboniferous to Early Mesozoic Orkhon-Selenga-Khanka Arc. The huge Barguzin Batholith was emplaced behind this arc in Transbaikalia. The northern edge of this arc is present in Transbaikalia and its eastern edge overlaps the Bureya Terrane, but this latter structure belongs to the Kazakhstan-Khingan Domain.

Kazakhstan-Khingan Domain

In Russia, the Kazakhstan-Khingan Domain includes the Urals and buried continuation of the Kazakh uplands. Outside Russia, it is exposed in Kazakhstan, western China, and southern Mongolia, appearing again in the Russian Far East.

The Urals is a 2000 km long isolated Variscan Orogen (Figure 9). In the Urals, there was no 'Caledonian' deformation. Its tectonic cycle started in the Late Cambrian-Ordovician. The Urals is traditionally subdivided into western and eastern slopes. The western slope consists of a deformed passive margin, which records the Urals history from Cambrian to Devonian times.

The eastern slope consists of ophiolite sutures and island arc terranes of Ordovician to Carboniferous age. The Sakmara suture separates the East European Craton and Magnitogorsk and Tagil arc terranes thrust westward onto its deformed passive margin. Due to this position, the island arc sequences represent one of the world's best preserved ancient arcs, with well pronounced internal zonation and famous volcanic-hosted massive sulphide deposits.

To the east is the East Uralian Megazone, considered as a north–south trending Late Palaeozoic 'granite axis of the Urals'. However, its Middle

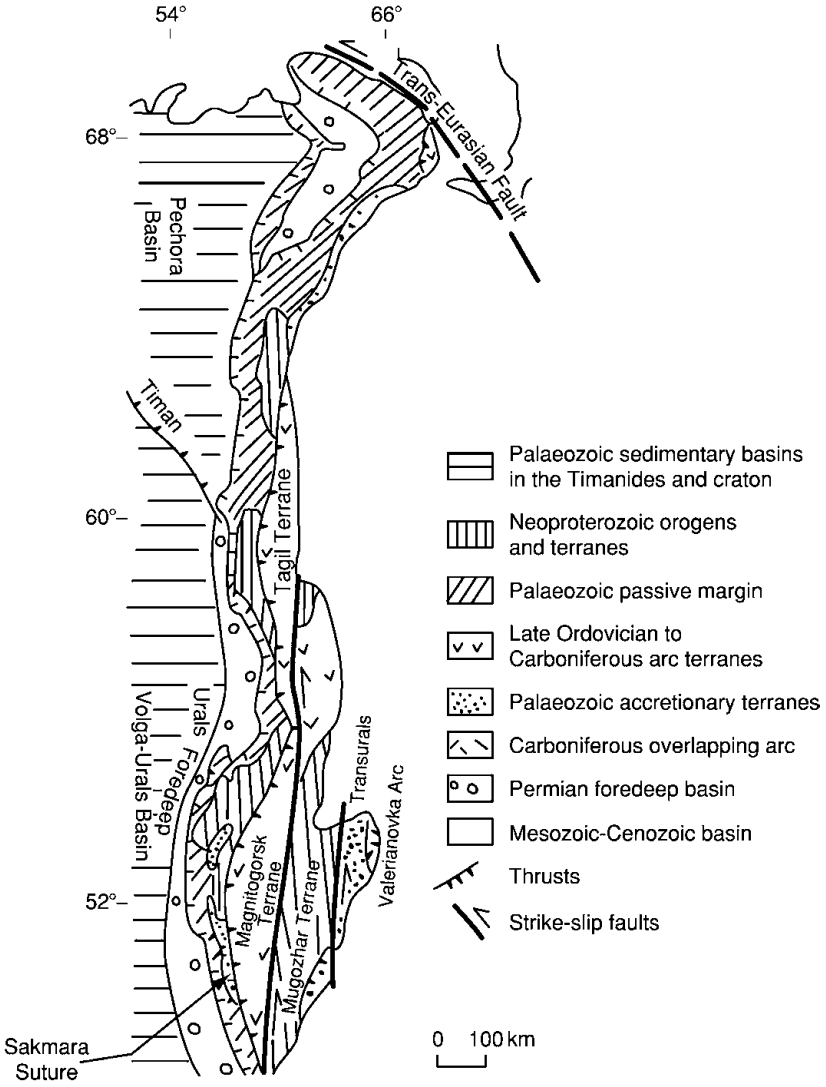


Figure 9 Tectonics of the Urals Orogen (compiled by Yakubchuk, using data by Puchkov and Sobornov). The traditional East Urals Zone consists of several NE-trending terranes truncated in the west and east by strike-slip faults. The sinistral amplitude of their displacement is at least 300 km.

Palaeozoic pre-granitic terranes strike obliquely relative to its western and eastern fault limits. These are two Precambrain metamorphic terranes and one Palaeozoic island arc terrane inbetween. The latter is principally similar to the Magnitogorsk Terrane, but it is offset sinistrally for 300 km. This suggests that the boundary faults are likely to be strike-slip faults. Seismic studies revealed that they dip to the west, producing an apparent divergent structure of the orogen.

Further to the east is the Trans-Uralian Terrane, consisting of Ordovician to Devonian accretionary wedge rocks. The easternmost structure in the southern Urals is the poorly exposed Carboniferous Valerianovka Volcanic Arc. It has been considered as an accreted arc, but regional airborne magnetic maps suggest that it can be traced to the Kurama Arc of Uzbekistan. This indicates that the Uralides can be interpreted as the result of arc-arc-continent collision.

This collision was accompanied by the formation of flysch in the Famennian to Early Carboniferous, which continued on into the Middle Carboniferous. This collision progressed northward, because in the northern Urals the accumulation of flysch and thrust deformations started in the Early Visean and ended by Kungurian time, when accumulation of salt and molasse took place. The collisional deformations terminated at the end of the Permian. In the Early Triassic, graben structures filled with clastic rocks and basalts formed in the Urals. They were deformed in Early Jurassic times.

The northern continuation of the Urals is a long-lasting debate. Some researchers suggested that Pai-Khoi and Novaya Zemlya represent its 'degraded continuation'. However, airborne magnetic data clearly suggest that magmatic arcs in the Polar Urals turn to the south-east under Mesozoic-Cenozoic sediments and might also be exposed in eastern Kazakhstan, forming a giant orocline. In the core of this orocline are Early and Middle Palaeozoic accretionary, magmatic arc and 1.0 Ga metamorphic terranes, identified on the basis of geophysical and drilling data. They are well exposed in the Kazakh uplands. The bulk of the Palaeozoic orogenic structures in Russia lie under the West Siberian sedimentary basin.

Mesozoic-Cenozoic Sedimentary Basins of Western Siberia

The West Siberian Basin is one of the three world's largest supergiant oil and gas reservoirs. It occurs between the Urals Orogen and the Siberian Craton (Figure 1). Its basement consists of various Neoproterozoic to Palaeozoic terranes that were amalgamated in the Permian. It has steep eastern and western borders and a flat bottom that dips to the north.

At the Permo-Triassic transition, all western Siberia experienced orogenic collapse which was followed by Triassic rifting and voluminous basalt magmatism of the Siberian Superplume. Rifts formed between the Urals and the Siberian Craton and in the western Kara Sea Basin. Urengoi is the largest rift, which was drilled down to a depth of more than 7 km. At the Triassic-Jurassic transition was a weak inversion and uplift, followed by post-rift subsidence for 4–7 km in the Jurassic-Cenozoic.

In Early-Middle Jurassic times, continental sedimentation dominated in the south, whereas there was a shallow-water marine environment in the north. The shoreline frequently migrated north or south. The dominant rocks are clastic alluvial, limnic, and marine sediments. In the Late Jurassic, the proportion of marine sandy and clayey sediments increased. At the Jurassic-Cretaceous transition, West Siberia subsided for 500 m, accompanied by the accumulation of the Bazhenovo oil-bearing facies which are enriched in organic matter. This subsidence was possibly facilitated by synchronous dextral strike-slip faulting along the western flank of the Siberian Craton.

In the Neocomian it became a relatively deep basin, which was quickly filled with sand-clay clinoforms. The main flux of clastic rocks was from the Siberian Craton and the Altai-Sayan Orogen. In the Aptian-Albian, the sedimentation environment became shallow-water. In the Late Cretaceous, Paleocene, and Eocene, the typical rocks are clay and sand with some cherty sediments. At the end of the Eocene, the northern part of the basin was uplifted and its link with the Arctic Seas was terminated. The Oligocene-Neogene in the central and southern parts of the basin is represented by shallow-water marine, limnic, and alluvial facies. Quaternary sediments are widely distributed. These are various sandy-clay facies which accumulated during several episodes of glacial and interglacial events and development of limnic-alluvial systems.

Circum-Pacific Orogenic Collages

The Circum-Pacific orogenic collages started to form in the Palaeozoic and their growth has continued until the present time. They occupy the Russian North-east and Far East. In the Russian North-east is the Verkhoyansk-Chukotka Orogenic Collage, bordered by the Nipponides.

Verkhoyansk-Chukotka Orogenic Collage The Verkhoyansk-Chukotka Orogenic Collage (Figure 10)

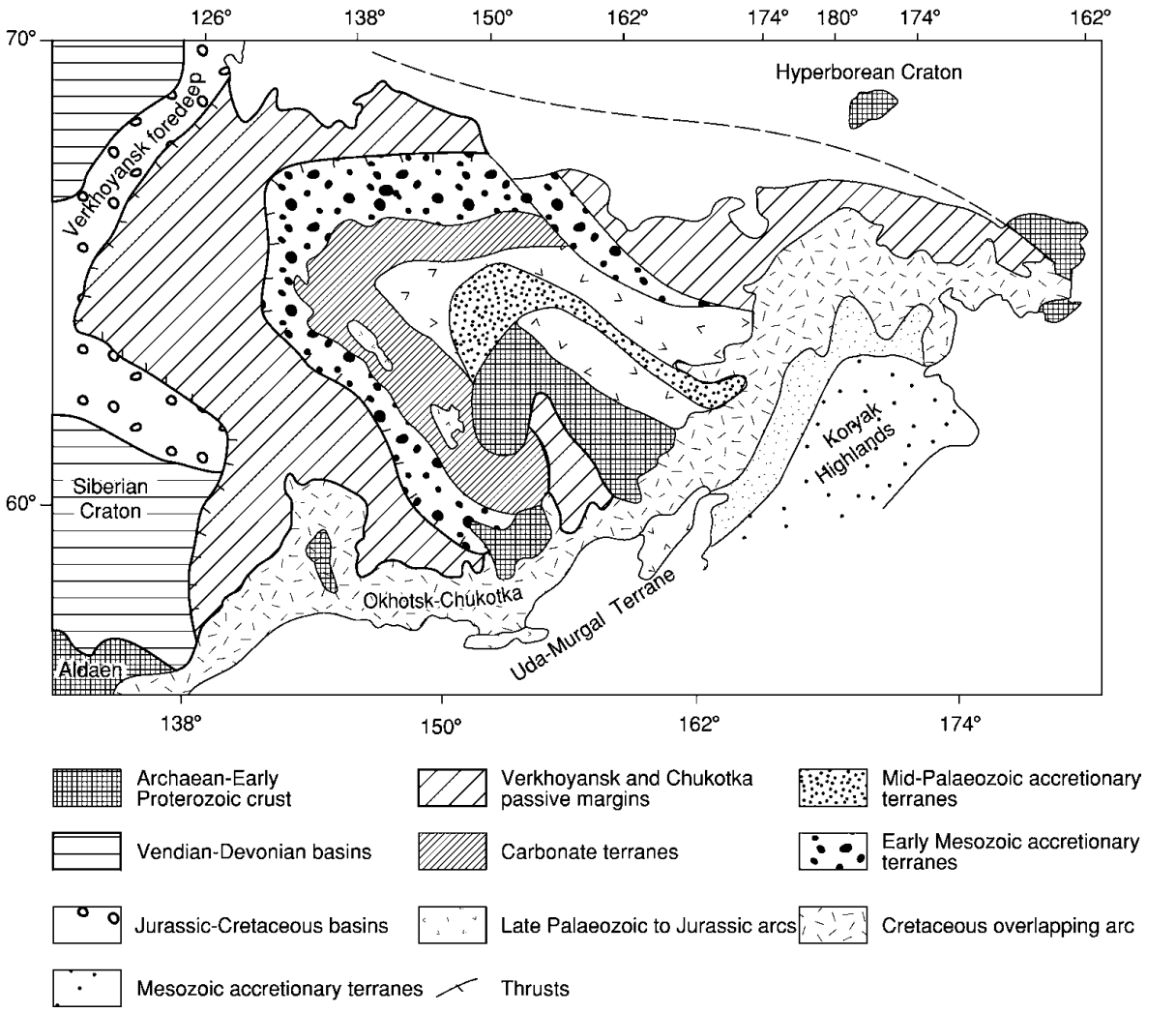


Figure 10 Tectonics of the Verkhoyansk-Chukotka Orogenic Collage (compiled by Yakubchuk and Nikishin). Note oroclinal structure of the collage. Mid-Cretaceous Okhotsk-Chukotka Arc stitched the terranes that constitute the collage. Terranes of the Koryak Highlands represent an accretionary wedge that formed in front of this arc.

evolved from the Early to Middle Palaeozoic until the Cretaceous. In the north, it is bounded by the Hyperborean Craton. To the south-west is the Siberian Craton. The eastern boundary of the collage coincides with the Okhotsk-Chukotka Andean-type arc that stitched various terranes in the Late Cretaceous.

The Hyperborean Craton is a possible fragment of the North American Craton. It was incorporated into the collage after spreading in the Canadian basin in the Late Jurassic to Early Cretaceous. Plate tectonic reconstructions suggest that its arrival in the Early Cretaceous was the major reason for the amalgamation of the Verkhoyansk-Chukotka Collage. This craton occurs almost completely in the shelf zone, and its small fragments are exposed in eastern Chukotka, Novaya Sibir, and the Wrangel Islands. In the south, it is bounded by the deformed passive margin terrane of Chukotka and Arctic Alaska. It consists of Triassic-Jurassic terrigenous rocks deformed in the Early Cretaceous. In the south, the Chukotka Terrane is bounded by the South Anyui Suture, hosting Late Palaeozoic to possible Early Mesozoic ophiolites and accretionary wedges.

The Verkhoyansk passive margin is a mirror image of the Chukotka Terrane, but it includes Ordovician to Middle Jurassic turbidite sequences which are present in the imbricated structure thrust onto the Siberian Craton. These assemblages extend along the north-east margin of the Siberian Craton and can also be traced along the western flanks of smaller cratonal terranes of Okhotsk, Omolon, and Prikolyma, forming the Kolyma Orocline.

The Verkhoyansk passive margin rocks are structurally superimposed by the Jurassic accretionary terranes. Inside them, in the core of the Kolyma Orocline, are the Early Palaeozoic carbonate terranes

and Middle to Late Palaeozoic terranes of the Alazeya Arc which are overlain by the Middle–Late Jurassic to Early Cretaceous Uyandina-Yasachnaya magmatic arc. The opposite ends of this arc collided with each other in the Early Cretaceous to form the Oloy Suture which extends to the east towards the Nipponides.

The 2500 km long Okhotsk-Chukotka continental arc stitched this collage in the mid-Cretaceous. This arc was active from the Albian until the Eocene due to subduction of the oceanic crust under the Siberian Craton. It reveals classic transverse zonation and Andean-type retreat of its magmatism towards the continent in time.

Nipponide Collage The term Nipponides was proposed to describe the structures whose tectonic style is similar to the Japanese islands. In Russia, the Nipponide Collage occupies Sikhote-Alin, Sakhalin, Sea of Okhotsk, Koryak Highlands, Kamchatka, and the Kurile Islands, where it is growing due to subduction of the Pacific oceanic plate (Figure 11).

In Sikhote-Alin, these are Middle Palaeozoic to Early Cretaceous accretionary terranes and forearc

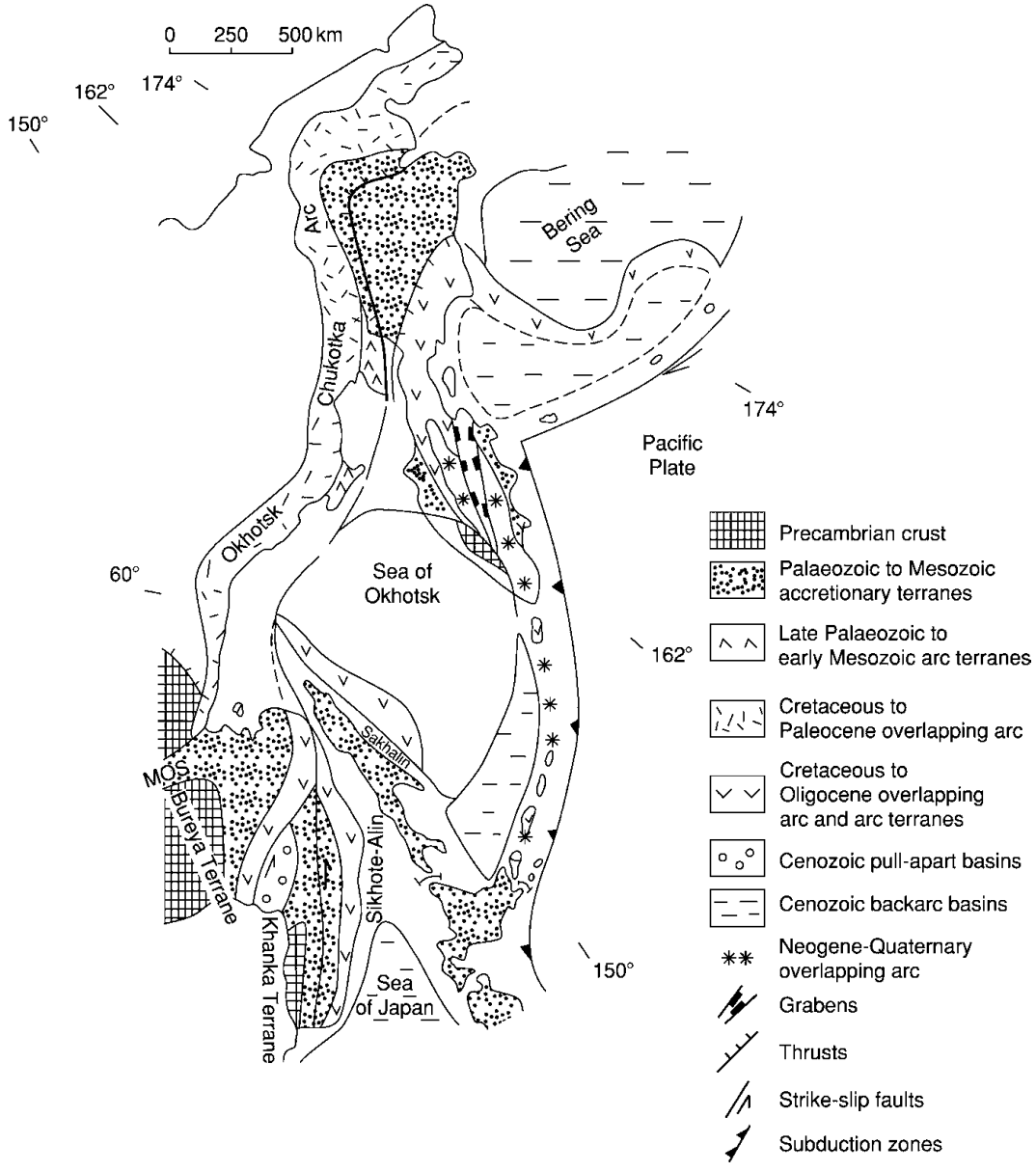


Figure 11 Tectonics of the Nipponides (compiled by Yakubchuk). Oroclinal bending and strike-slip duplication is typical for the Nipponides. Behind the active subduction zone of the Pacific plate, new oceanic crust formed in the Sea of Japan and southern Sea of Okhotsk basins. The Central Kamchatka Graben represents back-arc rifting.

turbidites. Similar accretionary terranes occur in front of the Okhotsk-Chukotka Arc in the Koryak Highlands. In Sikhote-Alin, they are overlain by another Late Cretaceous to Oligocene magmatic arc that extends towards the Sea of Okhotsk. Magnetic data and outcrops on the islands indicate that this arc turns to the south, so that the Mesozoic accretionary terranes of Sakhalin are squeezed between its two branches. The arc then turns to the north-east and continues to Kamchatka, where it forms several island arc terranes, continuing into the Aleutian island arc.

Inside this orocline is the Sea of Okhotsk. In the 1980s, a Precambrian cratonal terrane was proposed as its basement, but more recent work suggests that this is an accreted Cretaceous oceanic plateau. Whichever is correct, this structure is a major undeformed block inside the Nipponides.

Along the eastern flank of the Nipponides is the Kurile-Kamchatka chain of active volcanoes that have been superimposed on the older arc since the Neogene. In front of them are the actively growing accretionary wedges in the deep subduction-related trenches. Behind this arc are active back-arc basins with newly formed oceanic crust in the south of the Okhotsk Sea and Sea of Japan.

Alpine Belt

The Alpine Orogenic Belt intersects Russian territory south of the East European Craton, in the Scythian Platform located near the Sea of Azov and the northern Caucasus (Figure 1). Its sedimentary cover is 3–5 km thick. Its basement is intersected in numerous drillholes, but because of poor dating there are many problems in the interpretation of this region.

In its basement are the Karpinsky Swell and Scythian Orogen. The Karpinsky Swell is a buried eastern continuation of the Donets Basin that accumulated up to 15 km of Carboniferous shales. Devonian rifting was followed by formation of a Carboniferous back-arc deepwater basin was proposed to explain the Karpinsky Swell. This basin was deformed and thrust northward onto the Peri-Caspian Basin in the Early Permian. The structure of the Palaeozoic Scythian Orogen is poorly known, but in the Palaeozoic it was a continuation of the Greater Caucasus Orogen. The basement of the Scythian Orogen consists of Devonian and Carboniferous schists, Precambrian terranes, Palaeozoic ophiolites, and island arc terranes. Some interpretations suggest that this area was a Cordilleran-type orogen framed by the Palaeotethys Ocean to the south at the end of the Palaeozoic. In the Carboniferous to Early Permian, along the southern flank of the orogen in the Greater Caucasus, an Andean-type magmatic arc accompanied by molasse basins was formed. The other fragments of the orogen are in the Dzirula Massif in Georgia, the Eastern Pontides in Turkey, and Dobrudja in Romania. It was proposed to call it the Euxinian Orogen. To the west, this orogen continued into the Variscides of western Europe.

In Mesozoic to Cenozoic times, the Caucasus-Black Sea region had a complex history of a supra-subduction continental marginal system. In the Triassic, a back-arc basin represented by shales in Georgia and Dobrudja opened. In the Late Triassic, a volcanic arc formed in the north of the Greater Caucasus. At the Triassic-Jurassic transition, the entire north Caucasus region was deformed and its rifts were inverted. In the Early Jurassic, a Greater Caucasus deep back-arc trough was formed on the rifted and thinned continental crust. This trough was inverted in pre-Callovian times. In the Callovian to Late Jurassic, a new deep back-arc rift on the continental crust was formed along the present southern slope of the Greater Caucasus. This basin survived as a deep-water structure until the end of the Eocene, but it also experienced several phases of extension and compression. In the Middle Cretaceous, continental back-arc rifting occurred in the Black Sea and transformed into spreading in the West and East Black Sea Basins in the Late Cretaceous. At the Eocene to Oligocene transition, the main collision in the Greater Caucasus took place. Since that time Cis-Caucasus foredeeps – Kuban in the west and Terek-Caspian in the east – began to form. The collisional shortening in the Greater Caucasus is estimated to be at least 200–300 km, and the orogen was uplifted by 5 km. This elevation was a result of Arabian Plate collision, and it was accompanied by intrusive and explosive volcanism in such volcanoes as Elbrus. In contrast, the molasse basins subsided by 5–7 km.

Modern Plate Tectonics

Russian territory occupies two major modern plates. These are the Eurasian and North American Plates, and their boundary goes from the Arctic Ocean towards Sakhalin Island (Figure 12). In the ocean it is a spreading ridge, whereas in the continent it is a system of diffused continental rifts. It behaves as a transform (strike-slip) fault boundary near Sakhalin island. In the east, the North American Plate is bounded by a subduction zone of the Pacific Plate that produced the active volcanoes of the Kurile-Kamchatka Arc.

Some researchers subdivide the North American Plate into smaller plates, suggesting identification of

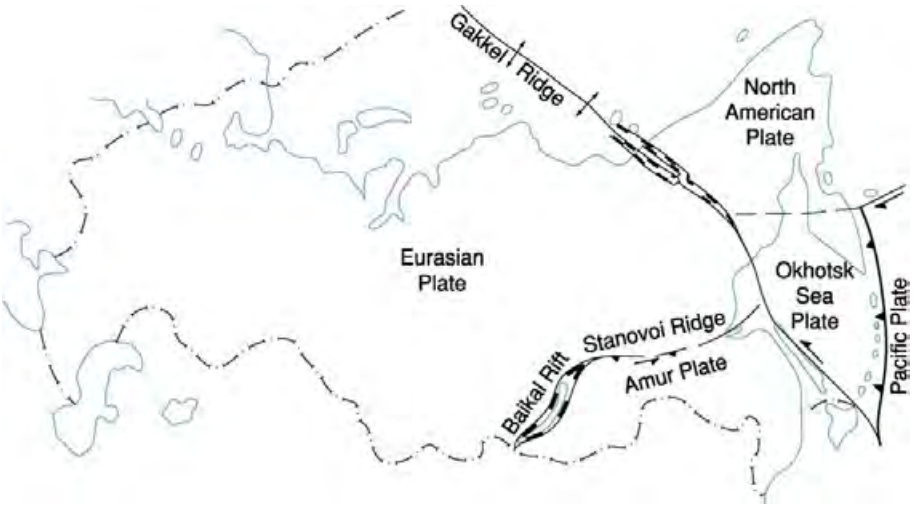


Figure 12 Modern plates in the territory of Russia (compiled by Yakubchuk and Nikishin). Parts of the two major plates of Eurasia and North America can be separated as smaller plates of Okhotsk and Amur.

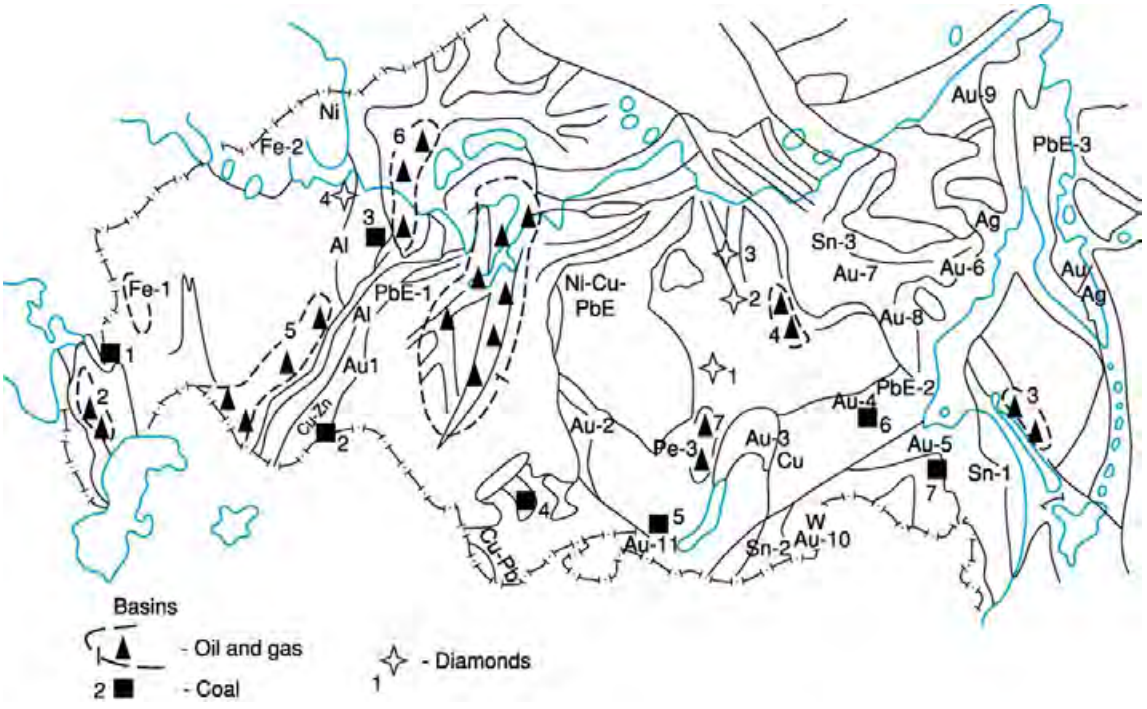


Figure 13 Mineral resources of Russia (compiled by Yakubchuk). Commodities and deposits: U – Transbaikalian province, Fe – iron ore deposits and their numbers in the text, Ni – Pechenga Ni deposit, Ni-Cu-PGE – Norilsk-Talnakh deposits, Cu – Udokan deposit, Cu-Zn – Urals VMS deposits, Cu-Pb – Rudny Altai VMS deposits, Au – gold deposits and their numbers in the text, PGE – platinoid deposits and their numbers, Al – bauxite deposits, Sn – tin deposits and their numbers in the text, Au/Ag – emerging epithermal gold province in Kamchatka. Solid lines correspond to the tectonic units as shown on **Figure 1**.

a separate Okhotsk Sea Plate. Following the active seismic belt extending along Lake Baikal and the Stanovoy Ridge, it was proposed to recognize the area to the south as the Amurian Plate that includes most of (China *see* **China and Mongolia**). The Eurasian-Amurian plate boundary in this case is a zone of active continental rifting produced in response to India-Asia collision.

Mineral Resources

Russia has a large variety of mineral resources, many world class. Numbers on Figure 13 will be referred to as a basis for the following brief description.

Energy Resources

Russia is the world's second largest producer of oil and the world's largest producer of gas. Its supergiant oil and gas fields are in the Mesozoic-Cenozoic West Siberian Basin (Oil basin N1 in Figure 13). Smaller Cenozoic provinces are in the Scythian Basin (2) and in Sakhalin (3). Mesozoic oil is known in the Verkhoyansk foredeep (4). Middle Palaeozoic provinces are in the Cis-Urals (5), Pechora (6), and Lena (7) basins. Russia's significant coal resources occur in several basins. Palaeozoic coal is in the Donets (1) and Kuznetsk Basins. Most of the Mesozoic coal is in the Urals (2), Pechora (3), Kuznetsk (4), Bureya (5) and Aldan (6) provinces. The supply for generation of nuclear energy comes from a domestic source in Transbaikalia as well as from other CIS countries.

Mineral Deposits

Russia has enormous mineral wealth due both to the size of its territory and world-class quality of some of its deposits. Russia is ranked fourth or fifth in the world in terms of iron ore reserves. Its most productive deposits are in the Early Proterozoic Kursk Magnetic Anomaly (Fe-1), Palaeozoic intrusions in Karelia (Fe-2), and Triassic plume-related pipes in the Siberian Craton (Fe-3). Russia is the world's largest producer of nickel, whose ores are mined from the Early Proterozoic Pechenga deposit in the Kola Peninsula and lateritic deposits in the Urals, but most of the nickel production (as well as 50% of Russian Cu, 60% of the world's Pd and 20% of the world's Pt) is extracted from the Norilsk-Talnakh supergiant deposit formed during the activity of the Siberian Superplume.

Russia's other producing base metal (Cu-Pb-Zn) deposits are volcanic-hosted massive sulphides in the Middle Palaeozoic arcs in the Urals and Altai. Despite significant aluminium production, Russia has only small- to medium-size bauxite deposits, such as in Timan. Domestic tin production is almost suspended, but in recent years Mesozoic tin was extracted from Sikhote-Alin (Sn-1), Transbaikalia (Sn-2), and Yakutia (Sn-3). Russia is famous worldwide due to its precious metals and diamond production. 25% of the world's raw diamonds are extracted annually from the Palaeozoic and Mesozoic kimberlite pipes in the Siberian Craton, such as Mir (1), Udachnaya (2), Aikhal, and Ebelyakh (3).

In terms of gold resources, Russia is second in the world after South Africa. Russia's first gold mine was established in the Urals at the Berezovskoe deposit (Au-1) in the nineteenth century. Today most of production comes from the Yenisei Ridge (Au-2), Patom Highlands (Au-3), Aldan (Au-4), Far East (Au-5), Kolyma (Au-6), Indigirka (Au-7), Sette-Daban (Au-8), Chukotka (Au-9), Transbaikalia (Au-10), and eastern Sayan (Au-11). Russia's largest silver mine is at Dukat in the north-east. A new epithermal gold–silver province was discovered in Kamchatka. In addition to Norilsk, Russian Pt production comes from placer deposits in the Urals (PGE-1), Kondyor (PGE-2), and the Koryak Highlands (PGE-3).

See Also

Asia: Central. **China and Mongolia**. **Europe:** East European Craton; Timanides of Northern Russia; The Urals. **Precambrian:** Vendian and Ediacaran.

Further Reading

Khain VE (2001) *Tectonics of the Continents and Oceans*. Moscow: Nauchnyi Mir(in Russian).

Milanovskiy EE (1996) *Geology of Russia*. Moscow: Moscow University Press(in Russian).

Nikishin AM, Ziegler PA, Stephenson RA, *et al*. (1996) Late Precambrian to Triassic history of the East European Craton. *Tectonophysics* 268: 23–63.

Nikishin AM, Ziegler PA, Abbott D, Brunet M-F, and Cloetingh S (2002) Permo–Triassic intraplate magmatism and rifting in Eurasia: implications for mantle plumes and mantle dynamics. *Tectonophysics* 351: 3–39.

Mezhelovsky NV, Morozov AF, Gusev GS, and Popov VS (eds.) (2000) *Geodynamics and Metallogeny: Theory and Implications for Applied Geology*. Moscow: Ministry of Natural Resources of the Russian Federation, Inter-Regional Center for Geological Cartography (GEOKART).

Rozen OM (2003) Siberian craton: tectonic zonation and evolution stages. *Geotektonika* 3: 3–21.

Sengor AMC and Natal'in BA (1996) Palaeotectonics of Asia: fragments of a synthesis. In: Yin A and Harrison MT (eds.) *Tectonic Evolution of Asia*, pp. 486–640. Cambridge: Cambridge University Press.

Yakubchuk A (2002) The Baikalide-Altaid, Transbaikal-Mongolian and North Pacific orogenic collages: similarity and diversity of structural patterns and metallogenic zoning. In: Blundell DJ, Neubauer F, and von Quadt A (eds.) *The Timing and Location of Major Ore Deposits in an Evolving Orogen*, pp. 273–298. London: The Geological Society.

Yakubchuk AS and Nikishin AM (2004) Noril'sk–Talnakh Cu–Ni–PGE deposits: a revised tectonic model. *Mineralium Deposita* 39: 125–142.

Zonenshain LP, Kuzmin MI, and Natapov LM (1990) *Geology of the USSR: A Plate Tectonic Synthesis*. Washington, DC: American Geophysical Union.

SATURN

See **SOLAR SYSTEM: Jupiter, Saturn and Their Moons**

SEAMOUNTS

S M White, University of South Carolina, Columbia, SC, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Maps of the seafloor reveal a surface that is dotted with a vast number of cones and truncated cones, arranged in lines, grouped in clusters, or isolated. These are known as seamounts, the name given to any steep-sloped more-or-less conical feature on the seafloor. The overwhelming majority of seamounts are volcanoes. In fact, seamounts are the most common volcanic landform on Earth, but one of the least studied, owing to their wide dispersal and relative inaccessibility. As a consequence, much of the information presented here is derived from a small number of well-studied seamounts and applied to the general population of seamounts.

Seamounts are officially defined as volcanic cones on the seafloor with at least 1 km of relief. This distinction arose during the days of seafloor mapping with wide-beam echosounders when only these relatively large features could be unambiguously identified. Technological advances in remote sensing have improved our ability to locate and image seamounts of increasingly smaller size. Modern shipboard multibeam echosounders provide high-resolution maps showing the morphology and distribution of small ($>$50 m high) seamounts. Deep-towed sonars reveal tiny ($<$20 m high) seamounts along the axes of mid-ocean ridges. Satellite altimetry has enabled complete mapping of the ocean basins, but is capable of imaging only relatively large ($>$2 km high) seamounts. There is little scientific justification for a 1 km height cut-off for seamounts. All volcanoes start growing from the seafloor, regardless of their current size. For the purpose of this work, seamounts are defined broadly to include submarine volcanic edifices of any size.

Seamounts, ocean islands, and guyots form a natural continuum in the process of submarine volcanic construction from submerged, through emergent, to erosional. This article focuses on the processes that build and erode seamounts and their distribution in the oceans. The quasi-conical shape and volcanic origin of seamounts distinguish them from other relief-forming features on the seafloor, such as abyssal hills and carbonate reefs, which are not dealt with here. One exception is the process of serpentine mud volcanism, which builds seamounts in convergent-margin settings.

Geochemical and Geophysical Characteristics

Nearly all seamounts, aside from some in island arcs, are composed of basalt. Geophysical studies also suggest that seamounts have densities, magnetizations, and seismic velocities that are consistent with porous basalt. Most seamounts consist of a base layer of alkali basalt that grades into voluminous edifice-building tholeiitic basalt, which is capped by alkali basalt. Occasionally differentiation products of alkali basalt (hawaiite, mugearite, benmoreite, and trachyte) are observed in the latter stages of growth. However, not all seamounts exhibit all these growth stages, and small seamounts in particular may cease activity before reaching the main tholeiite stage. Smaller seamounts located near mid-ocean ridges have a composition that strongly resembles that of the lava erupted at the mid-ocean ridge but have a tendency to be both more primitive and more depleted than ridge-axis basalt.

Seismic, geodetic, and gravity studies have identified magma chambers beneath several ocean islands. Geophysical studies of the Kilauea volcano on Hawaii provide perhaps the clearest picture of a magma system anywhere. These studies define a partly molten conduit that rises to about 5 km beneath the summit caldera, with large blade-like dikes running laterally down the rift zones. Large positive gravity anomalies centred beneath many of the islands of French Polynesia suggest large (100 km^3) frozen magma chambers at shallow levels (2–4 km). At Krafla on Iceland, attenuation of seismic shear waves suggests a large

magma body (10–100 km^3) at a depth of 3 km below the surface. Small melt bodies that have been detected seismically in the uppermost mantle near the EPR (East Pacific Rise) may be evidence that even very small near-ridge seamounts have ephemeral magma chambers.

Distribution

Global Distribution and Spatial Arrangement

The earliest studies of seamount distribution and abundance in the ocean basins relied on estimates of height and volume obtained from single-beam echosounder crossings of seamounts. Since single-beam data provide only a cross-sectional view of the topography, with no guarantee of crossing the summit, approximations were required to estimate true seamount height and shape. Nevertheless, these early studies provided valuable indications of the distribution of seamounts in the ocean basins. Modern multi-beam echosounders have revealed the complexity and variety of seamount shapes and distributions, but still only cover regional areas (a few hundreds of square kilometres). The development of satellite altimetry has allowed the complete mapping of the larger seamounts in the ocean basins (Figure 1). Studies of seamounts suggest that satellite altimetry reliably detects seamounts that are at least 2 km high.

The global distribution of seamounts is best approximated by power-law models. Power-law models predict that there are in the order of 10^5 seamounts over 1 km high in the ocean basins and in the order of 10^7 seamounts including seamounts that are much less than 1 km high. Over 50% of all seamounts are found in the Pacific Ocean. Statistical studies of seamount distribution find very clear differences in the distributions of seamounts in different oceans and in different tectonic settings (Table 1).

Seamounts are distributed in several spatial patterns. The most widely known pattern is the seamount chain (Figure 2). In seamount chains, several very closely spaced individual volcanoes occur in a line. The Hawaiian–Emperor Seamounts are a classic example of a seamount chain. In this chain, as in many others, the bases of individual seamounts nearly touch or may completely overlap, but each volcano creates an individual peak. When seamounts are even more closely spaced, so that individual seamounts along the chain are difficult to recognize, the arrangement is called a volcanic or aseismic ridge. Aseismic ridges often occur along the track of a hotspot, for example the Ninety

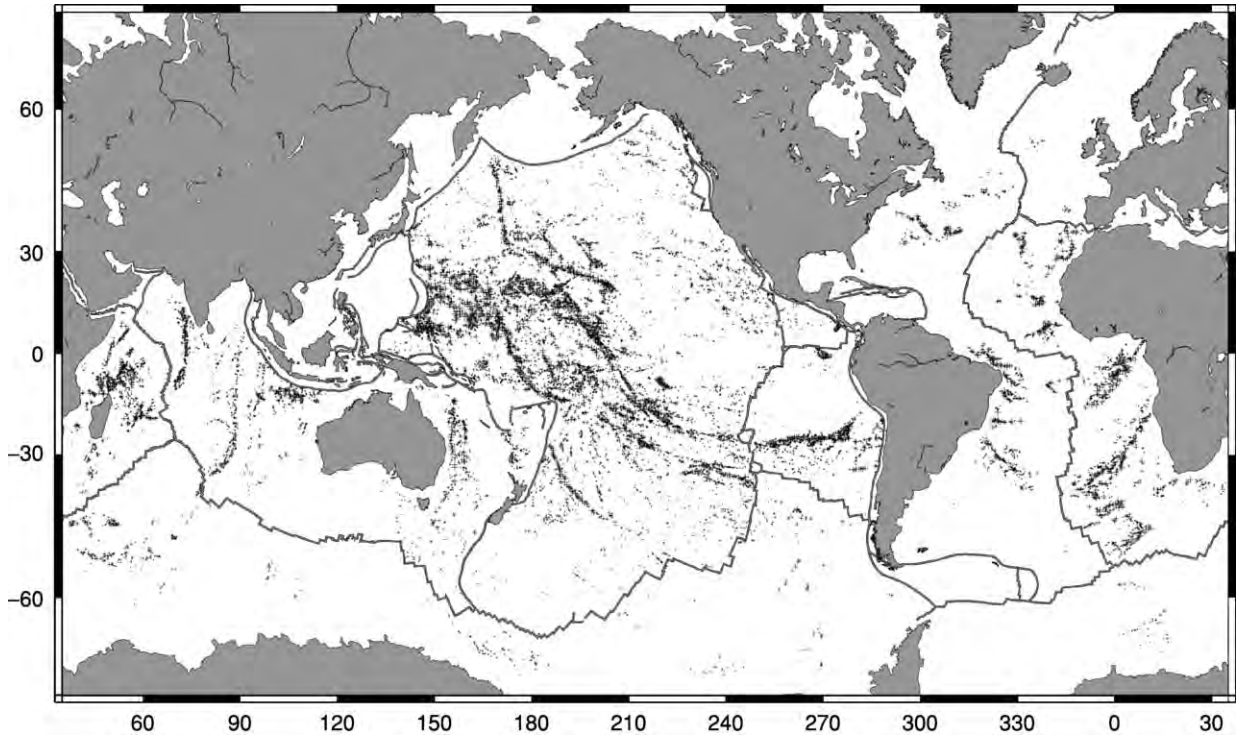


Figure 1 Global distribution of seamounts (indicated by plus symbols scaled according to the size of the seamount edifice) detected by Geosat/ERS-1 altimetry. Approximately 14 675 seamounts ranging in size from approximately 1 km to 7 km were identified. Reproduced by permission of American Geophysical Union from Wessel P (2001) Global distribution of seamounts inferred from gridded Geosat/ERS-1 altimetry. *Journal of Geophysical Research* 106: 19 431–19 441. Copyright 2001 American Geophysical Union.

Table 1 Seamount abundances in intraplate and mid-ocean ridge settings

Region	Inclusive area (km^2)	Lithospheric age (Ma)	Observed seamount abundance ($10^3\,km^{-2}$)	Minimum height cutoff (m)	Reference
Mid-Atlantic Ridge, 24° – 30° N	6027	0–0.5	79.8	50	Smith and Cann (1992)
Reykjanes Ridge, 57° – 62° N	5075	0–0.6	106	50	Magde and Smith (1995)
Mid-Atlantic Ridge, 25° – 27° N	73 000	0–28	18.35	70	Jaroslow *et al.* (2000)
Galapagos Ridge, 91° – 97.6° W	14 100	0–0.5	18.4	50	Behn *et al.* (2003)
East Pacific Rise, 8° – 17° N	200 400	0–2.9	0.9	200	Schierer and Macdonald (1995)
East Pacific Rise, 15.5° – 20° S	113 000	0–1.9	9.5	50	White *et al.* (1998)
Easter Seamount Chain	243 000	0–9.8	1.6	200	Rappaport *et al.* (1997)
Pacific plate	Pacific plate	0–169	Total = 8882	~1500	Wessel and Lyons (1997)
Global	All ocean basins	0–180	Total = 14 675	~1500	Wessel (2001)

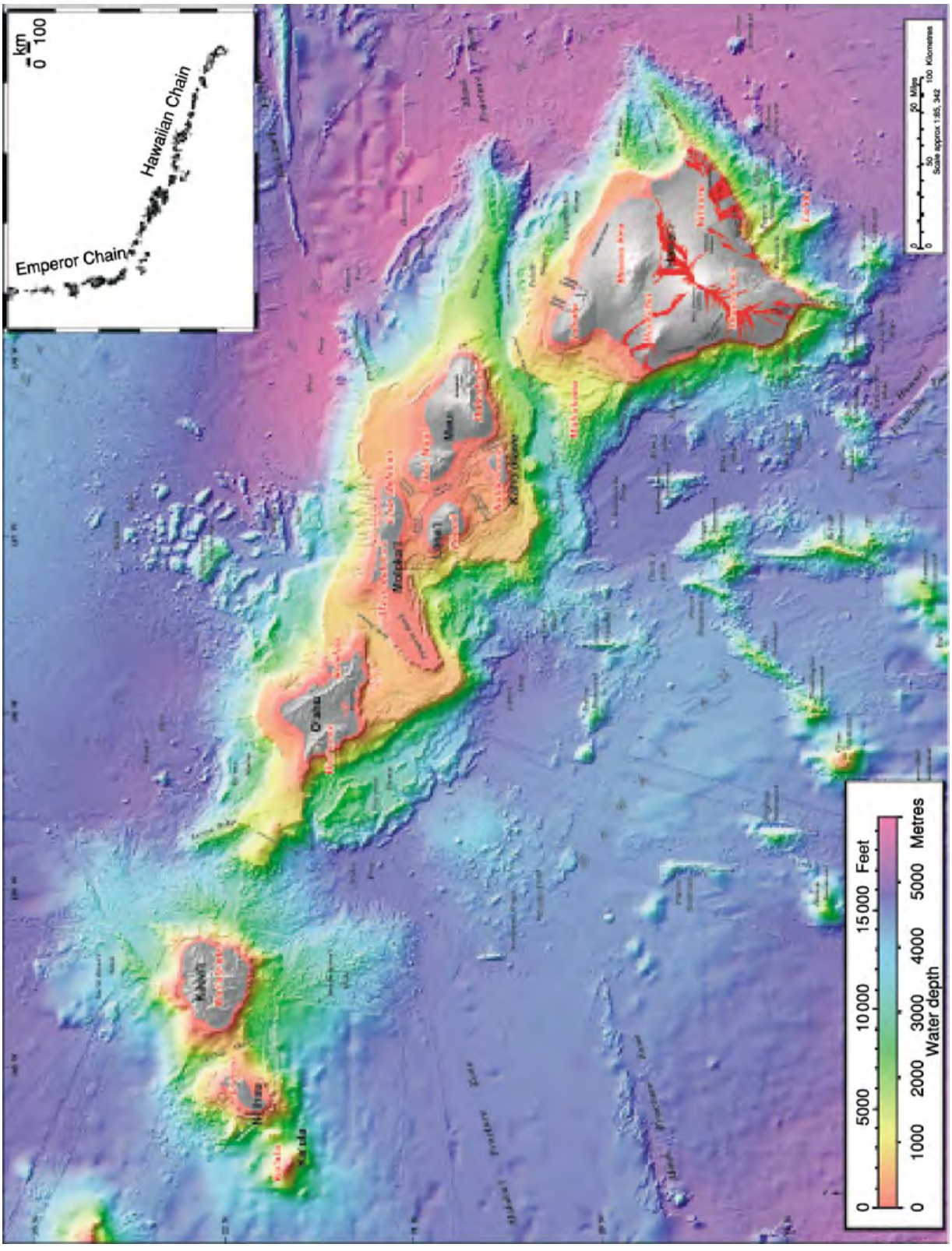
Data based on raw seamount counts and seafloor survey areas selected on the basis of complete seafloor coverage with modern mapping technology in the following references: Behn MD, Sinton JM, and Detrick RS (2003) Effect of the Galapagos Hotspot on seafloor volcanism along the Galapagos Spreading Center (90.9° – 97.6° W). *Earth and Planetary Science Letters*, vol 217, pp. 331–347 (2004); Jaroslow GE, Smith DK, and Tucholke BE (2000) Record of seamount production and off-axis evolution in the western North Atlantic Ocean, 25°25 – 27°10 N. *Journal of Geophysical Research* 105: 2737–2760; Magde L and Smith DK (1995) Seamount volcanism at the Reykjanes Ridge: relationship to the Iceland Hotspot. *Journal of Geophysical Research* 100: 8449–8468; Scheirer DS and Macdonald KC (1995) Near-axis seamounts on the flanks of the East Pacific Rise, 8° – 17° N. *Journal of Geophysical Research* 100: 7871–7885; Smith DK and Cann JR (1992) The role of seamount volcanism in crustal construction at the Mid-Atlantic Ridge (24°30 N) *Journal of Geophysical Research* 97: 1645–1658; Rappaport Y, Naar DF, Barton CC, Liu ZJ, and Hey RN (1997) Morphology and distribution of seamounts surrounding Easter Island. *Journal of Geophysical Research* 102: 24 719–24 728; Wessel P (2001) Global distribution of seamounts inferred from gridded Geosat/ERS-1 altimetry. *Journal of Geophysical Research* 106: 19 431–19 441; Wessel P and Lyons S (1997) Distribution of large Pacific seamounts from Geosat/ERS-1: implications for the history of intraplate volcanism. *Journal of Geophysical Research* 102: 22 459–22 475; White SM, Macdonald KC, Scheirer DS, and Cormier M-H (1998) Distribution of isolated volcanoes on the flanks of the East Pacific Rise, 15.3° – 20° S. *Journal of Geophysical Research* 103: 30 371–30 384.

East Ridge in the Indian Ocean. However, some aseismic ridges seem to emerge from mid-ocean ridges, such as the Sojourn Ridge and the Pukapuka Ridges in the southern Pacific. Seamounts also appear in clusters. Classic examples of the cluster pattern include the Mid-Pacific Mountains and the Hawaiian North Arch volcanic field. Finally, isolated seamounts are fairly common around fast-spreading mid-ocean ridges, where large numbers of very small seamounts (less than 1 km high) occur randomly on the seafloor.

Tectonic Setting: Intraplate Seamounts

The largest seamounts and the longest seamount chains known on Earth are intraplate seamounts. In the Atlantic, several intraplate seamount chains can be traced back to the edge of the continent, indicating 80–120 Ma of continuous activity. Probably the most famous example of an intraplate seamount chain, the Hawaiian–Emperor chain, stretches for 5000 km from its source to 80 Ma seamounts at its end (Figure 2). The oldest parts of this chain may have been subducted, so its true longevity is unknown. The Hawaiian chain also boasts the largest volcano on Earth: Mauna Loa (with an estimated volume of 42 500 km^3). What causes this great volume of volcanic activity is still a subject of debate.

The most widely accepted hypothesis for the origin of intraplate seamounts is that they are formed above hotspots supplied by mantle plumes (*see* **Mantle Plumes and Hot Spots**). The widely accepted model for hotspots states that they are fixed with respect to the Earth's centre of mass and thus leave a trail of consecutively older seamounts as the plate moves over the stationary hotspot. The trend and age progression of the hotspot seamount chain should record the absolute (relative to the Earth's centre of mass) plate-motion vector. The strongest evidence supporting the hotspot model is the linear arrangement and age progression along seamount chains. When a hotspot occurs underneath a mid-ocean ridge it can send out seamount chains or ridges on both sides of the plate boundary. The Carnegie and Cocos ridges from


km
0 100
Hawaiian Chain
Emperor Chain
Oahu
Molokai
Maui
Kahoolawe
Hawaii
Kauai
Niihau
0 1000 2000 3000 4000 5000 Metres
0 5000 10000 15000 Feet
Water depth
Scale approx 1:85, 342
100 Kilometres

the Galapagos hotspot, and the Walvis Ridge and Rio Grande Rise from the St Helena hotspot are two examples of this dual-ridge phenomenon.

However, many intraplate seamount chains fail to show the simple age progression predicted by the hotspot model. Cocos Island on the Cocos Ridge has lavas that are much younger than predicted by a simple age-progression model consistent with a hotspot origin. Likewise, the Line Islands, the Cameroon Line, and the Pratt–Welker chain show no regular age progression. The lack of age progression in many, even most, seamount chains has resulted in the advancement of several hypotheses to the explain them. Recent palaeomagnetic studies suggest that the hotspot source of the Emperor seamount drifted south at a rate of approximately 30 mm a^{-1} before becoming stationary at 43 Ma. Several other seamount chains in the Pacific exhibit changes in direction near 43 Ma (the Gambier–Tuamotus, Gilbert–Marshall, and Austral chains) but the significance of these kinks in the seamount chains is not yet clear. Lack of age progression and hotspot fixity make it unlikely that all intraplate seamounts result from mantle plumes rising from the deep mantle. Other hypotheses for the origin of seamount chains are extensional stress across zones of pre-existing weakness in the lithosphere, feedback between plate flexure and tectonic stress, and small-scale convective rolls in the upper mantle.

Tectonic Setting: Mid-Ocean Ridges

Seamounts are most abundant near mid-ocean ridges (*see* **Tectonics:** Mid-Ocean Ridges). These seamounts are predominately small; thus, high-resolution bathymetric mapping is necessary to detect them. Most of the studies mapping seamounts near mid-ocean ridges rely on morphometric assumptions that they have an aspect ratio of 2 : 1 or less and that their heights are several times the resolution of the mapping instrument to distinguish them from abyssal hills or other tectonic seafloor topography. The near-ridge population consists mostly of seamounts less than 1 km high, with the overwhelming majority being small isolated seamounts less than 100 m high (Table 1). On slow-spreading ridges, such as the Mid-Atlantic Ridge, seamounts occur directly over the spreading axis within the rift valley of the ridge, while on fast-spreading ridges, such as the East Pacific Rise, there is no rift valley and no seamounts along the spreading axis. The largest on-axis seamount yet discovered (more than 5 km high) is on the ultra-slow spreading Gakkel Ridge, while recent surveys have found tiny (less than 20 m high) pillow-lava mounds as the largest edifices on the axis of the fast-spreading East Pacific Rise. Thus, seamounts over the spreading axis grow larger but become less abundant as the spreading rate decreases. However, small (less than 200 m high) seamounts are ubiquitous away from the axis of fast-spreading ridges and increasingly rare on slower-spreading ridges. Small but ubiquitous seamounts form on lithosphere younger than 0.2 Ma, regardless of spreading rate. This suggests that the thickness of the brittle lithosphere controls the near-ridge seamount population.

Many seamount chains originate at mid-ocean ridges. Near-ridge chains are typically shorter and composed of smaller volcanoes than the intraplate seamount chains. Near-ridge chains become more common as the spreading rate increases, with the largest seamount chains being associated with locations where the ridge axis exhibits signs of a high magma budget. For example, the Axial seamount on the Juan de Fuca ridge is a place where excess magma has produced the Cobb–Eikelberg seamount chain. The highest concentration of seamount chains is adjacent to the most inflated and shallowest part of the super-fast-spreading southern East Pacific Rise (Figure 3). Rarely seamount chains are found near fracture zones, but these are smaller lower-volume seamounts than those found near the middle of ridge segments. Unlike the hotspot chains discussed above, most near-ridge seamount chains are orientated between the direction of absolute plate motion and that of relative plate motion at the ridge. Most of the chains are aligned with the direction of shallow asthenospheric flow beneath the ridge flanks.

Tectonic Setting: Island Arcs

Seamounts occur in island arcs as a product of subduction-zone volcanism. Seamounts are distributed in a linear or curved pattern following the curvature of the trench. Tectonic segmentation of the arc by cross-faults creates discontinuities in the seamount distribution, illustrating the strong tectonic control on the distribution of volcanoes in island-arc settings. Seamounts in island arcs are more diverse in both morphology and composition than seamounts in the other two tectonic settings. The Izu–Ogasawara Arc

Figure 2 Bathymetry surrounding the Hawaiian Islands, a prominent intraplate seamount chain stretching across half of the Pacific plate (see inset), reveals several long rift zones and smaller seamounts. Volcanically and hydrothermically active Loihi, to the southeast of Hawaii, is the youngest volcano in the chain. Reproduced with permission from Eakins BW, Robinson JE, Kanamatsu T, *et al.* (2003) *Hawaii's Volcanoes Revealed.* Geological Investigations Series I-2809. US Geological Survey.

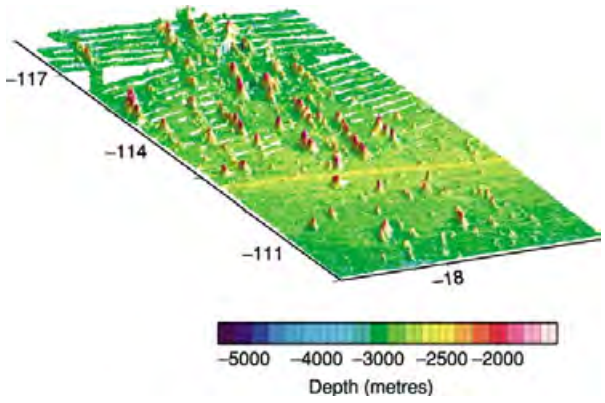


Figure 3 Seamount chains in the Rano Rahi Field adjacent to the southern East Pacific Rise. The shaded-relief bathymetry reveals a remarkable abundance of seamounts near the part of the spreading axis with the greatest inferred magma budget in the region.

contains eight submarine calderas that have erupted large (more than 100 km^3) volumes of rhyolite pumice. In contrast, Kick-Em-Jenny, the southernmost island in the Lesser Antilles, is an island-arc seamount that has erupted 10 times since 1939, forming lava domes of basaltic magma within the summit caldera of its very steep-sided edifice. Almost every island arc has a number of seamounts that are smaller versions of the emergent volcanic islands making up the arc.

Mud volcanoes are a unique amagmatic type of seamount found in the fore-arc regions of several subduction zones. The widely accepted explanation for mud volcanism is the upwelling of overpressured serpentinite diapirs from the subduction décollement through normal faults in the fore arc. One of the largest and best-studied mud volcanoes is the Conical Seamount (25 km in diameter and 2 km high), which is located in the fore arc of the Marianas Arc. As more seafloor mapping is done along outer fore arcs, more examples of mud volcanoes are discovered.

Distribution through Time: Cretaceous Seamounts

During the Cretaceous (85–120 Ma) incredibly intense seafloor volcanism occurred throughout what is now the western Pacific (Figure 1). Not only were many of the large oceanic plateaus emplaced in the Pacific during the Cretaceous (Ontong–Java, Manihiki, Shatsky), but also the topography of the seafloor reveals a striking abundance of large seamounts (Mid-Pacific Mountains, Magellan Seamounts, Wake Guyots, etc.). Both the highest density of seamounts per unit area and the highest density of large (more than 3.5 km high) seamounts occur in the Pacific. Satellite altimetry data reveal that the concentration of Cretaceous seamounts is almost double that of Cenozoic seamounts.

Morphology

The vast majority of seamounts are predominately constructed of effusive lava flows. Thus, the overall shapes of seamounts are primarily controlled by the geometry of the magma plumbing system, the eruption rate, the lava type, and the local stresses on the volcano edifice. The cooling effect of water leads to the construction of much steeper edifices than those on land. Most seamounts have steep (10° – 30°) sides but nearly flat summit areas, in contrast to the low slopes of shield volcanoes (5° – 10°). As a seamount grows, it is subject to more complex stresses and tends to develop a more complex shape (Figure 4). In map-view, seamount shape evolves from nearly circular small seamounts to the more complex stellate forms of large seamounts and ocean islands. As a seamount grows larger, mass wasting from gravitational instability or wave erosion becomes more important in modifying its morphology.

Seamount Growth and Development

Deep-water stage Basaltic magma erupted under high hydrostatic pressure produces non-explosive eruptions forming edifices built from lava flows (Figure 4A). Repeated eruptions build extrusive edifices composed of pillow lavas with minor sheet lava flows. Intrusion of magma into the pelagic sediments on the seafloor produces a mixture of lava and wet sediment called peperite. Small seamounts (less than 1 km high and less than 10 km^3 in volume) usually form edifices that can best be described as low lava domes. Many seamounts even of this small size have relatively steep slopes and flat summits, unlike subaerial lava shields. Studies of small isolated seamounts near mid-ocean ridges show that a few seamounts of this size develop complex shapes, including relatively large summit craters. The seamount shape becomes generally more complex as the seamount grows. When seamounts reach a somewhat larger size (*ca.* 1 km high and 10–100 km^3 in volume) they more often have an irregular outline and an 'overturned soup bowl' or truncated-cone shape.

Shoaling stage Large seamounts (more than 2 km high and with volumes of over 200 km^3) may develop a stellate shape as a result of long rift zones concentrating lateral eruptions of lava from a central feeding system (Figure 4B). Not all large seamounts develop rift zones (e.g. Galapagos), but the rift zones on some seamounts become very pronounced ridges extending for more than 100 km (e.g. Hawaii). The presence of a magma chamber above the level of the seafloor and

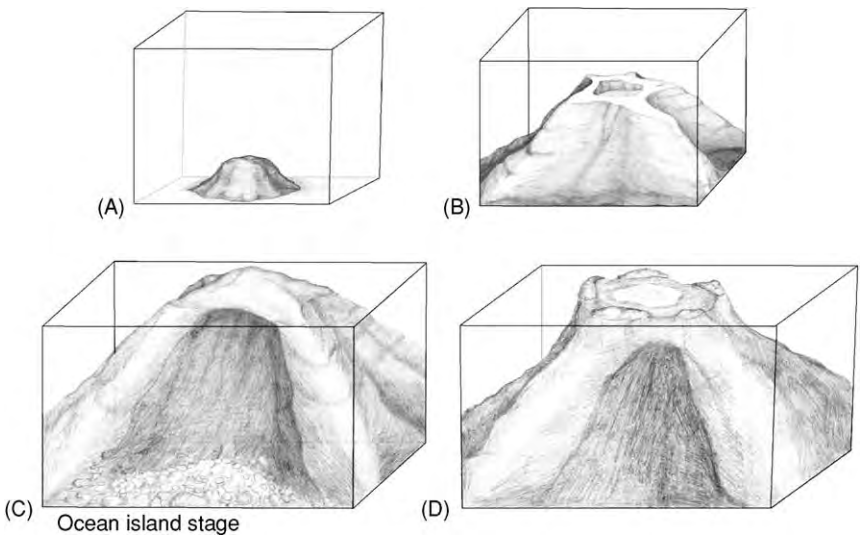


Figure 4 The growth stages of seamounts from (A) early deep-water stage, through (B) shoaling stage, showing the development of rift zones, and (C) ocean island stage, including the development of a subaerial shield volcano and mass wasting, to (D) an atoll in the guyot stage.

within the edifice itself aids the development of rift zones on large seamounts by enabling lateral dikes and flank eruptions in focused zones of weakness. Within about 1 km of sea-level, reduced hydrostatic pressure allows exsolution of gas, and explosive eruptions become frequent. Lava flows of scoriaceous pillows and lapilli breccia become important lava types.

Emergent stage At sea-level, eruptions become surtseyan in type, producing fine-grained base surge and air-fall tephra deposits. Besides the famous and well-described eruption of Surtsey off Iceland (1963–1967), the only recorded observations during the emergent stage are from Graham Island (1831), Metis Shoal (1968, 1979), Myujin-sho (1952–1953), and Shin-Iwo-Jima (1986). Phreatomagmatic eruptions form tuff rings and cones consisting of easily eroded deposits of loose lapilli and ash. If eruptions stop at this stage, wave action will quickly erode the seamount. The small islands of Myujin-sho disappeared within months of formation.

Ocean island stage Once the volcanic vent rises above sea-level, phreatomagmatic eruptions become rare and effusive eruptions resume (Figure 4C). Pahoehoe and aa are the important lava types, and in most cases a lava shield is created from numerous low-relief lava flows. Giant landslides may cover the flanks of the island with debris from slumping or debris avalanches. Some of the largest landslides on Earth (involving volumes of 5000 km^3) have occurred around the Hawaiian Islands (Figure 2), and some studies speculate that up to 50% of the original volume of the older Hawaiian Islands has been lost owing to mass wasting.

Guyot stage Thermal subsidence and erosion combine to drown all islands after volcanic activity ceases. Wave erosion planes off a flat summit plateau, but the sides of the volcano remain steep, creating the distinctive guyot cross-section. Atolls (including the Marshall, Tuamotu, Kiribati, and Maldive Islands) are a type of guyot found in tropical oceans, where coral forms a fringing reef around an ocean island. If the upward growth of the coral reef keeps pace with the subsidence of the seamount, a ring of coral islands form, encircling a lagoon, and a thick layer of carbonate sediment develops, which caps the old seamount basement rocks (Figure 4D).

Why Seamounts Have Flat Tops

The characteristic steep-sided flat-topped shape of seamounts is a morphology rarely seen in subaerial volcanoes. The flat tops of guyots can be explained by wave erosion, but many seamounts with flat tops have never been near the wave base. Several models have been advanced to explain why so many seamounts have a summit plateau or flat-topped shape. It is still unclear whether one model can explain all the occurrences of flat-topped volcanoes, or whether each of various models is appropriate under certain conditions.

In the ring-dike model proposed by Batiza and Vanko, eruptions occur from a series of circumferential dikes encircling the summit area rather than from

a central feeder vent. The steep slopes and flat summits of the Galapagos volcanoes, which essentially have the same form as flat-topped seamounts, are attributed to lateral extension by ring-dike intrusion. Circumferential fractures observed in the summit areas of some near-ridge seamounts support the ring-dike model. Ring dikes are thought to form when magma stalls within the crust, creating a small laccolith. Ring dikes shoot upwards from the edge of the laccolith as it inflates. A variation on this model proposes dipping cone sheets instead of more vertical ring dikes as the intrusive body.

In the caldera-fill model proposed by Clague and others, large calderas form early in seamount growth and are filled by later eruptions. A repeated sequence of caldera collapse and re-filling allows the seamount to grow upwards and outwards while maintaining a flat top. This model is based on new high-resolution mapping that suggests that lava shields (erupted from central vents) formed on the summits of flat-topped volcanoes in the north-eastern Pacific.

In the neutral-buoyancy model proposed by Barone and Ryan, low-density basaltic crust counterbalances the buoyancy of the magma so that eruptions will rise only to a limited height. Seamounts will grow only to the height that balances the magma pressure in their magma chambers and will then grow laterally instead of upwards, forming a flat top. Other studies, however, have questioned why this model would lead to circular volcano outlines and what continues to drive eruptions once the critical height is reached.

Broader Effects of Seamounts

Subduction Asperities

Seamounts being subducted erode the fore arc and may also cause erosion at the base of the overriding plate. Huge furrows, mass wasting, and structural disruption of the fore arc are the effects attributed to seamount subduction (Figure 5). There are many areas that exhibit the morphological signs of seamount subduction, but proving the existence of buried seamounts depends on interpreting seismic and magnetic data. However, seamounts being subducted are sometimes directly visible, for example Daiichi-Kashima Guyot in the Japan Trench and the Bannock (Seamount) Structure in the Mediterranean. Seamount subduction may also be responsible for seismic asperities on the subduction thrust-fault interface. Seamount tunnelling, the process by which a seamount going down on the subducting plate scrapes material off the overriding plate, has been inferred to have an important role in crustal development and fluid flow in the margin. The subduction of large seamount chains and aseismic ridges coincides with a reduced dip angle in the subducting plate, for example in the case of the Juan Fernandez chain and the flat slab in northern Chile.

Hydrothermal Circulation

Active seamounts may host high-temperature focused hydrothermal flow. The Loihi seamount in the Hawaiian chain hosts high-temperature hydrothermal vents within its summit caldera, which are very similar to vents found at mid-ocean ridges (Figure 2). Most seamounts have neither the magma supply volume nor the longevity of Loihi, but sulphide hydrothermal deposits have been found on the summits of several other seamounts during submersible dives. These hydrothermal deposits indicate that seamount and mid-ocean ridge hydrothermal systems are broadly similar and contribute the same ions to seawater. Possibly more significant is the role that inactive seamounts play as basement outcrops that provide easy paths for the escape of fluid and heat.

Oceanographic Circulation

The topography created by seamounts has a large effect on the local water masses, primarily generating upwelling and enhancing currents. Anticyclonic circulation over the tops of seamounts creates a so-called ‘cold dome’ as a result of enhanced upwelling along the sides of the edifice. Also, long aseismic ridges and seamount chains are barriers to deep- and mid-water circulation. Aseismic ridges standing 2–4 km above the seafloor and stretching for hundreds of kilometres across the Pacific deflect north–south currents at depth.

Critical Habitat

Seamounts play a key role in ocean environments by providing habitats for fishes and suspension feeders. The enhanced upwelling combined with the potential for a reef environment in very shallow seamounts make large seamounts sites of high primary productivity. Seamounts have been known as productive fishing grounds for centuries, but their role in oceanic biodiversity has been appreciated only in the last 50 years. Seamounts host a relatively large percentage (estimated at 15–35%) of endemic species and may be important sites of speciation for deep-sea fauna. Seamounts in the south-west Pacific show highly localized species distributions and fairly limited recruitment among seamounts, even between closely spaced seamounts. The formation of atolls by coral growth on subsiding seamounts creates an important habitat niche for filter feeders and reef fishes. Finally, many humans find ocean islands and

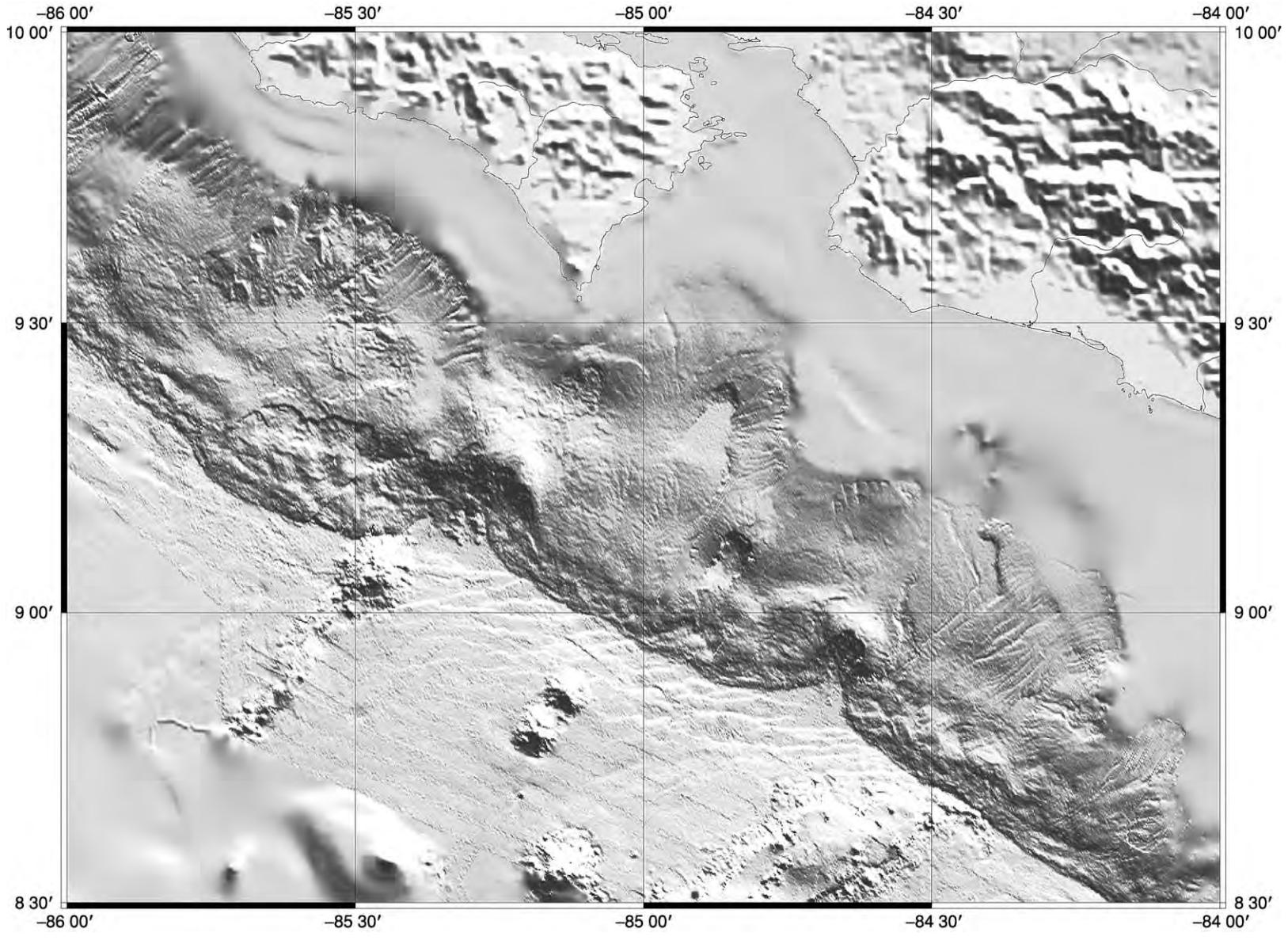


Figure 5 Seamount subduction has produced huge furrows and re-entrants in the fore-arc margin of Costa Rica. Shaded-relief bathymetry reproduced from Dominguez S, Lallemand SE, Malavieille J, and von Huene R (1998) Upper plate deformation associated with seamount subduction. *Tectonophysics* 293: 207–224, copyright 1998, with permission from Elsevier.

atolls highly desirable locations for settlement and vacation.

Glossary

Abyssal hill A low-relief elongate hill formed by horst-and-graben style faulting around a mid-ocean ridge.

Alkali basalt A variety of basalt that has relatively high sodium and potassium contents.

Atoll A ring of carbonate reef islands enclosing a lagoon built atop a subsiding seamount in tropical latitudes.

Pillow lava A type of subaqueous lava flow consisting of spherical or elongate tubes formed by the rapid quenching of lava. Tubes seen in cross-section often resemble pillows, with a flat base and a convex upper surface.

Scoria Glassy porous pyroclastic lava.

Surtseyan A type of volcanic eruption, based on observations at Surtsey in Iceland, which is the phreatomagmatic equivalent of a Hawaiian or quiescent lava flow eruption.

Tholeiitic basalt A variety of basalt characterized by the presence of orthopyroxene in addition to clinopyroxene and calcium-plagioclase.

See Also

History of Geology Since 1962. **Large Igneous Provinces**. **Lava**. **Mantle Plumes and Hot Spots**. **Sedimentary Environments:** Reefs ('Build-Ups'). **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes. **Tectonics:** Hydrothermal Activity; Mid-Ocean Ridges. **Volcanoes**.

Further Reading

Barone AM and Ryan WBF (1990) Single plume model for asynchronous formation of Lamont Seamounts and adjacent Pacific Rise terrains. *Journal of Geophysical Research* 95 N. B7: 10801–10827.

Batiza R (1989) Seamounts and seamount chains of the eastern Pacific. In: Winterer EL, Hussong DM, and Decker RW (eds.) *The Geology of North America, Volume N, The Eastern Pacific Ocean and Hawaii,* pp. 289–306. Boulder: Geological Society of America.

Batiza R and Vanko D (1984) Volcanic development of small oceanic central volcanoes on the flanks of the East Pacific Rise inferred from narrow beam echo-sounder surveys. *Marine Geology* 54: 53–90.

Clague DA, Reynolds JR, and Davis AS (2000) Near-ridge seamount chains in the north-eastern Pacific Ocean. *Journal of Geophysical Research* 105: 16 541–16 561.

Davis EE (2000) Earth science: volcanic action at Axial Seamount. *Nature* 403: 379–480.

Johnson HP and Embley RW (eds.) (1990) Axial Seamount: an active ridge axis volcano on the central Juan de Fuca Ridge. *Journal of Geophysical Research* 95: 12 689–12 966.

Keating BH and McGuire WJ (2000) Island edifice failures and associated tsunami hazards. *Pure and Applied Geophysics* 157: 899–955.

Keating BH, Fryer P, Batiza R, and Boehlert W (eds.) (1987) *Seamounts, Islands, and Atolls*. Geophyiscal Monograph Series Volume 43. Washington DC: American Geophysical Union.

Mitchell NC (2001) Transition from circular to stellate forms of submarine volcanoes. *Journal of Geophysical Research* 106: 1987–2004.

Moore JG, Normark WR, and Holcomb RT (1994) Giant Hawaiian landslides. *Annual Review of Earth and Planetary Sciences* 22: 119–144.

Ranero CR and Von Huene R (2000) Subduction erosion along the Middle America convergent margin. *Nature* 404: 748–752.

Schmidt R and Schminke H-U (2000) Seamounts and island building. In: Sigurdsson H, Houghton B, McNutt SR, Rymer H, and Stix J (eds.) *Encyclopedia of Volcanoes,* pp. 383–402. San Diego: Academic Press.

Takahashi E, Lipman PW, Garcia MO, Naka J, and Aramaki S (eds.) (2002) *Hawaiian Volcanoes: Deep Underwater Perspectives*. Geophysical Monograph Series Volume 128. Washington DC: American Geophysical Union.

Thorarinsson S (1967) *Surtsey: The New Island in the North Atlantic*. New York: Viking.

Wessel P (2001) Global distribution of seamounts inferred from gridded Geosat/ERS-1 altimetry. *Journal of Geophysical Research* 106: 19 431–19 441.

SEDIMENTARY ENVIRONMENTS

Contents

Depositional Systems and Facies
Alluvial Fans, Alluvial Sediments and Settings
Anoxic Environments
Carbonate Shorelines and Shelves
Contourites
Deltas
Deserts
Lake Processes and Deposits
Reefs ('Build-Ups')
Shoreline and Shoreface Deposits
Storms and Storm Deposits

Depositional Systems and Facies

J Collinson, John Collinson Consulting, Beech, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The concept of facies as applied to sediments is a means of classifying and grouping deposits in such a way that objective differences, usually with genetic significance, are highlighted. In present-day settings, the relationships between processes and environments are self-evident, but in sequences of ancient sediment various analytical procedures are needed to reconstruct environments. 'Facies analysis' is such a suite of techniques and is based on an assemblage of facies and the relationships of the facies to one another. The results of facies analysis may be expressed as facies models, which summarize and, to some extent, idealize the relationships and allow the facies of a stratigraphical interval to be predicted in areas where it cannot be directly observed. Facies analysis also provides a major input to 'sequence stratigraphy' (*see* **Sequence Stratigraphy**), whereby facies successions are placed within a stratigraphical framework defined by key stratal surfaces that are related to changes in relative sea-level or base-level.

Facies

The concept of facies dates back to the early nineteenth century, when A Gressly introduced the term to summarize those descriptive features of a rock that distinguish it from other rocks. Historically, the concept has been applied at a wide range of scales and, therefore, with varying significance. At the largest scale, 'facies' has been applied to major stratigraphical units in particular tectonic contexts. 'Flysch' and 'molasse' are examples of this and are used particularly in the description of the Tertiary fold belts of Europe. At a large scale, 'facies' is also used to describe regional differences of depositional environment within a particular stratigraphical interval. For example, in Britain, one might contrast the 'basinal facies' of the Silurian of Central Wales with the coeval 'shelf facies' of the Welsh Borders. These usages are generally accepted and understood but, in order to arrive at an interpretation of the depositional environment, it is necessary to go through a process of facies analysis. This involves establishing an appropriate facies scheme for the succession being studied and then analysing the spatial relationships of those facies.

Facies schemes are established using a variety of physical, chemical, and biological criteria, and there is no such thing as a universally applicable facies system. Different sedimentary successions demand different schemes because of their contrasting character or because the intended end products of facies analysis may differ. A simple facies scheme may suffice for a rather basic interpretation, whilst a more elaborate scheme may be appropriate for a more refined interpretation. The facies scheme therefore depends on the scale and detail of the study.

In successions of dominantly clastic sediments, a facies scheme would commonly be based largely on

the physical properties of the sediments. Grain size is a common starting point for subdivision. Not only dominant grain size but also secondary features, such as sorting, might be considered. Any differences in composition would also be used for primary subdivision. A subdivision based on grain size may be refined by considering sedimentary structures (e.g. styles of lamination), so that facies such as 'fine-grained, ripple cross-laminated sandstone' or 'horizontally laminated siltstone' might result. Facies defined in this way are called lithofacies, and they can commonly be interpreted in terms of depositional processes by applying knowledge of the relationships between hydrodynamics, particle size, and sedimentary structures. However, many depositional processes occur in a range of sedimentary settings, and, therefore, few lithofacies are fully diagnostic of depositional environment.

Other features, formed soon after deposition, permit more refined facies schemes. Bioturbation by animals or plant roots can provide clues about the environment, and highly distinctive facies such as coal may also allow a specific environment of deposition to be deduced. Where burrowing is important, the assemblage of burrow types and the intensity of burrowing may necessitate schemes of ichnofacies. Early-formed concretions and textures in palaeosols can indicate the drainage conditions of emergent surfaces soon after deposition and can lead to pedofacies. If sediments are fossiliferous, the types and diversities of the fossils may lead to biofacies, and consequent facies analysis may incorporate palaeoecological considerations. Analysis of very fine plant debris or spores may lead to palynofacies.

For carbonate sediments, the nature of the constituent grains is important in establishing the depositional setting, and their characterization through petrographical description may lead to microfacies. In complex sequences, combinations of these different approaches may be appropriate.

Where two contrasting lithologies are interbedded, it may be appropriate to recognize event deposition (Figure 1). In many successions of interbedded muds and sands, sand beds commonly have sharp bases and may show internal grading and vertically changing styles of lamination. They were clearly deposited by sudden high-energy events in otherwise low-energy settings. Establishing the nature of the events may depend on interpreting quite subtle features of both the sand beds and the finer interbeds. Where the muddy component lacks evidence of higher energy and the sands show only current-generated structures and lamination, the events are likely to be turbidity currents in a deep-water setting. If the finer sediments are strongly bioturbated and show signs of fluctuation energy and the sand beds show evidence of wave action, perhaps reworking their tops, then storm events on a shelf may be the appropriate inference. If the finer sediments show evidence of emergence, with soil textures, and the sands are current dominated, then river floods (crevasse events) on a floodplain may be the appropriate interpretation. In all these cases, it would be valid to regard the finer and coarser components as separate facies, but it may prove practical to use composite facies based on the thicknesses and proportions of the coarse and fine components. This approach is particularly applicable to turbidite successions.

Figure 1 Interbedded sandstones and mudstones showing the distinction between the quiet background sedimentation of muds and the high-energy events that delivered the sands. In this case, the high-energy events were turbidity currents in a deep-water marine setting. Paleocene, Zumaia, Spain.

Facies Analysis

Most facies defined according to the criteria outlined above may be interpreted in terms of processes of deposition, but these are seldom unique to a

particular environment. In order to identify the depositional settings of ancient sediments, it is necessary to understand the processes active in present-day environments and the ways in which sediments are deposited and preserved there. In other words, it is necessary to interpret the patterns of facies seen in the rocks by comparison with the patterns of processes established in modern settings. Understanding the interrelationships of facies and processes allows environments of deposition to be inferred with greater precision.

Process-based facies analysis, which began in earnest in the early 1960s, concentrated initially on understanding the vertical succession of facies observed in a borehole core or stream section. Key to this approach was an appreciation of the importance of the nature of the contacts between vertically adjacent facies. In some cases, the contact is a gradational transition, whilst in other cases it is sharp or even erosive. These distinctions are important in interpretation, which is usually based on the application of Walther's Law of the Correlation of Facies, which was first formulated in 1884, but has been applied critically only since the late 1950s. This law, simply stated, points out that facies that are laid down side-by-side in their depositional environment will come to lie vertically on top of one another in the same order as their lateral distribution, or, conversely, only those facies that are superimposed one upon the other without a break can have been deposited in laterally adjacent settings (Figure 2). The importance of the nature of the vertical facies contacts is apparent in Figure 2, which shows that an erosion surface can allow the superposition of facies deposited in widely separated settings. Sedimentary hiatuses can have the same result. A rapid transgression can lead to shallow-water facies being directly overlain by deep-water facies, with facies representative of intermediate environments missing across the contact because of a lack of deposition or condensation, though not necessarily requiring a temporal break in sedimentation.

Some facies successions show systematic trends of vertical change, expressed through grain size or other properties. In other cases, a less organized succession may occur. Trends of vertically increasing and decreasing grain size (coarsening-up and fining-up units) are amongst the most common patterns encountered in clastic successions, and these may, for example, be equated with prograding shorelines and laterally migrating channels, respectively (Figure 3). Such facies trends may occur as single examples or as vertically repeated patterns. Where repetition occurs, successions are commonly described as cyclic, and the idea of cyclic sedimentation has a history extending back to the early nineteenth century. In cyclic successions, it is important to distinguish between interpreting the facies sequence within cyclothems (i.e. the environment) and explaining the cyclicity (i.e. the repetition). Facies patterns may repeat vertically because of inherent properties of the depositional environment, such as the tendency for river or distributary channels to switch positions on a floodplain or delta top. Such processes are termed autocyclic. Where external controls, such as fluctuating sea-levels or subsidence rates, are thought to cause the repetition, the processes are termed allocyclic.

In cyclic successions, it is rare for all cyclothems to show an identical facies sequence. Rather, they are variations on a theme (Figure 4). Various statistical techniques can be employed to extract a 'complete' or 'ideal' cyclothem, the interpretation of which can be developed into a facies model. However, it is also instructive to consider the variations in the facies successions of cycles and their relationships with the ideal cycle, since the variability itself may help to refine the model. For example, a series of coarsening-up cycles may record successive progradations of a delta, whilst the differences between individual cycles may reflect differences between subenvironments of the delta, stacked in vertical succession through autocyclic delta-lobe switching. Comparison of the facies in the upper parts of several

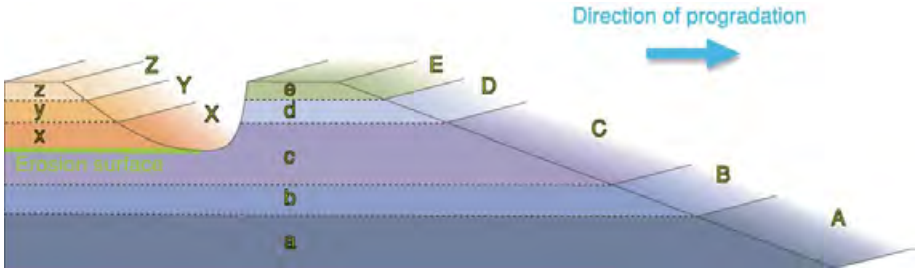


Figure 2 An illustration of the operation of Walther's Law of the Correlation of Facies. The subenvironments A–E lie side-by-side and, through progradation, generate a vertical facies succession a–e. The development of an erosion surface through the migration of a channel triggers a new facies sequence, x–z, which is independent of the earlier one. The erosion surface separates facies that were generated in settings that have no predictable lateral relationship.

Figure 3 A large-scale coarsening-up unit generated by the progradation of a major delta. The unit ends abruptly at a flooding surface (arrowed), above which are stacked smaller-scale coarsening-up units (parasequences). Namurian, County Clare, Ireland.

cycles may lead to inferences about the variability of the delta front and delta plain and hence allow a more refined characterization of delta type. The validity of such an approach depends, of course, on the successively prograding deltas having been of broadly similar character.

In successions where no cyclicity is apparent and no patterns of vertical facies change can be identified, mathematical investigation such as Markhov chain analysis may help to detect an underlying pattern beneath the random noise. However, the demonstration of statistical significance may be limited by insufficient data.

Architectural Elements and Bounding Surfaces

Where sedimentary successions are very well exposed in extensive desert or coastal cliffs, it is clearly not sensible to rely solely on the vertical sequence of facies to reconstruct the depositional environment. Lateral facies relationships, where they can be

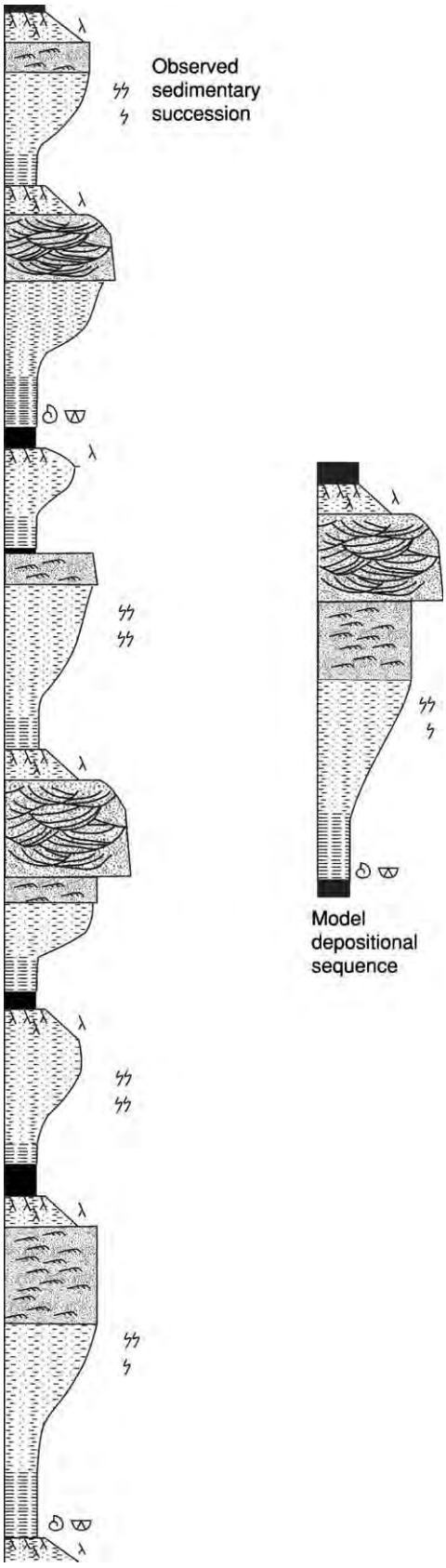


Figure 4 The facies succession of an idealized model cyclothem compared with the observed facies successions upon which it is based.

observed, are highly instructive, and the two- and three-dimensional shapes and sizes of facies units may be highly diagnostic. Lateral changes in facies or in the thicknesses of facies units (architectural elements) may help to localize particular elements of the depositional environment. For example, the thickening of crevasse splay sands within a floodplain succession may be linked to the proximity of a channel sandstone.

Laterally extensive exposures may reveal subtle low-angle dipping surfaces that would remain undetected in one-dimensional sections such as borehole cores. Such surfaces may be confined within channel sand bodies, indicating lateral accretion of the channel or downstream accretion of a channel bar. They may also occur on a larger scale within coarsening-up units such as progradational clinoforms deposited in a delta-front setting. Such surfaces contribute to facies definition and, more importantly, they aid interpretation by constraining the context of the facies.

Within complex channel sandstones, the usefulness of Walther's Law is often limited by abundant erosion surfaces. In a narrow one-dimensional section, the significance of a particular erosion surface may be difficult to establish. The basal erosion surface of a channel sand body, where it rests on contrasting, probably finer-grained, facies, is likely to be extensive, but erosion surfaces within the sand body, perhaps picked out by pebble-lag or intraformational conglomerates, may have more local significance. Within large complex channel units, laterally extensive exposures commonly show that erosion surfaces fall naturally into a hierarchy of scale, lateral extent, and genetic significance. The bodies of sediment that they bound are termed architectural elements, facies building blocks of differing genetic significance that are linked to the geomorphology of the channel system (Figure 5). Within aeolian dune deposits, similar hierarchies of bounding surfaces are critical to the discrimination of dunes and larger compound bedforms.

It is clear that the more fully the three-dimensional patterns of facies are established, the more precise can be the environmental interpretation. Where only subsurface data are available, the one-dimensional facies information provided by boreholes may sometimes be suitably supplemented by three-dimensional seismic data. However, in most cases, the lateral extent of facies units and lateral facies relationships must be predicted intuitively using Walther's Law and judiciously chosen analogues, usually from extensive outcrop exposures.

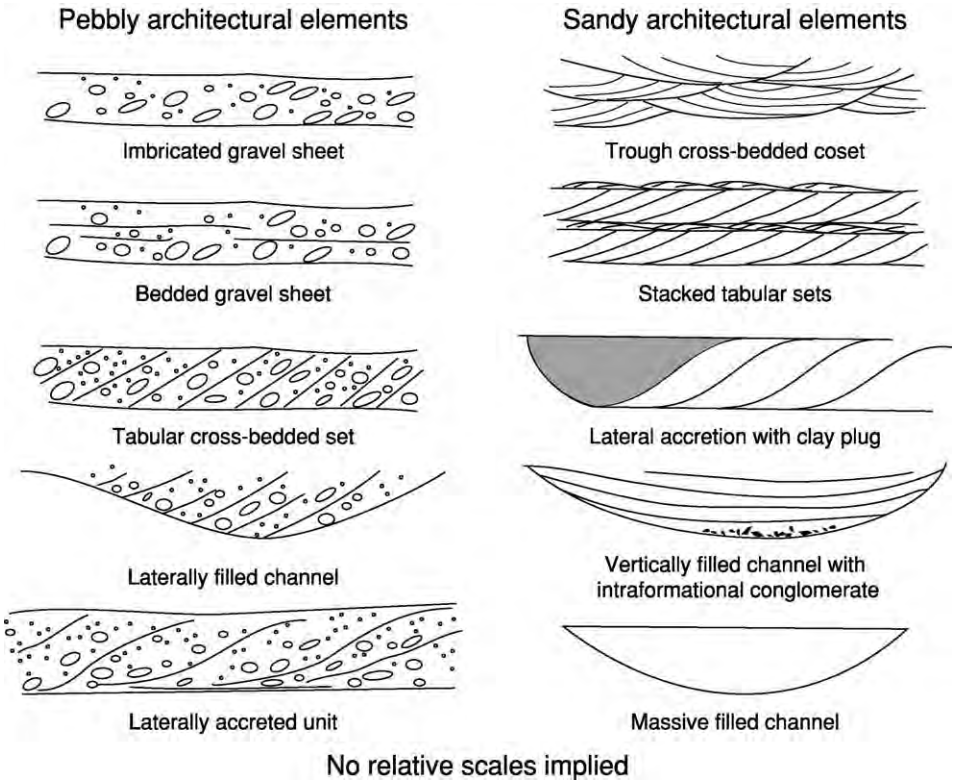


Figure 5 Examples of the various 'architectural elements' that can commonly be identified in alluvial channel deposits. The architectural elements are the products, usually only partially preserved, of various morphological features (bedforms, bars, channel fills) that occupied the channel floor.

Facies Models

Facies models are attempts to summarize the essential aspects of facies sequences and relate them to inferred depositional environments. They are usually most effectively conveyed graphically as three-dimensional block diagrams that relate the environment and its behaviour to the facies pattern. Facies models have been proposed for most major depositional environments, and there is a large measure of consensus about their general aspects, particularly for fluvial, aeolian, deltaic, and other shallow-water settings. For deep-water sand settings, differences of opinion about what constitutes a 'submarine-fan model' have largely been resolved by recognizing that there is a whole spectrum of fan types and, accordingly, a spectrum of valid models. General rather idealized models of major depositional settings have their place as educational tools. They also form a basis with which to compare an actual sedimentary succession, where differences between the ideal and the actual may be important in refining the interpretation. However, it is important to recognize that some settings are inherently more organized and predictable than others and that lateral variation takes place on a variety of scales. This is important when a model is used to predict the nature of a succession between or beyond known data points. In a well-organized setting, it may be appropriate to build a predictive model that is deterministic, and predicted interpolations and extrapolations can be expected to reflect reality closely. In more complex, less organized settings, the stochastic component means that such predictions must be generalized and expressed in probabilistic terms.

Sequence Stratigraphy and Key Surfaces

At outcrop or in a borehole section, the establishment of a sequence stratigraphical framework depends on distinguishing key surfaces of potentially widespread significance that can be related to episodes of rising or falling base level, which in marine settings is relative sea-level, in lacustrine settings is lake level, in fluvial settings is the graded fluvial profile, and in aeolian settings is the water table.

Erosion surfaces at the bases of channel units or elsewhere in a sequence are, in many cases, simply consequences of the depositional processes operating within the environment, for example the migration of a river channel on a flood plain. Such surfaces are autocyclic and have only local significance. Other channel bases may have much more widespread stratigraphical significance and may be related to valley incision caused by a fall in base level (i.e. allocyclic control) (Figure 6). These are one expression of sequence boundaries in sequence stratigraphical terminology. Distinguishing between the two types of erosion surface can be difficult, based on features of the erosion surface alone, especially

Figure 6 A major erosion surface (arrowed) at the base of a thick multistorey channel complex. The erosion surface is regional in extent and has a mapped relief of several tens of metres. It is inferred to be the base of an incised palaeovalley and would be regarded as a sequence boundary, related to a major fall in base level (relative sea-level). Namurian, County Clare, Ireland.

where local channel bases are nested within a larger-scale palaeovalley. However, deeper incision as a result of base-level fall may cause the erosion surface to cut down into facies that differ from those beneath autocyclic channel bases. Outside the confines of an incised valley, the long-term emergence of the sedimentary surface and the resulting nondeposition may allow highly mature soil profiles to develop, so that recognition of palaeosol facies in interfluve areas may have widespread stratigraphical significance. Non-erosive sharp contacts at the bases of coarser facies may also indicate that regressions were forced by falling base level. Falling base level or relative sea-level, therefore, may be manifested in a variety of different ways leading to uncertainty in discrimination and correlation.

Rises of base level or relative sea-level produce more uniform facies expressions, as they tend to be associated with transgression and a cut-off of sediment supply to more distal settings owing to sediment being trapped closer to its source. The results are flooding surfaces (**Figure 3**) characterized by condensation, often to the point of nondeposition. In marine settings, such surfaces are commonly characterized by intense bioturbation, reflecting the increased time available for animals to rework the sediment before burial. When flooding occurred in humid near-emergent settings, peat developed as plant productivity balanced increasing accommodation space. Coal seams, the compacted products of peat, may be of local or regional significance. Where flooding occurred through autocyclic processes, for example by cut-off of sediment supply through distributary switching on a delta plain, the seam extent reflects the area of an interdistributary bay or floodplain lake. Where the control was allocyclic, coal seams may be regionally important stratigraphical markers. Flooding surfaces, however expressed in the sediments, commonly mark the bases of coarsening-up units, termed parasequences, that record the progradations that eventually filled the resulting bodies of water. Where flooding is associated with the transgression of a high-energy shoreline, the transgressive surface may be an erosional pebble lag ravinement surface, separating nearshore sediment below from more offshore sediment above.

See Also

Sedimentary Environments: Alluvial Fans, Alluvial Sediments and Settings; Deltas. **Sedimentary Processes:** Depositional Sedimentary Structures; Post-Depositional Sedimentary Structures; Aeolian Processes. **Sedimentary Rocks:** Mineralogy and Classification. **Sequence Stratigraphy**. **Soils:** Palaeosols. **Trace Fossils**.

Further Reading

Anderton R (1985) Clastic facies models and facies analysis. In: Brenchley PJ and Williams BPJ (eds.) *Sedimentology: Recent Developments and Applied Aspects*, pp. 31–47. Special Publication 18. London: Geological Society.

Galloway WE (1989) Genetic stratigraphic sequences in basin analysis. 1. Architecture and genesis of flooding-surface bounded depositional units. *Bulletin of the American Association of Petroleum Geologists* 73: 125–142.

Leeder MR (1999) *Sedimentology and Sedimentary Basins*. Oxford: Blackwell Scientific.

Nichols G (1999) *Sedimentology and Stratigraphy*. Oxford: Blackwell Scientific.

Reading HG (ed.) (1996) *Sedimentary Environments: Processes, Facies and Stratigraphy*, 3rd edn. Oxford: Blackwell Scientific Publications.

Reading HG (2001) Clastic facies models, a personal perspective. *Bulletin of the Geological Society of Denmark* 48: 101–115.

Scholle PA and Spearing D (eds.) (1983) *Carbonate Depositional Environments*. Memoir 33. Tulsa: American Association of Petroleum Geologists.

Scholle PA, Bebout DG, and Moore CH (eds.) (1983) *Sandstone Depositional Environments*. Memoir 31. Tulsa: American Association of Petroleum Geologists.

Selley RC (1996) *Ancient Sedimentary Environments and their Subsurface Diagnosis*, 4th edn. London: Chapman & Hall.

Tucker ME (1985) Shallow-marine carbonate facies and facies models. In: Brenchley PJ and Williams BPJ (eds.) *Sedimentology: Recent Developments and Applied Aspects*, pp. 147–169. Special Publication 18. London: Geological Society.

Walker RG and James NP (eds.) *Facies Models; Response to Sea Level Change*. Waterloo, Canada: Geological Society of Canada.

Alluvial Fans, Alluvial Sediments and Settings

K D Jäger, Martin Luther University, Halle, Germany

© 2005, Elsevier Ltd. All Rights Reserved.

The term alluvia (alluvions) comprises clastic sediments, deposited by the running water, on river bottoms as well as on the surface of adjacent riverside areas inundated temporarily by floods. This material has to be attributed to the particular catchment area. Within the respective boundaries the sedimentary material supply is dominated by the area shared by open ground where protection by the vegetation cover has been reduced or entirely removed. Consequently, forest clearance, as well as arable land, have caused soil erosion by the clearance, development, and use of this open ground.

The varying intensity of soil erosion during different historical periods following each other reflects – especially in manmade landscapes – the alternating intensity of settlement and agrarian landuse over the course of time. This interrelation may be exemplified by numerous archaeological datings of alluvial layers in stratigraphical sequences of riverside exposures. Frequently, the dated development of such layers is linked with events structuring the development of settlement and landuse.

The petrographic composition of alluvial deposits reflects the superficial substrates (as rocks or loose earth masses) of the particular catchment area. However, the composition is dominated by components from superficially exposed open ground.

Consequently, since the first occurrence of agriculture, more and more alluvial deposits came into being. However, regional chronological differences of the global advance of agriculture have to be considered. As a rule, alluvions have been originated and developed during the Holocene to date. Moreover, usually the deposits are still unconsolidated.

The composition of alluvions is determined substantially, but not exclusively, by the material supply from the particular catchment area. Other decisive factors consist in continuity, discontinuity, and velocity of water flow. Their effect is reflected by the granulation of the sediment: high flow velocity results in course granulation, less velocity in fine granulation. Therefore, mainly gravels and sands occur more often at the upper courses of running waters provided with steeper inclination, for example, in mountain terrains. On the other hand, courses characterised by low inclination, as in the lowlands and near to the mouth of rivers, result in clay deposition. Frequently, the granulation of the deposits is mixed, and this feature is reflected by the nomenclature of alluvions in many languages as German ('Auelehm'), Russian ('пойменная глина'), Czech ('nivní hlína'), whereas other terms try to avoid a concrete relation to the granulation of the deposit as English ('alluvium'), French ('alluvion'), or the proposed German term 'Klock' (adapted from popular use in Lower Lusatia). Independent of granulation, is the term alluvial substrates, soil types in pedological nomenclatures using the term 'vega' (from the Spanish meaning of 'riverside').

On riversides, a two-dimensional differentiation of the distribution of alluvial deposits frequently occurs and corresponds to regular patterns. The pattern of superficial relief differentiation is determined by the buildup of embankments and channels. The spatial differentiation of processes and patterns of the relief, as of the sediment distribution on riversides, has been modified frequently, especially as a consequence of extreme flooding of the respective rivers. Therefore, a high- and narrow-scale heterogeneity characterises the area of stratigraphical subdivision of alluvions. A temporal modification of the spatial pattern of the riverside has been exemplified in the case of the lower Vistula in Poland.

Frequently, the alluvions are linked with alluvial and colluvial fans. The former occur at the confluence of tributaries to their main valley, the latter mainly in slope foot positions of the riverside margins. Floods from tributaries can rearrange the superficial shape of the stratigraphy of alluvions in the respective main valley, too, which has been exemplified by detailed investigation in the region of Ulm (Germany), where the River Iller flows into the Danube (so-called 'Iller-Schwemmkegel'). The appearance of colluvial fans covering flood-plain soils, concealing alluvial deposits, has been proved recently by examples from the Upper Rhine Lowlands region surrounding the Kaiserstuhl massif: Endingen, Emmendingen. A common feature of these investigations, with other detailed stratigraphical studies devoted to the alluvions in the valleys, is provided by archaeological observations proving precise Holocene origin. These datings include two other important indications. They are evidence of man's presence and impact on the landscape by irrigation and drainage works, as well as soil erosion, as a consequence of land use. On the other hand, the datings prove the alternation of sedimentation and erosion, respectively, incision into previous valley bottoms with their alluvions, as well as into alluvial fans.

The same situation for determining datings and the sequence of processes and events is valid regarding the stratigraphy of alluvions in more extended riversides and main valleys. The detailed regional examination and the comparability of the results may be exemplified by the studies described in the further reading section at the end of this chapter.

Sometimes the archaeological datings of single layers within alluvial strata evidence the covering of ancient sites of human presence and settlement by subsequent floods, as is the case of urban centres connected with the medieval Slavonic state of Greater Moravia (actually in Czechia: Pohánsko, Mikulčice, Staré Město, and Úherské Hradiště).

Other chances of achieving more specific and precise dating of the sequence of Holocene events and processes in the stratigraphy of alluvions, are provided by trunks of trees engulfed by the fluvial deposits. Trunks have been found frequently, for instance, in many central European valleys, as Elbe and Spree, Rhine and Main, or the Danube with its tributaries. In the German literature, they are called 'Rannen'.

On a regional scale, there are numerous trunks within the sediment. In the brown coal seams of central Germany (Saxonia, Saxony-Anhalt, Lusatia) the respective fluvial deposits are frequently well-exposed. However, independent of the conditions of exposure, these trunks have provided the opportunity of dendrochronological datings. Using this method can be applied to examination of the valley development in many cases. This may be exemplified for some catchment areas in central Europe as that of the Danube and its tributaries, or that of the River Main.

The regional development of valley-net systems is reflected by the stratigraphy of alluvions. A detailed documentation of varying stratigraphical observations in different parts of one and the same valley-net illustrates the regional differentiation in the course of events. This differentiation has been exemplified carefully in the case of the River Vistula (Weichsel) with its tributaries in Poland. The chronological foundation of this investigation has been completed by radiocarbon datings and pollen analysis.

However, in other parts of temperate Europe, careful and detailed stratigraphical examinations have covered different sections of the same catchment area. Such investigations may be exemplified by the valley system of the River Severn in the British Isles (counties of Shropshire and Worcestershire, England).

Alternating layers in the stratigraphy of alluvions reflect temporally changing conditions of accumulation, as in single valleys or valley segments, such as completely examined valley-net systems. These changes record climatic consequences, with the varying amount of precipitation in the respective catchment area and its temporal peaks caused by floods. They also reflect the changing intensity of human impact influencing the sedimentary yield of the respective river or creek. This yield comprises effects resulting from the extension of open grounds, forest clearance, and utilisation of arable land in the respective catchment area.

The same causes have ruled the origin, development, and stratigraphy of alluvions in the Mediterranean. Evidence has been provided previously for the Mediterranean and more recently, in particular, from the Iberian Peninsula.

In all the investigated catchment areas, the effect of human impacts depends on the respective regional course of settlement and land-use intensity. The climatic causes take effect in extended regions of the European continent in a comparable way. However, the detection of palaeoclimatic fluctuations, and especially of oscillations of the water balance in the course of the Holocene, is based on other stratigraphic observations as in exposures of freshwater lime deposits. Among geostratigraphical, palaeontological (especially palynological and palaeomalacological), and archaeological evidence supporting the resulting concept concerning the palaeoclimatic development during Holocene, the stratigraphy of alluvions claims an important but not essential role.

Related to the youngest alluvions, historical record archives especially support the linking of distinct stratigraphical facts with particular palaeofloods.

However, irrespective of all other sources clarifying the palaeoenvironmental development as a whole and on the valley bottoms in particular, significant contributions to the reconstruction of previous conditions during the Holocene are provided by the alluvions and their stratigraphy.

See Also

Europe: Holocene. **Sedimentary Processes:** Fluvial Geomorphology. **Sedimentary Rocks:** Rudaceous Rocks.

Further Reading

Andrzejewski L and Juśkiewicz W (2003) Lithofacies diversification of the alluvia in the area of Kepa Dzikowska, Kepa Bazarowa, Kepa Strońska and of the Vistulian floodplain near Toruń. *Prace Geograficzne* 189 (Warszawa): 159–176.

Becker B (1982) *Dendrochronologie und Paläoökologie subfossiler Baumstämme aus Flußablagerungen – Ein Beitrag zur nacheiszeitlichen Auenentwicklung im südlichen Mitteleuropa.* Wien. Mitt. Komm Quartärforschung d. Österr. Akad. d. Wiss. No. 5.

Becker B (1983) Postglaziale Auwaldentwicklung im mittleren und oberen Maintal. In: Schirmer W. (ed.) *Holozäne Talentwicklung – Methoden und Ergebnisse.* Geol. Jb., Reihe A, 71: 45–59.

Brown AG (1983) Floodplain deposits and accelerated sedimentation in the lower Severn basin. In: Gregory KJ (ed.) *Background to Palaechydrology–A Perspective,* pp. 375–397. Chichester: John Wiley & Sons.

Dury GH (1983) Osage-type underfitness on the River Severn near Shrewsbury, Shropshire, England. In: Gregory KJ (ed.) *Background to Palaechydrology – A Perspective,* pp. 399–412. Chichester: John Wiley & Sons.

Graul H and Groschopf P (1952) Geologische und morphologische Betrachtungen zum Iller-Schwemmkegel bei Ulm. *Ber. Naturforsch. Ges. Augsburg* 5: 3–27.

Havlíček P, Galuška, and Poláček (2004) Die geologische Situation im Bereich des großmährischen Zentrums von Staré Mesto – Uherské Hradiště. *Studien zum Burgwall von Mikulče* 7: Brno, 1–17.

Heine K and Niller HP (2003) Human and climate impacts on the Holocene landscape development in southern Germany. *Geografia Polonica* 76(2): Warszawa, 109–122.

Jäger K-D (1962) Über Alter und Ursachen der Auelehmablagerung Thüringer Flüsse. *Praehist. Z.* 60(1/2): 1–59.

Jäger K-D (1967) Anthropogene Ablagerungen im Holozän der südöstlichen Thüringer Triasmulde. In: Kliewe H (ed.) *Probleme und Befunde der Holozänstratigraphie in Thüringen,* Sachsen und Böhmen, Berlin and Prague, pp. 31–51.

Jäger K-D (1986) Reliefmustertypen in Talauen im nördlichen Mitteleuropa. *Acta Univ. Nicolai Copernici, Geografia XXI (Nauki matermatyczno-przyrodnicze* 67: *Toruń*, 109–117.

Jäger K-D (1997) Anthropogene Ablagerungen im Holozän Mitteleuropas. *Brandenburgische Geowissenschaftl. Beiträge* 4(3): Kleinmachnow, 89–96.

Jäger K-D (2002) Oscillations of the water balance during the Holocene in interior Central Europe – features, dating und consequences. *Quaternary International* 91: 33–37.

Jäger K-D (2003) Die Oberfläche: Uferwälle und Rinnen im Oderbruch. In: Schroeder JH and Brose F (eds.) *Oderbruch – Märkische Schweiz – Östlicher Barnim.* Führer Geol. Berlin u. Brandenburg 9. Berlin, Techn. Univ., 66–77.

Kubiena W (1953) *Bestimmungsbuch und Systematik der Böden Europas,* p. 392. Stuttgart: Ferdinand Enke.

Limbrey S (1983) Archaelogy and palaeohydrology. In: Gregory K J (ed.) *Background to Paleohydrology – A Perspective,* pp. 190–121. Chichester: John Wiley & Sons.

Litt Th (1988) Stratigraphische Belege für anthropogen ausgelöste Bodenlagerungen vom Neolithikum bis zur frühen Eisenzeit im circumhercynen Raum. *Ethnographisch-Archäologische Z.* 29: 129–137.

Litt Th, Hiller A, and Eissmann L (1991) Zur Entwicklung der jungquartären Tieflandstäler im Saale-Elbe-Raum unter besonderer Berücksichtigung von C^{14}-Daten. *Eiszeitalter u. Gegenwart* 41: 26–46.

Litt Th (1994) Holozäne Talentwicklung am Beispiel der Elster-Luppe-Aue, Tagebau Merseburg-Ost. In: Eissmann L and Litt Th (eds.) *Das Quartär Mitteldeutschlands – Ein Leitfaden und Exkursionsführer,* pp. 333–337. Altenburger Naturwiss. Forschungen 7, Altenburg.

Mäckel R, Schneider R, and Seidel J (2003) Anthropogene Impact on the Landscape of southern Badenia (Germany) during the Holocene – documented by colluvial and alluvial sediments. *Archaeometry* 45(3): 487–501.

Pfister Chr (1999) *Wetternachhersage – 500 Jahre Klimavariationen und Naturkatastrophen,* p. 304. Bern: Stuttgart and Wien.

Pfister Chr, Brázdil R, Glaser R, *et al.* (1999) Documentary evidence on climate in sixteenth Europe. *Climatic Change* 43: 55–110.

Scamoni A (1954) Die Waldvegetation des Unterspreewaldes. *Arch. Forstwesen* 2: 232–244.

Scheffer F and Schachtschabel P (1988) *Lehrbuch der Bodenkunde,* 14th edn., p. 494. Stuttgart: Ferdinand Enke.

Schulte L (2000) Climatic and human impact on river systems in Southeast Spain. In: Díaz del Olmo F, Faust D, and Porras AI (eds.) *Environmental Changes during the Holocene,* pp. 45–49. Sevilla.

Schulte L (2002) Climatic and human influence on river systems and glacier fluctuations in southeast Spain since the Last Glacial Maximum. *Quaternary International* 93/94: 85–100.

Schulte L (2003) River-response and terrace aggradation in the Mediterranean Iberian Peninsula during historical times. In: Thorndycraft VR, Benito G, Barriendos M, and Lasat C (eds.) *Palaeofloods Historical floods and Climatic Variability: Applications in flood Risk Assessment,* Proc. of the PHEFRA Workshops, Barcelona 2002, pp. 67–72.

Starkel L (ed.) (1982 and 1987) *Evolution of the Vistula river valley during the last 15000 years.* 2 vols. Wroclaw etc.: Polish Ac. Sci., Geographical Studies, Special Issues.

Vita-Finzi Cl (1969) *The Mediterranean Valleys – Geological Changes in Historical Times,* p. 139. Cambridge: Cambridge University Press.

Vollrath H (1963) Die Morphologie der Itzaue als Ausdruckhydro-und sedimentologischen Geschehens. *Mitt. d. Fränkischen Geograph. Ges. (Erlangen)* 10: 297–309.

Anoxic Environments

P B Wignall, University of Leeds, Leeds, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Nearly all modern environments, on both land and sea, are well oxygenated, thanks to the ample levels of oxygen available in the atmosphere. This also seems to have been the case for much of Phanerozoic history. Lack of oxygen appears to have been a critical feature for life only during specific time intervals (see below). However, you do not have to look very far to find anoxic environments in many terrestrial and marine settings. Oxygen is consumed by respiration, particularly by bacteria that use this powerful oxidant to break down organic matter. Therefore, in settings where abundant organic matter accumulates (e.g. swamps, estuarine muds, deep lakes, and shelf sea sediments) the oxygen is rapidly consumed, producing an oxygen-free (anoxic) zone that may, in some sediments, be only a few centimetres beneath the surface. Within the anoxic zone anaerobic bacteria continue to oxidize organic matter by using a series of increasingly less efficient oxidants, starting with nitrates, then sulphates, and finally fermentation processes that use carboxyl groups in the organic matter as oxidants. In the marine realm, sulphate ions are the most readily available oxidant after oxygen, and a sulphate reduction zone is therefore often well developed. The by-product of this activity is hydrogen sulphide, which rapidly reacts with iron oxides in the sediment to produce the mineral pyrite, which therefore often provides tell-tale evidence of sulphate reduction zone activity. In non-marine anoxic environments, such as swamps and lakes, sulphate is often not available, with the result that the surface oxic layer of sediment is often directly underlain by a methane-generating (methanogenic) zone.

Modern Anoxic Environments

The development of open-water anoxia is a relatively rare phenomenon in the modern world. Anoxia develops where the supply of oxygen is exceeded by the demand for oxygen by decaying organic matter, with the result that the chemical zones summarized above move up out of the sediment and into the water column. This occurs in two distinct modern environments: silled basins and upwelling zones (Figure 1). Silled basins arise where a body of deep water is semi-isolated from the world's oceans by the presence of a shallow-water sill that can restrict deep-water circulation. This occurs particularly if the basin receives a lot of freshwater run-off and thus exports a low-density surface layer of water across the sill into the open sea. The basin is therefore said to have a positive water balance and to display an estuarine circulation. This can effectively seal the deeper waters from vertical mixing and oxygen supply. The present-day Black Sea is a giant example of a silled anoxic basin, although many fjords provide smaller examples. These modern examples owe their density stratification to the differences in density caused by variations in salinity. However, temperature contrasts can also produce stratification, since warmer waters are less dense than colder waters, and it is often mooted that some ancient examples of silled anoxic basins may have been thermally stratified.

Silled anoxic basins are essentially the product of a failure of oxygen supply to deeper waters. Upwelling-zone anoxia, in contrast, is the product of an overabundance of organic matter and therefore an excessive oxygen demand. West-facing continental margins are often characterized by prevailing offshore winds, which drive surface waters offshore (Figure 1B). This water is replaced by water that upwells from a considerable depth and has usually originated in polar areas. Because this deep water has not resided in the photic zone for some considerable time, it is often nutrient rich, and its arrival at the surface of the ocean stimulates very high plankton productivity. It is when this plankton dies and descends that decay processes rapidly consume oxygen, producing an oxygen-minimum zone (OMZ) within the water column. Intense OMZs are known off the coasts of California and South America, although truly anoxic conditions are not developed. The zone of oxygen-poor deposition beneath an upwelling zone is at an intermediate water depth, where the OMZ intersects the continental slope (Figure 1B). This contrasts with the silled-basin environment, where the anoxic deposition is in the deepest-water locations.

Identifying Ancient Anoxic Environments

Recognizing anoxic environments in sedimentary rocks is relatively straightforward because of the distinct range of attributes produced by anoxic conditions. Probably the most diagnostic feature of

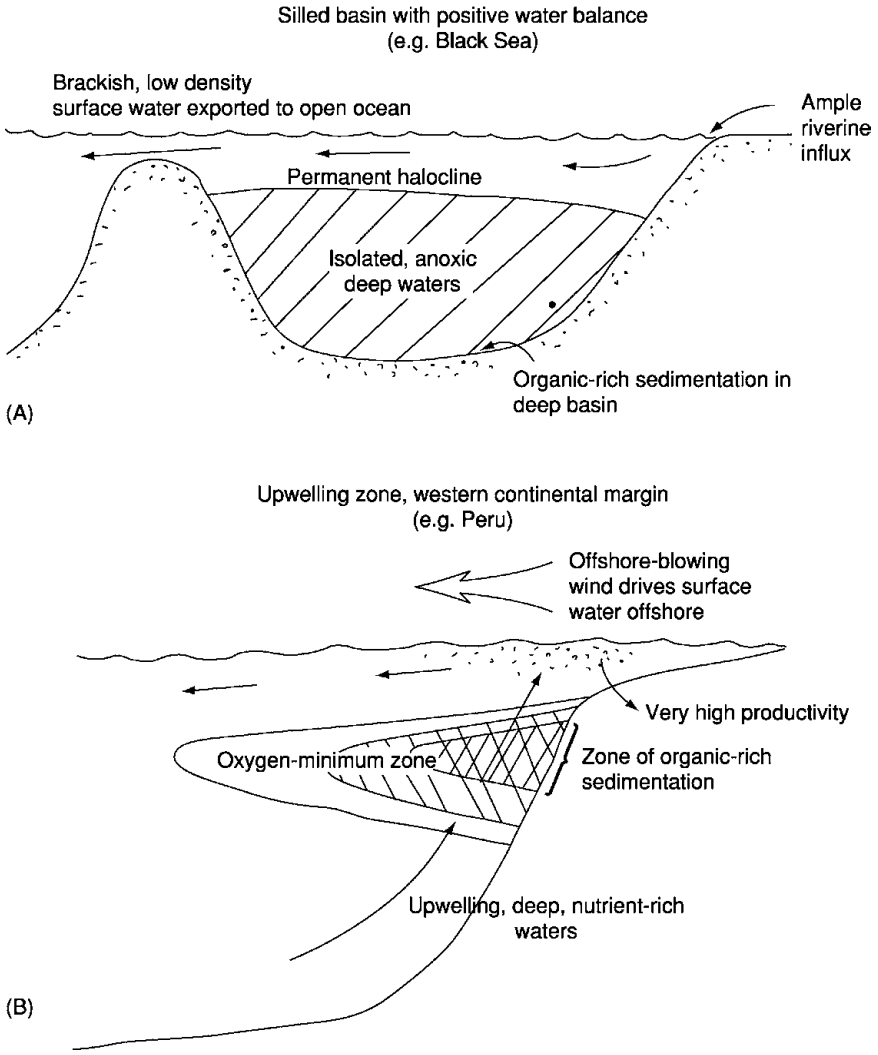


Figure 1 Models of the production of open-water anoxic environments. (A) The modern Black Sea is the type example of silled-basin anoxia, whilst (B) the high-productivity upwelling environments are encountered along the western seaboard of several countries, including Peru and the USA. Note that in silled basins the site of organic-rich deposition is in the deepest waters, whereas beneath upwelling zones organic-rich deposition occurs at intermediate water depths, often on continental slopes, as depicted here, or in deep shelf locations.

marine anoxia is the presence of abundant pyrite – particularly its framboidal form – produced in the sulphate reduction zone. Framboids are named after the French word for raspberry (because that is what they look like, at least superficially); they are spherical aggregates of tiny pyrite crystals (microcrysts) a few microns in size (Figure 2). The framboids themselves range from a few microns to a few tens of microns in size, although a diameter of 5–6 μm is typical. Framboids form today in weakly anoxic conditions above well-developed sulphate reduction zones where reduced iron reacts rapidly with the hydrogen sulphide produced in the zone beneath. When these conditions are developed within the water column only small (up to 6 μm in diameter) framboids are able to form before they sink out of the water column. Populations of tiny framboids are therefore produced when sulphide is present in the lower water column. This is sometimes called a euxinic environment. Note that very minor amounts of crystalline (non-framboidal) pyrite also form in some sediments beneath oxygenated bottom waters, but pyrite-rich sediments are highly diagnostic of anoxic environments.

The change from oxygenated to anoxic conditions causes a valence change in most metals with the result that anoxic sediments contain different concentrations of these elements from normal shales. Many metals (e.g. vanadium, molybdenum, and nickel) often precipitate as impurities in pyrite with the result that they are relatively enriched in black shales. Other elements occur in a more soluble state in anoxic conditions (e.g. phosphorus and manganese) and are therefore lost from the anoxic environment, although they often precipitate at the anoxic–oxic boundary.

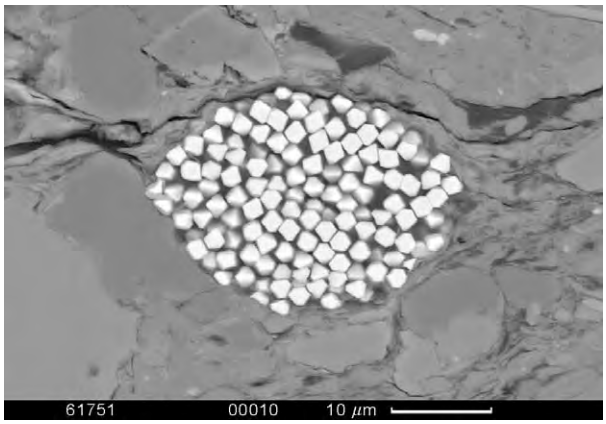


Figure 2 Scanning electron microscope image of a pyrite framboid, from the *annulata* Shale (Upper Devonian) of northern Germany. Such spherical accumulations of tiny pyrite microcrystals (microcrysts) form in prolific numbers in anoxic marine environments.

This is particularly the case for manganese, which is often concentrated as a carbonate on the margins of anoxic areas.

Although a subject of controversy, anoxic conditions are thought by many to favour the preservation of organic matter, and many anoxic sediments are organic-rich. This is manifest in the dark colour of the sediments, and the epithet 'black shale' is often used to describe anoxic sediments. Such sediments often also display very fine lamination, and they characteristically weather into thin sheets to produce a rock known as a paper shale. The preservation of fine laminae is testimony to the fact that animals cannot live in anoxic conditions. Thus, the worms and other creatures that normally disrupt delicate laminae by their burrowing activities are absent. On the face of it, fossils should also be absent from anoxic sediments because the anoxia inhibits all but microbial life. However, on the contrary, many black shales contain abundant and well-preserved fossil remains. Many of these belong to creatures that swam, such as fishes (*see* **Fossil Vertebrates:** Fish) and ammonites (*see* **Fossil Invertebrates:** Ammonites), and so their presence is readily explained: they presumably lived in the oxygenated surface waters and sank into the anoxic deeper waters only after death (Figure 3A). However, many fossils in anoxic sediments belong to ostensibly bottom-living forms, notably bivalves (*see* **Fossil Invertebrates:** Bivalves) and brachiopods (*see* **Fossil Invertebrates:** Brachiopods) (Figure 3B). For a long time such occurrences have been a subject of debate, but most workers now agree that they are dysaerobic fossils. The term 'dysaerobic' was first coined for modern forms found living on the seafloor within the OMZ of the Californian Borderland. Low-oxygen bottom waters provide harsh conditions for animals with the result that only a low-diversity, but often abundant, community of creatures can survive. Low diversity–high abundance is a common attribute of black-shale fossil assemblages too, and they are generally regarded as representing a community tolerant of low oxygen. Ancient dysaerobic assemblages are often dominated by bivalves that display a shell morphology of broad, flat, often circular valves, sometimes with dense, fine radial ribs. These are loosely termed 'paper pectens', and this characteristic morphology has evolved again and again in black-shale depositional environments (Figure 3B).

By implication, the presence of fossils in a black shale suggests the presence of oxygen during deposition, although levels need not have been very high. However, the geochemistry of anoxic sediments implies that there was no oxygen at all during deposition. For example, uranium is concentrated in sediments only in the complete absence of oxygen; otherwise it occurs as a highly soluble ion that is not precipitated. Thus, there is often a discrepancy between geochemical and palaeontological evidence for oxygen levels. This can be resolved if it is appreciated that black shales record a range of depositional environments in which the average depositional conditions were anoxic but seafloor oxygen was present during some, probably brief, intervals. These oxygenation events would allow transient colonization by rapidly dispersing species such as paper pectens. This is a distinctly different concept from that originally proposed for dysaerobic faunas, which envisaged persistently low seafloor oxygen levels, and as a result the alternative names 'episodically dysaerobic' and 'poikiloaerobic' have been proposed. However, neither has really caught on in the literature, and the term dysaerobic continues to be used. In order to get away from these generic terms, a simple descriptive oxygen-restricted biofacies scheme has been provided based on fossil and sediment attributes (Figure 4). Based on British Jurassic examples, the scheme recognizes a gradient of features thought to record improving seafloor oxygen levels. Oxygen-restricted biofacies 3 and 4 are very common, and they typically combine attributes of anoxic deposition (lamination and trace-metal enrichment) and dysaerobic deposition (presence of low-diversity benthic fossils); they undoubtedly formed in the variably oxygenated conditions described above.

Oceanic Anoxic Events

The modern oceans are extremely well ventilated, and deep waters are everywhere supplied with high levels of oxygen, thanks primarily to a vigorous thermohaline circulation regime. It therefore came as

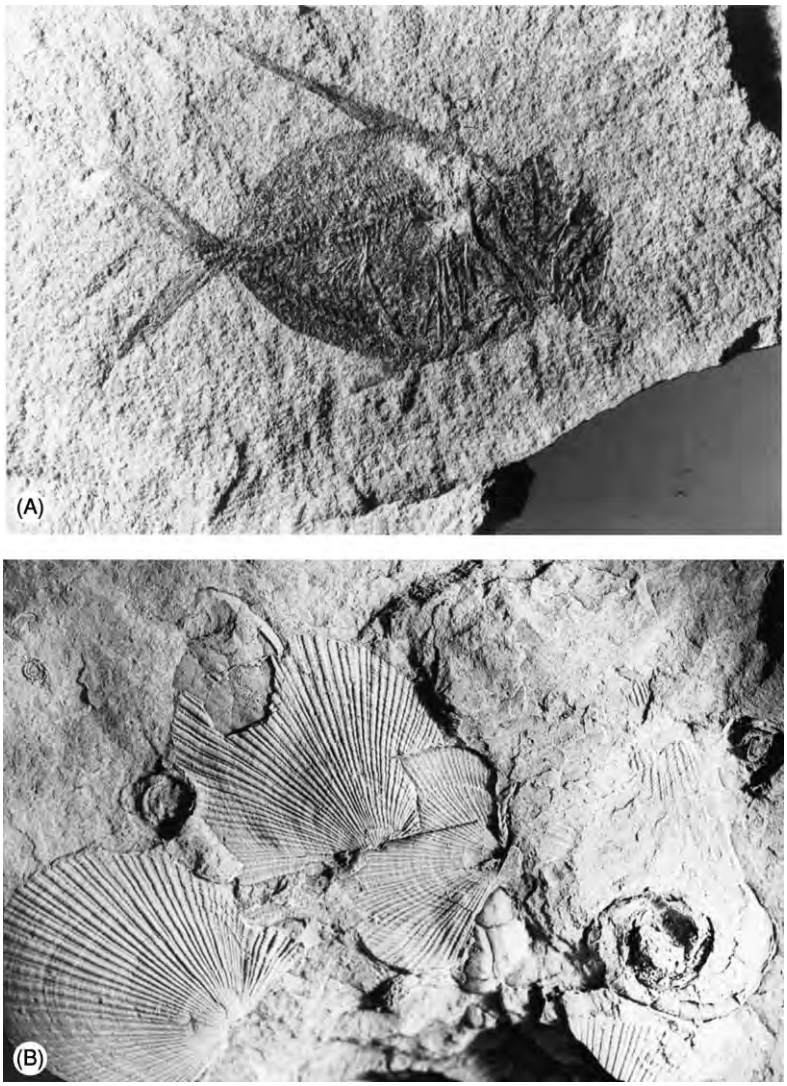


Figure 3 Fossils from ancient anoxic environments. (A) Well-preserved skeleton of the deep-bodied fish *Dorypterus*, from the Upper Permian Marl Slate of north-east England. The lack of predation in anoxic environments often favours the preservation of fragile skeletons such as this magnificent example. The fish is 10 cm long. (B) Typical dysaerobic bivalve assemblage dominated by a single species of the radially-ribbed bivalve *Dunbarella* from the *Reticuloceras subreticulatum* zone, Upper Carboniferous, Fisherstreet Bay, County Clare, western Ireland. Flat-valved bivalve morphologies such as this are common in black shales from the Devonian to the Jurassic. The circular impressions on this slab are poorly preserved specimens of the zonal goniatite *R. subreticulatum*. The left-hand bivalve is 3 cm high. Both specimens were coated in ammonium chloride for photography.

something of a surprise in the 1970s when the cores of the Deep Sea Drilling Program revealed the presence of thin black-shale layers sandwiched between the more normal oxygenated deep-ocean sediments. Many of these oceanic black shales were found to be contemporaneous with black-shale horizons encountered in shelf sections. Based on these observations, the concept of global oceanic anoxic events (OAEs) was proposed. Several are known from the Cretaceous, with the Bonarelli Event (named after a shale in Italy) providing the best-known and best-studied example. This event occurred at the transition between the Cenomanian and Turonian stages in the Late Cretaceous and saw anoxic or oxygen-poor waters develop over a broad area of the globe and over a broad range of water depths from the ocean floor in the Atlantic to many deep shelf sections. Ample evidence indicates that this event and probably most other OAEs coincided with phases of extreme global warmth (the Bonarelli Event may coincide with the peak of Cretaceous warmth). Many models for the origin of OAEs are therefore linked with this climatic observation. Potentially, warming of waters in high latitudes can shut down the supply of cold

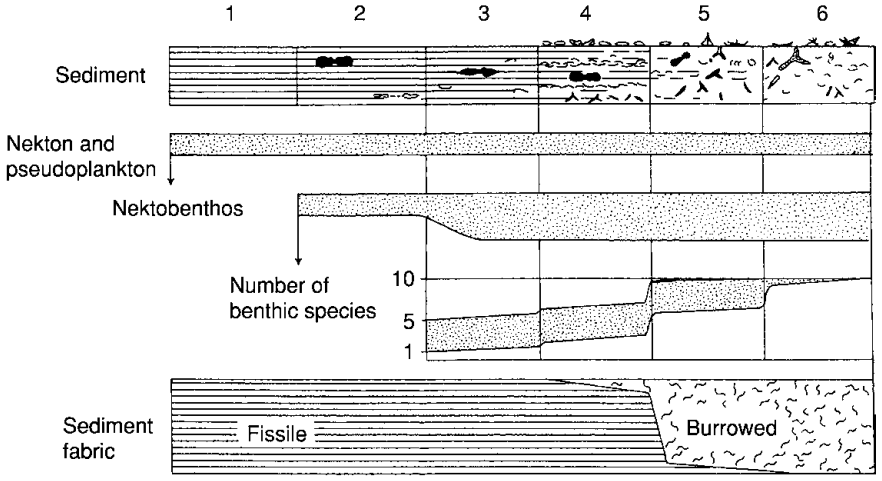


Figure 4 Oxygen-restricted biofacies based on fossil content and sediment properties. The fossils have been divided into three categories. First, nekton are free-swimming animals and pseudoplankton are forms that attach to floating objects, such as driftwood, in the surface waters. Such species are not affected by bottom-water oxygen levels and so are found in a broad range of sediments. Second, nektobenthos are swimming forms, such as ammonites, that probably lived near the seafloor and so were absent from environments developing bottom-water anoxia. Third, benthic species live on the seafloor and cannot tolerate bottom-water anoxia. Reproduced from Wignall PB (1994) *Black Shales*. Oxford Monographs in Geology and Geophysics. Oxford: Oxford University Press.

dense water that drives much of the present-day oceanic circulation. Oceanic deep waters are therefore envisaged to have been replaced by dense, warm, saline waters generated in evaporitic tropical-shelf seas. Warm water holds considerably less dissolved oxygen than colder water, and this factor alone will tend to encourage deep-sea anoxia. Some modelling experiments suggest that an ocean with warm saline deep waters will circulate more rapidly than one with colder waters, with the result that upwelling may have been more vigorous. This in turn would stimulate more plankton productivity, thereby intensifying the mid-water OMZ. Intense oceanic volcanism, which is also correlated with the OAEs, may also have supplied increased levels of nutrients to the oceans and fostered further productivity.

The Cretaceous OAEs were originally intended to denote intervals when oxygen-poor marine deposition was widespread, without implying that all oceans, beneath the surface waters, were simultaneously anoxic. In essence they refer to intervals when anoxic environments were very widespread. However, there is evidence to suggest that true global oceanic anoxia may have in fact happened, with the most notable example occurring at the transition from the Permian to the Triassic. Appropriately enough, this has been termed a superanoxic event, and it is probably no coincidence that this interval is also marked by the greatest marine mass extinction of all time (*see* **Palaeozoic:** End Permian Extinctions). Like the Bonarelli Event, the Permo-Triassic superanoxia coincides with an extreme greenhouse climate and with the eruption of a giant volcanic province. In fact, the marine anoxia–global warming–massive volcanism triumvirate is seen several times in Phanerozoic history, and invariably coincides with extinction events, although of considerably different magnitudes (the Bonarelli Event is marked by only a minor extinction event). The Early Jurassic provides another classic example, with volcanism in the Karoo region of South Africa being correlated with a significant marine extinction event and widespread deposition of black shales, particularly in north-west Europe, where they are variously known as the Jet Rock (England), Schistes Cartons (France), and the Posidonienschiefer (Germany). Linking these various phenomena into a cause-and-effect scenario is a key goal of much current geological research.

Productivity versus Preservation

All black shales appear to have formed in anoxic environments, and black shales are the source all the world's oil and much of its natural gas; therefore understanding and predicting the occurrence of anoxic environments is a key goal in hydrocarbon exploration. However, it is remarkably difficult to constrain the key attributes of anoxic-environment development. As noted above, study of modern environments indicates that there are two routes to producing organic-rich sediments, and they are distinctly different. In high-productivity settings, supplied with abundant nutrients, the seafloor is overwhelmed by the flux of organic matter, and there is insufficient oxygen to decay it all. As a consequence oxygen-poor or anoxic conditions develop, but this is merely due to

the abundance of organic matter. The low oxygen levels do not cause the organic enrichment. Herein lies the key difference between the productivity and preservation models. In the latter case, the development of anoxic conditions in, for example, a silled basin with a positive water balance causes enhanced preservation of organic matter because less is lost to efficient degradation by oxygen-consuming microbes. The two models also differ in terms of their water circulation: high-productivity models require a good supply of nutrients to the surface waters from, for example, vigorous upwelling, whereas high-preservation models imply stagnation or, at best, sluggish circulation.

Distinguishing between the two alternatives is considerably more difficult than it might at first appear. Both models produce anoxic conditions; therefore merely proving that anoxia occurred is not particularly useful. Similarly, high levels of organic carbon in the sediment cannot be held as evidence of originally high organic-matter input from the water column because these high levels can be achieved by the high-productivity or the high-preservation route. Indirect proxies of organic-matter flux to the sediments have been developed, but these are contentious. Barium occurs as barite in trace concentrations in organic matter. Therefore the flux of barite to the sediment is held as a proxy for organic productivity. This assumption has the advantage that barite is considered to be less labile than organic matter. Therefore, whilst only a few percent of the organic carbon settling on the seafloor is likely to be preserved within the buried sediment (even beneath upwelling zones), perhaps the majority of the barite flux is preserved. However, the total barium content of the sediments also includes a detrital component, which has to be removed from the total to leave the biogenic component. Unfortunately, this detrital barium concentration can only be guessed at (it is controlled by factors such as sediment grain size and sediment source). A typical barium content for sediments is often assumed, but this is a rather unsatisfactory approach. Furthermore, during early sediment burial it appears that barium is often lost owing to dissolution with the result that maybe less than half the flux is preserved in the sediment. Not surprisingly, barium contents have been used successfully (i.e. convincingly) to calculate productivity levels only in young sediments not subject to significant burial.

Biogenic silica produced by certain types of plankton, particularly diatoms and radiolarians, also has a higher preservation potential during sedimentation and burial than organic matter, with perhaps in excess of 90% surviving to be incorporated into the geological record. Therefore, concentrations of biogenic silica have been used as a proxy for organic-carbon productivity, and it is significant that many organic-rich sediments forming beneath modern upwelling zones also have high levels of biogenic silica. This is primarily thanks to the abundance of diatoms in upwelling settings. The problem for geologists is that diatoms have become a common component of plankton populations only in the past 50 Ma, whilst several other factors also control siliceous plankton productivity, not least the water temperature. High-latitude plankton populations have proportionately more siliceous plankton than lower-latitude ones, and unravelling this cold-water constraint from productivity constraints in ancient sediments is extremely difficult.

Many sediments accumulating beneath high-productivity upwelling zones are rich in phosphates (much of this material is derived from fish bones sourced by the prolific fisheries encountered in these productive waters), and many phosphatic black shales are therefore assumed also to have formed beneath ancient upwelling zones. The Phosphoria Formation, a thick Upper Permian rock unit seen in Idaho and adjacent states in the western USA, is a classic example of a black shale that probably formed beneath an upwelling zone. However, the simple relationship between phosphate content and productivity is confounded by the fact that phosphate-rich accumulation also develops under conditions of very low sedimentation. This is because, in the absence of any other form of sediment accumulation, biogenic phosphate is often one of the few sources of sediment. Black shales are classically very slowly accumulated sediments.

It is something of a paradox for geologists that, whereas it is very easy to identify ancient anoxic environments, it is one of the most difficult endeavours to determine the cause of this oxygen deficiency. The recent history of the eastern Mediterranean provides a classic example of this problem. Organic-rich sediments, called sapropels, have formed in the deepest waters of this region within the recent geological past. The youngest sapropel, labelled S1, ceased forming only around 6000 years ago, almost within historical times. Unfortunately, early civilizations have left little clue to contemporaneous water conditions in the eastern Mediterranean, and evaluating the cause of the anoxia has proved almost as hard as for more ancient examples. All the evidence indicates that the eastern Mediterranean water column became truly anoxic during sapropel deposition. These periods of deposition appear to coincide with glacial intervals, when increased rainfall in the Nile's hinterland

caused increased run-off into the Mediterranean. The consequences of this were probably two-fold. First, it may have generated a positive water balance, like that in the present-day Black Sea, with low-density lowered-salinity surface waters lying on top of denser saltier deep waters. This would have resulted in the lower water column becoming relatively isolated from surface waters and more prone to oxygen deficiency. Therefore, the sapropels could be a product of increased preservation. Second, the increased run-off would also have increased the nutrient flux into the eastern Mediterranean and stimulated higher plankton productivity. Therefore, the productivity model is generally favoured by sapropel workers. However, the most recent research suggests that cause-and-effect relationships are not always clear cut when it comes to understanding sapropel generation. In some instances, anoxic deposition appears to predate slightly the onset of organic-rich sediment accumulation. This is not the correct way around if the anoxia is caused by the increased flux of organic matter into the bottom waters. Within some of the thicker sapropels, palaeoproductivity indices show substantial fluctuations from high to low levels, which again suggests that productivity is somehow decoupled from the oxygenation history. Finally, some of the older sapropels are developed in the western as well as the eastern Mediterranean, a long way removed from the regional effects of the Nile's discharge.

See Also

Fossil Invertebrates: Brachiopods; Bivalves; Ammonites. **Fossil Vertebrates:** Fish. **Mesozoic:** Jurassic; Cretaceous. **Minerals:** Sulphides. **Palaeozoic:** End Permian Extinctions. **Petroleum Geology:** Overview.

Further Reading

Fenchel T and Finlay BJ (1995) *Ecology and Evolution in Anoxic Worlds*. Oxford Series in Ecology and Evolution. Oxford: Oxford University Press.

Jenkyns HC (1980) Cretaceous anoxic events: from continents to oceans. *Journal of the Geological Society* 137: 171–188.

McManus J, Berelson WM, Klinkhammer GP, *et al.* (1999) Geochemistry of barium in marine sediments: implications for its use as a paleoproxy. *Geochimica et Cosmochimica Acta* 62: 3453–3473.

Meyers PA and Negri A (eds.) (2003) Paleoclimatic and paleoceanographic records in Mediterranean sapropels and Mesozoic black shales. *Palaeogeography Palaeoclimatology Palaeoecology* 190.

Rhoads DC and Morse JW (1971) Evolutionary and ecologic significance of oxygen-deficient marine basins. *Lethaia* 4: 413–428.

Wignall PB (1994) *Black Shales*. Oxford Monographs in Geology and Geophysics. Oxford: Oxford University Press.

Wilkin RT and Barnes HL (1997) Formation processes of framboidal pyrite. *Geochimica et Cosmochimica Acta* 61: 323–339.

Carbonate Shorelines and Shelves

D W J Bosence, Royal Holloway, University of London, Egham, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Shorelines and shelves that accumulate sediment composed of calcium carbonate form a large, but not major, part of the world's continental-shelf seas. The best-known examples of these sediments are tropical coral reefs and the white sandy beaches of tropical and temperate coasts (Figure 1).

Most carbonate sediments are formed by the accumulation of skeletons and shells constructed by marine organisms through the precipitation of calcium carbonate (e.g. corals, molluscs, and foraminifera). These are generally known as skeletal, or bioclastic, carbonate sediments. Skeletal sediments occur in both the warm and the cold waters of the world's shelf areas (Figure 2). They are commonest on those continental shelves where they are not diluted by large amounts of eroded and transported siliciclastic sediment, as is the case near major river mouths and mountainous coastal regions. The composition of skeletal carbonates varies because of the different ecologies of the organisms that build skeletons and shells. Tropical waters (Figure 1A and B) are characterized by seafloor communities of corals (including shallow-water coral reefs), green calcareous algae (seaweed), molluscs, and foraminifera. Cold-water carbonate-forming environments (Figure 1C) are dominated by molluscs, bryozoans, red calcareous algae, and deep-water coral reefs. Other carbonate sediments are formed by the direct precipitation of crystals of calcium carbonate from seawater in the warm saline waters of some tropical seas. This

Figure 1 Examples of commonly seen environments of carbonate shorelines and shelves. (A) Coral reef on the Great Barrier Reef, Australia, constructed by the prolific growth of calcium carbonate secreting corals (coral colonies in the foreground are approximately 1 m across). (B) Tropical beach in southern Java composed of sand derived from the breakdown of the carbonate skeletons of organisms living in nearby coral reef communities. (C) Carbonate beach from the cold temperate climate of Connemara, western Ireland, composed of the skeletal debris of molluscs, coralline algae, bryozoans etc. that live in these shallow coastal waters.

process forms calcium carbonate mud and spherical sand-sized grains known as ooids (see below). These are called non-skeletal carbonate sediments and are found only in warm-water areas (Figure 2).

Shelf seas are a major environment of carbonate-sediment formation owing to the environmental conditions, which favour both the organisms that precipitate calcium carbonate skeletons and the direct precipitation of calcium carbonate from seawater. For skeletal carbonates to be produced, most organisms require well-lit oxygenated marine waters of normal salinity for the energy-consuming process of precipitating a skeleton. For non-skeletal carbonates, the chemical conditions must favour the precipitation of calcium carbonate from seawater; this occurs under conditions of increased temperature and salinity and reduced levels of dissolved carbon dioxide. All these conditions favour the faster accumulation of carbonate sediments in shallow tropical waters than in cooler temperate waters (Figure 2).

Tropical shelves typically reach rates of accumulation on the seafloor of about $1\,m\,year^{-3}$ for both skeletal and non-skeletal carbonates. Cold-water shelf environments accumulate only skeletal carbonates, and these accumulate at generally lower rates ($0.1–1\,m\,year^{-3}$) than their warmer-water counterparts. Both these rates of sediment accumulation are several orders of magnitude higher than rates of continental-margin subsidence ($0.01–0.1\,m\,year^{-3}$), which means that carbonate sediments generally infill shallow-water areas and build up to sea-level. In ancient limestone successions this is often demonstrated by the occurrence of hundreds of metres of vertically stacked limestone beds that were all deposited in water depths of a few meters. Because the shallow-water areas are infilled, any additional carbonate accumulation must then take place in deeper more offshore regions. By this mechanism the carbonate shelf builds outwards into deeper waters, or progrades. This is well illustrated by the seismic section of the carbonate shelf area of Great Bahama Bank in the Caribbean (Figure 3), where carbonate sediments have accumulated during the last 10 Ma and are still forming today. The sedimentary layers imaged in the seismic profile (*see* **Seismic Surveys**) show how the sediments have built up to the sea surface on the Bahama Bank to the east-north-east and are building out (prograding) into the deeper waters of the Santaren Channel to the west-south-west. Thick accumulations of shelf carbonate sediments with this characteristic flat-topped morphology are known as carbonate platforms, and these are the major sites of shelf carbonate accumulation both today (Figures 1 and 3) and in the geological past (Figures 4 and 5).

The term carbonate platform is both a stratigraphical term for thick deposits of shallow-water carbonate rocks and used to describe the following carbonate shelf morphologies (Figure 4). The terms carbonate platform and carbonate shelf are often used synonymously.

- Attached carbonate platforms are carbonate shelves that are tied to a nearby continental landmass, which may mean that at times land-derived siliciclastic sediments are mixed in with the carbonate sediments.

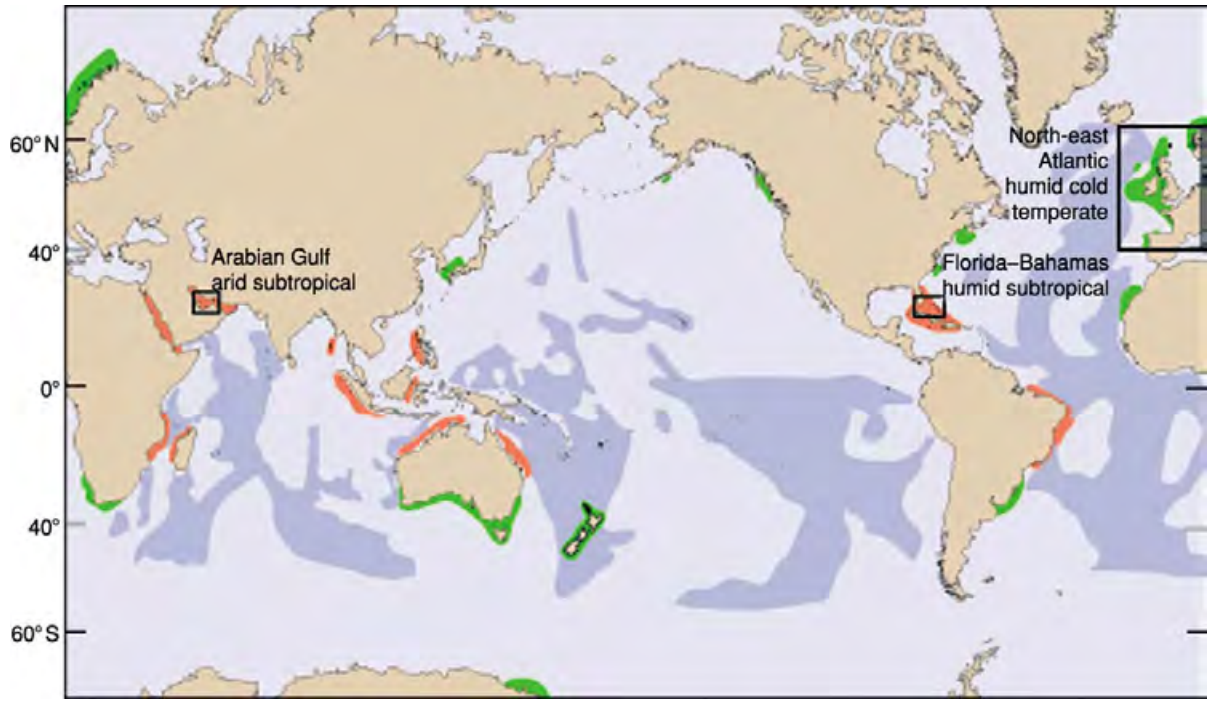


Figure 2 The distribution of calcium carbonate-rich sediments in the world's oceans, and the locations of the shelf areas described in this article. Red, warm-water shelf carbonates; green, cool-water shelf carbonates; dark blue, pelagic carbonates; brown, land; pale blue, ocean. (After Bosence DWJ and Wilson RCL (2003) Carbonate depositional systems. In: Coe A (ed.) *The Sedimentary Record of Sea-level Change*, pp. 209–233. Milton Keynes and Cambridge: The Open University and Cambridge University Press.)

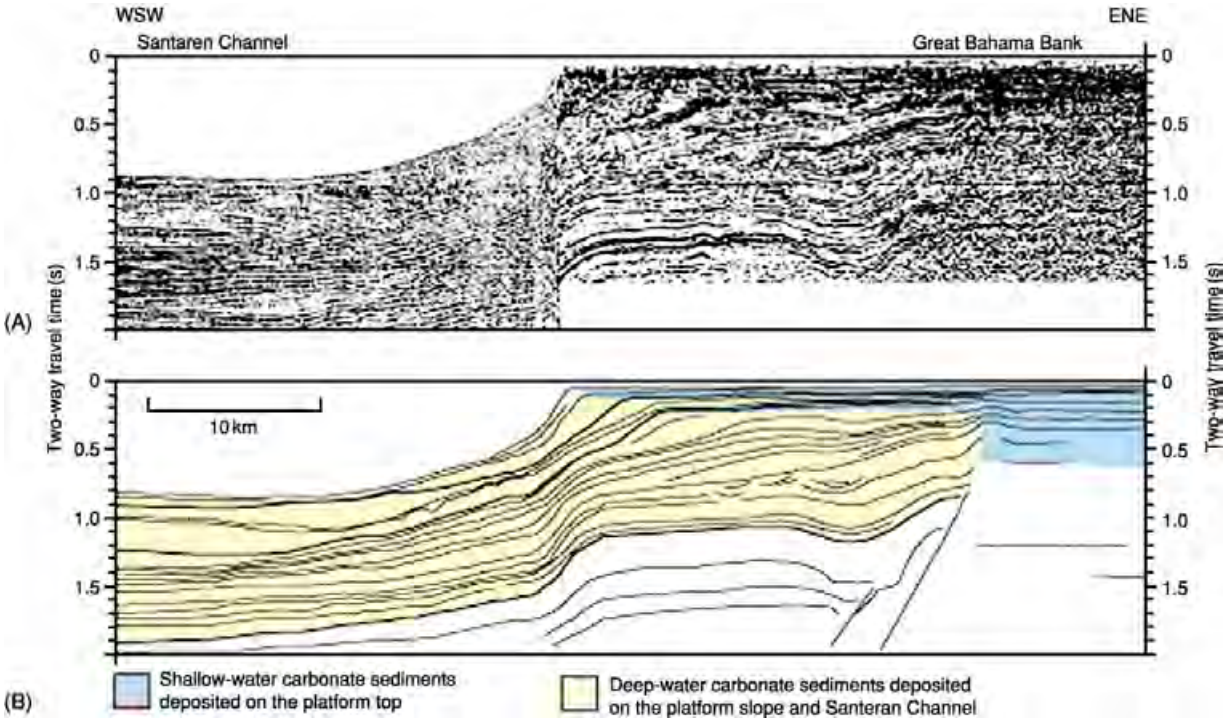


Figure 3 Seismic section through the carbonate shelf of the Great Bahama Bank (for location of section see **Figures 2** and **6**). (A) Seismic reflectors within the bank. These are where sound waves, sent from a source towed behind a boat, bounce off the sediment layers within the bank. (B) An interpretation of these layers, which are all carbonate sediment. The horizontal (blue) layers were deposited on the platform top and the inclined (yellow) layers were deposited on the platform slope and channel. Note how the reflectors preserve the morphology of the platform back through time. (Reproduced from Anselmetti FS, Eberli GP, and Ding Z-D (2000) From the Great Bahama Banks into the Strait of Florida: a margin architecture controlled by sea-level fluctuations and ocean currents. *Geological Society of America Bulletin* 112: 829–844.)

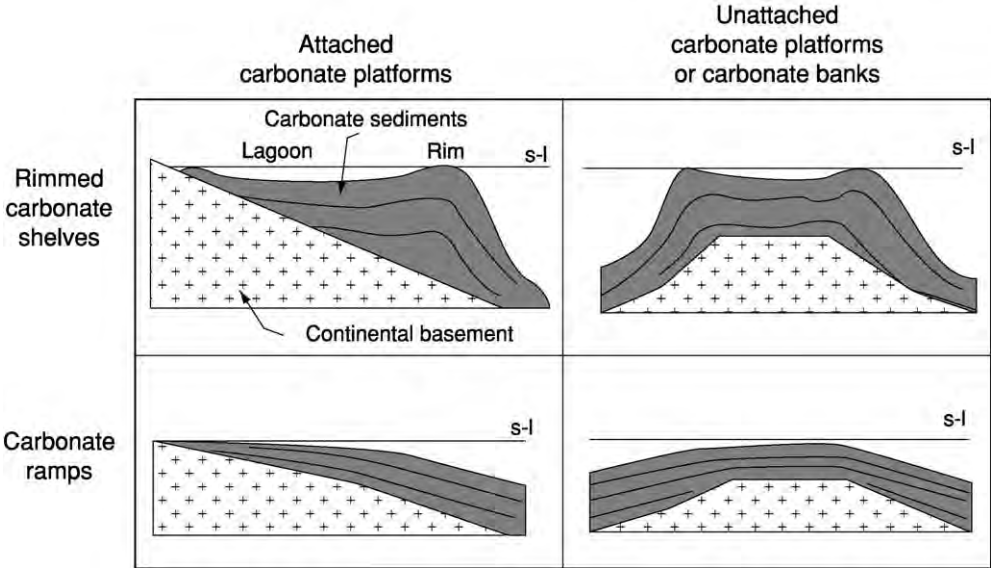


Figure 4 Morphological classification of carbonate shelves and platforms; s-l, sea-level.

- An isolated platform (sometimes also referred to as a carbonate bank, e.g. Great Bahama Bank) is isolated from continental landmasses (Figure 4), and an atoll is a small isolated platform that is commonly formed over a subsiding volcano.

Attached and isolated platforms may have different types of morphological profile, which fundamentally affect the accumulation of sediment on the shelf.

- A rimmed carbonate platform has a shelf margin rim or barrier, such as a reef or sand shoal, that partially isolates an inner platform or lagoon (Figure 4).
- A carbonate ramp is a platform or shelf that is gently inclined (<1°) towards an open sea or ocean with no major build-up of reefs or steep slopes (Figure 4).

When deposited in ancient shelf seas and lithified over time, carbonate reefs and sediments become limestones (*see* **Sedimentary Rocks:** Limestones) and dolomites (*see* **Sedimentary Rocks:** Dolomites), which are known collectively as carbonate rocks. The abundance of carbonate rocks formed in shelf-sea environments has varied considerably over geological time (Figure 5). Not surprisingly they were most abundant when global sea-levels were high and large areas of continents were flooded by shallow shelf seas. This occurred several times in Earth history when there were prolonged periods with no major ice-caps covering the polar regions as there are today. This meant that sea-level was higher all over the globe. These generally warmer waters formed extensive shelf seas in the past and were ideal environments for the formation of carbonate sediments that later formed limestones and dolomites. Much of this article focuses on the sedimentary processes that operate on the different morphological types of carbonate platform, and the different types of sediment that accumulate on them. This information comes from studies of present-day carbonate shelves and is used

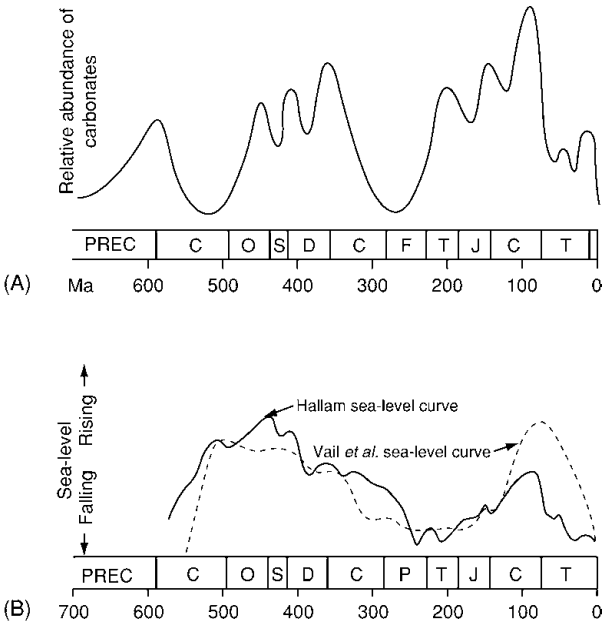


Figure 5 (A) The variation in the abundance of carbonate rocks through time, and (B) how this correlates positively with global sea-level through time. (After Tucker ME and Wright VP (1990) Carbonate Sedimentology. Oxford: Blackwell Scientific Publications.)

extensively in the interpretation of how and where ancient limestones and dolomites accumulated. These carbonate rocks have important economic uses as they host oil and gas reservoirs and heavy-metal accumulations; they are also widely used in building and in the pharmaceutical industry.

Principal Types of Carbonate Shelf

Attached Rimmed Carbonate Shelf

South Florida is an attached carbonate platform and shelf with a large low-lying hinterland (including the Everglades Swamps) set in a semi-humid subtropical climate (Figure 6). Carbonate sediments accumulate throughout this shelf area, and deep boreholes in the region tell us that this has been the case since Jurassic times. The platform comprises four main elements: a semi-enclosed lagoon (known as Florida Bay) that is partially isolated from open-marine waters by a chain of islands (the Florida Keys), a back-reef area of normal marine waters protected from high wave energy by the rimmed margin, the reef and sand shoal margin of the platform, and the fore-reef slope that descends into the deep waters of the Starit of Florida (Figures 6 and 7).

The semi-humid climate is seasonal, with a wet season from July to December (100–150 cm of annual precipitation). During this time there is considerable runoff of freshwater into the shallow-marine coastal environments, which can reduce salinities to 6‰ in Florida Bay (normal marine waters have on average a salinity of 34–38‰). By contrast, in the warmer conditions of spring and summer, temperatures increase and salinities of up to 60‰ are found in the restricted waters of Florida Bay. South Florida is within the north-easterly tradewind belt and, because coral reefs preferentially grow in shallow-water open-marine windward settings, rimmed shelf margins are found on the east-facing margin open to the waters of the Atlantic Ocean (Figures 6 and 7). The region has a very low tidal range, reaching 0.8 m^{-1} in open-water settings and reducing to zero in the semi-enclosed lagoons of Florida Bay. The reef areas experience normal marine salinities of 35–38‰, and water temperatures vary from 18°C to 30°C. In shallow-water settings Holocene carbonate sediments, which are locally lithified, rest on the eroded and cemented Pleistocene Key Largo Limestone (Figure 7). Many of the islands of the Florida Keys are limestones formed from reefs or carbonate sand shoals that accumulated during the last interglacial period, when the sea-level was slightly higher than it is today, and have since been cemented by diagenetic processes to form a rocky carbonate shoreline. The protected shorelines of Florida Bay are dominated by mangroves. These

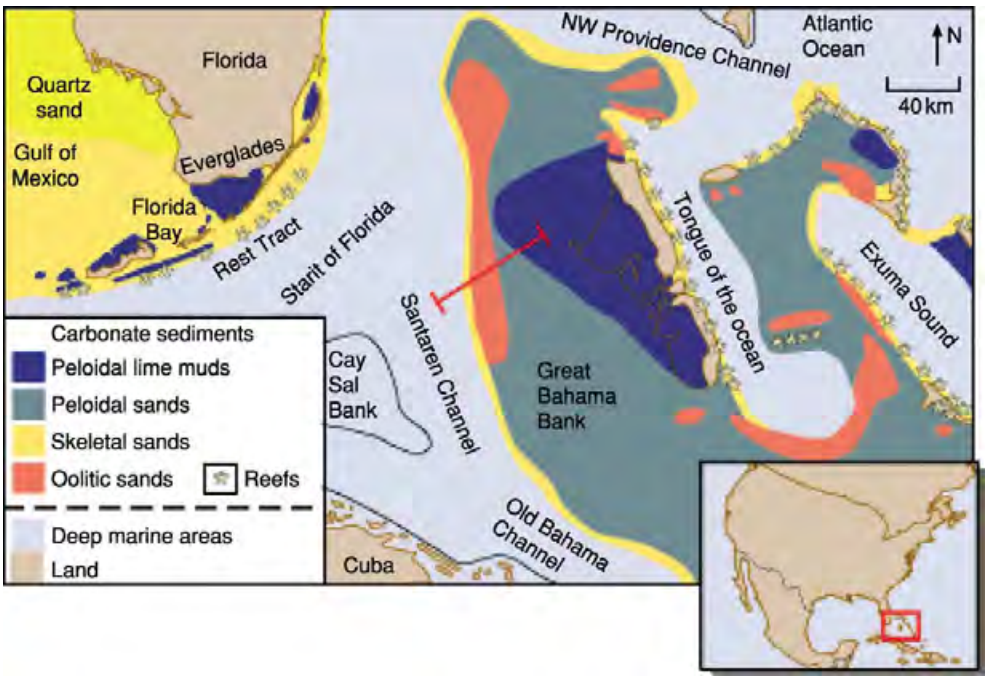


Figure 6 The distribution of major carbonate sediment types in the Florida–Bahamas shelf region (for location see Figure 2). (After Jones B and Desrochers A (1992) Shallow platform carbonates. In: Walker RG and James NP (eds.) *Facies Models Response to Sea Level Change*, pp. 227–301. St Johns: Geological Association of Canada.)

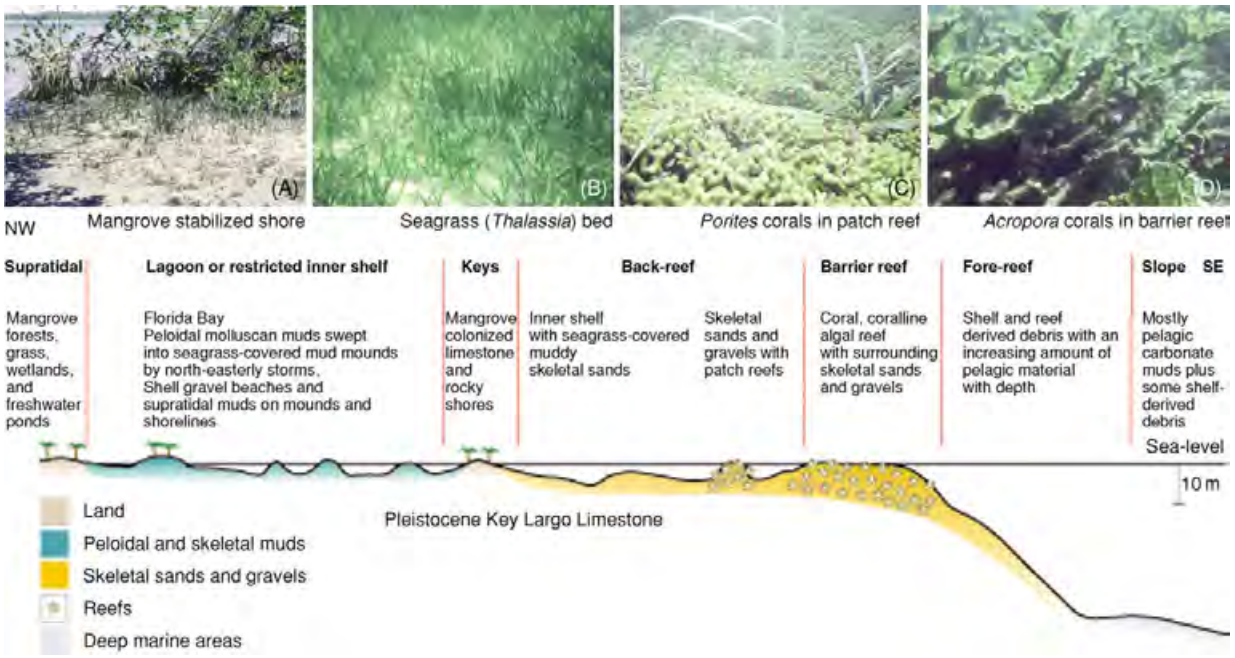


Figure 7 Cross-section through the Florida shelf region normal to the coast and shelf margin, showing the main sedimentary environments. (After Bosence DWJ and Wilson RCL (2003) Carbonate depositional systems. In: Coe A (ed.) *The Sedimentary Record of Sea-level Change*, pp. 209–233. Milton Keynes and Cambridge: The Open University and Cambridge University Press.) The photographs illustrate the fundamental control that organisms have on carbonate-sediment production and accumulation. (A) Mangrove roots stabilizing the shoreline of Florida Bay (foreground 4 m across). (B) *Thalassia* seagrass bed; the grass leaves (20–30 cm high) act as a substrate for carbonate-secreting organisms, whose debris is trapped and stabilized within the grass beds. (C) *Porites* corals (branches are 1 cm across) are common reef builders in the back-reef zone. Note also the encrusted seagrass blades. (D) The shallow zone of the shelf-margin barrier reef is mainly constructed by the elk horn coral, *Acropora palmata* (branches are 15–20 cm across).

trees are adapted to live in coastal waters of variable salinity and are unusual in having aerial root systems, which have a useful sedimentological function in that they trap and bind sediment so as to resist coastal wave erosion (Figure 7A).

The occurrence of fully marine organisms in semi-isolated shallow coastal waters of Florida Bay is restricted by the variable salinities, dissolved oxygen levels, and fine carbonate sediment suspended within the bay waters. Seagrass (e.g. *Thalassia*) beds (Figure 7B) and the calcareous organisms that encrust the grass blades (foraminifera, serpulid worms, and bryozoans) are common, as are other skeletal organisms such as molluscs, foraminifera, and calcareous green algae (*Penicillus* and *Udotea*). When these organisms die their skeletal debris accumulates as shelly muds, which would form bioclastic mudstones and wackestones (*see* **Sedimentary Rocks:** Limestones) if lithified into limestones. Shells are concentrated on beaches and storm ridges as bioclastic gravels and sands. Identifying the origin of the carbonate mud is a complex problem because lime mud may be produced by the breakdown of the lightly calcified tissue of green algae, by organisms that graze or bore into rocky carbonate shorelines and skeletal carbonate material (known as bioerosion), or by chemical precipitation directly from seawater. Mud production has been extensively studied in South Florida, and mineralogical (calcite and aragonite) and geochemical (calcium, magnesium, and iron levels, etc.) analyses of the muds show that they are similar to the carbonate minerals of the skeletal organisms. This implies that most of the mud is produced by the breakdown of skeletons of algae and benthic invertebrates. Once formed, the mud is reworked by burrowing organisms (mainly shrimps and worms) that digest what they can from the mud and defecate the remainder as sedimentary faecal pellets or peloids.

Within Florida Bay the peloidal muds are swept by storm waves into shallow-water mud mounds (piles of mud that accumulate from the seafloor to sea-level), which are stabilized by seagrass (*Thalassia*) (Figure 7B). These mounds have distinct windward erosional margins on their north-east sides and accretionary *Thalassia*-stabilized leeward (south-westerly) margins. Mounds therefore migrate in a south-westerly direction towards the Gulf of Mexico. Between the mounds there may be a thin cover of bioclastic (molluscan, algal, and foraminiferan) sands and muddy sands or alternatively areas may be bare of

sediment and expose the underlying Pleistocene Key Largo Limestone.

Back-reef areas are of normal marine salinity and are more exposed to waves, which generate well-oxygenated waters. These conditions favour a more diverse community of carbonate-secreting organisms, such as corals, encrusting coralline algae, calcified plate-like green algae (*Halimeda*), molluscs, echinoderms, and foraminifera. The corals are only abundant enough to form small patch reefs, which are built up by corals and coralline algae overgrowing each other to form a rigid reef framework (Figure 7C). This framework forms an attractive habitat for many other organisms, including those that eat or bore into the coral, such as sponges, worms, parrot fishes, echinoids, and molluscs. The coral debris they generate forms much of the skeletal carbonate sand and gravel that accumulates around the patch reefs. These, if lithified, would form skeletal or bioclastic packstones or grainstones (*see* **Sedimentary Rocks:** Limestones) in the geological record.

The rimmed margin of this platform comprises a barrier reef (because the reefs are detached from the shoreline and have back-reef and lagoonal areas behind them) with intervening high-energy shallow-water shoals of coarse skeletal sand and gravel. The patchiness of the reefs arises because Florida is near the northern limit of warm-water or tropical carbonates (Figure 2) and reef growth is not as prolific as it is further south in the Caribbean, where more continuous reefs occur. The shallowest zone of the reef is formed by the branching coral *Acropora palmata* (Figure 7D). The reef is built by a series of spurs and grooves that face out into the ocean. Spurs are built by seaward growth of the branching *A. palmata* and by some round-shaped (or head) corals and encrusting coralline algae. The grooves are gravel- and pebble-floored erosional channels through the reef. Deeper down the ocean-facing front of the reef, more massive and rounded coral growth forms are found, which give way to deeper-water flattened or dish-shaped corals. Lower areas of this fore-reef slope are characterized by coarse-grained reef talus deposits formed of broken and bioeroded debris from the reef. Similar reef morphologies and coral zones are preserved in ancient reef limestones, where they are called coral boundstones and are composed mainly of fossil corals.

Attached rimmed carbonate shelves are common in the geological record. In the Oligocene and Miocene of the Mediterranean, reef-rimmed shelves built out from coastlines to landmasses generated in the Alpine Orogeny. Mesozoic examples are found in the Jurassic limestones of the Jura Mountains (France and Switzerland), and Palaeozoic examples include reef-rimmed shelves (the Capitan Limestone) in the Midland Basin (USA) of West Texas and New Mexico in North America. In the Midland Basin these carbonate shelves are buried to a depth of several kilometres and contain oil and gas reservoirs in the porous limestones and dolomites that formed originally on a Permian carbonate shelf.

Unattached Rimmed Carbonate Shelf

The Great Bahama Bank is the best-studied present-day example of a large unattached or isolated carbonate platform with a rimmed platform margin. The early research of the 1950s and 1960s concentrated on the shallow platform-top environments and sediments; later work in the 1970s and 1980s studied the deeper slopes, and more recently the internal structure is being studied with seismic surveys (e.g. Figure 3) and drilling.

The climatic setting of the Bahamas is similar to that of Florida, and they are also dominated by north-easterly tradewinds. Therefore reef-fringed margins form on the east-facing windward platform margins (Figures 6 and 8). As in South Florida, the distribution of sediments across the platform is influenced by water depth, wave and tidal energy, and variations in salinity. The waters adjacent to the Bahamas Banks are open-marine waters (36‰ salinity and 22–31°C) that descend to oceanic depths. The large shallow bank-top environments are seasonally arid, and salinities up to 46‰ are recorded. The area is microtidal, with a range of about 0.8 m at the ocean margin, diminishing in the platform interior.

Many of the bank-top environments (Figure 8) are similar to those of South Florida, but there are also some important differences. The large sheltered interior of the platform accumulates an aragonite-rich mud (unlike Florida Bay mud, which is of mixed mineralogy) whose origin has long been disputed. Current theories focus on two interpretations, which are not mutually exclusive. Surveys of the skeletal marine communities of Little Bahama Bank (just north of Providence Channel; Figure 6) indicate that most of the mud is produced by the breakdown of calcareous green algae, as is the case in Florida Bay. The other proposal is that high salinities in the bank interiors may result in direct precipitation of aragonite from seawater. This is similar to the process of lime mud production in the lagoons of the Arabian Gulf (see below). In the leeward environments of Andros Island extensive mud flats are developed, which are cut by sinuous tidal channels. The channels and the mud flats are both active sites of sediment accumulation as they build out into the peloidal muds of the platform interior (Figure 8).

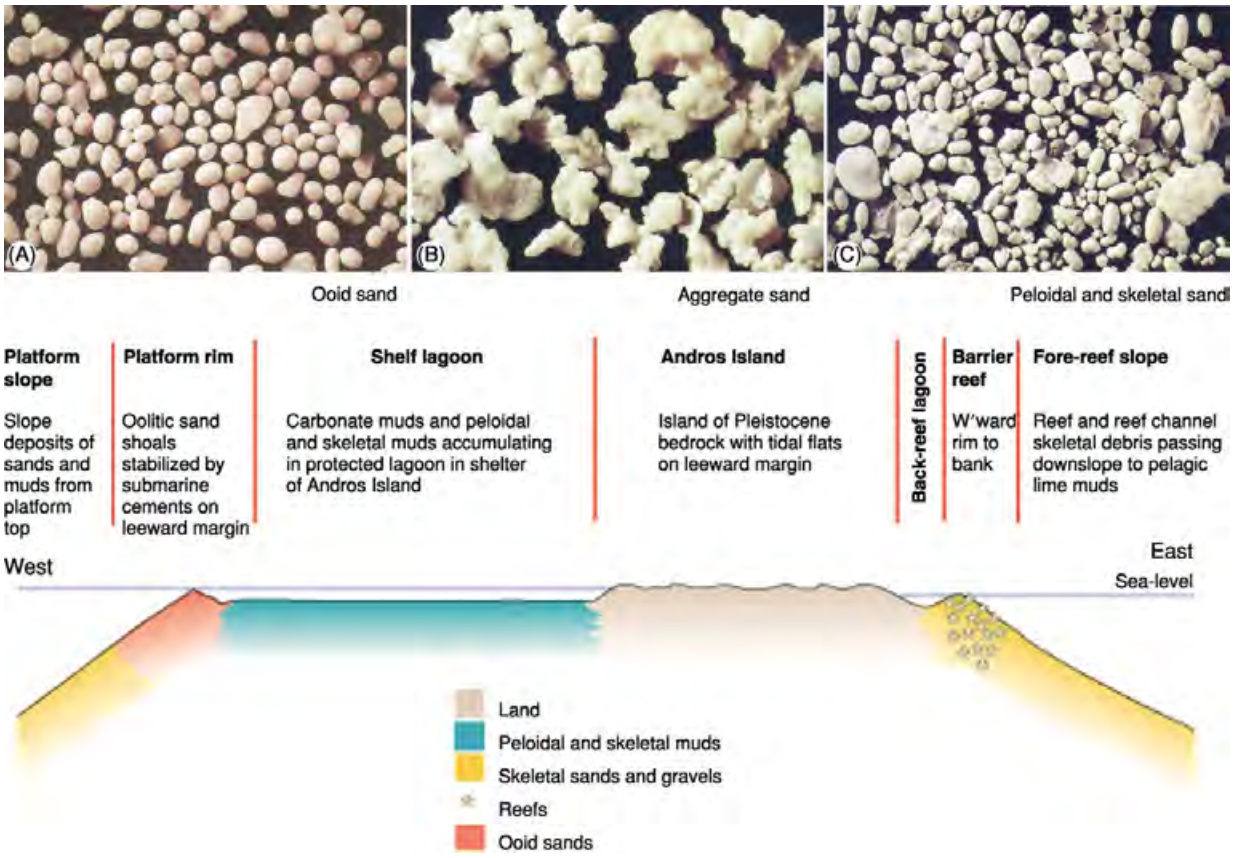


Figure 8 East–west cross section through the Great Bahama Bank (for location see Figure 6) showing the main sedimentary environments on this rimmed platform. Photographs of sediment types (courtesy R. Till) illustrate the range of non-skeletal grains on this isolated bank: (A) ooids; (B) aggregate grains; and (C) peloidal and skeletal grains (note foraminifera bottom left). All photographs are 2 cm across. (After Bosence DWJ and Wilson RCL (2003) Carbonate depositional systems. In: Coe A (ed.) *The Sedimentary Record of Sea-level Change*, pp. 209–233. Milton Keynes and Cambridge: The Open University and Cambridge University Press.)

Whilst most windward margins are reef dominated and, as such, are similar to those described above from south Florida, other margins are rimmed by sand shoals. These may be formed of oolitic, peloidal, or bioclastic sands. Aragonite precipitation onto fine sedimentary particles forms ooids. These are spherical concentrically structured grains of carbonate sand (Figure 8A). The precipitation of aragonite crystals to form the concentric layers is favoured by the high salinity of platform interior waters, which are swept off the banks by tidal currents, and by the degassing of carbon dioxide from deeper waters off the platform margin. Tidal currents sweep the ooid sands into large submarine dunes cut by channels and interdune areas. The sand grains may also become cemented together on the seafloor by crystal growth to form grain aggregates (Figure 8B). Large areas may become cemented to form what are known as hardgrounds. This early cementation helps to stabilize the sand shoals and the steep margins of the platform.

The Bahaman platforms are bounded by slopes, some almost vertical, that plunge to depths of 4 km on the ocean-facing margins. These slopes may be indented by gulleys or canyons formed by slope failure followed by abrasion by sediment gravity flows. Whilst there is observable sediment production down to around 200 m on these slopes from deep-water coral and coralline algal communities, much of the sediment on the slopes and moving downslope is of shallow-water origin. In addition some is derived from erosion on the slopes themselves, and, in areas of little sediment redeposition, accumulation from pelagic organisms is important. The pelagic component, characterized by calcitic muds derived from pelagic organisms (in contrast to the aragonitic muds generated on the platform top), also increases from the toe of the slope towards the oceanic environments of the Atlantic Ocean (Figure 2).

The broad pattern of sediment distribution shown in Figure 8 has probably prevailed for many millions of years, but the leeward margin of the Bank has

prograded many kilometres westwards into the Florida Strait (Figure 3). The windward (eastern) side (not illustrated in Figure 3) has remained as a steep reef-rimmed escarpment for much of this time.

Unattached rimmed shelves or carbonate banks are known from many parts of the geological record. Tertiary examples are best known from the subsurface studies of the foundations of the Bahama Banks, as discussed above and in Figure 3. Mesozoic examples come from the ancient Tethys Ocean and now form extensive Jurassic limestones in the Betic–Alpine–Apennine mountain chains of Europe. In particular, the Jurassic limestones of the Apennines of Italy once formed large unattached platforms that were isolated from the margins of the Tethys Ocean. Palaeozoic examples come from the Silurian and Devonian inland seas of Canada. Here, communities of corals and stromatoporoids helped to build reef-rimmed margins to unattached platforms known locally as pinnacle reefs. In the subsurface these platforms are encased in shales or evaporites and contain oil and gas reservoirs hosted in the porous and permeable carbonate rocks.

Attached Carbonate Ramp in an Arid Tropical Environment

The southern coast of the Arabian Gulf offshore from the Gulf Coast States (Abu Dhabi, Dubai, and Qatar) is a fine example of a present-day carbonate ramp. The shelf slopes gradually from the low-relief desert of the coastal plain through the coastal waters down to a maximum depth of about 100 m over a distance of a couple of hundred kilometres (Figure 9). This gently sloping shelf morphology is the distinctive feature of a carbonate ramp, and there are no major reef systems or rimmed shelf margins. The prevailing winds ('Shamal') blow onshore from the north-west, which makes this a storm-wave dominated coastline (Figure 9). Owing to the restricted opening to the Indian Ocean through the Strait of Hormuz and the arid climate, salinities are elevated in the Gulf to 40–45‰ and are even higher in coastal areas, where evaporite minerals are precipitated from these waters.

The depth of the wave base, the water depth at which wave-generated currents affect the seafloor, has an important control on sediment texture in

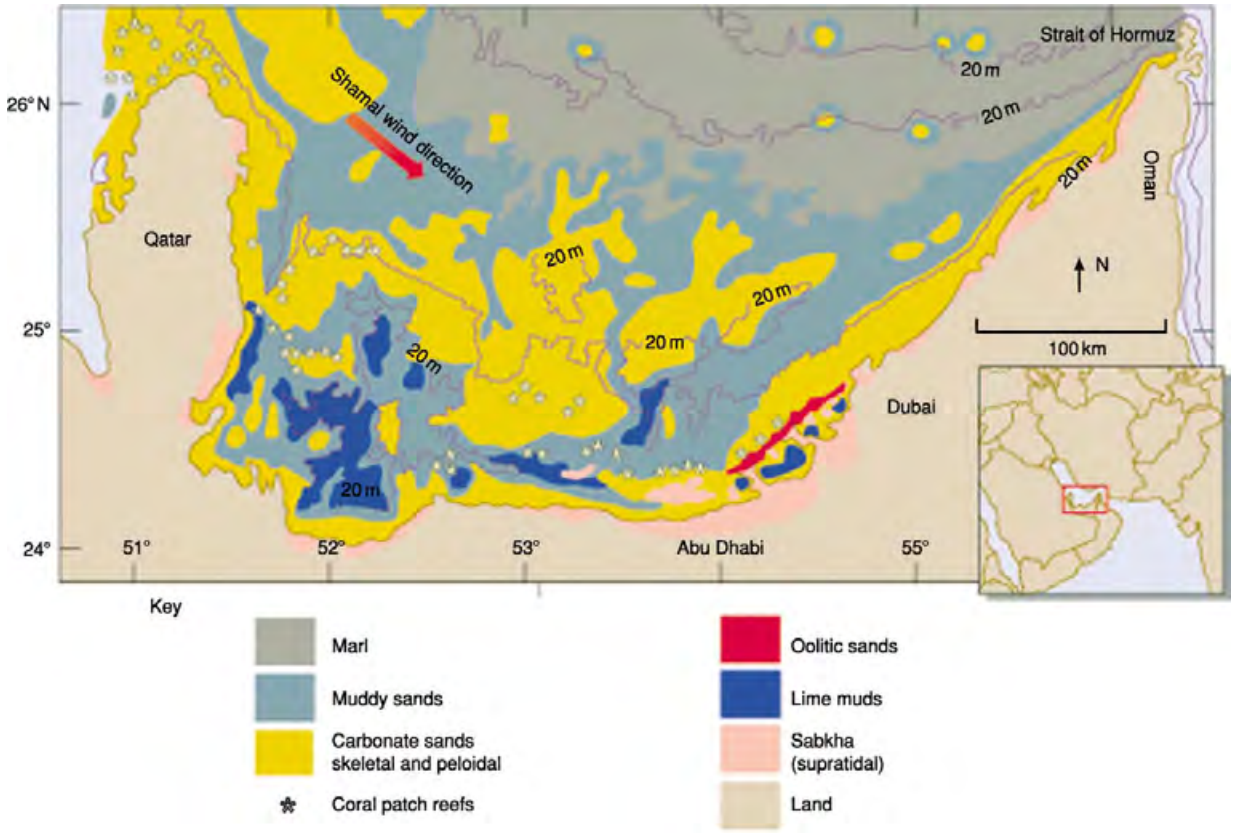


Figure 9 The south-eastern region of the Arabian Gulf, showing bathymetry and the distributions of the main carbonate sediment types. (After Bosence DWJ and Wilson RCL (2003) Carbonate depositional systems. In: Coe A (ed.) *The Sedimentary Record of Sea-level Change*, pp. 209–233. Milton Keynes and Cambridge: The Open University and Cambridge University Press.)

gently sloping ramps and is used to subdivide ramps into zones (Figure 10). Above the fair-weather wave base (the depth at which waves impinge on the seafloor during normal or fair-weather conditions) is known as the inner ramp (Figure 10). The zone between the fair-weather wave base and the storm wave base (wave base for storm-generated waves) is the mid-ramp, and the outer ramp is the deep-water zone below the effect of waves.

In the Gulf, the outer-ramp areas accumulate muds with a mixed composition (called marl), which are part carbonate of pelagic origin and part siliciclastic material brought into the Gulf by rivers. These marls grade up the gentle ramp slope into muddy and then clean skeletal and oolitic sands at the fair-weather wave base at around 10–20 m (Figures 9 and 10). Areas of clean skeletal sand are found in inner-ramp environments, and these are commonly cemented on the seafloor to produce extensive hardgrounds. Small patch reefs (Figure 10) occur but the coral fauna is of low diversity because of the elevated salinities and sometimes low winter temperatures.

The inner ramp is dominated by skeletal sand beaches or ooid barrier beaches in the Abu Dhabi area (Figure 9). Aragonitic ooids are precipitated in the relatively high-salinity marine waters with a moderate tidal and wave-energy regime. Ooid barrier islands are situated up to 20 km from the shoreline, and broad shallow lagoons have developed behind the barriers (Figure 9). The lagoons have elevated salinities of 40–50‰ and so have a reduced diversity of marine faunas comprising gastropods and ostracods that occur in peloidal muddy sands. The lagoons accumulate aragonitic lime muds, and both the

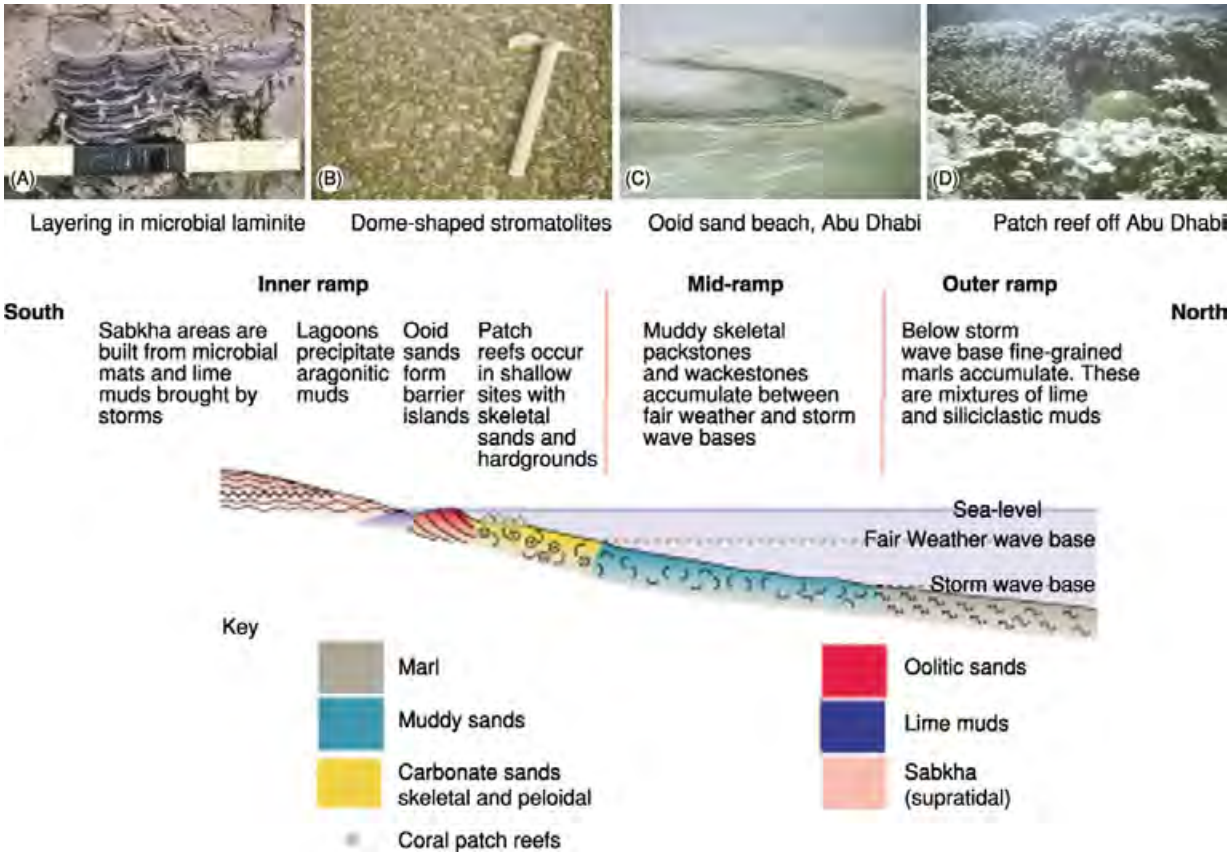


Figure 10 Cross-section through the south-eastern margin of the Arabian Gulf to illustrate depth-related sediment types on a carbonate ramp. Note the changes in sediment type at the storm wave base and fair-weather wave base. (Modified from Bosence DWJ and Wilson RCL (2003) Carbonate depositional systems. In: Coe A (ed.) *The Sedimentary Record of Sea-level Change*, pp. 209–233. Milton Keynes and Cambridge: The Open University and Cambridge University Press.) The photographs (courtesy R. Till) illustrate the diversity of inner-ramp sediment types in this arid climate. (A) Vertical slice though intertidal flats, showing alternating microbial (dark) and sediment (light) layers. White patches within the layering are precipitates of the calcium sulphate salt anhydrite (scale in inches). (B) Dome-shaped stromatolites in the intertidal sabkha. These result from the upward growth of microbial mats and bound sediment, with lighter-coloured sediment accumulating between the domes. (C) Light-coloured beaches of ooid sand occur on a string of barrier islands off the coast of Abu Dhabi. The dark area by the water's edge is microbial mat. (D) Patch reefs offshore of the barrier islands have a low diversity compared with reefs of the nearby Indian Ocean because of the high salinity of the Gulf waters.

absence of calcified green algae (such as those present in South Florida and the Bahamas) and geochemical evidence from the lagoon waters and the deposited muds indicate that these muds were precipitated chemically within the lagoon rather than having an organic origin, as is the case in Florida Bay.

The sabkhas (Arabic for area of low-lying salty ground) are broad intertidal areas and coastal plains (Figures 9 and 10) that may be flooded by lagoon waters when storm winds blow landwards. Intertidal areas have extensive microbial populations on the tidal flats growing on and within sediment deposited on the sabkhas when they are flooded by storms. These processes give rise to sediments with alternating microbial-rich and sediment-rich layers, known as stromatolites (Figure 10). High aridity in the area leads to a net evaporation of floodwaters and the drawing up of saline groundwaters into the sabkha. This leads to the precipitation of evaporite minerals, such as dolomite, gypsum, and anhydrite, within the microbial mats, including the stromatolites.

With time all these depth-related sedimentary zones will build out or prograde into the Gulf, forming a sedimentary succession that has been recognized in many ancient limestones throughout the world. Ancient carbonate ramp successions are known from the Carboniferous of north-western Europe and the USA, where they build out carbonate shelves with facies similar to those described from the present-day Arabian Gulf. However, in addition, the deep-water areas are characterized by carbonate mud mounds that do not have an exact modern-day analogue. Very extensive Jurassic and Cretaceous carbonate ramps are known from the subsurface of the eastern part of the Arabian Plate. Here they host the rich oil and gas reserves of the Middle East in prograding units of inner ramp carbonate sand shoals and patch reefs. The oil and gas is preserved in these porous carbonate rocks and sealed by overlying evaporites, which formed in sabkha environments similar to those that occur today in the southern Arabian Gulf.

Attached Carbonate Ramp in a Cold Temperate Environment

The shelf regions of the north-eastern Atlantic form a large area of cold-water shelf carbonate sediments (Figures 2 and 11). These extend from the narrow Spanish shelf to broad shelf regions within the Arctic Circle near Spitzbergen. Similar extensive areas of cold-water carbonates are found in the southern hemisphere, particularly off the south coast of Australia (Figure 2). Cold-water carbonates differ from their warm-water counterparts in that they are formed only of skeletal material: there are no precipitated grains (ooids etc.) in these cooler waters. In the north Atlantic temperatures vary from an average of 7°C in winter to 15°C in the summer, and salinity levels are relatively constant at around 35‰. Within these waters skeletal sands and gravels are formed from a variety of different seafloor communities that contrast with those of the tropics.

The shelves of north-western Europe form what are termed distally steepened ramps in that they have very gentle gradients from the shoreline to a shelf margin at around 200 m water depth (Figure 11). Here, the seafloor drops away more steeply on what is known as the continental slope. The shelf environment is subject to a number of different currents that have an important effect on sediment types and their distribution. The North Atlantic Drift is an oceanic current that flows north-eastwards along the continental slope and is strong enough to transport sand along the upper slope in large seafloor sand dunes. The shelf is also affected by westerly winds that have blown across the Atlantic Ocean and generate significant storm waves (i.e. a maximum of 30 m during a 50 year period) that can rework the seafloor sediments

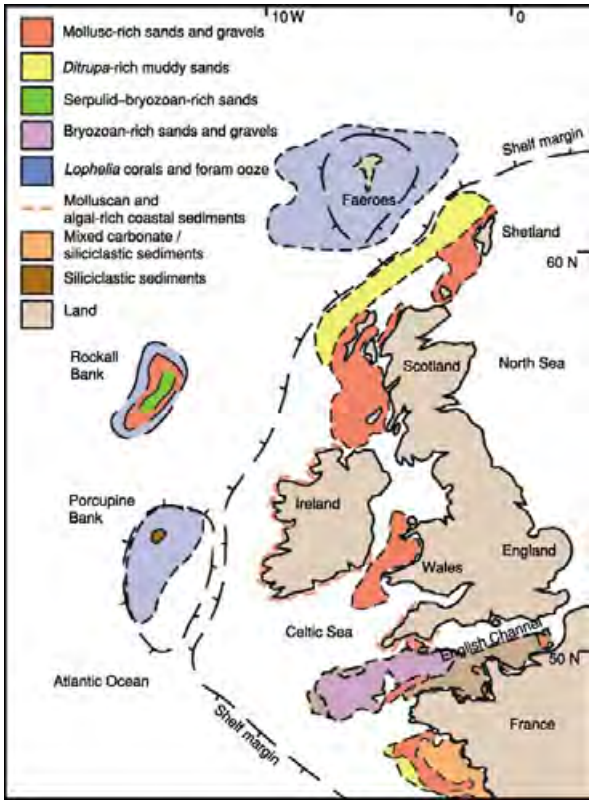


Figure 11 The north-east Atlantic Ocean, showing the distributions of the main types of shelf and slope carbonate sediments. Note that carbonate sediments are abundant on the attached shelves of north-western Europe and on the isolated platforms of the Faeroes, Rockall Bank, and Porcupine Bank. White areas of shelf region are those that have not been studied in detail.

right down to the shelf edge at 200 m depth. In addition, tidal currents are strong and can rework sediments and prevent mud deposition over much of the shelf. Therefore the entire shelf is above storm wave base and much of it is within the mid-ramp zone as the nearshore areas slope steeply off the coastal cliffs. Locally, for example in western Ireland and north-west Scotland, carbonate beaches are formed of skeletal sands and gravels (Figure 1C). By way of example, a shelf profile is shown across the shelf north of Scotland (Figure 12). Inner to mid parts of the shelf are swept by tidal currents that restrict any mud sedimentation, and these areas are characterized by mollusc-rich sands and gravels (Figure 12). Such facies are common throughout this carbonate province (Figure 11). Locally they are enriched in siliciclastic sand grains that have been reworked from nearby rocky cliffs and islands. These shell gravels not only represent the remains of a diverse molluscan-rich seafloor community but also form a substrate for encrusting calcareous organisms such as tube worms and bryozoans.

The outer part of the shelf and the shelf margin are inhabited by a quite different seafloor community dominated by an unusual free-living or non-encrusting tube worm known as *Ditrupa* (Figure 12). These conical tubes occur in great abundance (upto 1000 individuals per square metre) in the outer shelf, associated with minor contributions from molluscs, bryozoans, and solitary corals that use the *Ditrupa* tubes as a hard substrate to grow on. With increasing depth, pelagic foraminifera become more important and the sediments have a muddy texture below wave base and the shelf edge. Many areas of the deep shelf and continental slope are covered by a colonial coral known as *Lophelia* (Figure 12). Although not present on the selected transect, it is common around the unattached carbonate shelves of the Faeroes, Rockall, and Porcupine Banks of the north-east Atlantic (Figure 11) where it forms extensive thickets on the seafloor. On the Norwegian shelf it forms deep-water reefs tens of metres high and hundreds of kilometres long. Living at depths of 200–500 m, these coral thickets exist by suspension feeding. They are not

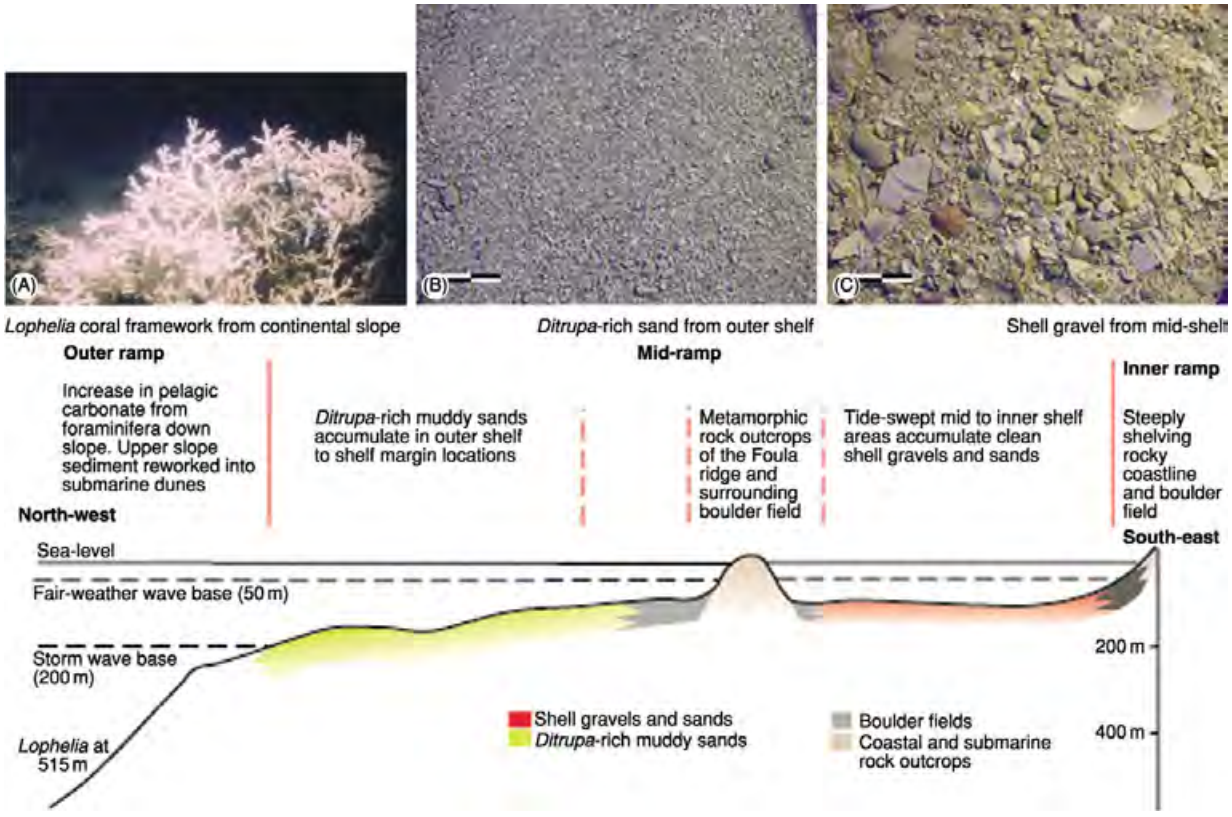


Figure 12 North-west–south-east section across the shelf to the west of Shetland, illustrating the occurrence of the main carbonate facies with respect to shelf regions. Note the considerable depth of the fair-weather and storm wave bases compared with the sheltered Gulf (Figure 10) and the importance of the rocky shoreline and islands (Foula) in this region. Photographs illustrate the skeletal nature of the shelf carbonates: (A) scale: foreground 50 cm across; (B, C) scale bars = 2 cm. (After Light JL and Wilson JB (1998) Cool-water carbonate deposition on the West Shetland Shelf: a modern distally steepened ramp. In: Wright VP and Burchette TP (eds.) *Carbonate Ramps*, pp. 73–105. Special Publication 149. London: Geological Society.

restricted to the photic zone like their warm shallow-water counterparts such as those from the Caribbean described above. However, they have an analogous complex associated fauna. The extent and thickness of these cold-water carbonates in the northern and southern hemispheres has become known only in the last 20 years or so, and the full geological impact of their discovery has yet to be established. Many Miocene limestones from southern Europe are now believed to be of temperate origin as are some Late Palaeozoic limestones of North America and southern Australia. Because of their more limited extent, cold-water carbonates have fewer economic uses than their tropical counterparts. In the coastal areas of north-western Europe the calcareous algae are dredged for use as agricultural and horticultural fertilizer and for liming the acidic soils of the region. However, because of the low growth rates of the carbonate producers, this is a non renewable resource.

Conclusions

From this review it can be seen that continental shelf seas have extensive areas of carbonate sediment accumulation. The nature and distribution of these areas are mainly controlled by climate and morphology. Although pelagic sediments forming in deep ocean waters are more abundant today, their usual geological fate is to disappear into subduction zones. This explains why limestones formed in shelf environments are the commonest type of limestone in the geological record. If we are to understand how limestones formed and what controls their distribution and composition, we must study their modern counterparts.

To conclude, carbonate shorelines and shelves are important features of the world's sedimentary and biological systems; they contain the biologically diverse and economically important habitat of tropical coral reefs, analogues for the interpretation of ancient limestones, and some of the most attractive coastlines of the world.

See Also

Diagenesis, Overview. **Fossil Invertebrates:** Corals and Other Cnidaria; Bivalves. **Microfossils:** Foraminifera. **Minerals:** Carbonates. **Sedimentary Environments:** Reefs ('Build-Ups'). **Sedimentary Processes:** Karst and Palaeokarst. **Sedimentary Rocks:** Dolomites; Limestones. **Seismic Surveys**.

Further Reading

Bosence DWJ and Wilson RCL (2003) Carbonate depositional systems. In: Coe A (ed.) *The Sedimentary Record of Sea-level Change*, pp. 209–233. Milton Keynes and Cambridge: The Open University and Cambridge University Press.

Jones B and Desrochers A (1992) Shallow platform carbonates. In: Walker RG and James NP (eds.) *Facies Models Response to Sea Level Change*, pp. 227–301. St Johns: Geological Association of Canada.

Light JL and Wilson JB (1998) Cool-water carbonate deposition on the West Shetland Shelf: a modern distally steepened ramp. In: Wright VP and Burchette TP (eds.) *Carbonate Ramps*, pp. 73–105. Special Publication 149. London: Geological Society.

Purser BH (ed.) (1973) *The Persian Gulf: Holocene Carbonate Sedimentation and Diagenesis in a Shallow Epicontinental Sea*. Berlin: Springer Verlag.

Tucker ME and Wright VP (1990) *Carbonate Sedimentology*. Oxford: Blackwell Scientific Publications.

Wright VP and Burchette TP (1996) Shallow water carbonate sediments. In: Reading HG (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, pp. 325–394. Oxford: Blackwell Science.

Contourites

M Rebesco, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Italy

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Contourites are sediments deposited or substantially reworked by the powerful action of bottom currents.

The study of contourites is important in at least three main respects: palaeoclimatology, hydrocarbon exploration, and slope stability. The continuous and relatively high-resolution record included in these sediments is crucial for palaeoclimatology since it holds the key for priceless information on the variability in circulation pattern, current velocity, oceanographic history, and basin interconnectivity. Bottom currents are a crucial factor in oil and gas reservoir development since weak

flows may favour the accumulation of source rocks and robust flows may represent a viable mechanism capable of forming 'clean' sands in the deep sea. Low-permeability, fine-grained contourites have been found recently to play a critical role in slope stability by providing potential overpressured glide planes, either when their high-water-content is rapidly loaded or when their rigid biosiliceous microfabric is collapsed by diagenesis.

Despite its significance, little is still known about this group of sediments. There are three reasons for the long-lasting disregard: the inherent elusive nature of these complex deposits, the 50-year dominance of the turbidite paradigm, and the controversy that surrounds these sediments since they were first recognized. The elusive and very subtle characteristics of these slowly and continuously accumulated sediments and their occurrence within a spectrum of deep-water deposits (Figure 1) does not allow them to be easily recognized and decoded. The monumental efforts to promote the turbidite systems, the simplicity and predictability of turbidity current concepts, and the sense of confidence given by the routine use of widely applicable models has induced the geologic community to ignore alternative, more complex deep-water models. The never-ending disputes regarding the occurrence of sandy contourites versus reworked turbidites and the early errors in recognition of fossil examples have prevented the establishment of a widely shared consensus on valid diagnostic criteria for the identification of these deposits.

The growing level of interest and research in contourites is shown by the recent publication of several special volumes dealing with such systems, covering large parts of present ocean floors and continental margins. An increasing number of fossil occurrences in ancient sediments exposed on land have also been documented. Nevertheless, wide multidisciplinary approaches and the integrated work of different international specialists (e.g., deep-water sedimentologists, seismic interpreters, physical oceanographers, and palaeoclimatic modellers) are still needed to improve our knowledge of these systems and to help tackle the problems that remain.

History

The systematic study of deep-sea sediments began at the end of the nineteenth century. At that time the ocean floor was perceived as a tranquil realm receiving only pelagic clays. The possibility that thermohaline bottom currents may influence sedimentation in the deep oceans was suggested by the German oceanographer Wust in 1936. However, the contourite concept was not accepted in marine science until the second half of the 1960s after the American team of Heezen and Hollister had provided geological and oceanographic evidence of this process along the eastern North American continental margin. During the next two decades standard facies models for contourite sediments were developed from coring contourite accumulations. Other important steps forward were achieved by addressing the link between current strength and grain size and by confronting the problem of distinguishing contourites from other deep-sea facies like hemipelagites and fine-grained turbidites. Among the projects that have provided significant contributions to contourite research have been the HEBBLE

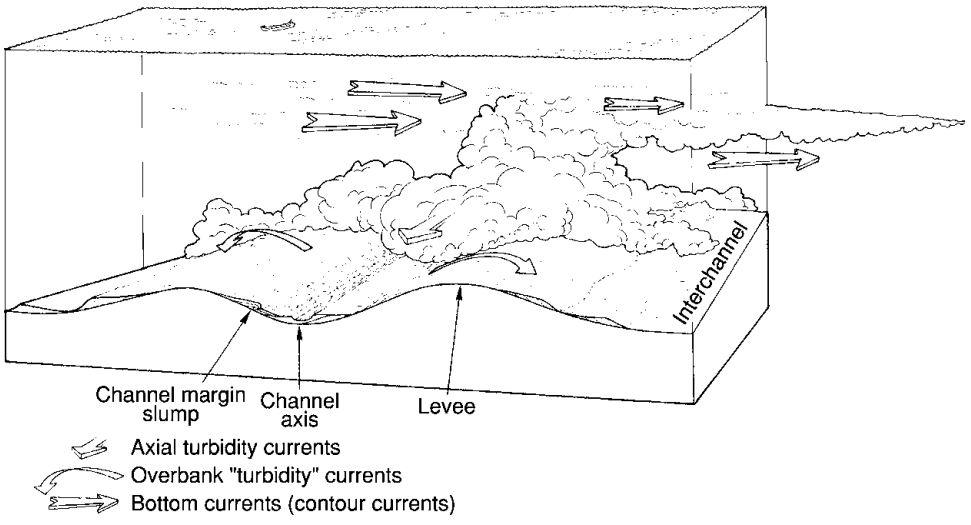


Figure 1 Schematic model of downslope gravity-driven currents and alongslope bottom currents interacting within the continuous spectrum of deep-water processes. Reprinted from Shanmugam (2000) 50 years of the turbidite paradigm (1950s–1990s): Deep-water processes and facies models—A critical perspective, *Marine and Petroleum Geology* 17: 285–342, with permission.

(High-Energy Benthic Boundary Layer Experiment) in the Nova Scotia Rise and the international drilling programs (ODP and predecessors) in most contourite deposits of the world's oceans (Figure 2). At the very end of the 1990s the International Geological Correlation Programme (IGCP) Project 432, named 'Contourite Watch', was launched to facilitate research on contourites and bottom currents. It included the establishment of a global network of contourite workers, the organization of international workshops, and the publication of selected special volumes, including an atlas of contourite systems.

Terminology

Most workers currently agree on very broad definitions and on the use of most terms related to contourite research. The need to maintain a certain degree of flexibility to allow for development in understanding has been acknowledged; nevertheless the present state of contourite terminology usage is possibly still too loose to allow for a clear description of the products (sediments) and consequently for a clear understanding of the depositional processes.

The term 'contourite' was originally introduced to define the sediments deposited in the deep sea by contour-parallel bottom currents. The term was subsequently widened to embrace a larger spectrum of sediments that at diverse depths are affected to various extents by different types of currents. Since both the depth and direction of the bottom currents that influenced nonrecent deposits are seldom precisely identifiable, a rigorous restriction of contourite to its original sense would severely limit its application. Such a long-established term should hence be maintained for nonspecific use, and the employment of modifying terms to qualify the deposit in terms of depth, current type and action, and interacting process is recommended (see Table 1).

The term 'contour current' refers to those thermohaline-driven, deep-water currents that flow approximately parallel to the bathymetric contours. It is widely used synonymously for 'bottom current', which in contrast includes all deep currents not driven by sediment suspension that are capable of eroding, transporting, and depositing sediment on the sea floor. Although bottom current is suitably applicable to different types of currents flowing alongslope as well as upslope and downslope (see below), the term contour current should be applied only to currents whose flow can be recognized as being parallel to contours.

Contourites were first described as being at great depths beneath deep-water bottom currents. Subsequently, contourites and bottom current-controlled deposits were recognized in several settings including shallow gateways, outer shelf and slopes, and even lakes. The depth ranges for a three-category subdivision was hence proposed: deep water ≥ 2000 m; mid-water $= 300$–2000 m; shallow water $= 50$–300 m. However, a rigid definition of the depth confine is of limited value, especially for nonrecent and fossil contourites for which the deposition depth is largely unknown. Nevertheless, the use of a modifying term to qualify the deposit in terms of depth is recommended when a distinction can be made.

'Sediment drift' is the generic term for a sediment accumulation that was appreciably controlled by the action of a bottom current. It is often used synonymously for 'contourite drift', which in contrast should be specifically used for sediment accumulations deposited principally (though not exclusively) by contour currents. Therefore, the most appropriate term for contourite accumulations is in most cases the wide-ranging designation 'sediment drift' or simply 'drift'. The occurrence of mixed system is the norm rather than the exception with sediment drifts. For this reason, a qualifying modifier like bottom current-controlled (modified, reworked, etc.) is recommended for turbidite systems and hemipelagic deposits that were significantly affected by interaction with bottom currents. Similarly, when mounds, levees, fans, lobes, channels, and other terms closely associated with downslope systems are applied to contourite systems, they must be preceded by the term contourite.

Bottom Currents

There is a wide spectrum of bottom currents that operate in deep water. The bottom currents sensu stricto (contour currents) are those that are part of either the thermohaline- or wind-driven major circulation patterns (Figure 3). They are usually persistent for long time intervals as testified by the thick drifts that took million of years to develop. Such currents generally have an overall alongslope flow, but in detail their velocity and direction are extremely variable in both time and space. In fact they are affected by seafloor morphology (obstacles, gateways, and changes in slope direction and steepness), Coriolis force, circular motions (gyres and eddies) unrelated to contours, and eddy kinetic energy changes (seasonal and at different scales). Substantial changes in kinetic energy at the seafloor may be produced by sea-surface topographic variations, resulting in episodic high bottom current velocities, referred to as 'benthic storms'. Such storms may resuspend large amount of sediment that nourish the bottom 'nepheloid layer' (several hundred metres thick, long-living, low-concentration (50 to 100 $\mu g\,l^{-1}$) clouds of sediment particles close to the seafloor).

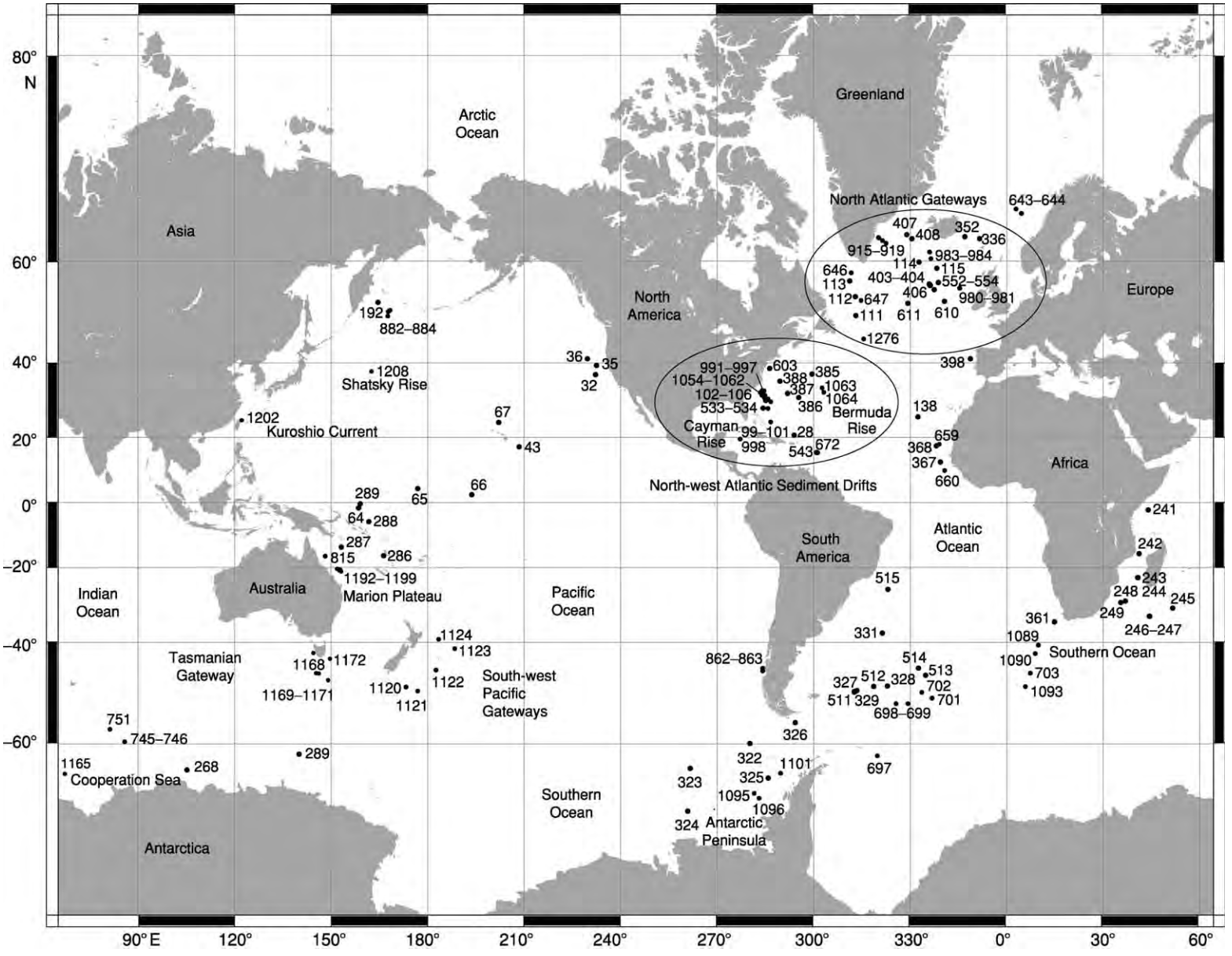


Figure 2 ODP, IPOD, and DSDP sites drilled in contourite deposits of the world's oceans. Data sources: 1998 to present from ODP Web site http://www-odp.tamu.edu/sitemap/sitemap.html; and before 1998 from Stow DAV, Faugères JC, Viana A, and Gonthier E (1998) Fossil contourites: A critical review, in: Stow DAV and Faugères JC (eds.) Contourites, turbidites and process interaction, *Sedimentary Geology* 115: 3–31.

Table 1 Examples of qualifying terms for contourite-related deposits

Depth	Deep water	Mid-water	Shallow water	Shelf	Slope	Lacustrine
Type of current	Contour-parallel	Thermohaline-driven	Wind-driven	Up- and downwelling	Up- and down-canyon	Internal waves and tides
Type of action	Resuspended	Pirated	Far-off transported	Condensed	Polished	Winnowed
Interacting process	Turbidite sourcing	Pelagic settling	Overbanking	Faulting	Creeping	Downslope resedimentation

N.B.: Not all kinds of terms are to be used concurrently (e.g., deep-water turbidites winnowed by wind-driven bottom currents, or mid-water turbidite-sourced contourites).

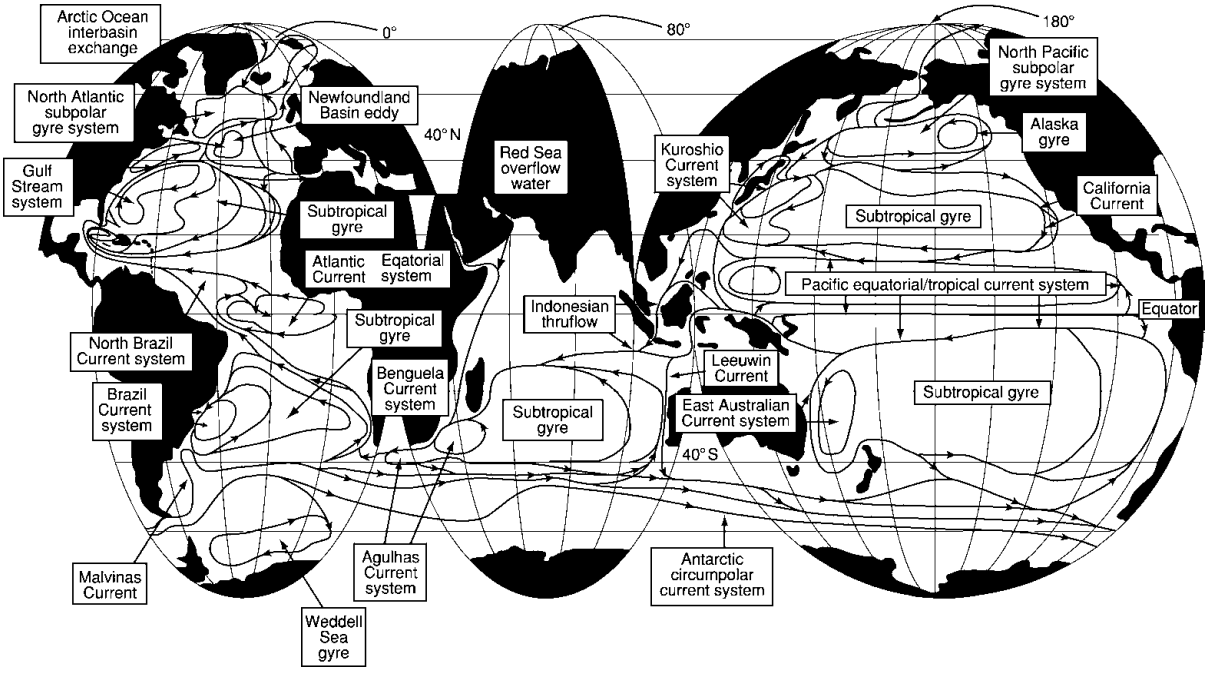


Figure 3 Global ocean circulation pattern. Reprinted from Steele JH, Turekian KK, and Thorpe SA (eds.) *Encyclopedia of Ocean Sciences*, Academic Press, London, with permission.

Additional types of deep (bottom) currents not driven by sediment suspension have been observed to flow mostly perpendicular to the slope. These include currents related to internal waves and tides (formed between subsurface water layers of varying density), canyon currents (frequently reversing flow of clear or very low-density turbid waters), and down- and upwelling flows (currents on the continental slopes that may be generated by density, wind, storms, or obstructions).

The present global thermohaline circulation of deep water is originated by the cooling and sinking of surface water masses in the polar areas. The Antarctic bottom water (AABW) forms beneath Antarctic floating ice shelves (especially in the Weddell Sea), descends the continental slope, and hence circulates clockwise in the Southern Ocean from which it escapes northwards into the world's oceans. The Arctic bottom water (ABW) forms in the Norwegian and Greenland Seas, flows intermittently south through basin gateways, and mixes with the North Atlantic bottom water and Labrador seawater. Subsequently, AABW and ABW gradually mix with other stratified water masses of their own density and circulation pattern (e.g., intermediate depth water produced by evaporation in semienclosed seas, like the Mediterranean overflow water—MOW).

The present global surface circulation is governed by dominant wind systems. In many cases the action of such currents affects the entire water column down to the seafloor. This is the case for the powerful Antarctic circumpolar current (ACC) and western boundary currents like the Gulf Stream, Brazil, and Agulhas and Kuroshio Currents.

All the above currents impinge upon the seafloor where they may erode, transport, and deposit sediment

in the form of contourites. The velocities of these bottom currents are generally very slow (few centimetres per second) and hence just able to slightly affect the action of other depositional processes (hemipelagic settling and low-density fine-grained turbidity currents). However, bottom currents accelerate when constrained by Coriolis force within western boundary undercurrents restricted against the continental slope (commonly with velocities ranging between 10 and 100 $cm\,s^{-1}$) or when constricted in narrow gateways (velocities may even exceed 200 $cm\,s^{-1}$). In many cases the bottom currents are strong enough to profoundly affect deep-water sedimentation by a number of processes including (in order of increasing current strength) the prevention of fine sediment deposition, the entrainment (pirating) and transport of suspended sediments within the nepheloid layer, the reworking of the coarse fraction introduced by episodic down-slope processes, the erosion and transport by traction of silt and fine sand, and sedimentary lag production by polishing and winnowing the seafloor. Such energetic processes take place chiefly beneath the core of the main bottom current flow. The velocity decreases away from the axis of the current, and deposition of various types of sediment drifts takes place in the relatively slack waters to the side of sediment-laden bottom currents. In most palaeocurrent reconstructions, the axis of the current is hence generally considered subparallel to the elongated direction of the drift and to its crest (if traceable). However, such a relationship between current and accumulation is in most case highly speculative since the associated time-scales are extremely dissimilar. In fact, the time required for a drift development is commonly one order of magnitude larger than the glacial–interglacial cyclicity, which is widely accepted to profoundly affect bottom-water circulation, and exceeds by over four orders the time length of our direct oceanographic observations.

Sediment Drifts

There is a large variety of contourites, ranging from those that occur closely interbedded with other deep-water facies to those that build up individually distinct bodies (mounded drifts). Due to their large dimensions, commonly ranging between a few tens and several hundred thousand square kilometres, drifts are essentially identifiable by means of seismic profiles. Although in some instances sediment drifts may be clearly recognized on the basis of seismic profiles alone according to a widely agreed-upon criteria, a rigorous identification generally requires independent supporting data. The large-scale features of the drifts (their morphology and overall geometry) are controlled by a number of interrelated factors including bathymetric setting (physiography, depth, morphology), current conditions (velocity, time length of activity, variability, frequency of changes), sediment availability (amount, type, source type, frequency, and variability of input), and the interaction with other depositional processes (in time and space, both collaboratively or antagonistically). Accumulation rates in sediment drifts are rather variable, from a few tens to hundreds metres per million years, depending on regional conditions, specific age, and evolutionary stage. However, for most drifts their average growth rate is commonly several tens of metres per million years.

Repeated attempts to categorize the essentially continuous spectrum of contourite accumulations into several types of drift have been made (**Figure 4**); however, the categorizations are not totally distinct such that some drifts may be classified as more than one type:

1. A 'sheeted drift' is characterized by a very broad low-mounded geometry. It has a very low relief (up to a few hundred metres) over a very large area (up to $1 \times 10^6\,km^2$) and shows a slight decrease in thickness towards the margins. Two kinds of sheet drift may be identified:
   a. 'abyssal sheet', which fills basin plains whose margins trap the bottom currents within a complex gyratory circulation, and
   b. 'slope plastered sheet', which is coated (plastered) against continental margins where a gentle and smooth topography favours a broad nonfocussed current.
2. An 'elongated, mounded drift' is characterized by distinctly elongated and mounded geometry (**Figure 5**). This kind of drift is the most spectacular and was among the first to be identified. It may have a relief of several hundred metres above the surrounding seafloor (total thickness up to 2 km) and a very variable extent (10^3–$10^5\,km^2$) with an elongation ratio of at least 2:1. Elongation is generally parallel to the margin, which means that the crest of the drift is parallel to the current axis. Nevertheless, elongation and progradation trends are variable, being controlled by at least three interacting factors: current system, bathymetry, and Coriolis force. Two types of mounded drift may be identified:
   a. A 'separated drift' is kept apart (separated) from the adjacent continental slope by a distinct moat in which the main core (axis) of the bottom current is flowing as testified by local nondeposition and/or erosion (**Figure 6**). Such drifts show an evident upslope progradation with a minor alongslope migration in the downstream direction.

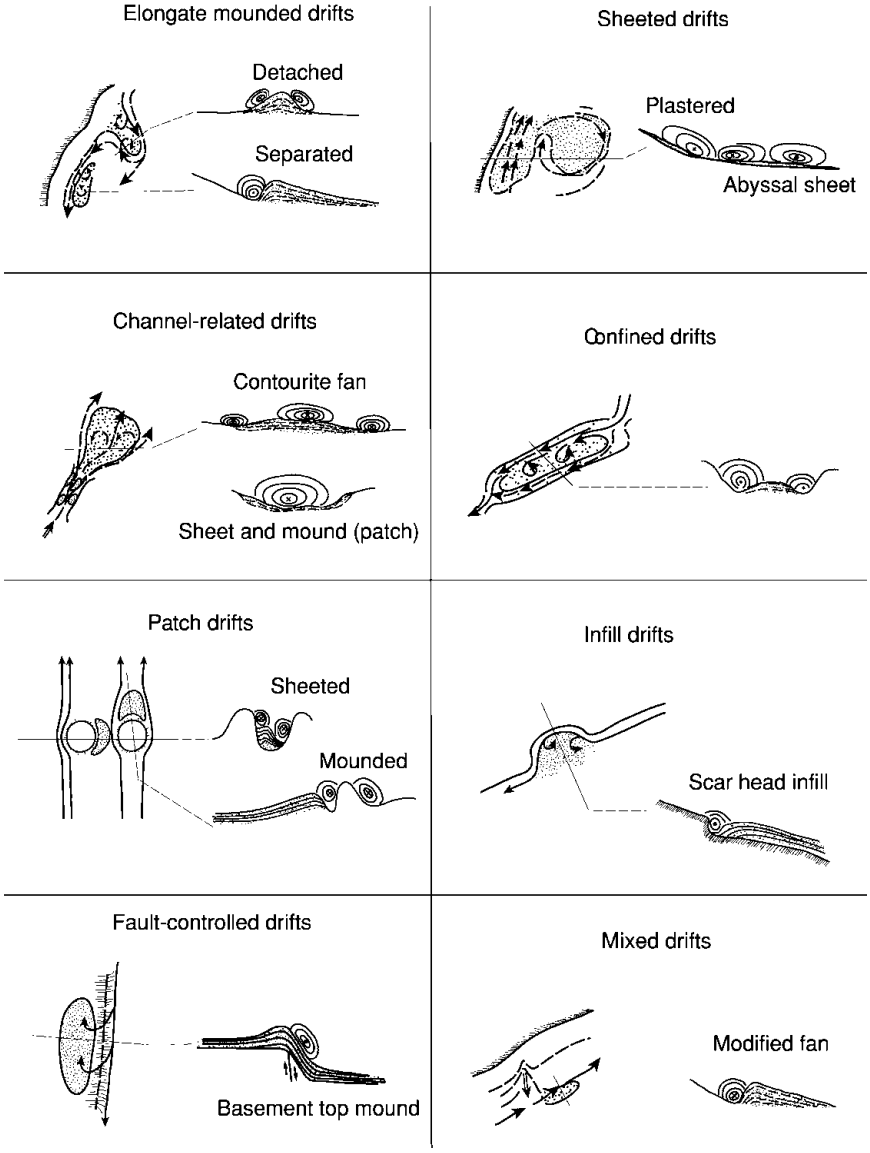


Figure 4 Sediment drift types and inferred bottom current paths. Adapted from Rebesco M and Stow DAV (2001) Seismic expression of contourites and related deposits: A preface, In: Rebesco M and Stow DAV (eds.) Seismic expression of contourites and related deposits, *Marine Geophysical Researches* 22: 303–308, with kind permission of Kluwer Academic Publishers; and: Stow DAV, Faugères JC, Howe J, Pudsey CJ, and Viana A (2002) Bottom currents, contourites and deep sea sediment drifts: Current state-of-the-art, in: Stow DAV, Faugères JC, Howe J, Pudsey CJ, and Viana A (eds.) Deep-Water Contourite Systems: Modern Drifts and Ancient Series, Seismic and Sedimentary Characteristics, *Geological Society, London, Memoir* 22: 7–20, with permission.

b. A 'detached drift' is removed (detached) from that part of the continental slope where it originally enucleated and has oppositely directed currents flowing along its two flanks. It is generated by the interaction of the bottom current with a distinct current system or with a prominent change in the slope orientation, and as a consequence in plain view it generally has an arched shape produced by the dominant downslope progradation.

3. A 'channel-related drifts' is characterized by its relationships with narrow conduits (deep channels, gateways, trenches, moats, etc.) where bottom currents are constricted and drastically accelerated. In addition to significant erosion/nondeposition, two quite distinct types of channel-related drift may develop:
   a. A drift deposited within the conduits, either as axial mounds on the floor or as lateral sheets on the flanks of the channel, is known as a 'patch drift' (though this name may cause confusion with patch drifts unrelated to channels) and 'subsidiary drift', where it occurs within the moat, separating a giant elongated drift

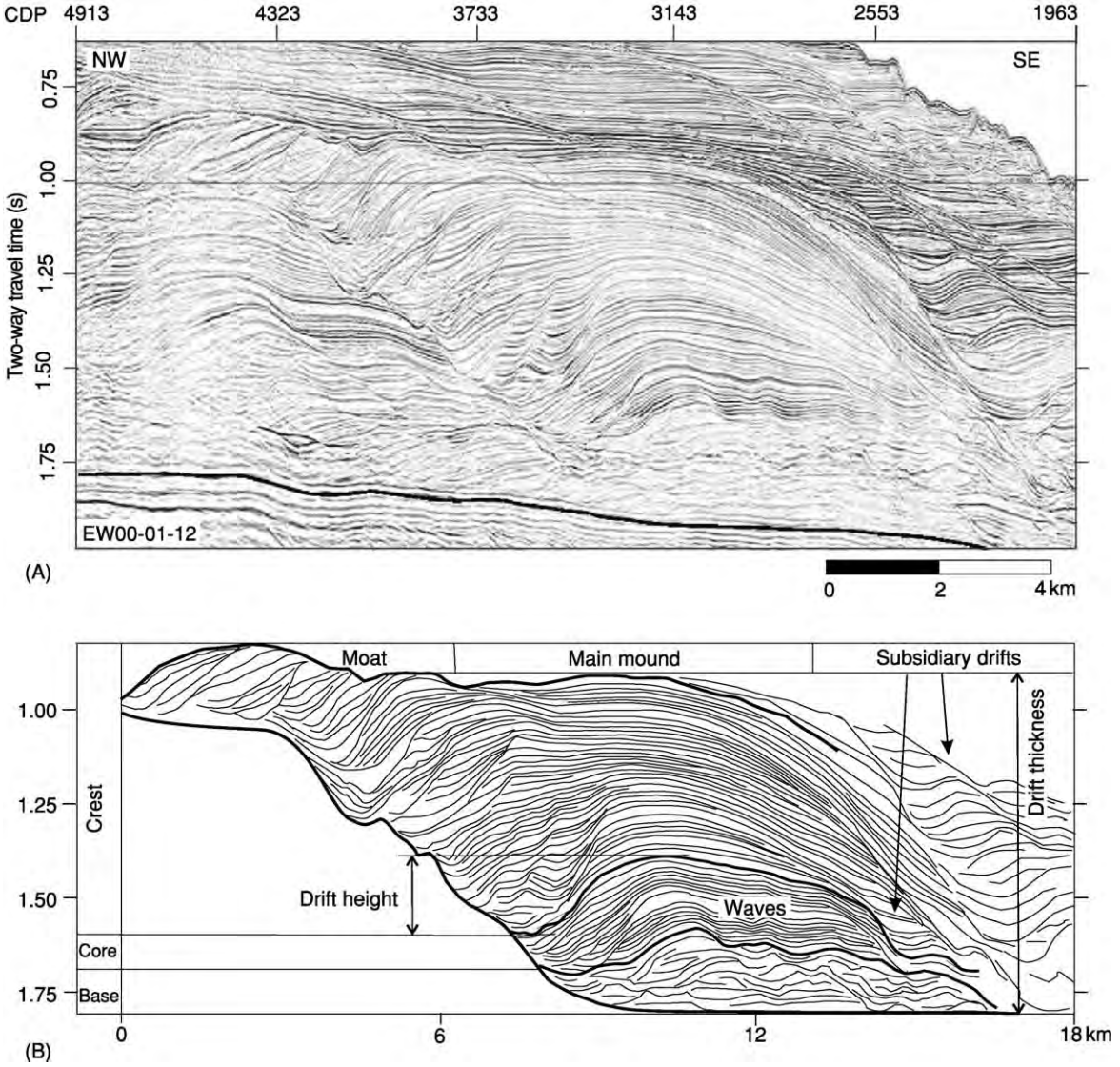


Figure 5 Seismic profile (A) and interpretation (B) showing an elongated, mounded drift in the north-eastern Canterbury Bight. Adapted from Lu H, Fulthorpe CS, and Mann P (2003) Three-dimensional architecture of shelf-building sediment drifts in the offshore Canterbury Basin, New Zealand, *Marine Geology* 193: 19–47, with permission.

from the continental slope. This irregular, discontinuous channel-related drift generally elongates in the flow direction.

b. A 'contourite fan' is a much larger, cone-shaped deposit developed at the downcurrent exit of the conduits. It usually overlays major basal erosive disconformities, but for the most part shares many characteristics with medium-sized turbidite fans.

4. A 'patch drift' is characterized by a random (patchy) distribution controlled by the intricate interaction of the bottom current system with a complex seafloor morphology. Such a small-scale (a few tens of square kilometres), elongated, irregular drift may be either relatively mounded or thinly sheeted. It occurs plastered against reliefs or within a trough where the irregular topography may modify both direction and velocity of the local current flow.
5. A 'confined drift' is characterized by a mounded shape elongated parallel to the basin axis and distinct moats along both flanks. It occurs within relatively small enclosed basins, eventually actively subsiding, and generally shows a complex staking of upward-convex lenticular depositional units.
6. An 'infill drift' is characterized by a mounded geometry that progressively infills the topographic depression or irregularity in which it occurs. It typically forms at the head of the scar or at the

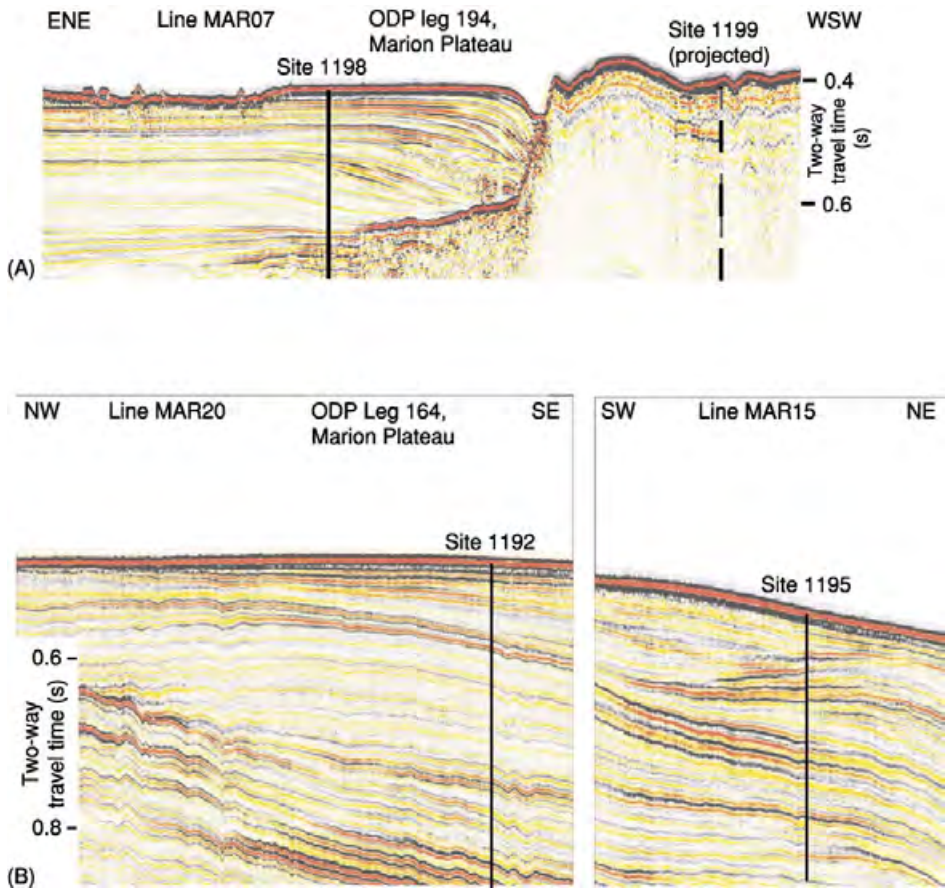


Figure 6 Seismic profiles showing a separated drift close to the Marion Platform margin (A) and a mounded drift complex, with location of some ODP leg 194 sites. Adapted from Shipboard Scientific Party (2002) Leg 194 summary, In: Isern AR, Anselmetti FS, Blum P, *et al.* (eds.), *Proc. ODP, Initial Reports*, Vol. 194. Available from: http://www-odp.tamu.edu/publications/194_IR/chap_01/chap_01.htm.

margin of the toe of slumps developed beneath the path of a bottom current, and generally shows moderate relief and extent, variable shape, and downcurrent progradation.

7. A 'fault-controlled drift' is characterized by a certain influence in its development from faulting. This recent addition lacks many well-documented examples and is too little known to allow any generalization regarding deposit geometry and bottom current nature. However, it appears to develop either at the base or at the top of a fault-generated basement relief in response to perturbations in the bottom current flow pattern. A supplementary characteristic may be (subsequently reactivated) syndepositional faulting that affects the relatively steeper side of such drifts.
8. 'Mixed drift systems' are characterized by the interactions of alongslope currents with other depositional processes. The most effective interplay is between contourites and turbidites, but drift development may be variously affected by the association with debrite, hemipelagite, and glacigenic systems. On the basis of a large variety of different types of interactions (interfingering, intercalation, imbrication, incorporation, winnowing, entrainment, etc.), a wide assortment of terms has been adopted, including: companion drift-fan, fan-drift, levee-drift, composite slope-front fan, etc.

Seismic Characteristics

Although much progress has been made in determining the combination of criteria that best represent contourite deposits, seismic data without additional supporting evidence should not be used to make a firm identification of contourites. This is especially true when contourites are closely interbedded with other deep-water sediment facies (e.g., turbidites, hemipelagites, debrites, or glacigenic deposits). In addition, the vital information derived from detailed bathymetric images (either by multibeam or 3D

seismic surveys) of contourite deposits is still relatively rare (Figure 7). Nevertheless, there is a wide consensus on a three-scale set of seismic diagnostic criteria for confident contourite identification.

Large Scale (i.e., Sediment Body)

1. Drift geometry (one of those described in the previous section) beneath a bottom current system;
2. Large dimensions (up to $1 \times 10^6\,km^2$), larger on average than those of turbidite channel–levee systems;
3. Asymmetric mound-and-moat geometry in contrast to the symmetric gull-wing geometry typical of turbidite channel–levee systems (Figure 8);
4. Overall downcurrent elongation; and
5. Widespread regional discontinuities, especially at the base of the drift.

Medium Scale (i.e., Unit)

1. Overall downcurrent migration of the stacked lenticular, upward-convex, seismic units;

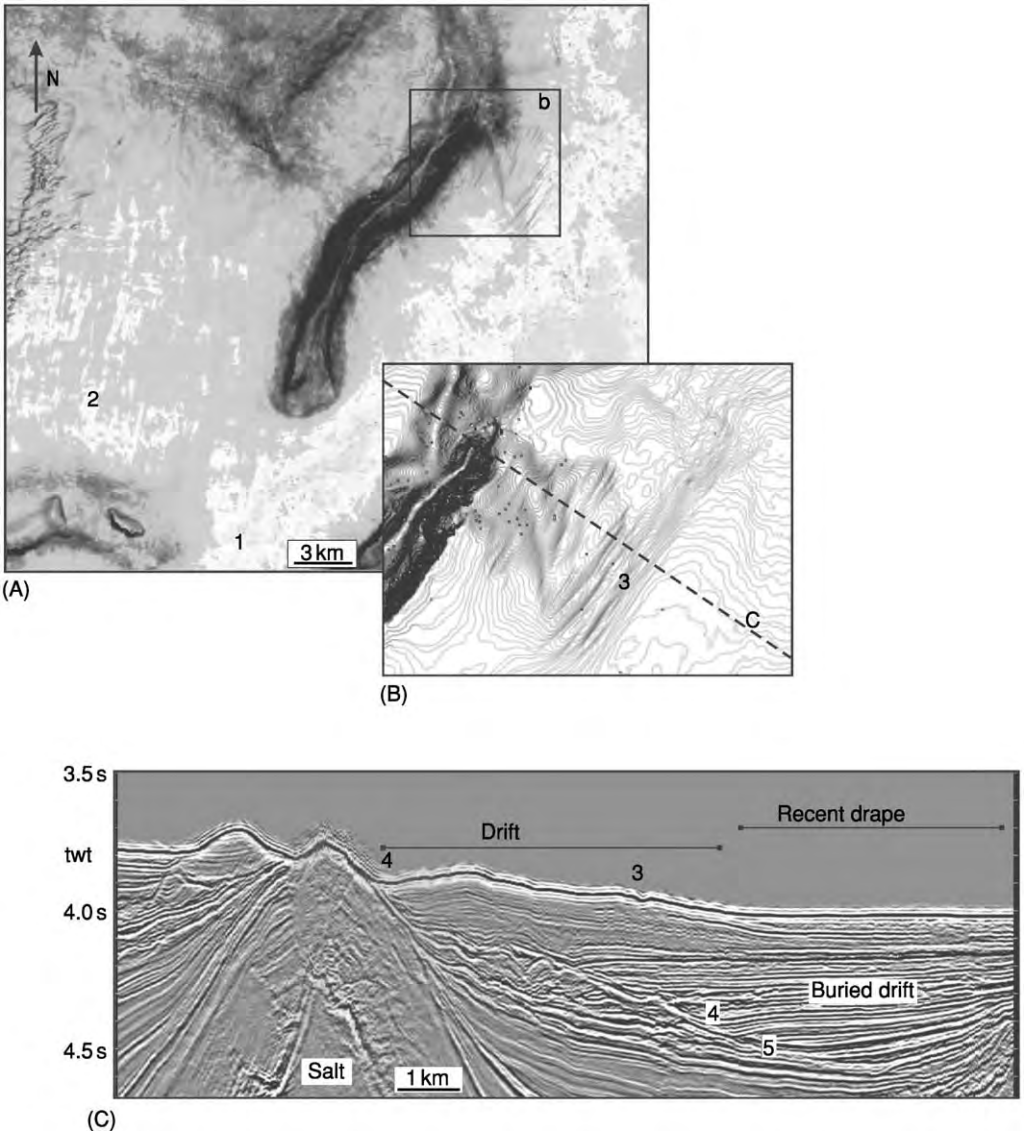


Figure 7 Grey-scale seafloor map derived from 3D seismic amplitude (A), detail of the contour map (B), and seismic line (C) showing a mounded drift and bottom current-related deposits adjacent to a NE–SW trending topographic high. (1) Sheet-like coarse-grained deposit; (2) nonconfined, string-like deposit; (3) erosional furrows; (4) moat between upslope-migrating separated drift and topographic high; (5) basal erosional discontinuity. Reprinted from Viana A (2001) Seismic expression of shallow- to deep-water contourites along the south-eastern Brazilian margin, In: Rebesco M and Stow DAV (eds.) Seismic expression of contourites and related deposits, *Marine Geophysical Researches* 22: 509–521, with kind permission of Kluwer Academic Publishers.

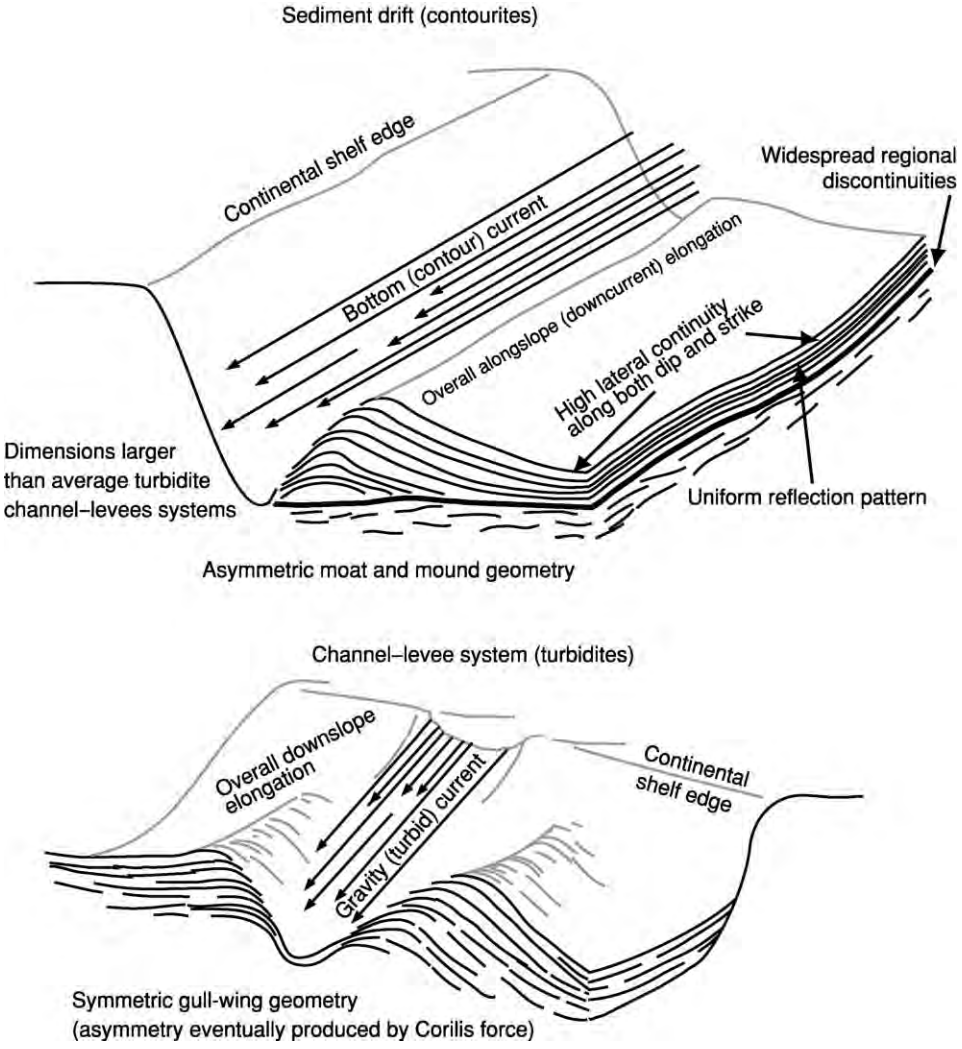


Figure 8 Schematic model showing some typical seismic characteristics of the sediment drifts and the ideal difference with a channel–levee system.

2. Downlapping (onlapping on steep slopes) and sigmoidal progradational reflector patterns where downstream and upslope migration occurred;
3. Uniform reflection pattern, commonly found with extensive subparallel moderate to low amplitude reflectors;
4. Well-layered units with high lateral continuity along both strike and dip; and
5. Distribution of the depositional/erosional areas and lateral migration influenced (significantly at high latitudes) by the Coriolis force.

Small Scale (i.e., Facies)

1. Transparent layers and low-amplitude parallel reflectors, though seismic characteristics are very dependent on acquisition and processing methods;
2. Continuous, discontinuous, chaotic, and wavy reflectors in contourites and in turbidites, though seismic facies exclusively associated with contourites have not been defined yet;
3. Sediment waves, in both contourite and turbidite systems, though initially considered diagnostic of contourites; and
4. Bedforms, including longitudinal furrows, depositional tails in the lee of obstacles, and dunes.

Facies Model

Most typically, contourites are composed of fine-grained, structureless, highly bioturbated mud. However, they show a wide range of grain-sizes, composition, and preserved sedimentary structures (Figures 9 and 10).

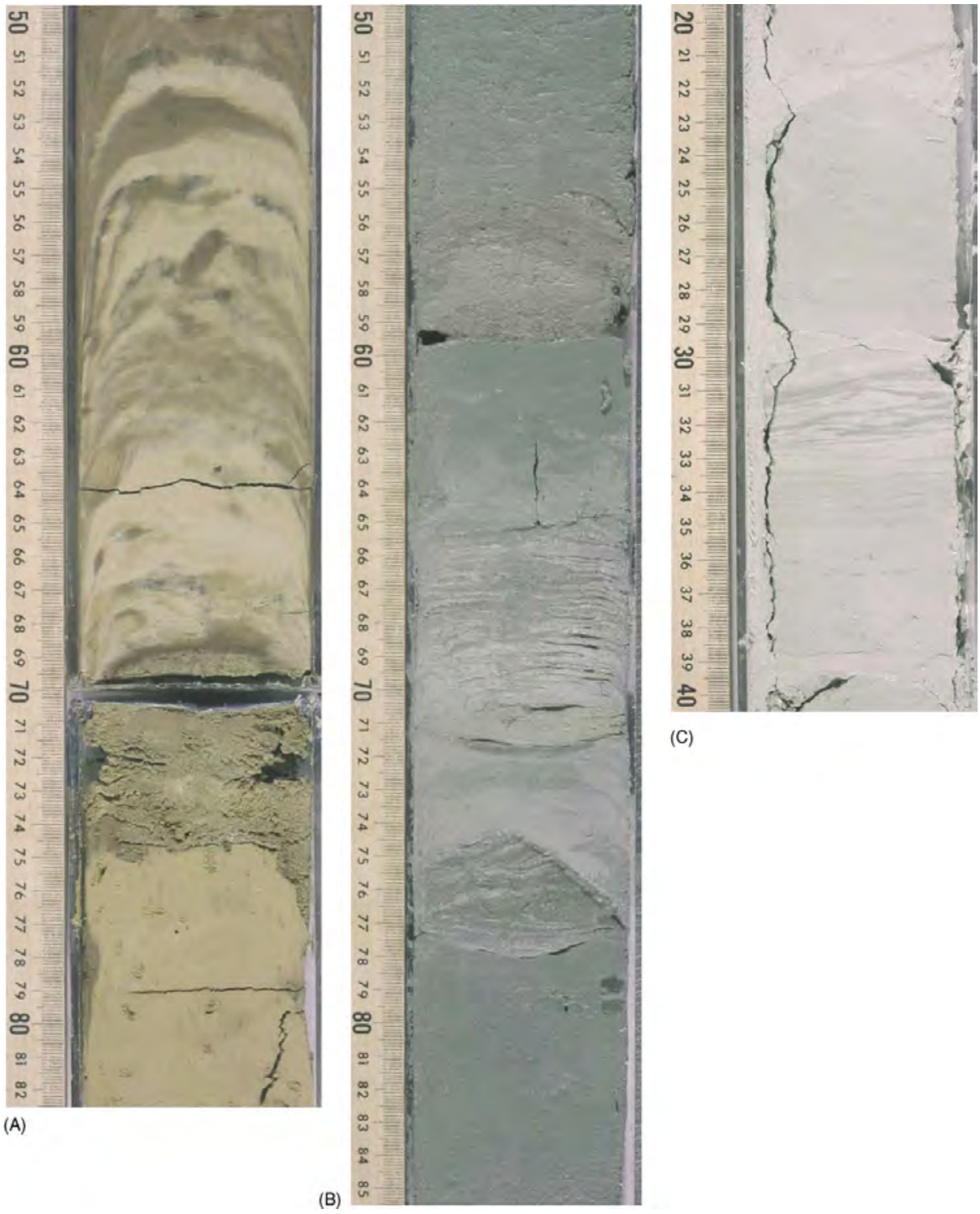


Figure 9 Core photos from ODP Leg 181 in the SW Pacific Gateway. From Carter RM, McCave IN, Richter C, *et al.* (eds.) (1999) *Proc. ODP, Initial Reports*, Vol. 181. Available from: http://www-odp.tamu.edu/publications/181_IR/181ir.htm. (A) Core 181-1121B-1H-CC, 50–82 cm from the Campbell Drift showing brownish yellow sand with extremely sharp bottom contact (75 cm) with the underlying yellow clay. This sand bed exhibits no grading and is interpreted as a contourite deposite left behind after that an intense winnowing removed the fine material. (B) Core181-1122C-44X-2, 50–85 cm in the left (north) bank levee of the abyssal Bounty Fan showing greenish grey fine sand and silt beds that commonly exhibit sharp scoured top and bottom contacts with the interbedded mottled, dark greenish grey pelagic/hemipelagic bioturbated silty clay. The conspicuous planar and cross laminations, representing concentrations of foraminifera and carbonate debris, suggest a stronger, episodic benthic flow regime, in contrast to that of the decelerating turbidity currents. (C) Core 181-1124C-37X-2, 20–40 cm from the Rekohu Drift showing light greenish grey clay-bearing nannofossil chalk with flaser-like interbeds and laminae suggesting the presence of sediment-moving bottom currents.

They normally occur as thick (tens to hundreds of metres), uniform, fine-grained sequences (including thin to medium coarser-grained beds), or interbedded with hemipelagites, turbidites, and other resedimented facies, or as coarse lags within gateways. They are dominantly homogeneous (Figure 9A), poorly bedded (Figure 10A), and mostly bioturbated (Figure 10B) and mottled throughout, with little primary structures preserved. Bioturbation is generally considered an essential characteristic of contourites based on the belief that an active bottom current would increase oxygen concentration in the water and in turn increase organic activity, thus allowing contourites to be distinguished from episodic turbidites. However, bioturbation potential is also inferred to be a function of current intensity and other life-favourable environmental conditions. As a matter of fact, nonbioturbated contourites have been described as well (Figure 10C). Well-developed but somewhat irregular, fine lamination with indistinct bioturbation may be evident both in shallow-water and in high-latitude facies (Figure 10C), possibly as a result of a hybrid turbidite–contourite deposition. Although cyclicity is common, primary silty parallel and cross lamination (Figures 9B, 10C), where present, shows no regular sequence of facies as in turbidites. Contourites may show reverse grading, with coarse lag concentrations and sharp or erosive contacts (Figures 9A and

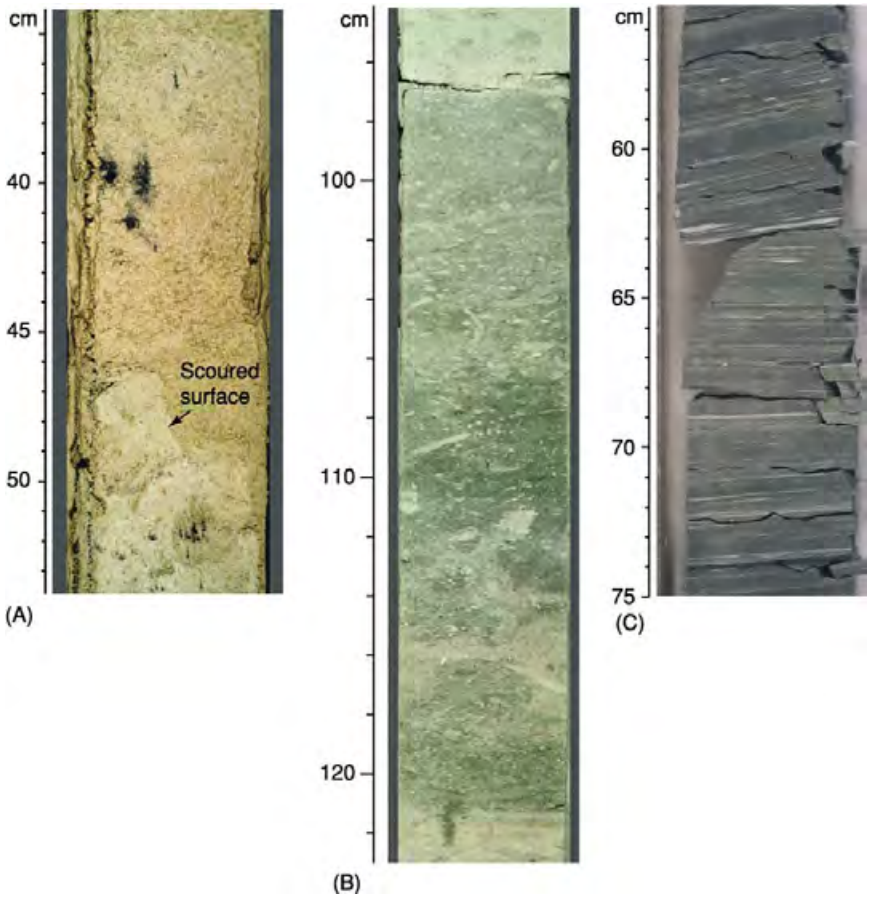


Figure 10 Core photos from two different ODP Legs. (A, B) ODP Leg 194 in the sediment drifts adjacent to Marion Plateau. From Isern AR, Anselmetti FS, Blum P, *et al.* (2002) *Proc. ODP, Initial Reports*, Vol. 194. Available from: http://www-odp.tamu.edu/publications/194_IR/194ir.htm. (C) ODP Leg 188, Wild Drift in Cooperation Sea. From Shipboard Scientific Party (2001) Site 1165, in: O'Brien PE, Cooper AK, Richter C, et al. (eds) (2001) *Proc. ODP, Initial Reports*, Vol. 188. Available from: http://www-odp.tamu.edu/publications/188_IR/VOLUME/CHAPTERS/IR188_03.PDF. (A) Core 194-1195B-21H-5, 34–54 cm showing a surface scoured into a light greenish grey wackestone by strong bottom currents in a high-energy hemipelagic setting. The sediments above the scoured surface consist of alternations between light grey silt- and very fine sand-sized wackestone and very fine to fine sand-sized skeletal packstone dominated by broken planktonic foraminifera. (B) Core 194-1195B-44X-1, 94–122 cm showing alternation of silt-sized light grey skeletal packstone and greenish gray mudstone with well-preserved *Chondrites* burrows and top and basal sharp scoured surfaces. (C) Core 188-1165C-3R-3, 55–75 cm showing a dark grey claystone with $<$1-cm beds and silt partings along bedding planes interpreted as contourites deposited by relatively slow-moving, near-continuous bottom currents entraining episodic turbidity currents, whereas bioturbation occurring in discrete decimetre-sized intervals reflect comparatively low deposition rates.

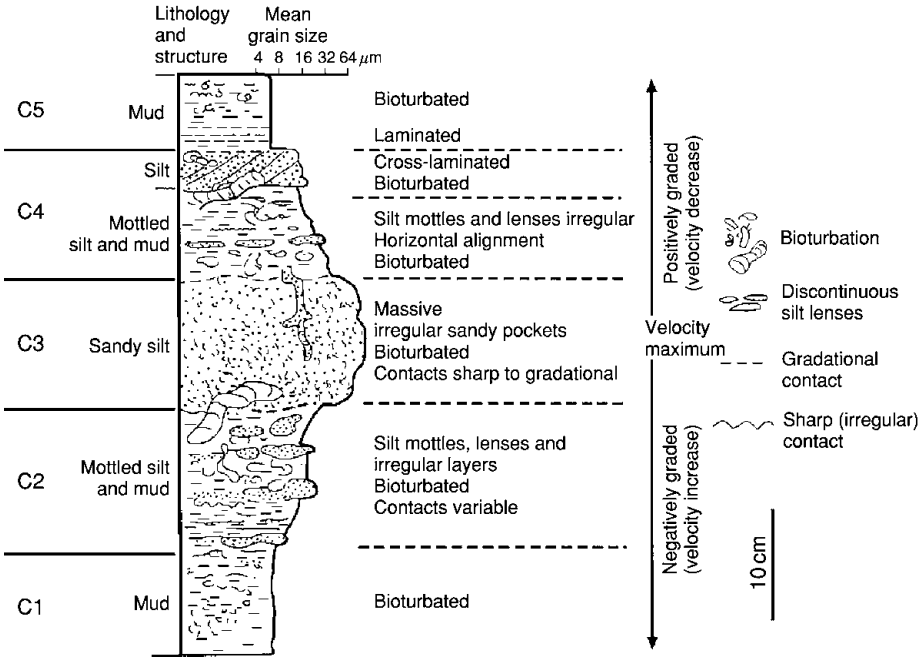


Figure 11 Conjectural standard composite (muddy + sandy) contourite facies model. Adapted with permission from Stow DAV, Faugères JC, Howe J, Pudsey CJ, and Viana A (2002) Bottom currents, contourites and deep sea sediment drifts: Current state-of-the-art, In: Stow DAV, Faugères JC, Howe J, Pudsey CJ, and Viana A (eds.) Deep-Water Contourite Systems: Modern Drifts and Ancient Series, Seismic and Sedimentary Characteristics. *Geological Society, London, Memoir* 22: 7–20, with permission.

9B, 10A and 10B). Grain size may vary from (silty) mud to sand. Gravel-rich contourites are common in glacigenic environments due to the presence of ice-rafted debris. Bottom current winnowing and erosion in narrow gateways may also produce gravel lags or shale-chip layers. Sorting is generally medium to poor, but some sandy contourites may be well sorted and relatively free of mud, showing low or negative skewness values. No offshore textural trend may be observed. Composition is generally mixed with a combination of biogenic and terrigenous material. High sand-sized content in muddy hemipelagic contourites is often formed by bioclasts (Figures 9B, 10B). The terrigenous component largely reflects the nearby local sources, but alongslope mixing and far-off contributions also occur. In contrast the composition is essentially pelagic in biogenic contourites developed in open oceans or in upwelling areas. Manganese dioxides or iron oxides may occur as micronodules, stains, coating, and metal-enriched crusts within the manganiferous contourites. Fabric (when not exceedingly altered by bioturbation) is indicative of alongslope grain orientation, parallel to the bottom current flow. Contourites (Figure 11) are normally arranged in few decimetre-thick sequences showing gradual grain size and/or compositional changes. Five divisions may be eventually identified: (C1) lower mud, (C2) lower mottled silt, (C3) middle sand, (C4) upper mottled silt, and (C5) upper mud. A complete C1–5 sequence is interpreted to represent a gradual long-term current velocity variation and/or sediment supply change. Partial base-only or top-only sequences are also common.

Facies Continuum and Distinguishing Criteria

Deep-sea sediments may be interpreted as the product of three main processes: gravity-driven downslope resedimentation, alongslope bottom current activity, and slow pelagic settling through the water column. However, such processes are in fact end members belonging to a continuum spectrum. Deep-sea sedimentary facies are often hybrids resulting from interactions within such process continuum (Figure 1) rather than a mere stacking of interstratified deposits produced by discrete processes. Imaginably, the distal part of very-low-concentration (0.025 to $2.5\,g\,l^{-1}$), low-velocity ($10–50\,cm\,s^{-1}$), fine-grained (silt and clay) turbidity currents may be easily deflected by Coriolis force from its natural downslope path to an alongslope direction and then entrained in the indigenous contour current. Concurrently, the tests of calcareous and siliceous planktonic organisms and the associated organic matter biosynthesized in the surface layers of the ocean may be significantly contributed by slow vertical

settling through the water column. The current originally driven by the gravitational effect of its terrigenous sedimentary load would hence imperceptibly grade into a thermohaline- or wind-driven, contour-parallel bottom current capable of depositing a hemipelagic, 'resedimented' contourite.

Although the traction effects of the bottom currents within an otherwise tranquil pelagic realm are fairly visible and may easily lead to the interpretation of the resulting deposits as contourites, a much more difficult and disputable work is required in trying to distinguish contourites from turbidites. Different sets of criteria have been proposed to characterize bottom current-reworked sandy sediments that occur where strong, permanent bottom currents have been active within a turbidite setting. In general, reworked turbidites are constituted by predominantly fine-grained sand and silt and are specially characterized by thin-bedded to laminated sand (usually less than 5 cm) in deep-water mud. Rhythmic occurrences of sand and mud layers are common (Figure 10C), with numerous sand layers (50 or more per 1 m of core). In many cases, the lower divisions of turbidites may be preserved and the upper ones entirely removed or partially modified by bottom current reworking. Reworked tops may show common bioturbation and burrowing, reverse grading, and lag concentrations. Horizontal and bidirectional low-angle cross-laminations and sharp erosive contacts may occur within the turbidite sequence. Cross-bedding is believed to be produced by steady and equlibrium flow conditions due to the bottom current action and, in contrast, is, envisaged with difficulty in episodic turbidity current events (Figure 9B). With respect to the adjoining turbidites, they may be cleaner, better sorted, more negatively skewed, and with a more bimodal fabric. In addition, the following diagnostic sedimentary structures have been proposed, although not universally agreed upon: lenticular bedding and starved ripples; mud offshoots; flaser bedding (Figure 9C); and the occurrence of sand layers with traction structures in discrete units, but not as part of a vertical sequence of structures, such as the Bouma sequence with basal graded division.

See Also

Petroleum Geology: The Petroleum System. **Sedimentary Environments:** Depositional Systems and Facies. **Sedimentary Processes:** Erosional Sedimentary Structures; Depositional Sedimentary Structures; Deposition from Suspension. **Sedimentary Rocks:** Deep Ocean Pelagic Oozes. **Seismic Surveys**. **Unconformities**. **Unidirectional Aqueous Flow**.

Further Reading

Faugères JC, Stow DAV, Imbert P, and Viana A (1999) Seismic features diagnostic of contourite drifts. *Marine Geology* 162: 1–38.

Gao ZZ, Eriksson KA, He YB, Luo SS, and Guo JH (eds.) (1998) *Deep-Water Traction Current Deposits*. Beijing/New York: Science Press.

Lu H, Fulthorpe CS, and Mann P (2003) Three-dimensional architecture of shelf-building sediment drifts in the offshore Canterbury Basin, New Zealand. *Marine Geology* 193: 19–47.

Maldonado A and Nelson HC (eds.) (1999) Marine geology of the Gulf of Cadiz. *Marine Geology* 155.

McCave IN and Tucholke BE (1986) Deep current-controlled sedimentation in the western North Atlantic. In: Vogt P and Tucholke B (eds.) *The Geology of North America, Vol. M, The Western North Atlantic Region,* pp. 451–468. Boulder, CO: Geological Society of America.

Mienert J (ed.) (1998) European North Atlantic margin (ENAM): Sediment pathways, processes and Flux. *Marine Geology* 152.

Nowell ARM and Hollister CD (eds.) (1985) Deep ocean sediment transport—Preliminary results of the high energy Benthic boundary layer experiment. *Marine Geology* 66.

Pickering K, Hiscott R, and Hein F (1989) *Deep Marine Environments: Clastic Sedimentation and Tectonics*, pp. 218–245. London: Unwin Hyman.

Rebesco M and Stow DAV (eds.) (2001) Seismic expression of contourites and related deposits. *Marine Geophysical Researches* 22(5–6).

Shanmugam G (2000) 50 years of the turbidite paradigm (1950s–1990s): Deep-water processes and facies models—A critical perspective. *Marine and Petroleum Geology* 17: 285–342.

Steele JH, Turekian KK, and Thorpe SA (eds.) *Encyclopedia of Ocean Sciences*. London: Academic Press.

Stoker MS, Evans D, and Cramp A (eds.) (1998) *Geological Processes on Continental Margins: Sedimentation, Mass-Wasting and Stability.* Special Publication 129. London: Geological Society.

Stow DAV and Faugères JC (eds.) (1993) Contourites and bottom currents. *Sedimentary Geology* 82(1–4).

Stow DAV and Faugères JC (eds.) (1998) Contourites, turbidites and process interaction. *Sedimentary Geology* 115(1–4).

Stow DAV and Mayall M (eds.) (2000) Deep-water sedimentary systems: New models for the 21st century. *Marine and Petroleum Geology* 17.

Stow DAV, Faugères JC, Howe J, Pudsey CJ, and Viana A (eds.) (2002) Deep-water contourite systems: Modern drifts and ancient series, seismic and sedimentary characteristics. *Geological Society, London, Memoir* 22.

Wynn RB and Stow DAV (eds.) (2002) Recognition and interpretation of deep-water sediment waves. *Marine Geology* 192(1–3).

Deltas

T Elliott, University of Liverpool, Liverpool, UK

© 2005, Published by Elsevier Ltd.

Introduction

Deltas form where sediment-laden rivers enter oceans, enclosed seas, lakes, or lagoons and supply more sediment than can be distributed by processes that operate within the receiving basin. They produce a protrusion of the shoreline and typically advance, or prograde, into the basin. Major deltas such as the Amazon, Ganges-Brahmaputra, Nile, and Mississippi are served by well-defined, often long-lived, continent-scale drainage basins that result in an alluvial trunk stream (Figure 1). Deltas are important sites of sediment deposition in their own right and they also create basin-scale depocentres, in that deltas in turn supply sediment to shelves, slopes, and deep-basin submarine fans. It is no accident that many present-day submarine fans occur basinwards of major deltas.

The social, cultural, and economic importance of deltas is considerable. They account for a significant proportion of the world's wetlands and are important in terms of habitation, agriculture, fishing, and navigation. They are extraordinarily sensitive to fluctuations in sea-level, whether caused by subsidence or eustatic sea-level change, and therefore feature prominently in studies of land loss, water management, and coastal stability. For example, in the United States, wetlands in Louisiana have been experiencing compaction-induced rates of sea-level rise in excess of $1\,\mathrm{cm\,year^{-1}}$ for several decades, causing land loss rates to exceed $100\,\mathrm{km^2\,year^{-1}}$. Current measurements and predictions of the rate at which eustatic sea-level is presently rising will have a major impact not only in Louisiana, but also on deltas globally. In view of their links to continent-scale drainage basins, deltas are also important when considering pollutant dispersal to the margins of oceans. Finally, the deposits of former deltas host significant fossil fuel reserves, providing both source and reservoir rocks in many oil- and gas-bearing basins.

Controls on the Variability of Deltas

Deltas are extremely diverse in appearance and behaviour. The formation and ultimate character of a delta are determined by variables that stem from the hinterland that is responsible for creating the sediment supply and from the fluvial system that serves the delta, and from the basin that receives the sediment. Major controls, such as climate and tectonics, act on both the hinterland and the receiving basin, and sea-level is critical at the interface between them. The means by which the variability of deltas is presently summarized involves a classification scheme based on the sediment calibre delivered to the delta and the interaction between fluvial, wave, and tidal processes that takes place in the shoreline or delta front sector (Figure 2). This allows interpretation of the Mississippi delta as a fine-grained, fluvial-dominated delta and the Ganges-Brahmaputra delta as a more sand-rich, tide-dominated delta. Despite their complexity and variability, deltas comprise two major components: the delta front, which includes the river mouth, adjacent shoreline, and seaward-dipping profile, and the delta plain, which is the low-lying, largely subareal part of the delta behind the delta front; the delta plain is dominated by

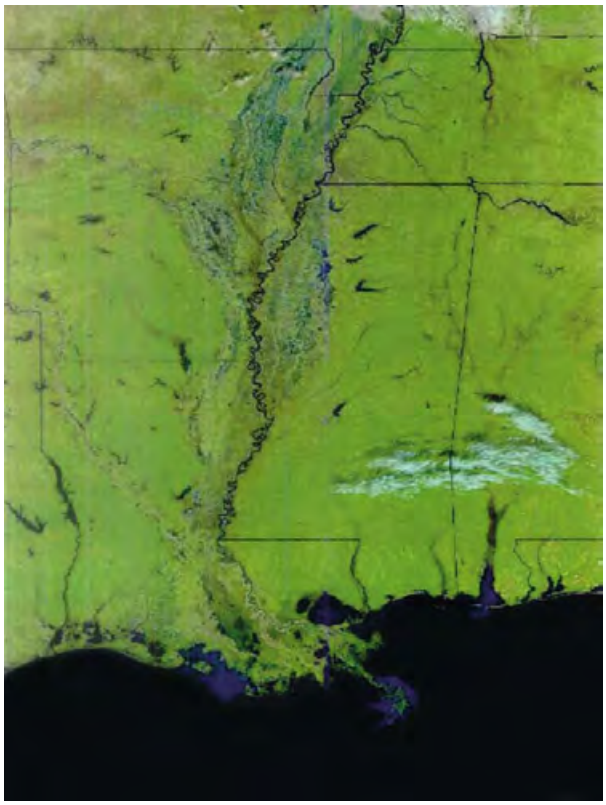


Figure 1 The Mississippi River and delta complex; the Mississippi River drainage basin extends over two-thirds of the United States and is the alluvial trunk stream of the delta. The modern birdsfoot delta is flanked by the now abandoned, premodern shoal water deltas. The light blue sediment-rich embayment west of the modern delta is where the new Atchafalaya and Wax river deltas are located.

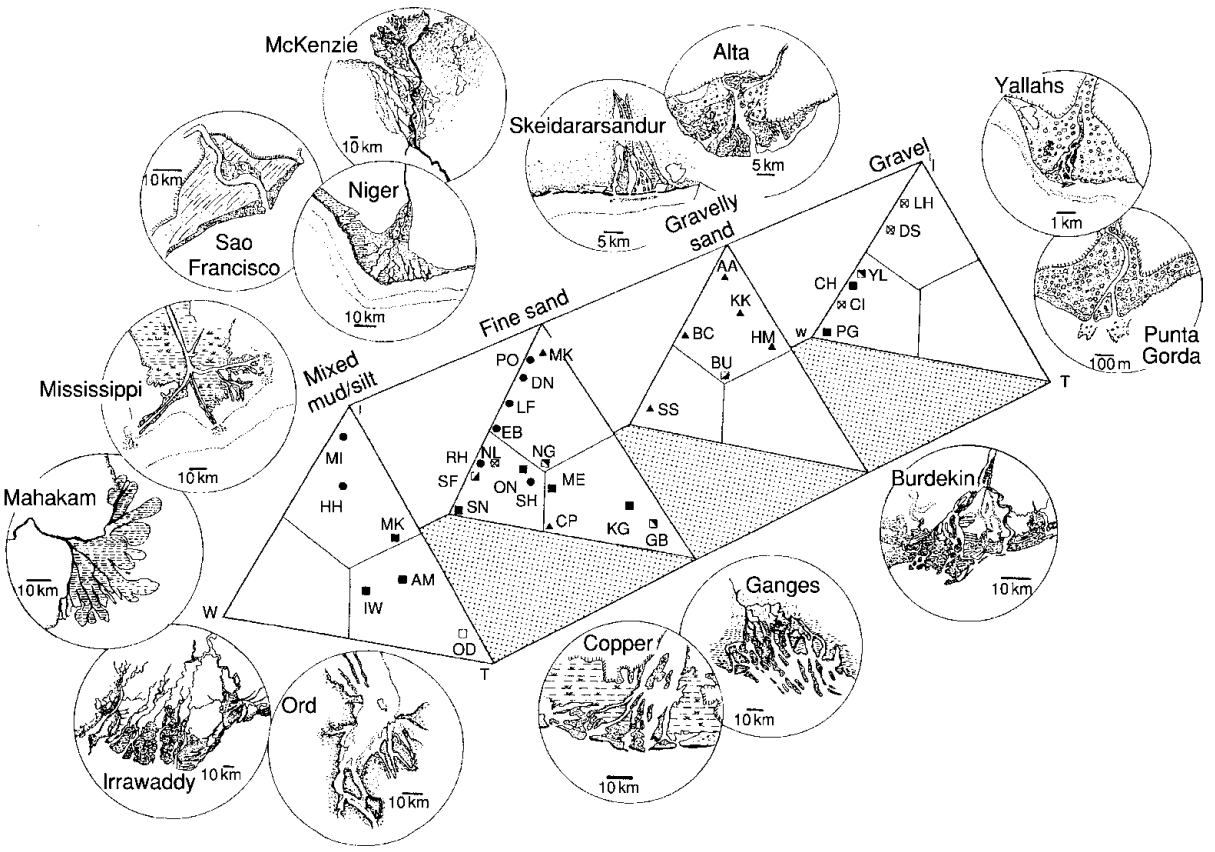


Figure 2 A classification scheme for deltas based on sediment calibre and the process regime of the delta front. In each triangular cross-section, the top portion represents an input-dominated (I) process, the left bottom portion represents a wave-dominated (W) process, and the right bottom portion represents a tide-dominated (T) process. Reproduced with permission from Orton GJ and Reading HG (1993) Variability of deltaic processes in terms of sediment supply, with particular emphasis on grain size. *Sedimentology* 40: 475–512.

distributary and crevasse channels, levees, bays, lakes, and tidal flats in tide-influenced deltas.

River Mouth Processes and Plumes in Deltas

The most important subenvironment of deltas is the mouth of the river, or the distributaries of the river, because this is where the interaction takes place between the sediment-laden river waters derived from the hinterland and the receiving basin. When river discharge reaches the mouth of a channel, it decelerates due to loss of gradient, expands due to loss of confinement by the channel margins, is required to mix with the waters of the receiving basin, and comes under the influence of basinal processes such as waves and tides. The outcome of this complex interaction will determine, first, whether a delta forms, and, if so, the characteristics the delta. The first-order control is the relative density of the incoming sediment-laden river waters and the waters of the receiving basin (Figure 3). Where the water masses are equal density (homopycnal flow), immediate mixing occurs at the river mouth and the sediment load is deposited locally at the river mouth. This can lead to the formation of localized, steep-fronted deltas referred to as 'Gilbert-type' deltas, which form commonly where bedload-rich rivers enter lakes. They can also form where gravel-grade rivers or alluvial fans enter the margins of marine basins, creating fan deltas. In this case, deposition results from the loss of competence to transport the gravel-grade material beyond the river mouth, rather than the density contrast of the water masses.

Where sediment-laden river waters are less dense than the basin waters (hypopycnal flow), then the outflow from the river is buoyantly supported and extends into the basin as a surface jet or plume (Figure 3). This process operates in major marine deltas because salt water is denser than fresh water is, and suspended sediment concentrations in river waters are typically insufficient to overcome this difference. Satellite images of major deltas invariably show murky, sediment-laden

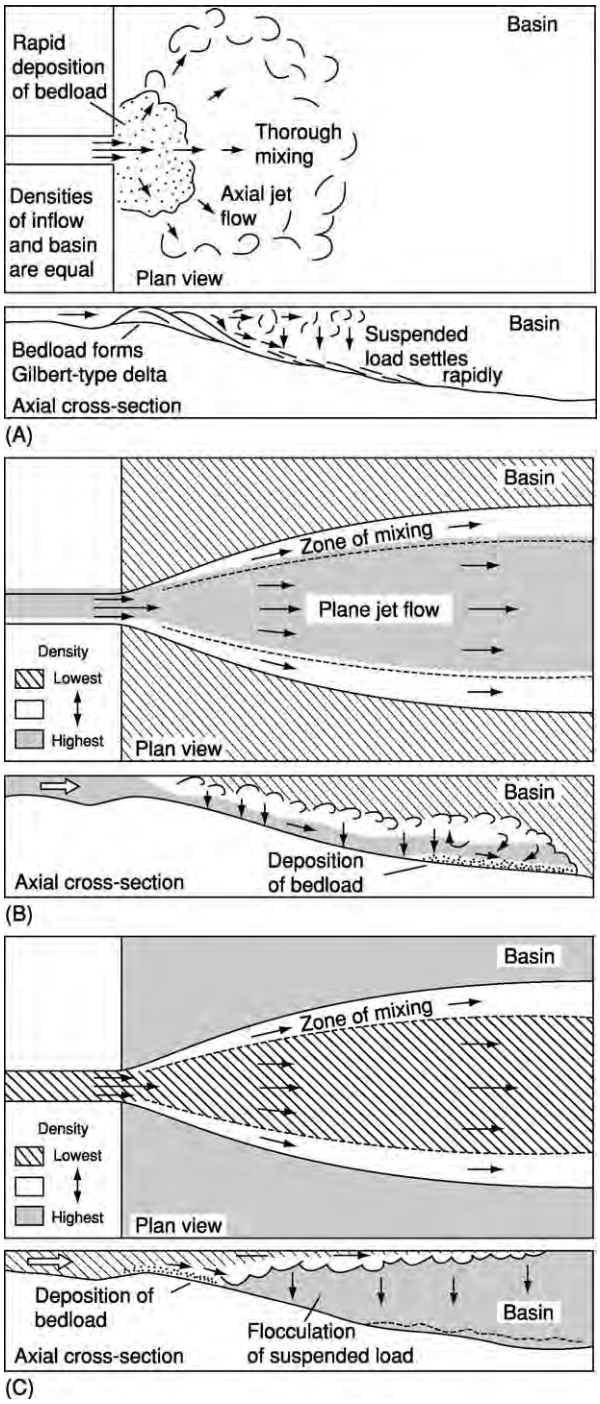


Figure 3 Fundamental processes that operate at the mouths of rivers or distributary channels based on the relative densities of the sediment-laden fluvial discharge and the water of the receiving basin, distinguishing homopycnal (A), hypopycnal (B), and hyperpycnal (C) flow. Reproduced with permission from Bates (1953).

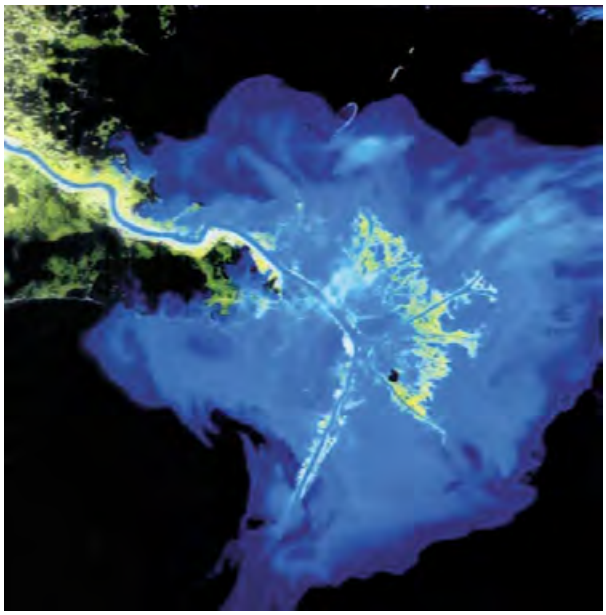


Figure 4 Plumes of sediment-laden water (light blue) issuing from the Mississippi delta during a high-discharge, river flood period. The plumes are buoyantly supported because the saline basin water is denser than the river water, despite the presence of suspended sediment in the latter.

plumes issuing from river and distributary mouths across the darker, sediment-free waters of the basin (Figure 4). In this situation, relatively coarse, sand-grade sediment is often deposited close to the channel mouth, either as bedload or from suspension, whereas finer material is transported further offshore in suspension. Turbulent mixing gradually reduces the density contrast between the water masses, causing the suspended load of the plume to be deposited with a trend of decreasing grain size offshore.

Where the sediment-laden river waters are more dense than the receiving basin (hyperpycnal flow), an underflow occurs, causing the sediment-laden flow to move discretely beneath the basin waters down the gradient of the basin margin (Figure 3). The extent to which this process operates in deltas is debatable. Clearly, if it operated exclusively, then a delta would not form, because all of the sediment load of the river would bypass the shoreline to be deposited subaqueously. However, the process may operate intermittently during major river floods. In a freshwater lake, suspended sediment concentrations of $\sim 1\,\mathrm{kg\,m^{-3}}$ can produce a hyperpycnal flow, but in marine settings, concentrations of 35–45 $\mathrm{kg\,m^{-3}}$ are required due to the greater density of salt water. The Huanghe River delta (China) regularly exceeds the critical value and produces hyperpycnal flows. Predictive modelling indicates that numerous small rivers entering marine basins also have the potential to produce hyperpycnal flows, because their catchments produce high sediment yields but limited water discharge. However, the modelling also argues that major river deltas cannot produce hyperpycnal flows,

Figure 5 The delta plain and upper delta front of the strongly tidally influenced Ganges-Brahmaputra delta. The distributaries are tidal estuaries, the lower delta plain comprises tidal flats and creeks, and tidal processes contribute to the seaward dispersal of the plumes of suspended sediment at the distributary mouths.

because their ratio of sediment yield to water discharge is too low, even during extreme floods. A residual question is the extent to which these arguments apply during periods of sea-level lowstand, when the physiography of the fluvial systems is different (see later, Shelf-Edge Deltas in Quaternary Lowstand Periods).

The dispersal of sediment at the river mouth is also influenced by processes within the receiving basin. Moderate to high levels of wave or tidal processes promote mixing of the two water masses and also influence the orientation of the resultant plume. Wave processes limit the offshore extent of the plume and extend shore-parallel transport. Continuous beach sands rim the delta front, connecting the distributary mouths. In contrast, tidal currents extend the offshore component of sediment transport and elongate the plume into the basin (Figure 5). The tidal currents produce a network of tide-dominated channels separated by long, linear sand ridges.

The Initiation and Abandonment of Deltas

Deltas have a natural life cycle of initiation, growth, and abandonment. As deltas prograde into the receiving basin, the length of the lower part of the river feeding the delta is increased and the gradient is reduced. Shorter, steeper gradients therefore develop adjacent to stretches of the coastal plain. During major floods, the banks of a river can be breached and a new channel can branch, or avulse, off the main channel. If the breach is deep and is reused during subsequent floods, the new channel, with its steeper gradient, develops at the expense of original channel by diverting increasing amounts of the river discharge into the new channel. On reaching the margin of the receiving basin, the new channel initiates a delta, providing that it supplies more sediment than the receiving basin can redistribute. The delta at the mouth of the original river subsides because it no longer receives sufficient sediment to counter compaction of previously deposited sediments. The waters of the receiving basin progressively drown the former delta. The original delta dies and new delta is born.

This life cycle is particularly pronounced where fluvial-dominated deltas form on broad, unconfined coastal plains, such as the Gulf of Mexico. During the past 6000 years, the site of the Mississippi delta has varied along the margin of the Gulf of Mexico. Lobes and major delta complexes have switched across the margin of the Gulf of Mexico to the west and, to a lesser extent, to the east of the modern delta (Figure 6). These lobes and complexes that have been abandoned sequentially during the past 6000 years are now in differing states of abandonment. The present-day birdsfoot delta was established 600–800 years ago and new deltas are presently forming at the mouth of the Atchafalaya River and the Wax River, west of the modern delta (Figure 7). The Atchafalaya River branches off the Mississippi River and has a course to the Gulf of Mexico that is more than 300 km shorter than that of the Mississippi River. The tendency for sediment to be diverted along the Atchafalaya River has been stemmed for several decades by man-made controls because of the dire consequences for land loss in Louisiana if the avulsion is allowed to proceed unabated. The continued progradation of the modern delta towards the shelf edge is therefore artificially maintained.

The Abandonment of Deltas

What happens when a delta is abandoned? In view of current concerns of rising global sea-level, this question is germane. Rising sea-level not only inundates low-lying delta plains but also reduces gradients in the lowest parts of the river course and promotes avulsion. Abandonment of a delta therefore involves a reduction in sediment supply, whether via a lowering of river gradients or diversion of the river. Compaction of previously deposited sediments and related subsidence continues and land loss results because the supply can no longer maintain the delta plain.

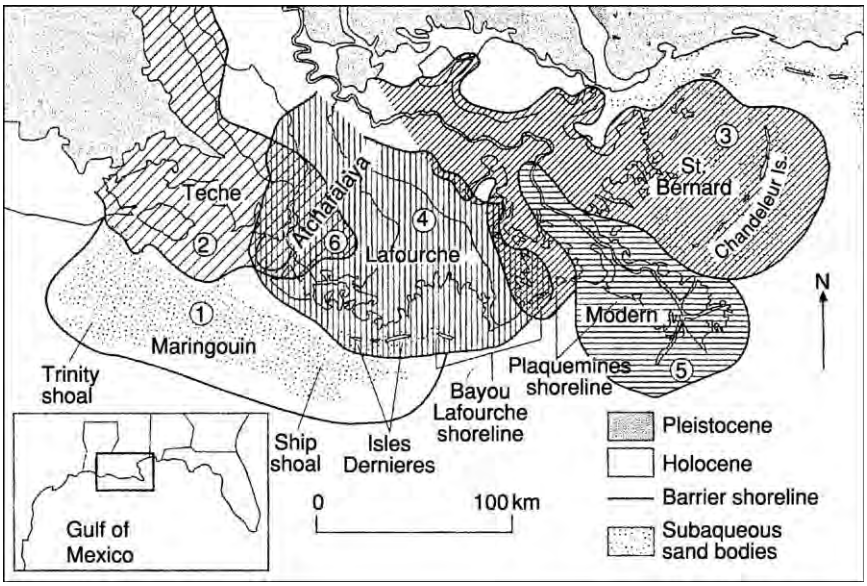


Figure 6 The major lobe complexes of the Mississippi delta system. The lower numbers indicate progressively older lobes that were active during the past 6000 years but are now in varying stages of abandonment (compare with **Figure 1**).

Figure 7 The Atchafalaya (right) and Wax River (left) deltas, forming in the Atchafalaya basin, are the new deltas of the Mississippi complex, west of the modern birdsfoot delta (see **Figure 1** for location of the Atchafalaya basin).

Additionally, with a reduction in sediment supply to the delta, the effectiveness of basinal processes is enhanced and these processes begin to rework the abandoned delta, particularly the former delta front. New shorelines are created and migrate landwards over the former delta plain as sea-level continues to rise. Precisely what happens during abandonment is therefore determined by the rate and magnitude of relative sea-level rise, the physical regime of the receiving basin, and the nature of the delta being abandoned.

In the Mississippi delta, the premodern, more lobate deltas experienced moderately high rates of relative sea-level rise related to compactional subsidence (**Figure 8**). Delta front areas are partially reworked initially into land-attached beaches (**Figure 9**) and subsequently into detached, arcuate barrier islands (**Figure 10**) that migrate landwards and are ultimately submerged as subsidence continues. The lower part of the former delta plain is converted into a shallow lagoon that progressively inundates the former delta plain (**Figure 11**), and the upper part becomes an extensive peat-forming swamp that blankets the underlying delta plain deposits. Offshore, the lower parts of the delta front accumulate slowly deposited, faunally rich muds or even carbonates as a consequence of the reduced sediment supply. In overall terms, the abandoned delta is draped by a suite of facies that record slow sedimentation rates by comparison with the facies that accumulate during active delta progradation. These slowly deposited abandonment facies produce distinctive marker horizons that are invaluable aids to identifying abandonment events and mapping the extent of former lobes and delta complexes.

Syn-Sedimentary Deformation in Deltas

Deltas are inherently unstable. This tendency has been known for some time in deltas worldwide, but as hydrocarbon exploration in the Gulf of Mexico extended onto the delta front area in the 1960s, the extent of these processes suddenly became clear. As soon as drilling rigs were placed on the delta front, they encountered problems. In 1969, as Hurricane

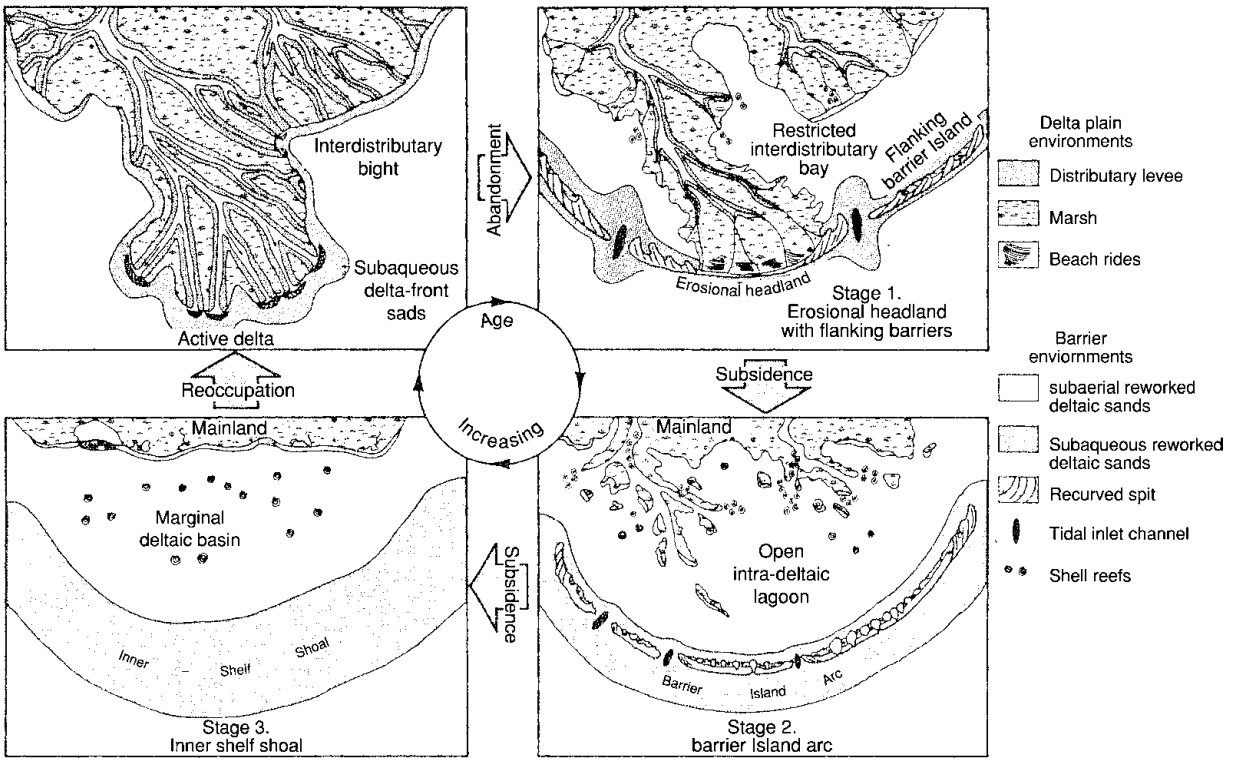


Figure 8 The evolution of the premodern lobe complexes as they are abandoned. In Stage 1 of abandonment, the former delta is reworked into a beach and barrier island complex as the effectiveness of the waves is enhanced and sediment supply diminishes. In Stage 2, transgression causes the initial beach to detach and migrate landwards and the former delta plain to be converted into a lagoon. In Stage 3, continuing transgression causes the barrier island to drown. At this stage, the only exposed part of the former delta is the upper delta plain, which is now blanketed beneath an extensive peat blanket (see **Figures 9–11** for illustrations of Stages 2 and 3).

Figure 9 Formation of land-attached, transgressive beaches during the early stages of abandonment of the Lafourche delta complex, Mississippi delta. Note the sealing of a former distributary channel by the beach sands and the storm-induced washover lobes that project onto the former delta plain (compare with Stage 1 of **Figure 8**).

Camille passed west of the delta, one rig was lost and another moved downslope. This prompted a thorough re-evaluation of the delta front, including seafloor mapping and renewed geotechnical investigations. During the 1970s, the diversity and areal density of deformational processes operating on the delta front were appreciated for the first time Figure 12). The Mississippi delta is still the most well-understood delta in this regard, but the processes are known to occur on many other deltas, and syn-sedimentary deformation should be regarded as an integral part of delta systems, particularly if the deltas have a significant fine-grained component to their sediment calibre. These processes are important in understanding patterns and rates of subsidence in deltas and also because they can have a profound effect on the final facies patterns that are preserved in deltaic deposits. The latter point is confirmed in studies showing that ancient deltaic successions commonly exhibit these features.

Aside from the possibility of seismic activity where deltas are located in tectonically active basins, there are three main mechanisms for syn-sedimentary deformation:

1. Oversteepening of the front of the delta related to high sediment accumulation rates can cause the delta to collapse, releasing large volumes of sediment as slides and slumps to deeper waters areas offshore and leaving a large detachment scar that is repaired by later sedimentation.

Figure 10 Chandeleur Island, a transgressive barrier island associated with abandonment of the St Bernard lobe, Mississippi delta complex. The lagoon on the landward side is established on the former delta plain; note the irregular margin on the barrier island on the lagoonal side due to multiple washover fans that are associated with the landward migration of the island across the lagoon (compare with Stage 2 of **Figure 8**; see also **Figure 1**, in which Chandeleur Island can be seen in an apparent NNE direction from the modern delta).

Figure 11 Levees bordering a former distributary channel in the delta plain of an abandoned delta, Mississippi delta complex. The remainder of the delta plain has been transgressed and is submerged, leaving only the highest former levee exposed.

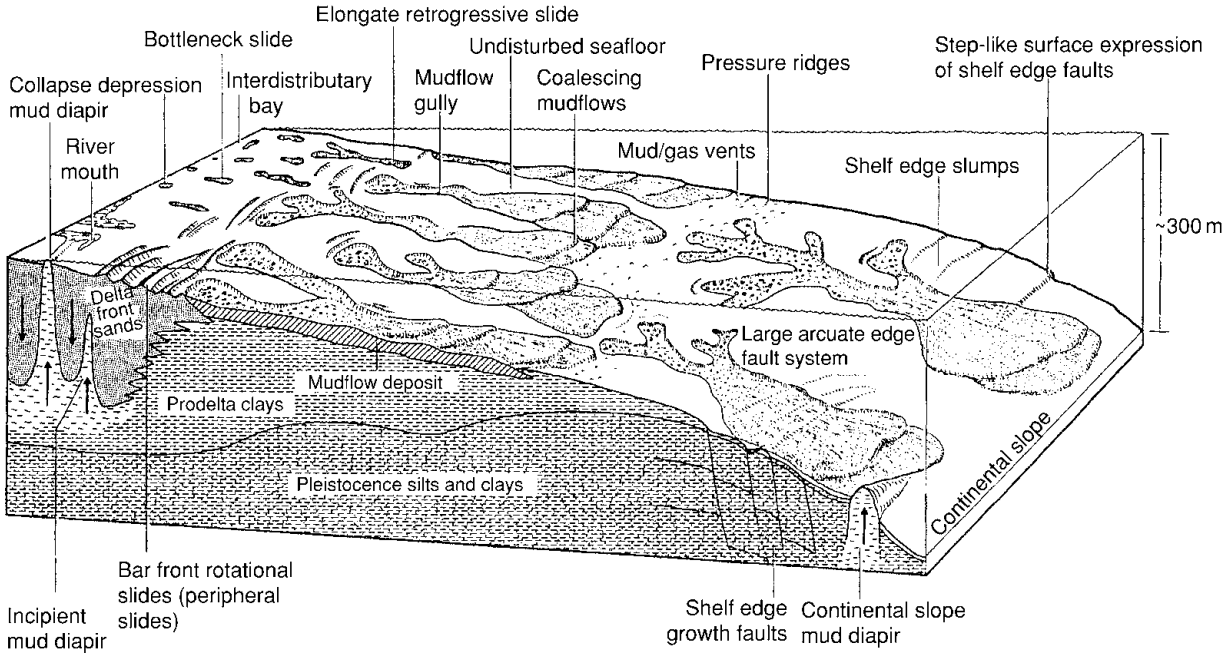


Figure 12 Summary of syn-sedimentary deformational features in the delta front of the modern Mississippi delta; note the distinction between superficial and deeper seated features. Reproduced with permission from Coleman JM, Suhayda JN, Whelan T, and Wright LD (1974) Mass movement of Mississippi River delta sediments. *Transactions of the Gulf Coast Association of Geological Societies* 24: 49–68.

2. Repetitive pounding of the delta front by storm or hurricane waves can reduce the shear strength of near-surface sediments, leading to a coupling between surface waves and bottom and to sediment surface waves that can penetrate up to 20 m below the sediment surface.
3. High rates of sediment accumulation lead to thick piles of sediment that are water rich and in some cases also include early methane and carbon dioxide gases. During early burial of these sediments, pore pressures often increase as the sediments begin to compact and permeability is reduced.

Eventually, pore pressures begin to support part of the sediment load, thereby slowing compaction and leading to the sediments being overpressured and undercompacted for their depth. These sediments are unstable in relation to further sediment loading, and they creep basinwards, creating a wide range of deep-seated deformational features.

Superficial Deformational Features

Superficial deformational features include circular collapse depressions, linear gullies and mudflows, and rotational slides (Figure 13). It is estimated that 50–60% of the front of the Mississippi delta has failed via these mechanisms and has been transferred basinwards in recent times. Collapse depressions are created by wave pounding of shallow upper delta front areas between distributary mouths during storm or hurricane periods. As waves cyclically pound the sediment surface, they cause sediments to expel water and early methane gas, to re-pack, and to locally collapse, creating depressions 100–200 m across and 1–3 m deep. The downslope margins of the depressions often fail and viscous sediment/water mixes flow downslope as mudflows. The mudflows shear a linear pathway that is enlarged by rotational failures of the margins after the mudflow has passed. This creates delta front gullies 1000 m wide and 3–20 m deep, at the mouths of which are lobate mudflows that can be 10–15 m thick and can include outsized blocks of sediment that were rafted in the viscous flows.

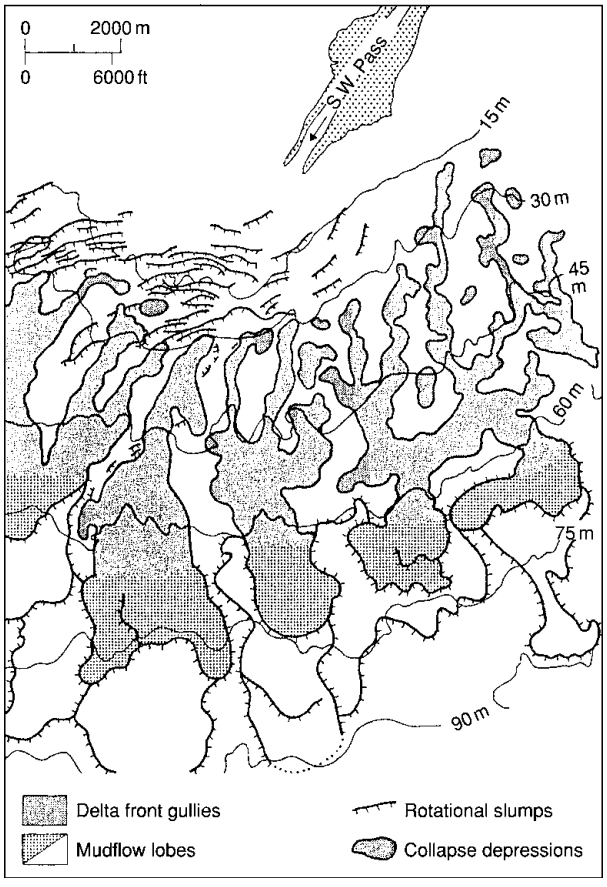


Figure 13 Map of collapse depressions, gullies, mudflow lobes, and rotational slumps in the delta front of the modern Mississippi delta. Reproduced with permission from Suhayda and Prior (1978).

Low-angle rotational slides, analogous to subareal landslides, are also common features, and it is this process that led to the loss of the drilling rig in 1969 (see earlier). Detachment surfaces dip at 1–4° in the headward regions then flatten into slope parallel surfaces. Blocks of sediment slide basinwards and may be preserved either as coherent, undeformed slides or as slumps that are deformed internally, particularly as they come to rest.

Deeper Seated Deformational Features

Basinward creep and gravitational sliding of overpressured/undecompacted fine-grained facies create an extensional regime along the basin margin and can lead to the formation of growth faults. Farther offshore in the toe regions of these structures, shale ridges and diapirs can occur (e.g., the Niger delta; Figure 14). The faults are listric extensional faults that mainly downthrow towards the basin, though counter-regional, landward-dipping faults also occur. Displacement varies from zero at the upper fault tip, to a maximum value in the middle part of the fault, decreasing again as the fault flattens with depth. In plan view, faults have curved traces that are concave to the basin and are of limited lateral persistence, though they may coalesce to form more extensive linked faults. The faults intersect the sediment surface and the hangingwall sides form preferred sites of deposition for delta front sediments. Thick, expanded, often sand-rich sections accumulate in the hangingwall of the faults and cause the faults to become self-perpetuating, because loading by the hangingwall section promotes further displacement. Anticlines offset by minor synthetic and antithetic faults form in the hangingwall succession. Through time, the site of active faulting migrates basinwards in concert with the overall migration of the delta system. Earlier faults experience a reduction in displacement and are eventually overlain by unfaulted strata. The vertical scale of the faults can vary from several tens of metres to kilometres. Sand-rich hangingwall anticlines are prime reservoirs in many deltaic hydrocarbon provinces (e.g., Niger delta).

Shale ridges and diapirs are compressional structures that form in the distal, offshore parts of deltas or the upper continental slope. They form where the toes

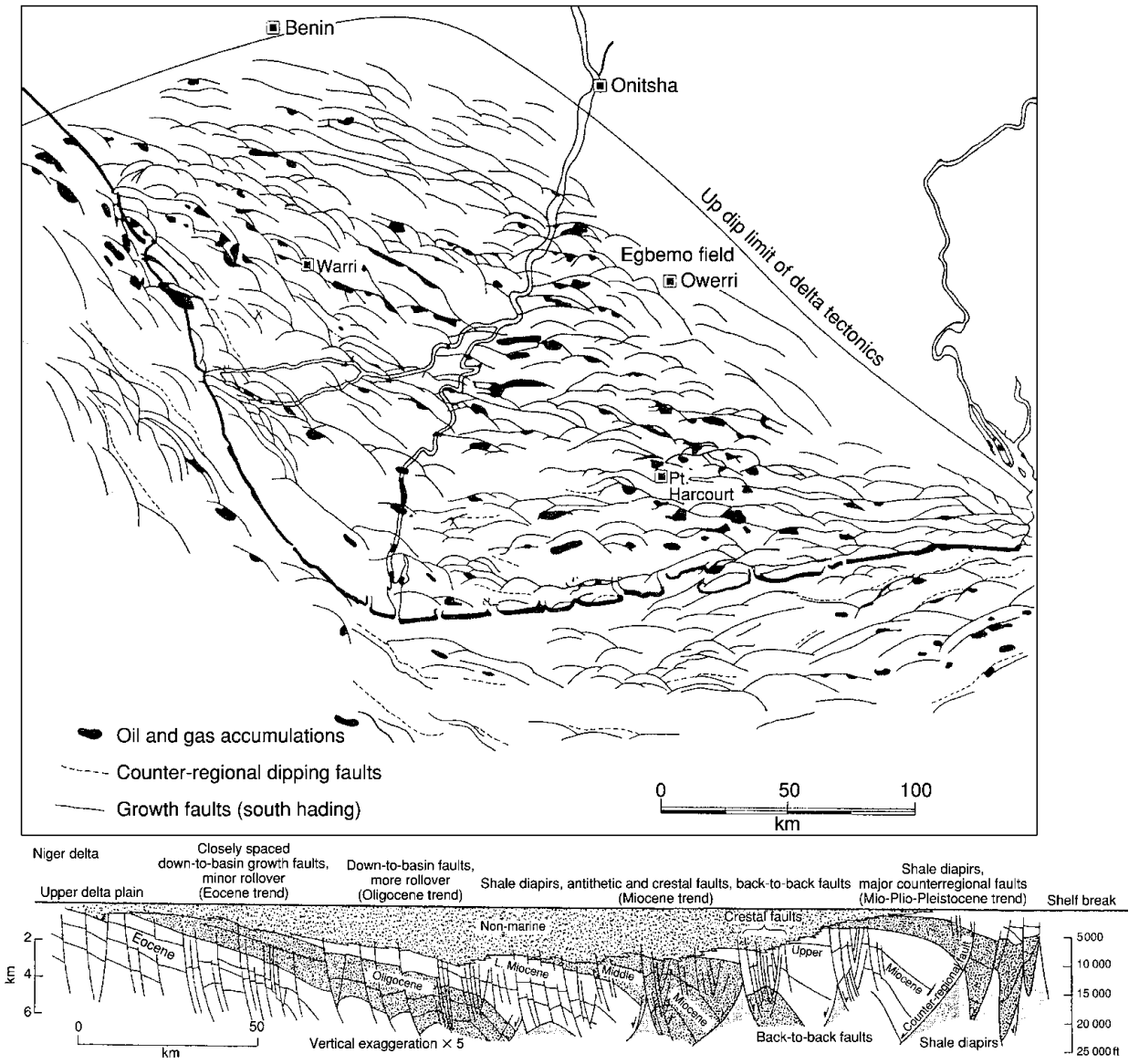


Figure 14 Map view of (top) extensional growth faults in the Niger delta and (bottom) a cross-section through the delta from the upper delta plain (left) to the shelf break (right). Note the lateral impersistence of the faults, their curved traces, and general parallelism to the delta front in the map. In the cross-section, note the down-to-basin displacement on most of the faults and the shale diapir province in the toe region of the deformation zone. Reproduced with permission from (top) Weber and Daukoru (1975) and (bottom) Winker CD and Edwards MB (1983) Unstable progradational clastic shelf margins. In: Stanley DJ and Moore GT (eds.). *The Shelfbreak: Critical Interface on Continental Margins. Special Publication the Society of Economic Paleontologists and Mineralogists*, Vol. 33, pp. 139–157. Tulsa, OK: Society of Economic Paleontologists and Mineralogists.

of listric faults cut up-section as they intersect the sediment surface, or where basinward-creeping overpressured shales intersect the surface. In the Mississippi delta, mud diapirs occur, somewhat unusually, in the delta front area, due probably to earlier shale ridges being point-loaded by distributary mouth bar sands. The diapirs appear as mudlump islands that project above sea-level by several metres. They are elevated shortly after major flood periods in response to high sediment accumulation rates and related sediment loading in the distributary mouth area and are eroded slowly by wave processes between flood periods. Mudlumps often survive between flood periods and are reactivated by the major flood, therefore becoming semipermanent features of the delta front. Sediment uplift of several hundred metres can be demonstrated and uplift rates of 100 m in 20 years have been documented. Subsurface investigations

Figure 15 A delta front gully (lower left) filled by mudflow deposits in an Upper Carboniferous deltaic succession in the Clare Basin, western Ireland. One margin of the gully is exposed on the right side, truncating earlier strata; the gully is 3–4 m deep and extends over at least 20–25 m laterally to the edge of the cliff.

Figure 16 A mud diapir in an Upper Carboniferous deltaic succession in the Clare Basin, western Ireland. Note that the bedded sandstones that overlie the diapir are thin above the crest and thicker on the flank of the diapir, indicating that the diapir grew during deposition of the sandstones. Cliff height is 25 m.

reveal that the mudlumps are growth anticlines related to reverse faults. The growth nature of the structures is depicted by thickness variations in the overlying mouth bar deposits that are thin above the diapir crest and thick in areas between diapers. Syn-sedimentary deformational features can also be observed in exposed, ancient deltaic successions (Figures 15 and 16).

Shelf-Edge Deltas in Quaternary Lowstand Periods

Presently, because of the Holocene transgression and the current highstand of sea-level, most major, ocean-facing deltas are located at the inner margin of a broad continental shelf. In contrast, during periods of sea-level lowstand associated with peak glacial in the Quaternary, these deltas were located at the outer edge of the continental shelf where they border on the continental slope. Falling sea-level forced the deltas into this position and they remained there until they retreated back across the shelf during sea-level rise associated with the onset of interglacial conditions. These shelf-edge deltas have been examined extensively in recent years using shallow, high-resolution seismic data that have demonstrated several distinctive features. The deltas are fed by widespread fluvial systems that are often incised into the former shelf (incised valleys). These systems supply large volumes of sediment to the shelf-edge deltas at a time when the accommodation space available for sediment accumulation is limited. As a result, the delta front slopes of lowstand deltas, seen as dipping clinoforms in seismic data, are typically steep (~2–4°) and frequently display evidence of large gravitational failures. Slump scars can cut 200–300 m of strata and extend laterally for several kilometres. Instantaneous bathymetric relief associated with the scars can be as great as 30 m and the scars may grow further by subsequent faulting. Slump scars therefore have the potential to release significant volumes of sediment to the deep basin as slides, slumps, debris flow, or turbidity currents, and to define minibasins that are repaired by subsequent deltaic sedimentation. Seismic data from these slump scar minibasins indicates that the repair of the scars is rarely simple. Healing by sediment accumulation alternates with renewed sediment failure, and the evolution of the scar fill can be critical in understanding the delivery of sediment to the deeper water setting during lowstand periods.

Economic Aspects

Deltaic deposits are commonly preserved in the geological record as thick, expanded sections of predominantly clastic sedimentary facies. As already mentioned, these deposits are of great economical importance. They contain vast reserves of hydrocarbons, both solid, in the form of coal, and fluid, as gas and oil (*see* **Petroleum Geology:** Overview). Because deltaic sediments are commonly rapidly deposited, with little reworking, organic matter is preserved in them. When buried and heated, this organic matter may generate oil and gas. In humid climates, the waterlogged surfaces of deltas are colonized by lush vegetation that is preserved in anaerobic marshes and swamps. These environments are sites of formations of peat that can accumulate to significant

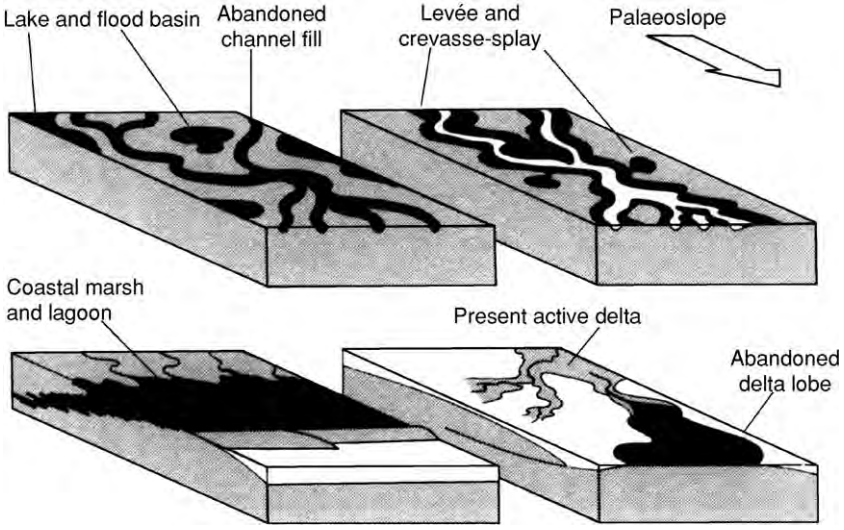


Figure 17 Sketches illustrating the common occurrence of beds of coal in deltaic settings. Note that to predict the geometry of an individual coal bed it is important to diagnose the subenvironment, rather than the overall depositional environment of the associated sediments. Channel abandonment coals, for example, occurs in fluvial, deltaic, and intertidal settings. Note also that coal geometry is controlled not only by its depositional environment, but also by subsequent erosion. Reproduced with permission from Selley RC (2000) *Applied Sedimentology*, 2nd edn. San Diego: Academic Press.

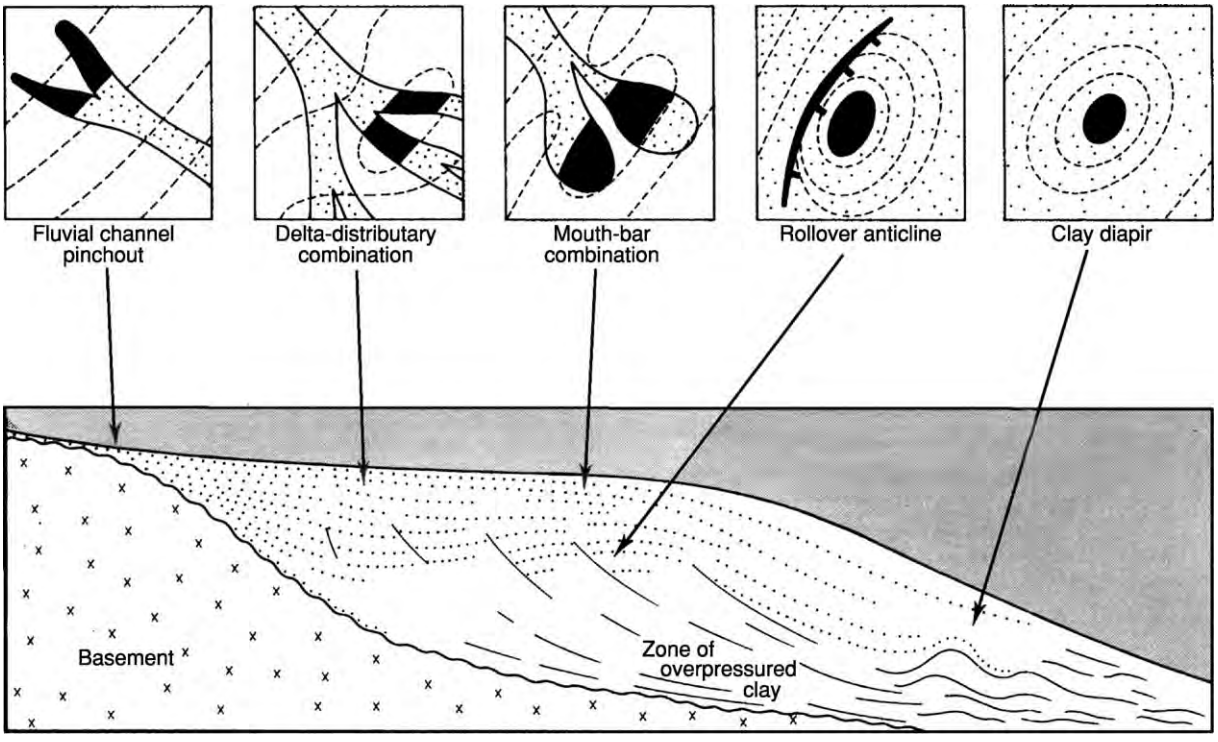


Figure 18 Diagrammatic cross-section to illustrate common trapping styles in deltaic petroleum systems. Reproduced with permission from Selley RC (2000) *Applied Sedimentology*, 2nd edn. San Diego: Academic Press.

thicknesses due to the high subsidence rates that prevail in deltas. When beds of peat are buried and heated, they metamorphose into lignite (brown coal), then bituminous coal, and finally anthracite. The conversion of peat into moderate- to high-grade coal involves a compaction ratio of ~10:1. Most coal beds are laterally continuous and relate to episodes of delta abandonment. In plan view, the geometry of coal beds may be irregular and is closely controlled by the depositional environment of the parent marsh (Figure 17). However, major coal beds in deltaic successions often relate to episodes of delta abandonment and are therefore laterally continuous coal beds that blanket underlying deposits. Most of the coal deposits of the world occur in deltaic deposits of Carboniferous age, and most of the lignite (brown coal) deposits of the world occur in Tertiary deltas. Deltas are often major petroleum systems (*see* **Petroleum Geology:** The Petroleum System). This is because a delta is a natural mechanism for depositing large volumes of sand (potential petroleum reservoirs) that prograde out over organic-rich muds (potential source rocks). Furthermore, deltas generate their own petroleum traps. These include the diapiric mud lumps, growth faults, and roll-over anticlinal structures. Deltas also contain a wide range of potential stratigraphic traps, such as sand-filled channels and delta front bars (Figure 18). The Tertiary–Holocene Niger and Mississippi deltas are major petroleum systems that illustrate these juxtapositions of source and reservoir rocks, and this range of trapping styles.

See Also

Petroleum Geology: Overview; The Petroleum System. **Sedimentary Environments:** Depositional Systems and Facies; Alluvial Fans, Alluvial Sediments and Settings.

Further Reading

Boyd R and Penland S (1988) A geomorphologic model for Mississippi delta evolution. *Transactions of the Gulf Coast Association of Geology Societies* 28: 443–452.

Coleman JM, Suhayda JN, Whelan T, and Wright LD (1974) Mass movement of Mississippi River delta sediments. *Transactions of the Gulf Coast Association of Geological Societies* 24: 49–68.

Morton RA and Suter JR (1996) Sequence stratigraphy and composition of shelf-margin deltas, northern Gulf of Mexico. *American Association of Petroleum Geologists* 80: 505–530.

Mulder T and Syvitski JPM (1995) Turbidity currents generated at river mouths during exceptional discharges to the World oceans. *Journal of Geology* 103: 285–299.

Nemec W and Steel RJ (eds.) (1988) *Fan Deltas: Sedimentology and Tectonic Settings*. London: Blackie.

Orton GJ and Reading HG (1993) Variability of deltaic processes in terms of sediment supply, with particular emphasis on grain size. *Sedimentology* 40: 475–512.

Reading HG and Collinson JD (1996) Clastic coasts. In: *Sedimentary Environments and Facies,* 3rd edn., ch. 6, pp. 154–231. Oxford: Blackwell Scientific Publications.

Roberts HH, Rosen NC, Fillon RH, and Anderson JB (2003) *Shelf Margin Deltas and Linked Down-Slope Petroleum Systems: Global Significance and Future Exploration Potential.* Proceedings of the 23rd Annual GCSSEPM Foundation Bob F. Perkins Research Conference. Available on CD ROM at www.sepm.org.

Winker CD and Edwards MB (1983) Unstable progradational clastic shelf margins. In: Stanley DJ and Moore GT (eds.) *The Shelfbreak: Critical Interface on Continental Margins. Special Publication the Society of Economic Paleontologists and Mineralogists,* Vol. 33, pp. 139–157. Tulsa, OK: Society of Economic Paleontologists and Mineralogists.

Wright LD (1977) Sediment transport and deposition at river mouths: a synthesis. *Bulletin of the Geological Society of America* 88: 857–868.

Deserts

N P Mountney, Keele University, Keele, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Deserts are arid regions in which the annual precipitation is less than half of the annual potential evapotranspiration (a measure of moisture loss). Approximately 30% of the present-day land surface of the Earth is characterized by desert conditions, with hyper-arid regions restricted primarily to parts of the Saharan, Arabian, Atacama, and Namib Deserts (**Figure 1**). The world's largest deserts, including the Saharan, Arabian, and Australian, are located in tropical tradewind belts that are characterized by the atmospheric circulation of relatively moisture-free air. Other deserts, for example the Sonoran of the

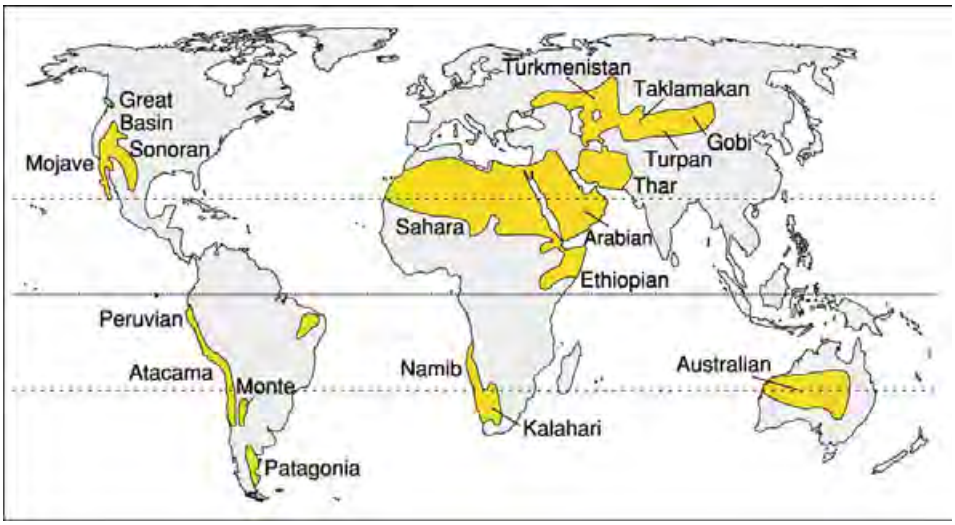


Figure 1 Distribution of the world's major climatic deserts.

south-western USA, are located in continental interiors away from humid maritime climates. The Namib and Atacama are coastal deserts in which aridity is promoted by cold offshore currents, whilst rain-shadow deserts, such as the Turpan of China, occur in the lee of mountain ranges. Deserts are not necessarily characterized by high temperatures; high-atitude polar deserts cover significant parts of Antarctica.

Deserts are composed of a variety of erosional and depositional geomorphic elements (Figure 2). Predominantly erosional mountainous and badland regions cover in excess of one-third of desert areas, whilst predominantly depositional regions composed of sand dunes, interdunes, sandsheets, playa lakes, sabkhas (salt flats), ephemeral and occasionally perennial rivers, and alluvial fans cover the remaining area. Where actively migrating sand dunes occur collectively, they form 'sand seas' or 'ergs', which cover approximately 20–30% of the surface area of the world's major deserts. The erosion, transport, and deposition of sediment in deserts are dominated by aeolian and fluvial processes, although chemical processes, such as salt precipitation, may be important locally.

The morphology and sedimentary characteristics of modern and recent desert environments, together with an understanding of the mechanisms by which these systems accumulate and become preserved, provide the basis for the recognition and interpretation of ancient outcrop examples. The architecture of ancient desert sedimentary systems can be summarized in the form of a depositional model.

Modern and Recent Desert Systems

Dunes and Sand Seas

Aeolian (wind-blown) dune bedforms are common in deserts in which the airflow is saturated with sediment. Dunes adopt a variety of morphologies (Figure 2) depending on factors such as the strength and directional variability of the wind, the size of sediment grains in transport, and whether the airflow undergoes downwind deceleration. Dune flanks and stoss (windward) slopes are characterized by wind-ripple strata, whilst steeper lee slope 'slipfaces', which are subject to periodic gravitational collapse, are characterized by grainflow (avalanche) and grainfall strata. Dunes occur either as individual, spatially isolated sand bodies, or they form collections ('trains') of bedforms that are arranged into sand sea accumulations, the morphologies of which are determined by the migration of hierarchies of bedforms of differing sizes and shapes, moving at varying rates and in varying directions relative to one another. Sand seas range in size from a few square kilometres to 560 000 km^2 in the case of the Rub al Khali erg in Arabia. Downwind changes in dune type across sand seas are common. For example, the Great Sand Dunes of Colorado are characterized by a zone of small, spatially isolated, partly cemented dunes at their upwind margin, a central zone of undulating barchan, parabolic, and transverse dunes with intervening interdunes, and a zone of large, actively accumulating transverse and star dunes separated by

Figure 2 The Sossusvlei region of the eastern Namib sand sea, Namibia. (A) Complex linear bedforms (draa) with superimposed transverse ribs. Net sand transport is from SSW to NNE. (B) Complex star (pyramid) bedforms with radiating arms and superimposed sinuous dunes. (C) Linear dune ridges. (D) Mosaic of star (pyramid) bedforms with isolated interdune hollows. White colour represents salt and calcrete deposits precipitated following the evaporation of ponded water from wet interdunes. (E) Salt flats in broad interdune corridor. (F) Ephemeral alluvial braid-plain. (G) Eroding basement with an alluvial fan apron. Image from NASA Earth Observatory Collection, with permission.

only minor interdune depressions at their downwind margin.

Interdunes

Interdune flats and hollows occur between aeolian dunes (Figure 2) and are classified as dry, damp, or wet. 'Dry interdunes' occur where the depositional surface exhibits no evidence of sedimentation that is influenced by moisture. Dry interdunes are dominated by wind-ripple strata, aeolian plane-beds, and granule and pebble lags left behind after aeolian 'deflation'. 'Damp interdunes' occur where the depositional surface is in contact with the capillary fringe of the water table and sedimentation is influenced by the presence of moisture. Damp interdunes are dominated by a range of adhesion structures where sand sticks to the damp surface and by minor salt precipitation structures. 'Wet interdunes' occur where the depositional surface is continuously or periodically inundated by water due to elevation of the water table, flash flooding within the dune-field, or overbank flooding from beyond the margins of the dune-field. Mud-, silt-, and sand-grade sediment may be supplied to these interdunes via a combination of aeolian and fluvial processes. Sedimentary structures associated with such settings include subaqueous current and wave ripples, wavy and contorted bedding, desiccation cracks, raindrop impressions, hardpan crusts, and mud flakes. Carbonate sediments are known to accumulate in some long-lived ponds. Near-surface moisture in damp and wet interdunes encourages plant and animal colonization, and a variety of bioturbation structures, including roots, burrows, and trackways, are common in these settings. The geometry of interdune elements is determined by the spacing and plan-view shape (morphology) of adjoining dunes, and varies from spatially isolated hollows that are completely surrounded by dunes, to narrow but elongate corridors that extend for several kilometres between rows of transverse or linear dunes, to extensive interdune flats completely surrounding isolated bedforms. Interdunes typically decrease in

size and degree of interconnectedness from the margin to the centre of a sand sea as surrounding dunes increase in size.

Sandsheets

'Sandsheets' are areas of wind-blown sand that lack high-relief bedforms, but are instead characterized by wind ripples, plane-beds, and sometimes by low-relief ridge- and dome-like features called 'zibar'. Sandsheets develop where aeolian sand supply and its availability for transport are restricted to the extent that dunes do not develop. Additionally, the presence of stabilizing factors, such as vegetation or a shallow water table, can act to restrict sand movement and promote accumulation. Sandsheets subject to erosion are often characterized by an armoured lag of granules and pebbles left behind following aeolian deflation of finer sand-grade sediment. Individual sandsheets vary in extent from localized sand patches ($<1\,km^2$) to major geomorphic features, such as the $>100\,000\,km^2$ Salima Sandsheet of the eastern Sahara.

Sabkhas, Playa Lakes, and Salt Flats

'Sabkhas' are low-relief flats in which accumulation occurs wholly or partly as a result of evaporite precipitation (and in some cases carbonate sedimentation). The term 'sabkha' was originally used exclusively for the description of salt flats in coastal desert settings, but is now also widely used for the description of inland salt flats (Figure 2), which often border evaporitic 'playa lake' basins. Sabkha sedimentation usually involves interactions between chemical (precipitate) and aeolian processes, and results in the generation of a variety of wavy and crinkly laminae that are often disturbed by salt growth structures such as 'teepees'. Salt precipitation in sabkhas requires periodic wetting and subsequent desiccation of the surface, and is often controlled by subtle water table changes which, in coastal settings, may be driven by sea-level change, as is the case in the Dhahran Sabkha of Saudi Arabia.

Desert Fluvial Systems

Ephemeral rivers that typically flow for no more than a few days each year are common in many desert systems and are usually fed by high-intensity but short-duration and low-frequency rainfall events. Flash flood events mean that the infiltration capacity of the substrate is rapidly exceeded, resulting in overland flow. Desert stream flood hydrographs are markedly peaked, exhibiting a rapid rise to peak discharge, followed by a rapid fall. High peak discharge rates, together with a lack of vegetation cover and a substrate that is often covered with loose sediment, mean that fluvial transport and the deposition of sediment in desert regions are important geological processes. A lack of vegetation and a high proportion of non-cohesive sandy sediment result in a low degree of bank stability, which, when coupled with high rates of bedload transport, tends to favour the development of unconfined, braid-plain fluvial systems, such as those encountered in Namibia (Figure 2). Many desert rivers exhibit a downstream reduction in discharge due to transmission losses through a porous, sandy substrate. 'Endogenic' desert streams have catchment areas within the boundaries of the desert region, whilst the catchment area of 'allogenic' streams lies beyond the confines of the desert region, and such rivers may therefore be perennial in nature and provide a valuable water resource in an otherwise arid environment.

Alluvial Fans

'Alluvial fans' are depositional landforms that occur where confined streams fed by mountain catchments emerge, often via a narrow feeder canyon, onto a low-relief plain. A decrease in gradient causes an abrupt reduction in stream power at the mountain front, which promotes deposition and the build-up of a radial fan or cone of sediment. Desert alluvial fans are ephemeral systems with sediment transport only occurring during and in the immediate aftermath of flash flood events. Sediment transport may occur either as a sheetflood or as channellized flow, in which case channels on the fan surface often develop a distributary (downstream splitting) architecture. Alluvial fan sediments are generally coarse grained and exhibit a proximal to distal fining that reflects a downstream loss of stream power and resultant transport capacity. Proximal fan deposits are typically chaotic and structureless, often being the product of debris flows, whilst more distal fan deposits usually form more organized cross-bedded units that reflect a stream flow origin. At any given time, alluvial fan sedimentation is typically restricted to one particular part of the fan surface called a 'depositional lobe'. However, over time, lobes become choked with sediment, and eventually it becomes more efficient for the sedimentation process to 'avulse' (jump) to a new lobe. Alluvial fans often develop adjacent to active normal faults, the movement of which provides the mechanism that maintains the high gradient necessary for fan development. Where faults control fan activity, as is the case in Death Valley, California, a line of overlapping fans develops into a linear apron or 'bajada'.

Accumulation and Preservation of Aeolian Strata

The preservation of desert aeolian systems in the rock record is a three-stage process that requires: (1) the construction of a sand sea, (2) its accumulation to form a succession, and (3) its long-term burial.

Sand Sea Construction

Aeolian sand sea construction is a function of sediment supply, sediment availability, and the transport capacity of the wind. The 'sediment supply' is the volume of sediment of a suitable grain size for aeolian transport generated per unit time. This sediment may form either a contemporaneous or time-lagged source of material with which to construct an aeolian system, and may be derived from a variety of pre-existing aeolian or non-aeolian sources. 'Sediment availability' is the susceptibility of surface grains to aeolian entrainment, and is influenced by stabilizing factors, such as the presence of vegetation, mud drapes, coarse-grained lags, an elevated water table, or surface binding and cementing agents. The 'transport capacity of the wind' is a measure of the potential sand-carrying capability of the wind, which increases with wind power. Where sediment availability is limited, the airflow will be undersaturated with respect to its potential sediment load, and the wind will be potentially erosional. Conversely, an airflow that is fully saturated with sediment and undergoes deceleration must drop some of its load, thereby encouraging aeolian bedform growth.

Accumulation

'Accumulation' to generate a body of strata occurs when an aeolian system experiences a positive 'net sediment budget', such that upstream sediment 'influx' exceeds downstream 'outflux' and the accumulation surface rises over time. Accumulation of migrating bedforms occurs as a consequence of 'bedform climb' with respect to the accumulation surface. The 'angle of climb' is determined by the ratio between the rate of downwind bedform migration and the rate of rise of the accumulation surface (Figure 3). For most aeolian dune systems, the accumulation rate is small compared with the migration rate and the resultant angle of climb is low, such that, as bedforms move through space, they truncate the upper parts of the preceding bedforms in a train, and only the basal parts of the bedforms accumulate to form sets of cross-strata, a process called 'subcritical climbing'.

Bounding Surfaces in Aeolian Strata

'Bounding surfaces' are erosional surfaces that are generated as an intrinsic product of aeolian dune migration and climb, such that bedforms (or parts thereof) scour into pre-existing deposits as they move through space. Three broad types of aeolian bounding surface are recognized to occur as a product of 'autocyclic' (intrinsic) migratory bedform behaviour. 'Reactivation surfaces' result from periodic erosion of a dune lee slope, followed by renewed sedimentation as a consequence of a change in bedform migration direction, migration speed, asymmetry, or steepness (Figure 4). These changes are common because the airflow on lee slopes is often subject to modification and is rarely steady. In some cases, the period of the flow fluctuation is regular and generates cyclic reactivation surfaces, for example due to diurnal or seasonal wind reversals. Nested reactivation surfaces on two or more scales occur when cyclic sets of cross-bedding separated by reactivation surfaces are generated by the interaction of two or more forcing parameters operating with different periodicities. 'Superimposition surfaces' result from either the migration of superimposed dunes over a larger parent bedform, or the

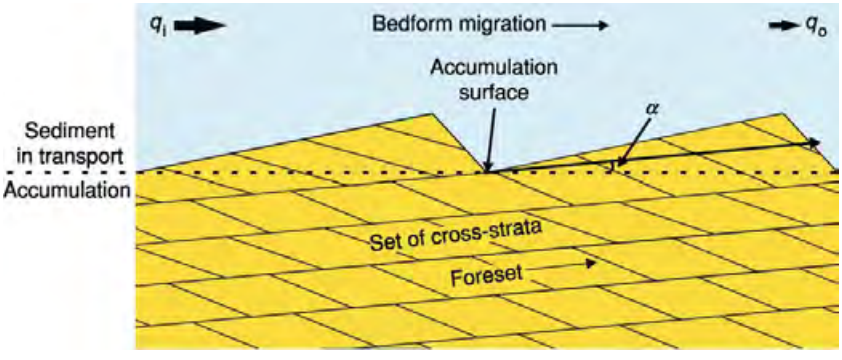


Figure 3 Generation and accumulation of sets of cross-strata by migrating and climbing bedforms. Accumulation requires a positive net sediment budget, whereby the influx (q_i) exceeds the outflux (q_o) and the accumulation surface rises through time. The ratio between the rate of rise of the accumulation surface and the rate of downwind migration of the bedforms determines the angle of climb (α) of the accumulated sets of cross-strata.

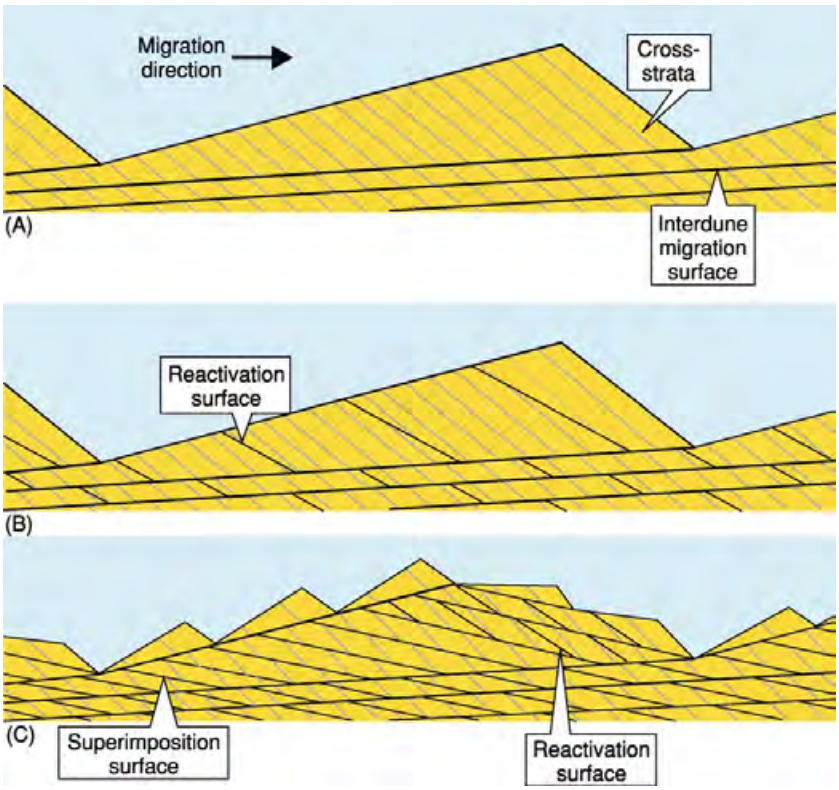


Figure 4 Generation of bounding surfaces in aeolian strata as a consequence of bedform migration. (A) Interdune migration surfaces. (B) Reactivation surfaces. (C) Superimposition surfaces arising from the migration of small bedforms over more slowly migrating parent bedforms. Reactivation surfaces may be nested within sets bounded by superimposition surfaces.

migration of scour troughs on the lee slope of a bedform (Figure 4). Although superimposed dunes and scour troughs can theoretically migrate directly up or down the lee slope of a parent bedform, oblique migration is more common because of secondary airflow along the lee slope of the parent bedform. Where both reactivation and superimposition surfaces are developed, the latter always truncate the former. 'Interdune migration surfaces' result from the migration of bedforms separated by interdunes (Figure 4). The surfaces are carved by the passage of an erosive scour that defines the interdune trough between successive bedforms. The depth to which the interdune trough scours as it migrates influences the extent to which deposits of the preceding bedform are eroded. Interdune surfaces, which truncate both superimposition and reactivation surfaces, typically extend continuously downwind for distances of hundreds of metres to several kilometres.

Types of Desert Aeolian System

In 'dry aeolian systems', the water table lies substantially below the accumulation surface, such that moisture plays no role in influencing sedimentation, and deposition occurs as a result of aerodynamic configuration alone. Bedforms in dry systems only commence climbing (i.e. accumulating) once they have grown to the point at which interdune flats have been reduced to isolated interdune depressions. In 'wet aeolian systems', the water table or its capillary fringe is in contact with the accumulation surface, such that moisture influences sedimentation, and deposition occurs as a result of both aerodynamic configuration and moisture content determined by the level of the water table. The angle of climb in wet aeolian systems is determined by the ratio between the rate of water table rise and the rate of downwind bedform migration, and such accumulations tend to be characterized by downwind climbing dune strata separated by climbing damp interdune units. In 'stabilizing aeolian systems', factors such as vegetation and surface cementation influence sedimentation, and deposition occurs as a result of both aerodynamic configuration and the degree and type of surface stabilization. For example, vegetation acts to disrupt the primary airflow, leading to a reduction in sand transport capacity and thereby a promotion of accumulation.

Supersurfaces and the Preservation of Aeolian Sequences

Aeolian accumulation ceases when the net sediment budget switches from positive to neutral or negative. Under such circumstances, a positive angle of bedform climb is replaced either by a zero angle of climb ('bypass'), when the accumulation remains static over time, or by a negative angle of climb ('deflation'), when an undersaturated airflow cannibalizes the dunes and the accumulation surface falls over time (Figure 5). Both bypass and deflation result in the generation of a 'supersurface', a fourth type of bounding surface that caps the underlying accumulation. The accumulation defines a 'sequence' and its bounding supersurface defines a 'sequence boundary'. Distinguishing supersurfaces in aeolian strata from other types of bounding surfaces is often difficult. Sedimentary features associated with supersurfaces include desiccation cracks, polygonal fractures, bioturbation, rhizoliths (calcified roots), and salt structures, all of which yield important palaeoenvironmental information regarding the nature of the desert surface at the time of supersurface formation. Supersurfaces often have great lateral extent and continuity and may bound entire sand sea successions. Thus, supersurfaces, which often result from changes in 'allogenic' (external) controlling factors, such as climate or tectonic setting, are considered to be of a higher order than the autogenically generated bounding surfaces that they truncate.

Long-term preservation of an aeolian succession requires that the body of strata is placed below some regional 'baseline of erosion', beneath which erosion does not occur, the principal agents promoting preservation being subsidence, water table rise, sea-level rise, and surface stabilization.

Ancient Desert Systems

Ancient aeolian desert systems exposed in outcrop are widely recognized from all seven continents, and a sizeable database of case examples now exists. In particular, the extensive Permian to Jurassic outcrops of the Colorado Plateau in the western USA (Figure 6) have been the focus of intensive study in recent years.

Dry Aeolian Systems

Dry aeolian systems, in which accumulation is controlled solely by aerodynamic configuration, are widely documented in the rock record. Accumulation in such systems occurs once dunes have grown to cover former interdune flats. The angle of climb of the system is determined by the ratio between the rate of accumulation (i.e. rise of the accumulation surface) and the rate of bedform migration. Steep angles of climb enable larger proportions of the original bedforms to be preserved, although climb angles rarely exceed 1° and typically only the basal 10–15% of a climbing bedform is preserved as a dune set (Figure 3). Thus, the facies preserved within aeolian dune sets record the processes that operated on the lower part of the dune lee slope at the time of accumulation, and are dominated by wind-ripple, grainflow, and grainfall strata. Additionally, superimposition and/or reactivation bounding surfaces may be present within climbing dune sets bounded by interdune migration surfaces. Supersurfaces within dry aeolian systems develop as a consequence of changed aerodynamic conditions, with deflation usually occurring when the airflow passing over a dune-field is not fully laden with sediment. The Page Sandstone of Utah and northern Arizona is a dry aeolian system represented by separate supersurface-bounded sequences, each composed of climbing dune strata and an

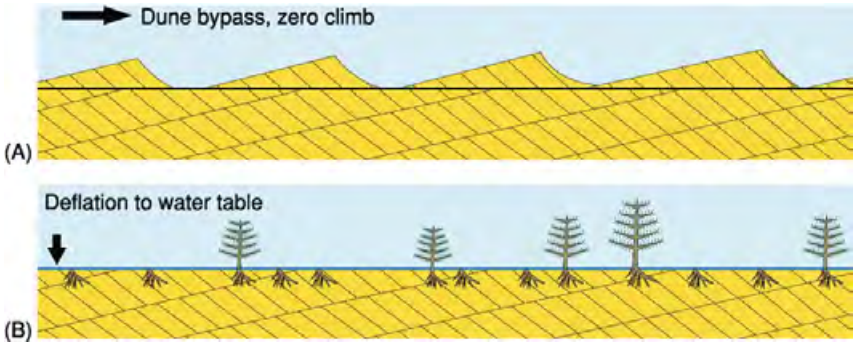


Figure 5 Models for the generation of supersurfaces in aeolian systems. (A) Bypass supersurfaces occur where the angle of dune climb reduces to zero, although dune migration may continue. (B) Deflationary supersurfaces occur as a result of erosion, often to the level of the water table. Such surfaces may exhibit evidence of widespread plant and/or animal colonization.

Figure 6 Sets of climbing aeolian dune strata bounded by interdune migration surfaces (I). Sets are composed of grainflow and grainfall strata that pass down into low-angle inclined wind-ripple strata at the set base. Numerous reactivation surfaces (R) are evident. Permian Cedar Mesa Sandstone, south-east Utah, USA. Rucksack is 0.8 m.

absence of extensive interdune flat strata. The aeolian succession interfingers with transgressive marine units of the neighbouring Carmel Formation, indicating a coastal dune setting. In many cases, the supersurfaces are overlain by sabkha and subaqueous strata and exhibit great lateral extent. They are characterized by polygonal fractures and a corrugated, erosional relief. The aeolian-dominated sequences record episodic, punctuated accumulation and, rather than being stacked vertically, occupy laterally variable palaeo-depocentres. Episodes of accumulation of dry aeolian strata occurred because of a plentiful sand supply during periods of marine regression when the exposed coastal shelf acted as the sand source. Deflation to the water table and supersurface generation occurred because of reduced aeolian sand supply during marine transgression.

Wet Aeolian Systems

Wet aeolian systems, in which accumulation is controlled by a progressive water table rise coinciding with ongoing aeolian activity, are less widely recognized in the rock record than dry systems. Ancient wet aeolian systems are characterized by cross-stratified aeolian dune sets that are separated by interdune units exhibiting evidence of accumulation at or close to the water table. A progressive water table rise that occurred synchronously with dune migration in such systems is indicated by intertonguing of the toesets of advancing dune strata and strata that accumulated in the adjacent interdune. Accumulation controlled by a water table rise is indicated by a downwind rise (climb) through the stratigraphy of both dune sets and intervening interdune units, as seen, for example, in the Jurassic Entrada Sandstone, Utah, USA (Figure 7). Changes in either the rate of water table rise or the rate of dune migration cause the angle of climb to vary. The relative level of the water table is determined by both absolute water table changes, as governed by the climatic regime, and by subsidence, whereby the accumulation can subside through a static water table. Increases and decreases in the rate of relative water table rise cause interdunes to expand, and contract, respectively, at the expense of dunes. Where climatic regimes fluctuate on a cyclical basis, interdune units preserved in wet aeolian successions reflect this change (Figure 7), as observed, for example, in the Triassic Helsby Sandstone of north-west England. As interdune units expand and contract through time, so conditions at the depositional surface vary from dry, through damp, to wet and back in response to subtle variations in the level of the water table relative to the accumulation surface. Bypass supersurfaces are generated in wet aeolian systems when the relative water table remains static, whilst deflationary supersurfaces may occur as a consequence of relative water table fall.

Stabilizing Aeolian Systems

Stabilizing aeolian systems, in which accumulation occurs because agents such as vegetation or cementation act to restrict the availability of sediment for

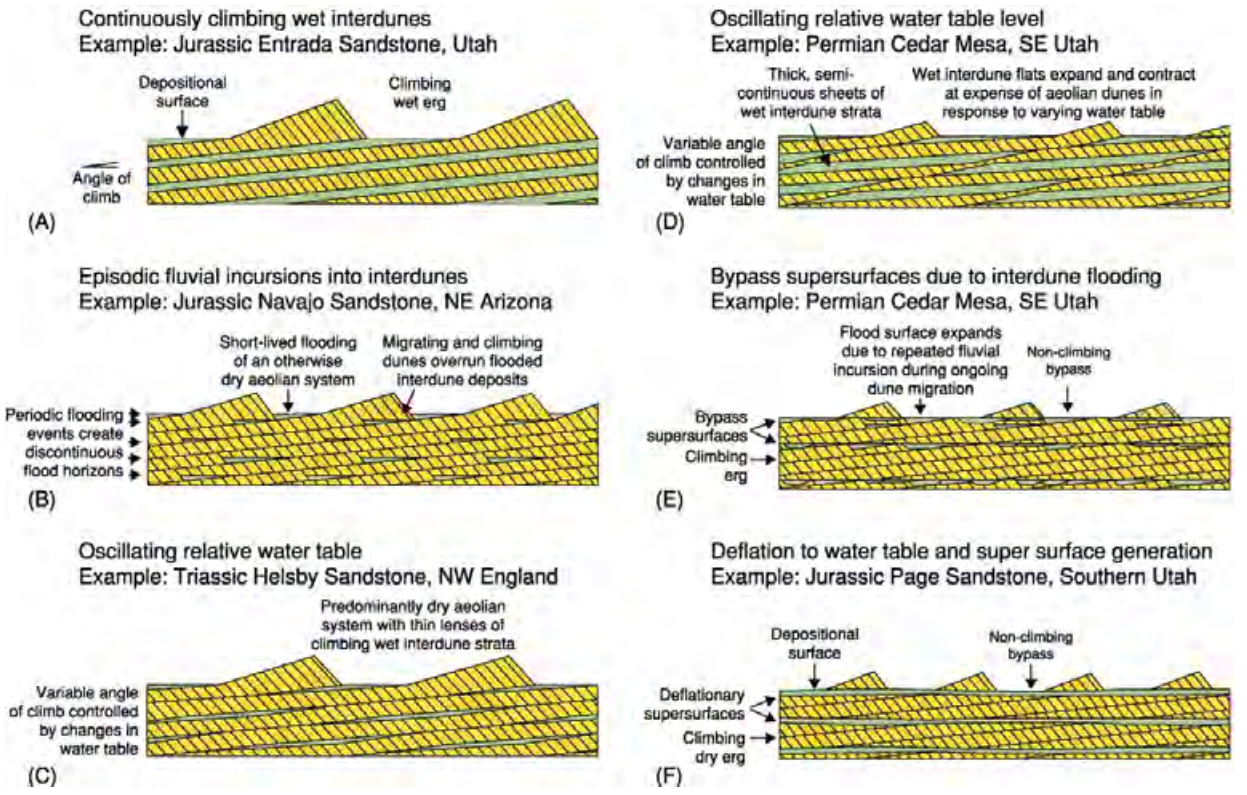


Figure 7 Common styles of dune–interdune architecture in preserved aeolian desert successions. Interactions can potentially occur on a variety of scales. Adapted with permission from Mountney NP and Thompson DB (2002) Stratigraphic evolution and preservation of aeolian dune and damp/wet interdune strata: an example from the Helsby Sandstone Formation, Cheshire Basin, UK. *Sedimentology* 49: 805–833.

transport, are not widely recognized in the rock record. Accumulation of such systems is determined by the rate of emplacement of the agent responsible for stabilization. Colonization of dunes by vegetation is the most common modern stabilizing agent and is also considered to be responsible for enabling the accumulation of some ancient successions. Part of the Tertiary Tsondab Sandstone that underlies much of the modern Namib Desert is characterized by sets of cross-stratified dune strata that contain abundant plant root structures, which are considered to have both decreased near-surface wind speed (thereby promoting sediment accumulation) and acted as a sand-binding agent (thereby preventing deflation). Prior to the development of widespread land-based vegetation in the Devonian, chemical and physical factors, rather than biogenic factors, were the main stabilizing agents. Surface and near-surface cementation by chemical precipitates, a high water table, periodic flooding, and the presence of coarse-grained lags all encourage accumulation in warm-climate stabilizing aeolian desert systems, whilst permafrost may play an important role in enabling accumulation in cold-climate systems.

Depositional Models for Desert Systems

Preserved desert accumulations exhibit complex transitions on a variety of scales between the various subenvironments that comprise the arid climate depositional system for two main reasons (Figure 8). Firstly, at a relatively small scale, the various neighbouring depositional elements present within desert systems often undergo contemporaneous accumulation, resulting in the generation of complex interfingering architectures. These types of interactions are examples of 'autocyclic' (intrinsic) behaviour within the sedimentary system, reflecting episodic switching of sedimentary processes. For example, aeolian erg margin systems often interfinger with ephemeral fluvial and alluvial fan systems, such that aeolian and fluvial processes are in direct interaction. Interdune corridors periodically act as conduits for flood waters and, in turn, dry river beds provide a source of sand for the construction of aeolian dunes. Secondly, at a larger scale, sedimentary architecture is governed by shifts in key controlling factors that influence

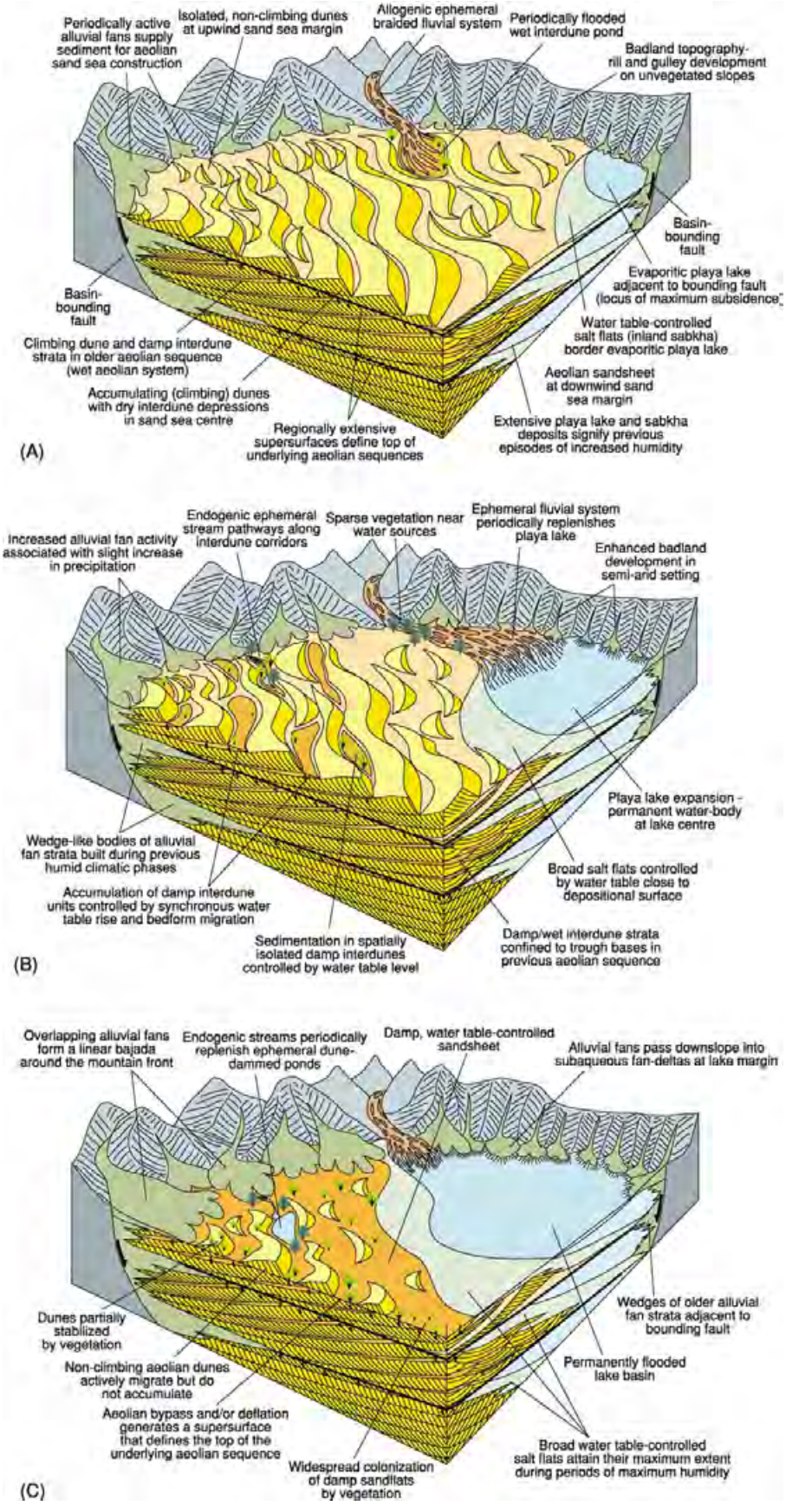


Figure 8 Depositional models for a conceptual desert system.

the gross-scale preserved architecture of the system. These 'allocyclic' (extrinsic) controls impose widespread change on the behaviour of the system. The major external controls on sedimentary architecture in desert systems are climate and tectonic basin setting. Together, these also determine the generation, supply, and availability of sediment for transport. Temporal changes in these factors will have different effects on different parts of the desert system (Figure 8). For example, a switch to a more humid climate may result in playa lake expansion due to a rise in water table. At the same time, elevation of the water table may restrict the availability of sediment for aeolian transport, causing dunes within an erg setting to stop accumulating and interdunes to expand or become flooded. Thus, various subenvironments within the desert system expand or contract, and accumulate or become eroded, over time as external controls vary. This is reflected in the preserved sedimentary architecture by large-scale shifts in the position of facies belts and the generation of surfaces of non-deposition and/or erosion (supersurfaces). In attempting to understand the complexity present within desert successions and to be able to differentiate between the extrinsic and intrinsic controls, preserved sedimentary architecture needs to be investigated on a variety of scales.

See Also

Arabia and The Gulf. **Palaeoclimates**. **Sedimentary Environments:** Depositional Systems and Facies; Alluvial Fans, Alluvial Sediments and Settings; Lake Processes and Deposits. **Sedimentary Processes:** Depositional Sedimentary Structures; Aeolian Processes. **Sedimentary Rocks:** Evaporites.

Further Reading

Alsharhan AS, Glennie K, Whittle GL, and Kendall CGStC (1998) *Quaternary Deserts and Climate Change*. Rotterdam: Balkema.

Brookfield ME (1992) Eolian systems. In: Walker RG and James NP (eds.) *Facies Models: Response to Sea Level Change*, pp. 143–156. St John's: Geological Association of Canada.

Brookfield ME and Ahlbrandt TS (1983) *Eolian Sediments and Processes. Developments in Sedimentology 38*. Amsterdam: Elsevier.

Cooke R, Warren A, and Goudie A (1993) *Desert Geomorphology*. London: UCL Press.

Frostick L and Reid I (1987) *Desert Sediments: Ancient and Modern. Geological Society of London Special Publication 35*. London: Geological Society of London.

Kocurek G (1991) Interpretation of ancient eolian sand dunes. *Annual Review of Earth and Planetary Sciences* 19: 43–75.

Kocurek G (1996) Desert aeolian systems. In: Reading HG (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, pp. 125–153. Oxford: Blackwell Science.

Kocurek G and Havholm KG (1993) Eolian sequence stratigraphy – a conceptual framework. In: Weimer P and Posamentier HW (eds.) *Siliciclastic Sequence Stratigraphy. American Association of Petroleum Geologists Memoir 58*, pp. 393–409. Tulsa, OK: American Association of Petroleum Geologists.

Kocurek G and Lancaster N (1999) Aeolian system sediment state: theory and Mojave Desert Kelso dune field example. *Sedimentology* 46: 505–515.

Lancaster N (1995) *Geomorphology of Desert Dunes*. New York: Routledge.

McKee ED (1979) *A Study of Global Sand Seas. Geological Survey Professional Paper 1052*. Washington: US Government Printing Office.

Mountney NP and Thompson DB (2002) Stratigraphic evolution and preservation of aeolian dune and damp/wet interdune strata: an example from the Helsby Sandstone Formation, Cheshire Basin, UK. *Sedimentology* 49: 805–833.

Pye K (1993) *The Dynamic and Environmental Context of Aeolian Sedimentary Systems. Geological Society of London Special Publication 72*. London: Geological Society of London.

Pye K and Lancaster N (1993) Aeolian sediments, ancient and modern. *International Association of Sedimentologists* 16. Oxford: Blackwell Science.

Rubin DM (1987) Cross-bedding, bedforms and palaeocurrents. *SEPM Concepts in Sedimentology and Paleontology* 1.

Lake Processes and Deposits

M R Talbot, University of Bergen, Bergen, Norway

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Lakes range in area from less than 1 km^2 to the *ca.* 374 000 km^2 of the modern Caspian Sea and in depth from less than 1 m to 1741 m in Lake Baikal. The salinity of lake waters ranges from fresh to salt concentrations several times that of seawater. Today, about 75% of large lakes (areas of over 500 km^2) are fresh; the rest are brackish to hypersaline. In view of this variety, it is not surprising that lakes share several sedimentary processes with the oceans but are also subject to other processes that, although not unique to lakes, are of much greater significance here than in marine environments. We can recognize a number of fundamental processes that directly or indirectly influence the types and distributions of deposits found in lakes.

- Physical processes are related to currents and waves generated by inflowing water or within the lake itself. Of particular importance are processes associated with rivers and with vertical and horizontal lake currents, and the physical structure of the water column. Physical processes, in particular the processes associated with slope failure and turbidity flows, are responsible for sediment mass movement and deposition.
- Biological processes may be directly responsible for the production of sediment or may play an important role in sediment accumulation and stabilization.
- Chemical processes: chemical sediments are common in some sorts of lakes, the most important being calcareous deposits and the evaporite accumulations characteristic of saline lakes.
- Hydrothermal processes: many lakes in tectonically active or volcanic regions are influenced by hydrothermal springs, which may affect the water chemistry or precipitate mineral deposits.
- Tectonic processes: most large long-lived lakes occur in basins of tectonic origin. In such basins tectonic effects may have a significant impact on patterns of sediment distribution and accumulation.

Although each process is listed separately, it should be borne in mind that some are closely interrelated. The production of biogenic sediments, for example, may depend on the presence of an essential nutrient, the supply of which is controlled by a lake's mixing regime. The inorganic precipitation of calcium carbonate often occurs during periods of high biological production and in some settings is probably controlled by it (see below).

Water-Column Structure, Currents, and Waves

Vertical Mixing

The degree to which surface and deeper waters mix is of fundamental importance to several aspects of lake sedimentation and biological production. Mixing is largely controlled by the turbulence associated with wind-driven surface waves and by the density contrasts between different parts of the water column. In shallow lakes wind stress may be sufficient to keep the water column more-or-less permanently well mixed, but in deeper lakes the degree of mixing between surface and bottom waters is largely determined by density differences. Although sediment and salt content influence the density of lake water, the primary control in freshwater lakes is temperature. Water becomes denser as it cools, achieving its maximum density at 4°C, so many temperate and high-latitude or high-altitude lakes go through an annual mixing cycle. In summer, when insolation is strongest, the surface water is warm and has a lower density than the deeper cooler waters. The water column is density stratified, with a warmer epilimnion and a cooler hypolimnion. These are separated by a relatively thin zone, the metalimnion, where there is typically a marked temperature gradient, the thermocline (Figure 1). During autumn and winter, surface waters cool until they become denser than the deep water, at which time the density stratification breaks down and vertical mixing occurs (Figure 1). In warm climates, particularly the subtropics and tropics, where temperatures are always high, surface waters may never cool sufficiently to become denser than the deep waters, so complete vertical mixing cannot occur. Such lakes are permanently stratified and are said to be meromictic. The role of climate in controlling mixing regime is clearly illustrated by the contrast between Lake Baikal and the deep East African rift lakes. The former is subject to a mid-latitude continental climate, with hot summers and frigid winters. Despite its great depth, Lake Baikal is well mixed. Lakes Tanganyika and Malawi, on the other hand, are meromictic tropical lakes, with an anoxic hypolimnion that constitutes a significant proportion of the total water mass.

Density stratification has several important sedimentological consequences. It determines the depth

at which inflowing water enters the lake (see below), determines the degree of oxygen renewal to the hypolimnion, and has a major influence on the nutrient supply to the surface waters, where most primary production takes place. A prolonged or permanent absence of complete vertical mixing may result in oxygen deficiency in the hypolimnion. This anoxia is inimical to higher forms of life (*see* **Sedimentary Environments:** Anoxic Environments). The hypolimnion can also become a sink for nutrients. Primary production is limited to the photic zone, which typically has a maximum depth of *ca.* 40 m and is often much less, depending on the clarity of the water. Dead organic matter produced in the epilimnion sinks into the hypolimnion where it is broken down by microbes; the nutrients released by this decomposition will be unavailable for new production unless they are returned to the surface waters. Recycling of nutrients in this way is of major significance to primary production and, ultimately, to economically important fisheries in many large lakes. Under anoxic conditions, some of the dead organic matter may survive decomposition to become incorporated into the bottom sediments. The deep-water deposits of stratified lakes are commonly notably rich in organic matter. The absence of higher organisms also means that the sediments can accumulate undisturbed by bioturbation. As well as having a high organic-matter content, sediments that accumulate in the hypolimnions of meromictic lakes are often characterized by a very regular fine lamination.

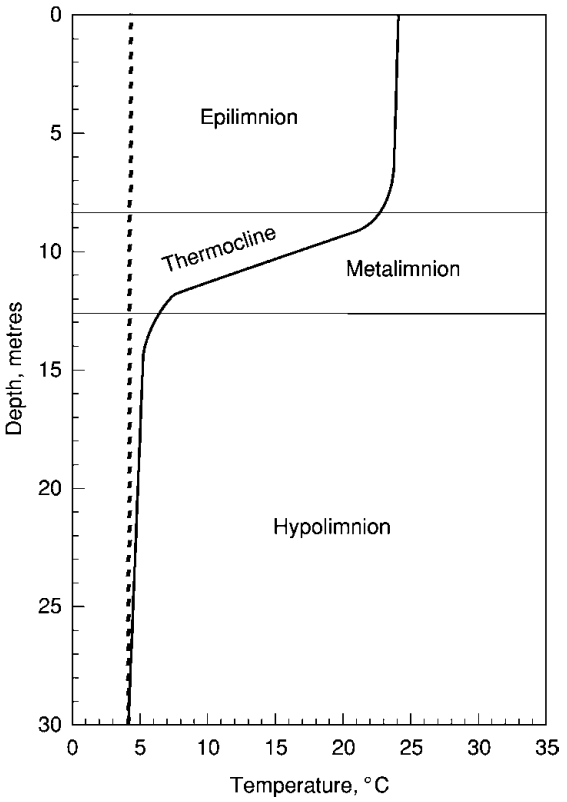


Figure 1 Temperature profiles through a typical temperate lake in summer (solid line) and winter (dashed line). Note how the thermocline separates the lake into two distinct water masses in summer. Isolation of the hypolimnion at this time can lead to severe oxygen deficiency or even anoxia. The winter profile shows the lake in an isothermal condition when it can mix to the bottom.

In addition to turbulent and density-related effects, other processes may enhance the mixing between deep and surface waters, even in lakes with permanent stratification. If a strong wind blows continuously from one direction, the surface waters can be pushed downwind, leading to the development of an epilimnetic wedge, which is separated from the hypolimnion by a tilted thermocline (Figure 2A). At the upwind end of this wedge, surface waters may eventually be completely displaced, to be replaced by upwelled water from the hypolimnion (Figure 2A). When the wind drops or changes direction, the thermocline tends to return to the horizontal; typically this does not occur in a single simple motion, but instead the thermocline oscillates in a seesaw fashion, setting up an internal wave or seiche (Figure 2B) along the length of the lake. The seiche will eventually decay, but, before doing so, may repeatedly bring cool nutrient-rich waters into the photic zone. Upwelling related to the development of an epilimnetic wedge and seiching is an important aspect of nutrient cycling in large tropical water bodies such as Lakes Malawi and

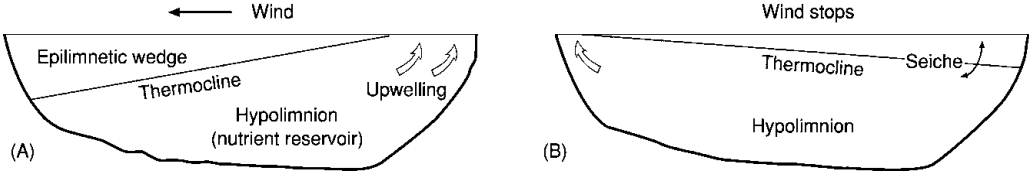


Figure 2 (A) Development of an epilimnetic wedge due to a strong unidirectional wind. Downwind displacement of the epilimnion causes the thermocline to tilt in an upwind direction and can eventually lead to the upwelling of cooler nutrient-rich water from the hypolimnion. (B) When the wind stops, the thermocline rebounds, setting up an oscillatory motion or internal wave, known as a seiche. The seiche eventually decays because of frictional effects, but while it persists it may cause periodic upwelling at both ends of the lake.

Tanganyika. The upwelling zones are typically areas of very high primary production and may be characterized by the accumulation of sediments that are notably rich in biogenic material (see below).

River Inflow

Virtually all lakes, apart from those in areas of karst, are maintained by surface inflow from rivers and streams, which are also the principal suppliers of detrital sediment. How the inflowing water interacts with the lake water depends on the relative densities of the two water masses. It is not uncommon for river water to be significantly cooler than the water at the lake surface, such that the inflow tends to sink, forming either an interflow or an underflow (Figure 3). High sediment loads or salt contents can also produce dense river water. When such waters enter a lake they will plunge below the surface until they reach a depth where the lake water has a similar density. Interflows and underflows can be very efficient distributors of detrital sediment to offshore areas of lakes (Figure 3). One of the best-known examples occurs in Lake Mead (USA), which was formed by the damming of the Colorado River. Relatively salty sediment-laden river water enters the head of Lake Mead as an underflow current, which persists until it reaches the dam *ca.* 150 km downstream.

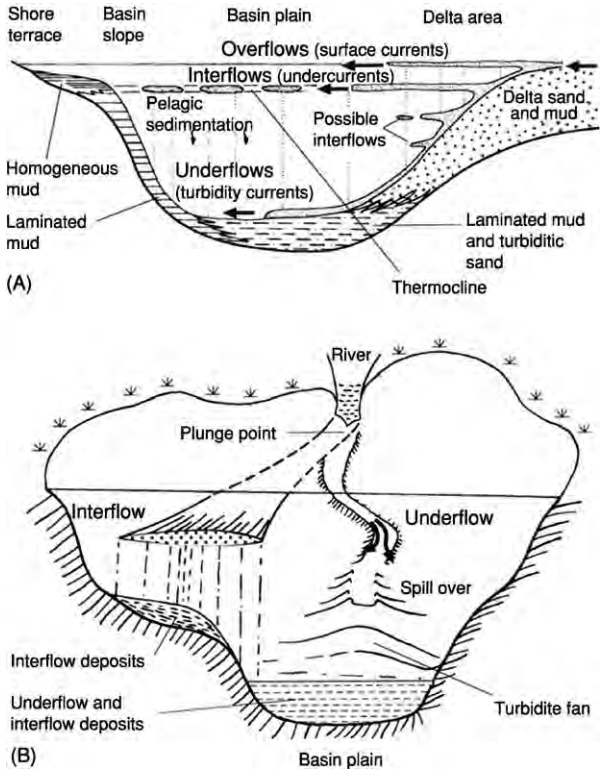


Figure 3 (A) Various types of density-driven currents and their characteristic deposits as they relate to the flow of water into a lake. Overflows occur when the inflowing water has a lower density than the lake water; interflows have a density similar to that of a particular level in the water column, and underflows have a density that is at least as great as that of the deepest water in the lake. (B) Sediment distribution as it relates to interflows and underflows. The former can transport sediment to areas of the lake floor that would otherwise be starved of sediment. Persistent underflows can lead to the development of turbidite fans with well-developed levées. Reproduced with permission from Talbot MR and Allen PA (1996) Lakes. In: Reading HG (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, pp. 83–124. Oxford: Blackwell.

The most characteristic deposits associated with river inflow are those that accumulate in deltas and fan deltas, which are a typical feature of many lake shorelines. Lacustrine deltas share many features with marine deltas, but the absence of tidal and, in many lakes, longshore currents means that sediment may not be as rapidly dispersed from the river mouth as it is in the ocean. These are some of the reasons why there is a tendency for lacustrine deltas to develop foresets with a large angle of repose, particularly where they build into relatively deep water (Figure 4). Such deltas are commonly referred to as Gilbert-type deltas after G K Gilbert, who first described these features in the abandoned shorelines of Lake Bonneville, a giant Pleistocene lake that formerly occupied the Great Salt Lake basin in Utah. The deposits of lacustrine deltas vary from coarse gravels to fine sands and silts, depending on flow conditions and the nature of the available sediment (Figure 5A). The relatively steep slopes of Gilbert-delta foresets can make them prone to collapse (see below), and, if underflows are frequent, the foresets can be reworked by them to produce ripples and imbricated clasts. Delta topsets merge with river-channel and floodplain deposits, and the bottomsets merge with the lake's deeper-water offshore deposits.

Waves

In addition to their role in the turbulent mixing of surface waters (see above), waves can be important sedimentological agents. In small or sheltered lakes where the wind's fetch is restricted, wave action may be limited. Indeed, shorelines that are little affected by waves may be marked by beds of reeds or other macrophytes, which act as very effective baffles, further reducing the potential impact of waves. On exposed shorelines, however, wave action can be significant, producing beaches and berms of well-sorted sand or gravel (Figures 5B and 6). In the absence of tides, lacustrine beaches may be relatively narrow in comparison with their marine counterparts and may show a rapid transition into offshore deposits. Nevertheless, seasonal or longer-term variations in lake level due to changes in water balance can cause large areas of the shoreline to be affected by wave action. In

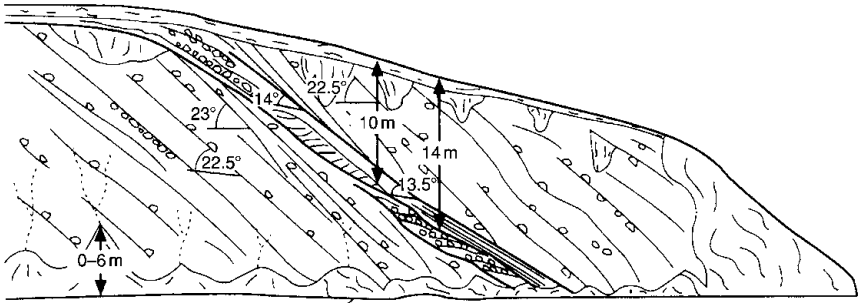


Figure 4 Cross-section through a Gilbert delta from the type area along the eastern shore of the Great Salt Lake. The unit with low-angle cross-bedding in the middle of the section is due to wave reworking of the delta foresets, possibly related to a fall in lake level. The length of the section is *ca*. 150 m. See **Figure 5A** for a close-up of the foresets showing the typical deposits. Reproduced with permission from Talbot MR and Allen PA (1996) Lakes. In: Reading HG (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, pp. 83–124. Oxford: Blackwell.

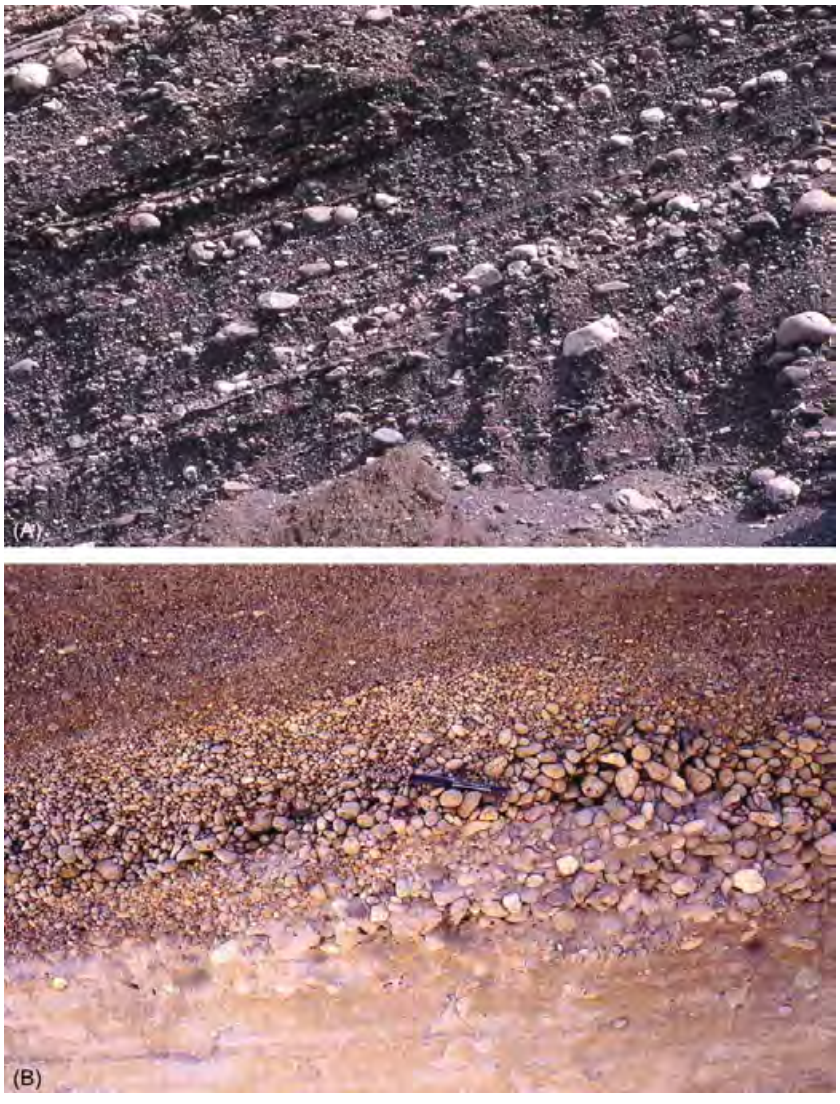


Figure 5 (A) Typical foreset deposits in a Gilbert delta. Note the variable grain size, due to changes in discharge, and the tendency to downslope imbrication in some of the layers of cobbles. The height of the section is *ca*. 2.5 m. Late Pleistocene, Willard, Utah. (B) Beach gravel consisting of well-rounded lava and pumice clasts. The coarser zones, in particular, show the typical open-framework texture (absence of matrix) characteristic of wave-washed beaches. Holocene, Lake Langano, Ethiopia.

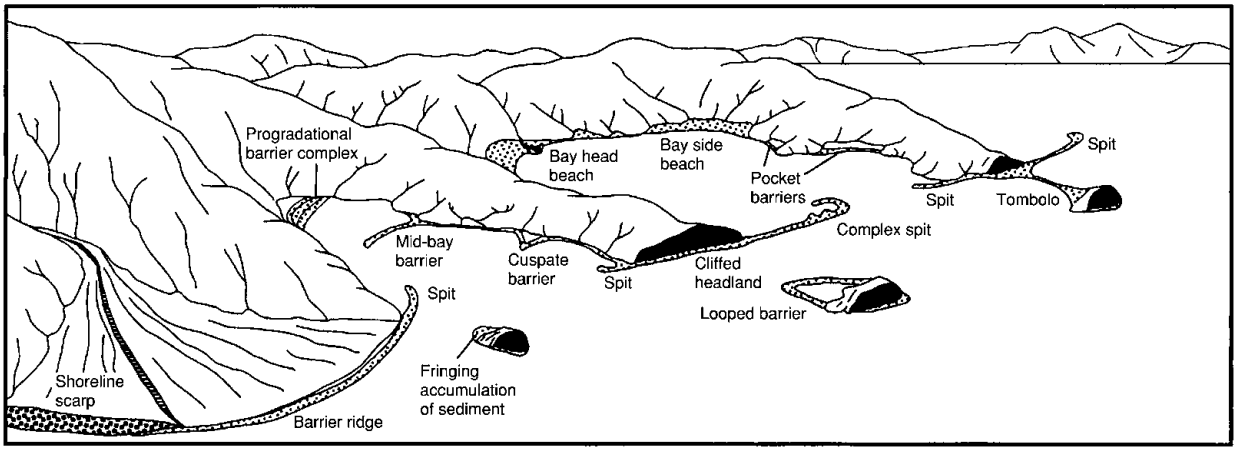


Figure 6 Depositional features formed along the shoreline of a wave-prone lake (see **Figure 5B** for an example of a gravelly beach deposit). Reproduced with permission from Adams KD and Wesnousky SG (1998) Shoreline processes and the age of the Lake Lahontan highstand in the Jessup embayment, Nevada. *Geological Society of America Bulletin* 110: 1318–1332.

some lake basins, stranded beach ridges provide important evidence of major changes in lake level related to regional climate change.

Surface Currents

Horizontal currents occur in all lakes, and in large lakes they can be a major feature of the water's circulation. Wind stress is the commonest cause of surface currents, but they can also occur in association with river inflow. Strong persistent winds can set in motion longshore currents of sufficient strength to transport significant volumes of sediment, leading to the formation of spits (Figure 6). Where surface flow is restricted by topographical obstacles such as islands, the resulting intensification of the current can also affect deeper waters – sandy current-reworked deposits have been observed at depths in excess of 200 m in Lake Superior, for example.

Subsurface Currents

As indicated above, cooler or sediment-laden river water may plunge beneath the surface of a lake to become an interflow or underflow; in this way river-derived sediment can be spread over large areas of a lake. It is also not uncommon for significant quantities of suspended sediment to be temporarily trapped at the equilibrium depth of the interflow (which is often the thermocline), to be released gradually through simple settling or rapidly as a result of a breakdown in thermal stratification. In both cases, large areas of the lake floor can be blanketed with silt- and mud-rich sediment (Figure 3). Persistent sediment-laden underflows can lead to the development of sublacustrine canyons and deep-water fans with well-developed channels and levees (Figures 3 and 7) and can deposit beds of graded sand over large areas of the deep lake floor. The accumulation of subaqueous fans from underflows and turbidity currents is a characteristic feature of deep-water sedimentation in lakes at high and low latitudes.

Underflows can also play an important role in the formation of one of the most characteristic lacustrine deposits, varves (Figure 8). These are composed of sand rhythmically interbedded with clay or silt and, as originally defined, preserve one annual cycle of sediment accumulation on a lake floor. Each sand lamina represents summer deposition, when abundant sediment is transported into the lake, and the cold sediment-laden river water forms an underflow that transports the sand to the lake floor. The clay or silt lamina accumulates in winter, when rivers are frozen and the lake is covered by ice. However, detailed studies of rhythmically laminated sand and clay or silt deposits have shown that by no means all such couplets are the product of an annual cycle of deposition. Some may be graded beds produced by turbidity currents during minutes to hours of sedimentation; others are the product of an individual flood event (the sand) and its waning flow (the clay or silt) with a duration of hours to days. Given that varve counting is widely used for chronological purposes, it is clearly essential to exercise care in the interpretation of varve or varve-like deposits. Ideally, the inferred annual cyclicity should be confirmed by an independent dating method.

A variety of rhythmically interlaminated deposits reflecting annual climatic variations other than those related to a freeze–thaw cycle are known from lakes. Some of these are described below.

Mass Failure

As indicated above, the foreset slopes of lacustrine deltas may be prone to collapse. In addition, many lake

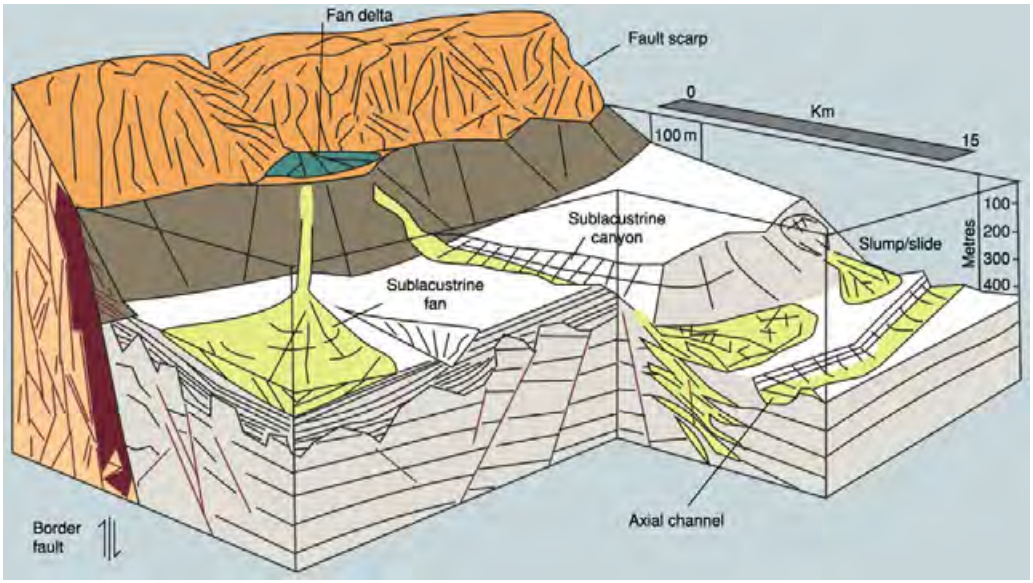


Figure 7 Sublacustrine depositional and erosional features along the margins of a deep lake subject to density flows and slumping. Based on the western margin of Lake Malawi, Africa. Adapted with permission from Wells JT, Scholz CA, and Soreghan MJ (1999) Processes of sedimentation on a lacustrine border-fault margin: interpretation of cores from Lake Malawi, East Africa. *Journal of Sedimentary Research* 69: 816–831.

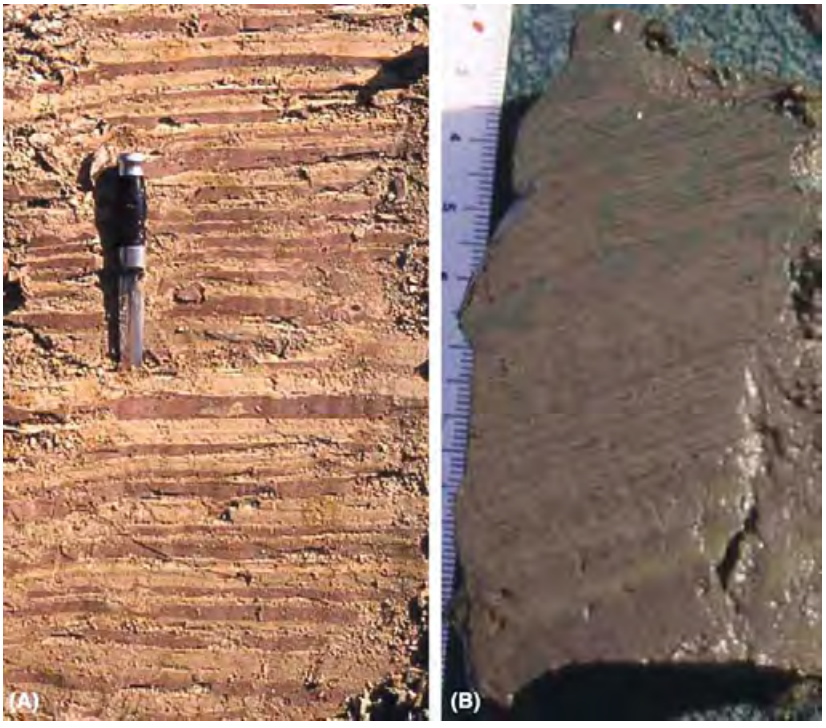


Figure 8 (A) Varves from a glacial lake. The darker brown layers are clay and silt (winter); the lighter layers are sand (spring–summer). Note that, in detail, some of the sand layers are distinctly laminated, suggesting that they accumulated during a number of discrete depositional events. Late Pleistocene, Bredåkra, Sweden. Photograph by Jan Mangerud. (B) Recent varves from Lake Malawi. The lighter laminae are rich in diatoms and accumulated during the dry windy season. The darker laminae contain abundant terrigenous plant and mineral material and were deposited from the abundant runoff generated during the rainy season. The thicker more homogeneous layers are thin turbidites.

basins have steep margins adjacent to high mountains. Both settings favour mass failure and the generation of slumps, slides, and turbidity currents (Figure 7). Such events can result in the redistribution of large volumes of sediment and leave extensive slump scars on delta tops (*see* **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes). In addition to producing characteristic deposits, such slumps can have catastrophic consequences for lakeside communities. Sediment cores and historical records document the occurrence of damaging surges (tsunamis) associated with major slumps in some Swiss lakes. Slumps may be triggered by earthquakes or major floods. Typically, they produce turbidite beds that are exceptionally thick in relation to the normal deepwater sediments and may be characterized by a relatively homogeneous texture due to the mixing and rapid deposition of sediment from a number of sources. Seismic shock may also deform lacustrine deposits through liquefaction, without causing mass movement. Zones of seismically induced deformation in the cores of lake sediments have been used to date prehistoric earthquakes and determine earthquake recurrence intervals.

Mass-flow processes are also of major importance in glacial lakes, where glaciers deposit large volumes of unconsolidated till and other sediment at the ice margin. Slumping of this material into deep water can be widespread. In addition, large masses of sediment can be dumped in the lake by the calving and overturning of icebergs.

Biological Processes

Organic remains can make a significant contribution to lake sediments; the most important types are calcium carbonate in the form of shells, opaline silica predominantly produced by diatoms, and organic matter.

Calcium Carbonate

In lakes there are no planktonic organisms equivalent to the calcium carbonate-secreting foraminifera and green algae (coccoliths) responsible for producing the extensive chalk deposits of Mesozoic to modern oceans. However, there are many species of freshwater mollusc, and, under suitable conditions, these can be important sediment producers. In Lake Tanganyika, for example, some shallow marginal areas are characterized by very extensive accumulations of gastropod shells, which are concentrated by the winnowing action of waves in this large and windy lake. Ostracod carapaces can also contribute significant amounts of biogenic calcium carbonate to lake sediments, but it is rare for them to be the dominant component of a sediment.

Some macrophytes actively precipitate calcium carbonate. Of particular significance are the charophytes (stoneworts), which can form dense 'meadows' in the shallows of hard-water lakes. Charophytes produce calcified reproductive bodies, and the plants themselves are often encrusted with calcite. Both forms of calcium carbonate can make important contributions to the sediments of shallow lakes.

Stromatolites are another type of calcium-carbonate accumulation that is widespread in modern and ancient lacustrine deposits. Stromatolites are formed through the development of mats and films of algae, bacteria (particularly cyanobacteria), and other microbes. In some situations the role of these micro-organisms has been primarily to trap and bind sediment particles, but many lacustrine stromatolites are calcified, giving the structures considerable rigidity and resistance to erosion. These calcified stromatolites are often referred to as lake tufa. Internally, the stromatolites typically show a fine wavy lamination, and in many the laminae are convex-up or hemispherical, reflecting the generally upward and outward growth of the structure. Some stromatolites reach spectacular proportions, and examples several metres high and tens of metres in diameter are known from the fossil record; in some cases they form reef-like accumulations around the margins of former lakes (*see* **Biosediments and Biofilms**).

Silica

During the Tertiary, diatoms emerged as important nonmarine algae, and from the Miocene onwards they have been major producers of lacustrine sediment. Diatoms secrete a porous skeleton (frustule) of amorphous opaline silica, the majority of which are between 5 μm and 200 μm in diameter. Where supplies of dissolved silica and other nutrients are abundant, diatoms often play a major role in primary production, even in offshore pelagic zones, thereby supplying abundant opal to the sediment. During the spring warming of temperate lakes, and at times of upwelling in stratified lakes, diatoms can occur in huge numbers, producing blooms that turn the surface waters cloudy. Deposits rich in diatoms are called diatomites and are characterized by a chalky appearance and very low density (some float when dry). Because of their low bulk density, it may seem surprising that so many diatoms sink to the bottom. It is likely that another biological process – grazing by zooplankton and the subsequent excretion of diatom-rich faecal pellets, which sink relatively quickly – plays an important part in the sedimentation of diatom remains. Variation in the rate of accumulation of biogenic silica has proved to be a useful proxy indicator of palaeoproductivity and thus palaeoclimate.

Where diatom blooms are a seasonal event, the resulting diatom-rich sediment layer may contribute to the development of varves, as is the case in some areas of Lake Malawi. Diatom production here is highest in the dry windy season, when turbulent mixing and upwelling recycle nutrients from the anoxic hypolimnion. At this time of the year runoff from the land is at a minimum, so a relatively pure diatomaceous lamina accumulates. In the rainy season, on the other hand, winds are slacker and productivity declines, but runoff is at a maximum, bringing abundant mineral and organic debris into the lake. Thus, the two seasons are represented by a light diatom-rich lamina and a dark detrital-rich layer (Figure 8B).

Organic Matter

Many lake sediments contain notable quantities of organic matter. This material has two principal sources.

1. It may consist of the remains of higher plants that populate the lake shallows, shore, and surrounding landscape. Accumulation of large quantities of such plant remains leads to the formation of freshwater peats and coals.
2. It may consist of the remains of planktonic organisms. In addition to diatoms, a number of planktonic algae and cyanobacteria can be major primary producers in lakes. Under some conditions these can make a major contribution to the sediment. In oxygenated environments, dead algal and bacterial-cell material quickly breaks down, but its chances of preservation are much greater if it accumulates in an anoxic environment, such as the hypolimnion of a stratified lake. This environment, allied with high primary production in the epilimnion, can result in the deposition of offshore muds with organic-matter contents in excess of 10%. Ancient organic-rich lacustrine shales are of major economic importance as hydrocarbon source rocks; in countries such as China, Brazil, and Angola a large proportion of the oil has been generated from such sources.

Chemical Processes

A number of important lacustrine sediments are of chemical origin, formed by direct precipitation from the lake water. The most important in terms of sediment production are calcium carbonate and evaporite minerals.

Calcium Carbonate

Lakes fed by rivers draining areas of limestone or other rocks rich in calcium, such as some basic lavas, can have high concentrations of dissolved calcium carbonate, which may precipitate as calcite, or, less commonly, aragonite. In temperate lakes, precipitation occurs during the warm season, but it may be a year-round phenomenon at lower latitudes. Although the precipitates typically comprise simple crystals, 10–100 μm long, of clearly inorganic origin, it is now recognized that their formation is often closely tied to organic processes. At times of high primary production, the photosynthetic consumption of dissolved carbon dioxide may greatly exceed the rate at which it can be replenished. Under these conditions, surface waters can rapidly become supersaturated with calcium carbonate, leading to widespread calcite or aragonite precipitation. Recently, it has also become clear that micron-sized plankton (picoplankton) may play a crucial role in the initial nucleation of the calcite crystals. During phytoplankton blooms precipitation can occur over large areas, resulting in events known as whitings (Figure 9). In Europe, deposits rich in calcium carbonate of this sort are sometimes referred to as lake marl. Seasonal precipitation of significant amounts of calcium carbonate can result in the formation of varved sediments, the light carbonate-rich lamina representing spring and summer seasons, while a darker detrital-rich layer is deposited in autumn and winter. A number of Swiss lakes have accumulated such deposits through much of the Holocene.

Another form of inorganic calcium carbonate can occur along the shores and in the shallows of wave-agitated carbonate-precipitating lakes. Here characteristic concentrically laminated sand-sized grains known as ooids may develop and, under favourable conditions, produce extensive deposits. They are, for example, widespread in the modern Great Salt Lake and on some of the beaches of Lake Tanganyika. Calcite cement may also precipitate in the pores of these deposits to produce beach-rock.

Evaporites

Evaporite minerals are deposited from saline lakes, where they form as a result of the evaporative concentration of lake water (*see* **Sedimentary Rocks:** Evaporites). Definitions of what constitutes a saline lake vary, depending on the criteria used. The most practical definition is a biological one, based on the observation that typical freshwater organisms are generally absent from waters whose salinity exceeds 5‰. A bewildering number of different salts have been recorded from saline lake basins, including some, such as borates, nitrates, and sodium carbonate minerals, that are of major economic importance. The commonest lacustrine evaporites are gypsum ($CaSO_4 \cdot 2H_2O$), halite (NaCl), and a variety of sulphates, carbonates, and bicarbonates of sodium.

Figure 9 Satellite image of Lake Michigan taken on 20 August 1999. The surface waters over much of the southern half of the lake have a milky appearance ('whiting') due to the extensive development of phytoplankton blooms and the accompanying precipitation of calcite (Part of NASA Visible Earth image #ev4338_S1999254181550.png).

Which minerals are precipitated is determined by the chemistry of the inflowing water. This is in marked contrast to the formation of marine evaporites, where the uniform composition of seawater ensures that only a limited suite of minerals can occur.

Because the precipitation of salts requires the evaporation of large volumes of lake water, their formation and accumulation tend to occur in lakes that are hydrologically closed (i.e. they have no outflow) and in regions with a dry or arid climate. In these settings, water loss due to evaporation can exceed inflow. Such lakes tend to alternate between brief periods of flooding, following rainfall, and longer periods of saline or even completely desiccated conditions. Because of their large variations in volume, many saline lakes are surrounded by extensive salt flats, where salt precipitation may continue owing to groundwater seepage and loss through capillary effects (Figure 10).

Hydrothermal Processes

Lake Baikal and many lakes in the East African rift system, the western USA, and the Andes receive a portion of their inflow from the subsurface. These springs may enter the lake from above or below the lake surface, and their temperature may be anywhere between ambient and boiling. The water may be rich in dissolved ions, which can have a measurable influence on the chemistry of the lake water and, in saline lakes, the nature of any evaporite minerals that may precipitate. In addition, effects induced by changes in physicochemical conditions at the point where a spring vents into a lake may trigger mineral precipitation and the accumulation of hydrothermal deposits. Silica and carbonates are the commonest precipitates; the latter may form spectacular deposits (Figure 11). Sulphide mineralization has also been recorded in some of the East African rift lakes.

The spring deposits typically take the form of irregular masses of travertine and sinter, which in many cases are rather porous and show complex growth patterns. One reason for these features is that many springs, even those that emerge at close to boiling point, have dense growths of filamentous and encrusting microbial communities associated with them. These colonies clearly influence the manner in which the minerals accrete. The larger voids represent conduits through which fluids passed; some of the sublacustrine springs in Lake Tanganyika are marked by chimney-like structures of aragonite, many of them carrying fronds of milky white filamentous bacteria, swaying in the stream of hot water exiting the conduit.

Tectonic Processes

Although lake basins have a variety of origins, outside areas of glaciation most large long-lived lakes occupy basins formed by tectonic processes. The morphology of such a basin is to a large degree determined by its tectonic setting. The Lake Baikal and East African rifts, and many of the basins in the western USA, for example, are the result of crustal extension, which typically produces half-graben structures characterized by marked topographical asymmetry (Figure 12) (*see* **Tectonics:** Rift Valleys). The form of the basin determines the position of the lake, which will occur at the topographically lowest point of the basin; typically, this means that the deepest water is close to a border fault. Sediment accumulation is also greatest here, where maximum subsidence rates produce the space needed to accommodate new deposits. Continuous Miocene–Holocene subsidence has resulted in the accumulation of several kilometres of lacustrine sediment beneath Lake Baikal and in parts of the East

Figure 10 (A) Surface of desiccated salt lake, Death Valley, California. (B) Close-up view showing upturned edges of desiccation cracks with salt efflorescence precipitated from seeping groundwater. The salt is mainly halite (NaCl).

Figure 11 Mounds of calcium carbonate ('The Pinnacles') formed by the discharge of carbonate-rich water through the floor of a Late Pleistocene lake. The mounds are aligned along a fault zone. Searles Lake, California.

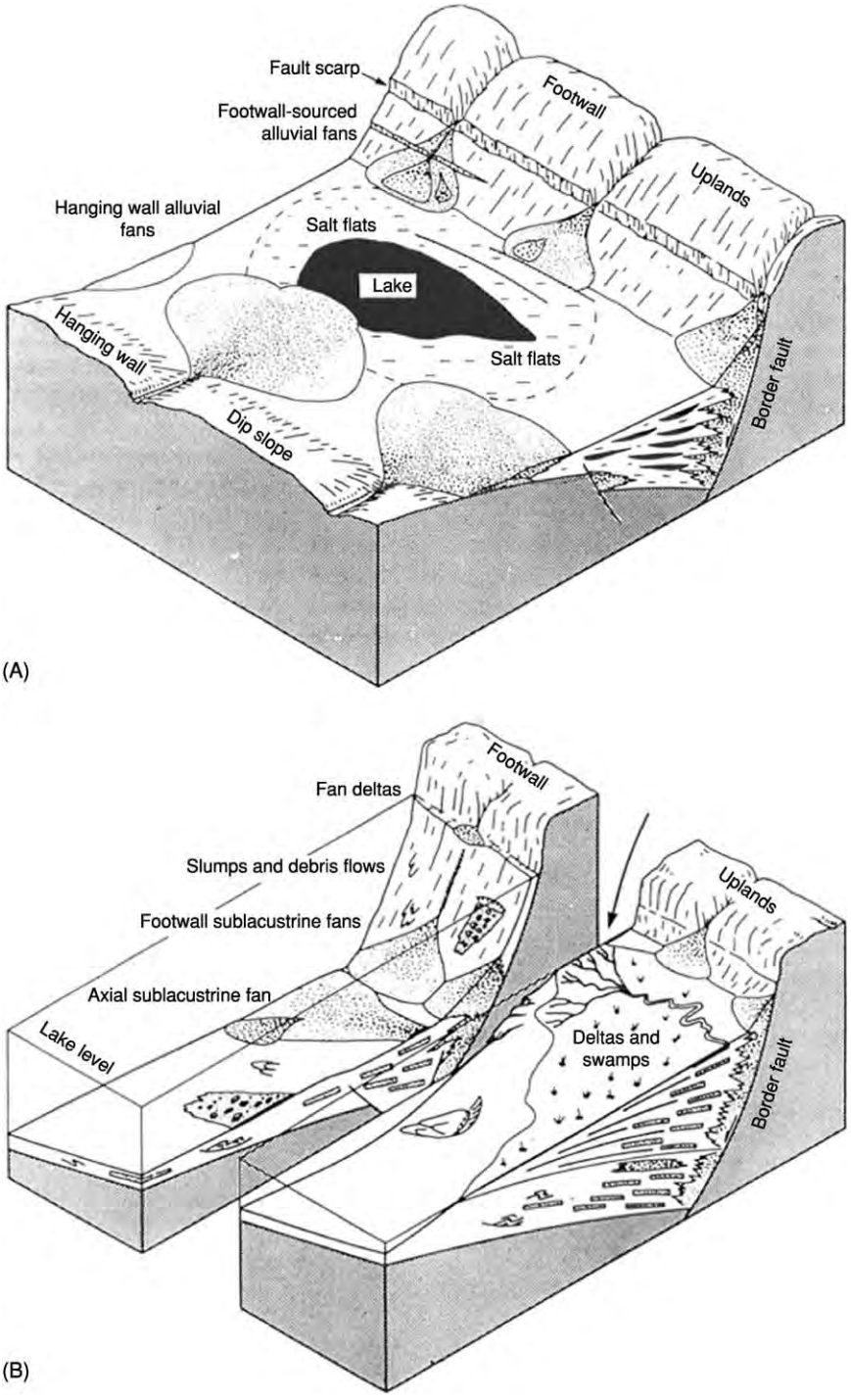


Figure 12 Schematic distributions of deposits in and around lakes developed in a half-graben (rift) basin under (A) dry to arid and (B) humid climatic conditions. (A) The basin is hydrologically closed and the lake is probably saline; extensive salt flats surround the lake and are flooded only following rare rainstorms. (B) A deep rift lake. In both cases the asymmetry of the basin has a strong influence on sediment distribution. This is particularly apparent in (B), where the deepest water is close to the border fault and the axial drainage system also tends to be located here owing to maximum subsidence in this part of the basin. Adapted with permission from Leeder MR and Gawthorpe RL (1987) Sedimentary models for extensional tilt block/half-garben basins. In: Coward MP, Dewey JF and Hancock PL (eds.) *Continental Extensional Tectonics*, pp. 139–152. London, Geological Society Special Publication No. 28.

African rift. The distribution of different sediment types closely reflects the asymmetry and varied bathymetry of the basin (Figure 12), while structurally determined sublacustrine topography exerts a major influence on deep-water sedimentation. Fault scarps form steep slopes, which tend to be unstable, generating debris and turbidite flows and leading to widespread sediment redeposition (Figures 7 and 12). Furthermore, studies of Lake Baikal and the East African rift lakes clearly demonstrate that many of the principal sublacustrine canyons along which sediment is channelled from shallow water to deep water are fault controlled. In addition to these structural effects, sedimentation in Lake Baikal has been strongly influenced by variations in the types and amounts of sediment delivered to the basin resulting from changes between glacial and non-glacial conditions, while the East African lakes have been subject to major climate-driven changes in sediment supply and lake level (on the scale of hundreds of metres). The impact of climate change and the active tectonism have resulted in complex depositional histories in all these basins.

See Also

Biosediments and Biofilms. **Sedimentary Environments:** Anoxic Environments; Deltas; Shoreline and Shoreface Deposits; Storms and Storm Deposits. **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes. **Sedimentary Rocks:** Evaporites. **Tectonics:** Rift Valleys.

Further Reading

Beadle LC (1981) *The Inland Waters of Tropical Africa*. London: Longman.

Cohen AS (2003) *Paleolimnology: The History and Evolution of Lake Systems*. Oxford: Oxford University Press.

Colman SM, Karabanov EB, and Nelson CH (2003) Quaternary sedimentation and subsidence history of Lake Baikal, Siberia, based on seismic stratigraphy and coring. *Journal of Sedimentary Research* 73: 941–956.

Eugster HP and Hardie LA (1978) Saline lakes. In: Lerman A (ed.) *Lakes: Chemistry, Geology, Physics*, pp. 237–293. Berlin: Springer.

Gawthorpe RL and Leeder MR (2000) Tectono-sedimentary evolution of active extensional basins. *Basin Research* 12: 195–218.

Gilbert GK (1885) *The Topographic Features of Lake Shores*. 5th Annual Report. Washington: United States Geological Survey, USA.

Johnson TC, Brown ET, McManus J, *et al.* (2002) A high-resolution paleoclimate record spanning the past 25 000 years in southern East Africa. *Science* 296: 113–132.

Katz BJ (2001) Lacustrine basin hydrocarbon exploration – current thoughts. *Journal of Paleolimnology* 26: 161–179.

Lezzar KE, Tiercelin J-J, De Batist M, *et al.* (1996) New seismic stratigraphy and Late Tertiary history of the North Tanganyika basin, East African rift system, deduced from multichannel and high-resolution reflection seismic data and piston core evidence. *Basin Research* 8: 1–28.

Lowenstein TK and Hardie LA (1985) Criteria for the recognition of salt pan evaporites. *Sedimentology* 32: 627–644.

Owen RB (2002) Sedimentological characteristics and origins of diatomaceous deposits in the East African rift system. In: Renaut RW and Ashley GM (eds.) *Sedimentation in Continental Rifts*, pp. 233–246. Special Publication 73. Tulsa: Society for Sedimentary Geology.

Rosendahl BR (1987) Architecture of continental rifts with special reference to East Africa. *Annual Review of Earth and Planetary Sciences* 15: 445–503.

Siegenthaler C, Finger W, Kelts K, and Sumin W (1987) Earthquake and seiche deposits in Lake Lucerne, Switzerland. *Eclogae Geologicae Helvetica* 80: 241–260.

Sturm M and Matter A (1978) Turbidites and varves in Lake Brienz (Switzerland): deposition of clastic detritus by density currents. In: Matter A and Tucker ME (eds.) *Modern and Ancient Lake Sediments*, pp. 147–168. Special Publication 2. London: International Association of Sedimentologists.

Talbot MR and Allen PA (1996) Lakes. In: Reading HG (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, pp. 83–124. Oxford: Blackwell.

TANGANYDRO (1992) Sublacustrine hydrothermal seeps in northern Lake Tanganyika, East African rift: 1991 TANGANYDRO Expedition. *Bulletin des Centres de Recherches Exploration-Production Elf-Aquitaine* 16: 55–81.

Thompson JB, Schultze-Lam S, Beveridge TJ, and Des Marais DJ (1997) Whiting events: biogenic origin due to the photosynthetic activity of cyanobacterial picoplankton. *Limnology and Oceanography* 42: 133–141.

Wells JT, Scholz CA, and Soreghan MJ (1999) Processes of sedimentation on a lacustrine border-fault margin: interpretation of cores from Lake Malawi, East Africa. *Journal of Sedimentary Research* 69: 816–831.

Wetzel RG (2001) *Limnology: Lake and River Systems*. San Diego: Academic Press.

Reefs ('Build-Ups')

B W Sellwood, University of Reading, Reading, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Reefs are biologically generated accumulations of carbonate (*see* **Sedimentary Rocks:** Limestones) that grow upwards from the adjacent seafloor and are entities of their own making. They are generally considered to be robust structures (i.e., rigid or wave-resistant), although biogenic mud mounds are treated as reefs by many authors. There have been periods in the geological past when the main contributors to biogenic mounds appear to have been organisms lacking a rigid frame, and their constructions have had lime mud (matrix-rich) compositions. Thus a wide spectrum of structures have been loosely classified as reefs. It has therefore generally been considered useful to draw a distinction between reefs with a rigid calcareous frame and those without (i.e., mud mounds). As communities growing on the accumulated debris of their own forebears, reefs are essentially unbedded constructions. The debris derived from the reef-dwellers fills the interstices between *in situ* constructors and also forms aprons of bedded resedimented material on the flanks of the edifice, and in adjacent hollows. Such aprons provide critical evidence that biogenic activity produced an elevated mound in the first place, and may indicate that upward growth of the mound had taken place into shallow waters, where strong currents and waves were active.

Growing reefs are often strongly zoned ecologically and usually have a core, the main carbonate factory within the system, showing a distinct vertical succession from initial colonization through to culmination. The vertical zonation can frequently be related to upward growth in response to the filling of accommodation space (i.e., formed in response to 'transgression' with the bulk of the reef pile accumulating during a 'highstand'). Reef successions are often stacked one on top of the other, separated by lowstand emersion surfaces (often 'karstic' in character).

Modern Reefs

Modern reefs (Figure 1), dominated by scleractinian corals (*see* **Fossil Invertebrates:** Corals and Other Cnidaria), reflect a balance between growth and destruction. Coral growth produces a rigid framework ('framestone'(Figure 2)), sheeted encrustations generate 'bindstone'(Figures 3 and 4), whereas meadows of upstanding but flimsy (generally noncoral) organisms may passively trap fines to form mud mounds ('bafflestone' after lithification (Figure 5)). Attached

Figure 1 Modern patch reefs with intervening grass-anchored carbonate sands as seen from the air. Great Bahamas Bank, field-of-view approximately 1 km across.

Figure 2 Framestone with *in situ* coral heads and unbedded character. Key Largo Limestone, Quaternary, Florida, USA. Height of outcrop about 2 m.

Figure 3 Bindstone predominantly composed of sheeted stromatoporoids, Devonian, Belgium. Height of outcrop approximately 40 cm.

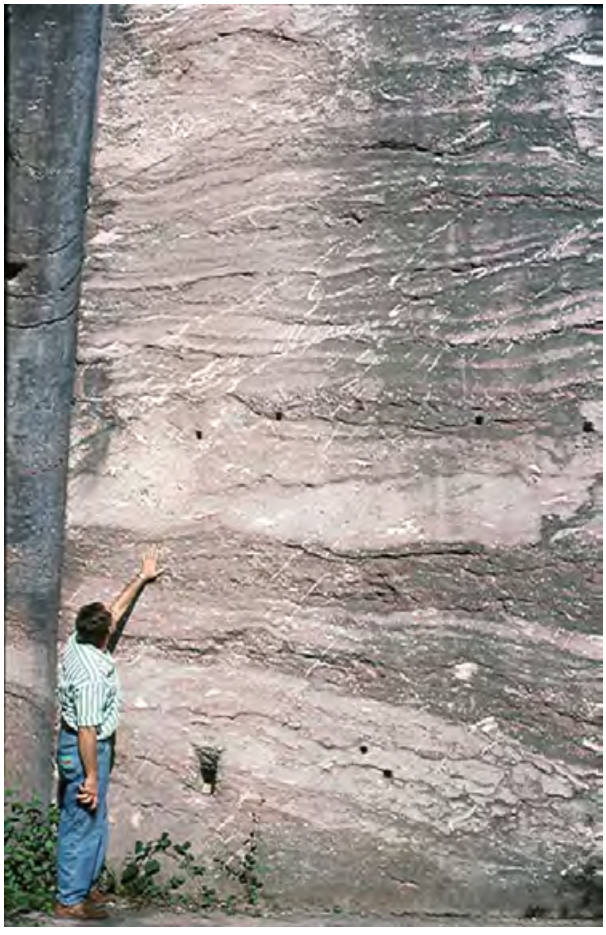


Figure 4 Bindstone sheets of stromatoporoids, Devonian, Belgium.

Figure 5 Bafflestone composed predominantly of lime-mud probably trapped by mangroves. Messinian Miocene, Mallorca, Spain.

organisms and those living in any cavities and crevices contribute to the overall community, and some will have fossilization potential, whereas many will not. Destructive processes may be both physical (waves and currents) and biological (raspers, borers, and grazers). *In situ* collapse (following bioinfestation) and downslope relocation, as gravel-grade rubble, produces 'rudstones'(**Figures 6, 7 and 8**), whereas mud slides and debris flows, with blocks entrained or embedded within slurries of carbonate fines, produce 'floatstones' (**Figure 9**). Sand-grade material generated by grazers may itself be rendered still finer by biological degradation.

Corals generally remain in place after death. Cores through coral heads reveal regular annual growth layers, which provide a precise year-by-year record ('sclerochronology') of climate-related, anthropogenic, volcanogenic, and other events that leave a geochemical signature within the coral frame. Bioinfestation by macro- and microborers degrades skeletal structures, leading to the obliteration of original, and often characteristic, textures in grains and generates micritized 'peloids'. These processes also contribute to reef sediment, particularly the fines, as

Figure 6 Rudstone of coral debris. Quaternary, Barbados. Scale at centre 10 cm.

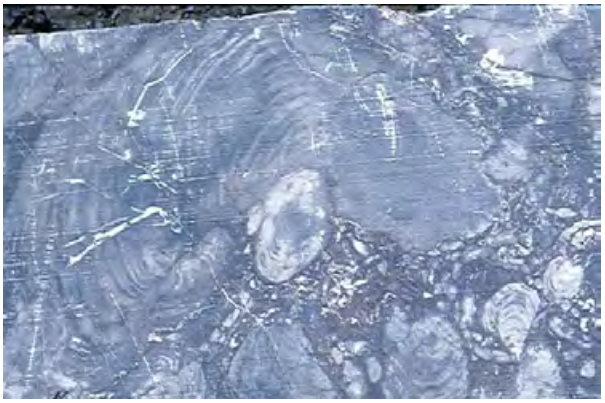


Figure 7 Rudstone of stromatoporoids debris, Devonian, Devon, England. Field width approximately 50 cm.

Figure 8 Rudstone of stromatoporoids, Devonian, Belgium.

do faecal and pseudofaecal material generated by the reef dwellers. However, many reef-dwelling organisms are microcarnivores and suspension-feeders highly sensitive to the amount of turbidity in the water, so reefs grow best on windward platform margins where wave surge cleanses the reef of the inhibiting fines. In such sites daily and seasonal temperature fluctuations are small, as are salinity fluctuations, the reef community generally thriving best in well-oxygenated waters within tight limits of normal marine salinities (~35‰ (range 27–40‰)), warm temperatures (18–36°C), and low nutrient content (phosphates, etc.). Investigations, using dynamite, have shown that the robustness of many reefs results from early marine cementation of the reef structure. Wave pumping forces vast quantities of seawater saturated with calcium carbonate through the interstices of the reef, and 'isopachous cements' of 'Mg-calcite' (and 'aragonite', which is often 'botryoidal' in form) grow actively and rapidly.

Figure 9 Floatstone. A reef-derived boulder set in deep-water mudstones and packstones, Permian Capitan Reef Complex, McKittrick Canyon, New Mexico, USA.

Reefs formed immediately adjacent to shorelines (*see* **Sedimentary Environments:** Carbonate Shorelines and Shelves) are termed 'fringing reefs'. Offshore reefs form a barrier, behind which is a 'lagoon'. Barrier reefs, generally circular in plan and with a central lagoon, are termed 'atolls', which, in the open waters of the Pacific and Indian Oceans, were initiated around volcanic islands and continued to grow upwards following the submergence of the original edifice. This mechanism was originally suggested by Darwin (1842) and was confirmed by the Funafuti boring, funded by the Royal Society of London. Lagoons behind barrier reefs may have patch reefs, which can form miniatolls, such as those behind the Great Barrier Reef off eastern Australia. Patch reefs (e.g., Figure 1) rise from depths of a few metres, and between them are usually extensive spreads of muddy lime sand anchored by sea grasses (e.g., *Thalassia*).

Platform margin reefs drop off, often precipitously, into deep water where the 'fore reef' comprises a zone predominantly receiving reef-derived debris, carbonate sand, and tumbled blocks (talus). The 'reef front' rises from this deep water. One hundred metres, in the clearest waters, is about the maximum depth in which coral colonization occurs but depths are usually less than this. The plate-like growth forms of the pioneers 'stabilizing' reef foundations in these gloomy waters are an adaptation for capturing the maximum amount of light. Coral polyps live symbiotically with zooxanthellae algae within their tissues, so enough light for their photosynthetic activities is an essential component in reef colonization and growth. Into shallower depths platy forms give way to 'colonization' by branching forms as architectures favouring food collecting take over in habitats where light is no longer a limiting factor. Massive, and finally encrusting, forms 'diversify' and proliferate in wave-activated depths approaching the reef crest. The reef crest is the highest energy zone impacted by breaking waves. Encrusting forms often 'dominate' there, but if wave energy is of moderate intensity they may be accompanied by forms with abbreviated branches and robust hemispherical shapes. The 'reef flat', often exposed at low tide and as such inhospitable to corals, is crossed by channels through which the surf surges. These are usually floored by sand that is carried lagoonwards. James (1983) suggests that reefs of many ages show a similar succession in growth forms (stabilization, colonization, diversification, domination) even though the organisms generating the reefs are often from phyla different from those constructing modern reefs. In contrast to the seaward zone the 'back reef' is a tranquil place, receiving much of the mud-grade carbonate generated by the reef. Abundant sand- and mud-producing organisms live there, particularly calcareous green algae (e.g., codiacians such as *Halimeda*). Stubby and dendroid growth forms dominate this habitat with species adapted to rapid upward growth in muddy environments.

Reefs through Time

This topic has been extensively reviewed. Prior to the advent of grazing metazoans in the Early Cambrian stromatolites were the prime reef builders. Stromatolitic reefs in Western Australia, formed between 3.0 and 3.5 Ga, and through the Archaean and Proterozoic large and increasingly complex structures and ecosystems, were formed by *Cyanobacteria* and other microorganisms ('microbes' such as noncalcified and calcified *Cyanobacteria*, Eubacteria, fungi, and other protozoans). These same communities are considered by many to be responsible, through their photosynthesizing activities, for changing the character of the atmosphere from its initial reducing (CO_2-rich state) to one with free oxygen (around 2.32 Ga).

The first metazoan reef mounds a few metres across were constructed in the Early Cambrian by 'Archaeocyathans' (sponge-like organisms) encrusted by calcitized algae and surrounded by carbonate sands rich in pelmatozoan debris. They have an apparently wide palaeolatitudinal spread (60° S to 30° N) and are particularly well described from Newfoundland, Labrador, and the Siberian Platform. Archaeocyathans became extinct at the end of the Early Cambrian and reef structures disappeared until the mid-Ordovician. At that time colonial corals and sponge-like 'stromatoporoids' made their first appearance and dominated palaeotropical reef associations, accompanied by an increasingly diverse array of sponges (*see* **Fossil Invertebrates:** Porifera), bryozoans, and calcareous algae, until the Late Devonian. Brachiopods (*see* **Fossil Invertebrates:** Brachiopods) and pelmatozoans contributed in a major way to these build-ups. Diversification continued through the Silurian and Devonian, a time of exceptionally high global sea-levels, which generated vast tropical epicontinental seaways within which grew abundant reefs dominated by sponges (stromatoporoids) and corals; reefs extended from 40° to 60° poleward of the equator. By Late Devonian times (**Figures 7 and 8**) reef systems probably rivalled modern reefs in terms of both their scale and complexity (some belts exceeding 2000 km in length). Many of the organisms contributing to these structures produced calcitic rather than aragonitic skeletons, and internal cavities within the reef systems were elongate lensoids with smooth flattish floors and complex, often crenulate ceilings. Termed 'stromatactis', and filled with radiaxial calcite spar, these cavities have long been a topic of controversy, as to whether they represent a long extinct organism group, dilation cracks caused by down-slope movement of partially cemented sediment, methane seepage, or strange burrow structures. That they were open cavities within the mounds, and preserved by early cementation processes, is confirmed by their internal sediment fillings and the lining cement phases.

A major extinction (*see* **Palaeozoic:** Devonian) event took place within the Late Devonian, considered by many to rival that at the End Permian in terms of its severity (*see* **Palaeozoic:** End Permian Extinctions). This caused the end of the coral-stromatoporoid reef systems, with mud mound systems apparently lacking rigid frame builders and dominated

by sponge-algal-microbial-foraminiferal communities continuing. The Early Carboniferous was a time of mud-mound communities (e.g., Waulsortian bank facies), the absence of major reef systems at this time being attributed to an overall decrease in global temperatures. The pattern of a relative absence of large frame builders continued through the Late Carboniferous and Early Permian when mound communities were dominated by calcareous algae, sponges, and bryozoans. During the Permian tropical and subtropical carbonate systems became dominated by aragonite, rather than calcite, a situation mineralogically similar to the Cenozoic and Recent.

Many mid- and Late Permian reefs, like the classic Capitan complex on the margins of the Delaware basin (**Figures 9 and 10**), most probably accumulated as a series of marginal mud mounds anchored by very early marine cements. Sponges, phylloid algae, and microbial precipitates compose the major components and debate has continued over many years about the architecture, palaeobathymetry over the reef, and other aspects of palaeoecology and geometry. It is clear that the evolution of the system was influenced significantly by sea-level changes, and considerable palaeobathymetry between reef and basin can be demonstrated by the presence of rudstones and floatstones that can be traced up to the reef from the basin floor. Another set of famous, but still controversial reefs are those of the Late Permian Zechstein basin of northern Europe. Communities within these reefs are highly zoned, containing bryozoans and stromatolites, with reef base, reef core, reef flat, reef crest, and patch reef communities all being recognized, the latter growing upon a distally steepened ramp. During the Late Permian reefs with corals (rugose corals that may have lived symbiotically with zooxanthellae and tabulates) are confined to the Australasian rim of the Pacific and the Middle East. Other biota within these reefs were calcareous algae, bryozoans, foraminiferans, and brachiopods. With the exception of coral communities in East Timor and the Salt Range of Pakistan, most of these build-ups occur within the palaeotropics.

There was a sudden and complete disappearance of high diversity reef biota at the end of the Permian (*see* **Palaeozoic:** End Permian Extinctions), and a global hiatus of about 10 Ma occurs before the renewal of reef development in the early Triassic. Triassic reefs were generated by a range of organisms including microbes; scleractinian corals (direct ancestors of modern corals); sponges; calcareous algae (dasyclads); bivalves; and serpulids. Microbial reefs dominate the earlier Triassic, sponge reefs the mid-Triassic, and coral reefs the Late Triassic. Throughout the Triassic great changes occur in the latitudinal distribution of reefs, expanding from an equatorial and tropical distribution in the mid-Triassic to a maximum of around 35° N and S in the Late Triassic. A mass extinction event at around the Triassic/Jurassic boundary caused a near extinction of reef biota and earliest Jurassic (Hettangian) reefs are extremely rare. Reef complexes, regarded now as classic locations, grew in the Tethys. Those in the Dolomites of northern Italy (**Figure 11**) were strongly controlled by sea-level changes, volcanic eruptions, and palaeobathymetry, and exhibit a wide range of facies and at a seismostratigraphic scale. Allochthonous (reef-derived) material is abundant; back-reef lagoonal and intertidal facies are also well developed. Some reefs have atoll-like geometries. The principle constructors in some earlier cases were encrusting and binding organisms, often termed 'problematica', of

Figure 10 Permian Capitan Reef Complex, McKittrick Canyon, New Mexico. Reef facies (massive face at upper left); fore-reef talus (comprising lower ground and area to the right).

Figure 11 Triassic reef-associated facies, talus, and bedded back-reef lagoonal sediments. Dolomites of northern Italy.

which one such is 'Tubiphytes'. Later reefs were made by calcareous sponges, corals, red algae, and problematic calcareous binders. Flügel emphasizes that although Late Triassic reefs exhibit some 'modern' traits as would be found in recent coral thickets, there are some important differences too. Modern corals are primarily constructors, whereas Triassic ones seem to have been bafflers and binders. There would have been fewer nooks and crannies in Triassic reefs, and there appears to have been significantly less by way of bioerosion, predation, and grazing. Finally, significant numbers of Triassic corals were adapted to colonizing unstable sandy and muddy substrates, whereas modern species require a firm seafloor.

Following their initial setback resulting from the end-Triassic mass extinction reefs, often with corals as a significant component, reappeared in the Early Jurassic, firstly in Morocco (Figure 12). As sea-level rose eustatically through the period, to reach a peak in the Oxfordian, so reef fortunes rose too. With the break-up of Pangaea and the formation of new seaways, reefs proliferated. Coral reefs are confined between about 30° N and S during the Early Jurassic, expanding to 35° N in the mid-Jurassic and to 45° N in the Late Jurassic. Resedimented coral boulders occur within Kimmeridgian deep-water debris flows on the margins of the North Sea in northern Scotland. Jurassic reefs were constructed by a range of organisms (sponges, microbes, bivalves (e.g., oysters and ancestors to the Cretaceous rudistids)), as well as corals.

The Cretaceous (Figures 13 and 14) saw the rise of rudists, heterodont bivalves that became very important reef builders that underwent a great proliferation after the Early Aptian (caprinids and caprotinids), particularly through the Tethyan region. They range overall from end-Jurassic to end-Cretaceous, once again tracking the general trend in eustatic change

Figure 12 Lower Jurassic reef facies, Morocco. Photograph by Professor A Hallam.

Figure 13 Cretaceous rudist reefs limestones, Mount Carmel, Israel.

Figure 14 View of *in situ* rudists within Carmel reef wackestones, Israel.

from Early Cretaceous lowstand to Late Cretaceous highstand and subsequent fall. Another group of rudists, the radiolitids, appear in the Aptian and underwent a major proliferation through the Late Cretaceous, despite a setback at the Cenomanian–Turonian boundary. From the Cenomanian there is a great expansion in the geographical range occupied by rudistids with a new group, the hippuritids, markedly pipe-like forms, contributing in a major way to build-ups. In the Early Cretaceous the rudistids are associated with corals, but there is an increasing tendency for ecosystems to become rudist-specific, to the extent that biological mechanisms have often been invoked to explain the dominance of single species within communities. Rudist reef complexes are associated with high-energy shelf grainstones and platform margin talus, and around the margins of the Gulf of Mexico and in the Middle East (Arabian/Persian Gulf) areas these build-ups, and their associated facies, are associated with prolific hydrocarbon production. Rudist reefs were confined to between about 30° N and S during the Late Cretaceous, whereas reefs with corals extended further polewards (by about 10° further of latitude). However, reef communities suffered badly at the Cretaceous–Tertiary boundary and, although mud-mound and bryozoan build-ups occur in the Danian (e.g., in Denmark) and through the Eocene, reef systems did not fully recover until the Oligocene when they took on a nearly modern character.

In the Tertiary reef systems reached their maximum northern extent during the Miocene (Burdigalian (**Figure 15**) to Messinian (**Figures 16, 17** and **18**)) when extensive reefs grew in the Mediterranean area. Tortonian reefs have a diverse coral biota, but species were progressively lost into the Messinian in this area where monospecific associations (dominated by the tolerant genus *Porites*) prevailed shortly

Figure 15 Fossilized fringing reef (wedge-shaped unit thickening seaward from below small tower). Reef prograded over talus and grew from a fossil cliff-line comprising Triassic Muschelkalk Dolomites (foreground). Burdigalian (Miocene), Banyalbufar, Mallorca, Spain.

Figure 16 Messinian (Miocene) prereef carbonate sands (white clinoforms just above sea-level); reef (massive brownish limestone in mid-section of cliff) and well-bedded postreef sediments above. Cap Blanc, Mallorca, Spain.

Figure 17 Finger *Porites* framestone; lower reef-front. Messinian (Miocene), Cap Blanc, Mallorca, Spain.

Figure 18 Massive *Porites* framestone, upper reef-front. Messinian, Cap Blanc, Mallorca, Spain.

before the onset of the 'salinity crisis'. Debate still continues as to whether lowering temperatures or rising salinity caused these restrictions, but following the resumption of normal marine sedimentation in the Pliocene coral reefs were never again developed at these latitudes. Today, Bermuda and the Gulf of Suez represent the most northerly outposts of shallow, warm-water reefs in the northern hemisphere, and in the latter reefs grow on the maritime fringes of alluvial fans, a situation which also occurred along this rift during the mid-Miocene.

In deep cold waters lacking a high turbidity, and with slow sedimentation rates, coral reefs do grow today, but these are made by the ahermatypic (non-zooxanthellate) coral *Lophelia*. These form major edifices along the Atlantic margins, but their growth rates are slow and they are extremely susceptible to disturbances from fishing and other offshore activities, which is why they are protected by nature conservancy measures.

See Also

Fossil Invertebrates: Brachiopods; Corals and Other Cnidaria; Porifera. **Mesozoic:** Triassic; Jurassic; Cretaceous; End Cretaceous Extinctions. **Palaeoclimates**. **Palaeozoic:** Cambrian; Silurian; Devonian; Carboniferous; Permian; End Permian Extinctions. **Sedimentary Environments:** Carbonate Shorelines and Shelves. **Sedimentary Rocks:** Limestones.

Further Reading

Bekker A, Holland HD, Wang PL, *et al.* (2004) Dating the rise of atmospheric oxygen. *Nature* 427: 117–120.

Bonney TG (ed.) (1904) *The Atoll of Funafuti*. London: Royal Society.

Cestari R and Sartorio D (1995) *Rudists and Facies of the Periadriatic Domain*. San Donato Milanese: Agip S.p.A.

Darwin C (1842) *Structure and Distribution of Coral Reefs*. Reprinted 1962 by University of California Press with foreword by HW Menard.

Embry AF and Klovan JE (1971) The Late Devonian reef tract on northern Banks Island, N.W.T. *Bulletin of Canadian Petroleum Geology* 19: 730–781.

Geldsetzer HHJ, James NP, and Tebbutt GE (1988) Reefs, Canada and adjacent areas. *Memoir of the Canadian Society of Petroleum Geologists* 13.

Heckel PH (1974) Carbonate buildups in the geological record: A review. In: Laporte LF (ed.) *Reefs in Space and Time*. Special Publication, vol. 18, pp. 90–154. Tulsa: Society of Economic Paleontologists and Mineralogists.

Hollingworth NTJ and Tucker ME (1987) The Upper Permian (Zechstein) Tunstall reef of north east England: Palaeoecology and early diagenesis. In: Peryt TM (ed.) *The Zechstein Facies in Europe*. Lecture Notes in Earth Science, vol. 10, pp. 23–50. Berlin: Springer-Verlag.

James NP (1983) Reef environment. In: Scholle PA, Bebout DG, and Moore CH (eds.) Carbonate depositional environments. *Memoirs of the American Association of Petroleum Geologists* 33: 345–440.

James NP and Ginsburg RN (1979) *The Seaward Margin of Belize Barrier and Atoll Reefs*. Special Publication of the International Association of Sedimentologists, 3.

Kiessling W, Flügel E, and Golonka J (eds.) (2002) *Phanerozoic Reef Patterns*. Special Publication. Tulsa: Society of Economic Paleontologists and Mineralogists.

Longman MW (1981) A process approach to recognizing facies of reef complexes. In: Toomey DF (ed.) *European Reef Models,* Special Publication, vol. 30, pp. 9–40. Tulsa: Society of Economic Paleontologists and Mineralogists.

Runnalls LA and Coleman ML (2003) Record of natural and anthropogenic changes in reef environments (Barbados West indies) using laser ablation ICP-MS and sclerochronology on coral cores. *Coral Reefs* 22: 416–426.

Scott RW (1990) *Models and stratigraphy of mid-Cretaceous reef communities, Gulf of Mexico*. Concepts in

Sedimentology and Paleontology, 2. Tulsa: Society of Economic Paleontologists and Mineralogists.
Sellwood BW (1986) Shallow-marine carbonate environments. In: Reading HG (ed.) *Sedimentary Environments and Facies*, pp. 283–342. Oxford: Blackwell Scientific.
Wood R (1999) *Reef Evolution*. Oxford: Oxford University Press.
Wright VP and Burchette TP (1996) Shallow-water carbonate environments. In: Reading HG (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, pp. 325–394. Blackwell Science.

Shoreline and Shoreface Deposits

J Howell, University of Bergen, Bergen, Norway

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

A shoreline is the interface between a standing body of water, typically the sea, and a landmass. Over geological time, shorelines are highly transient features and the depositional and erosional systems associated with them are extremely dynamic. The shoreface *sensu stricto* is defined as the zone that lies between the low tidemark and the fair weather wave base (the maximum depth at which everyday wave movements affect the seabed). The shoreface is part of a linked depositional system that includes all of the coastal region affected by wave action, from the top of the beach, seaward, to the mean storm wave base (the maximum typical depth at which the storm waves move sediment on the seafloor) (*see* **Sedimentary Environments:** Storms and Storm Deposits).

At the shoreline, wave processes are the most significant agents of erosion, sediment transport, and deposition, although fluvial, tidal, and biogenic processes are also important and may be locally dominant. Sediments deposited in the system may be siliciclastic, carbonate, or a mixture of the two, and grain sizes can range from those of clay particles to boulders. The position of the shoreline is highly sensitive to changes in sea-level and sediment supply, and understanding shoreline systems is central to sequence stratigraphy and basin analysis. Shoreface deposits form important hydrocarbon reservoirs and aquifers in many parts of the world, and some contain placer mineral and gem deposits. Modern systems are environmentally sensitive areas and many of the world's coastlines are heavily populated and susceptible to sea-level change. The study of shoreline processes and deposits includes aspects of geology, geomorphology, biology, and oceanography.

Both physical and biogenic processes (*see* **Sedimentary Environments:** Carbonate Shorelines and Shelves) occur at shoreline systems. All clastic, coastal depositional systems can be classified within a triangular diagram that summarizes the relative importance of fluvial, tidal, and wave processes (Figure 1A). Away from fluvial input points, the key controlling factor is the tidal range. In areas with a moderate to low tidal range, narrow beaches dominated by wave action occur. Conversely, in areas with a high tidal range, broad low-lying, muddy tidal flats are created (Figure 1B). The wave-dominated, beach-related systems that make up the majority of the world's coastlines are the focus here (but *see* **Sedimentary Processes:** Fluvial Geomorphology).

Wave Processes

Surface water waves are formed by the movement of the wind across a standing body of water. Wind shear effectively drags the upper layers of the water and horizontally compresses them, producing waves. The key features of a wave are its length and height (Figure 2). Within the water body, the vertical motion at the water surface sets up circular orbital motion that is translated downward. This downward translation of energy typical moves the water in a series of orbital motions to a depth that is approximately half of the wave length. This depth is termed the 'wave base' (Figure 2), and waves that occur in water that is deeper than the wave base will not move sediment on the seafloor. Because the depth to the wave base is controlled by the size of a wave, it varies both between basins and over time within an individual basin. In large, open bodies of water with a large fetch (such as the major oceans), the wave base may be in excess of 100 m, whereas in enclosed seas and lakes, it may be less than 10 m. Wave size and the depth to the wave base also vary within an individual setting on a daily and seasonal basis, depending on changes in the weather. During storms, short-term, larger waves result in a deep wave base, whereas during fair weather periods, the wave base is much shallower. It is common to consider marine depositional systems in terms of two hypothetical surfaces,

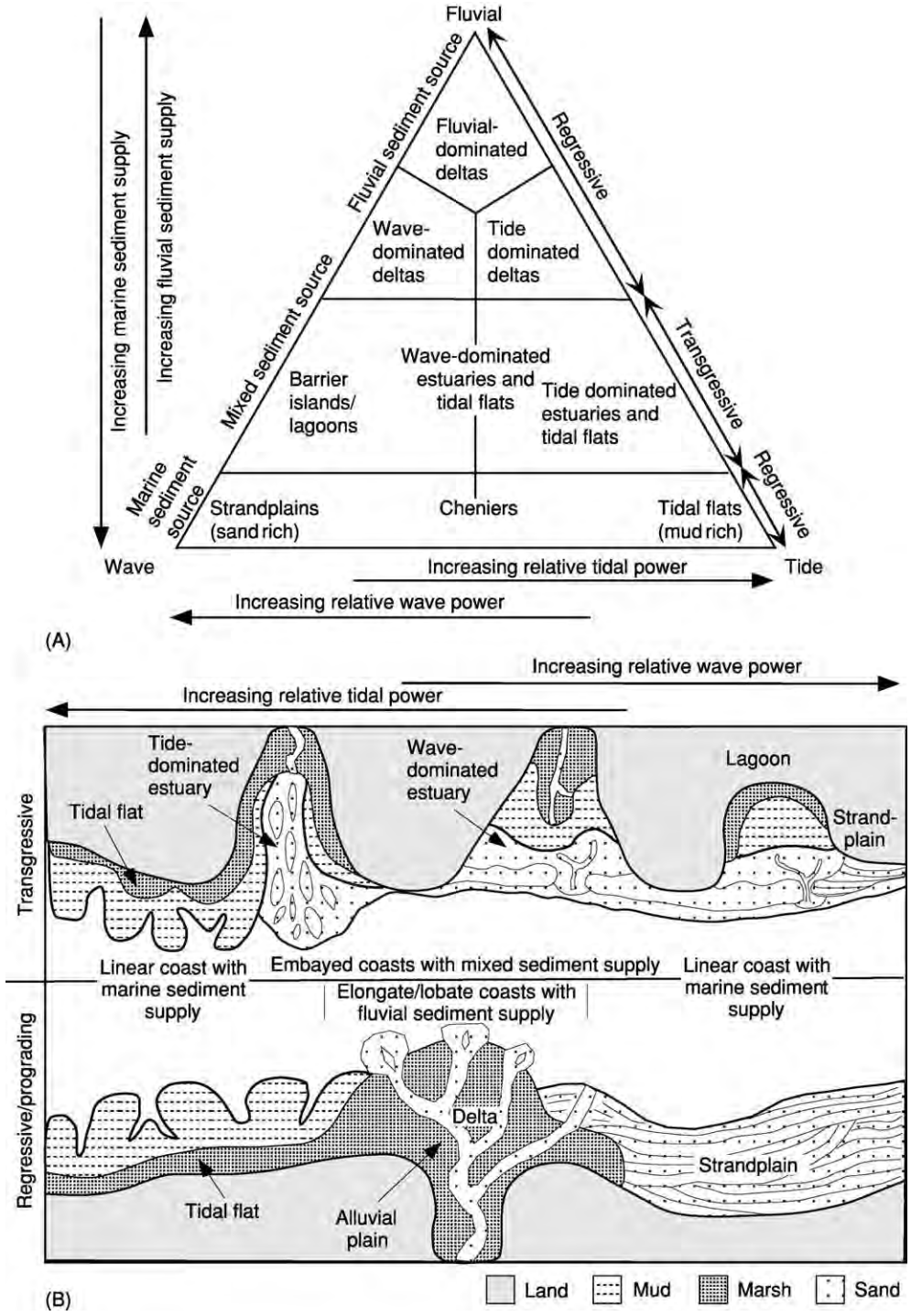


Figure 1 Classification of clastic shallow marine systems. (A) The relative importance of fluvial, tidal, and wave processes. (B) Transgressive and progradational coastal systems with varying degrees of fluvial, tidal, and wave influence. Reproduced with permission from Dalrymple RW, Zaitlin BA, and Boyd R (1992). Estuarine facies models: conceptual basis and stratigraphic implications. *Journal of Sedimentary Petrology* 62: 1130–1146.

the mean storm wave base (SWB) and the mean fair weather wave base (FWWB). The mean SWB is the depth at which typical storms affect the seabed for short periods of time (in the order of hours to days); the FWWB reflects the typical wave climate that is present within the basin between storm events.

Waves that are formed in the open ocean are pushed towards the land. As waves approach the coastline,

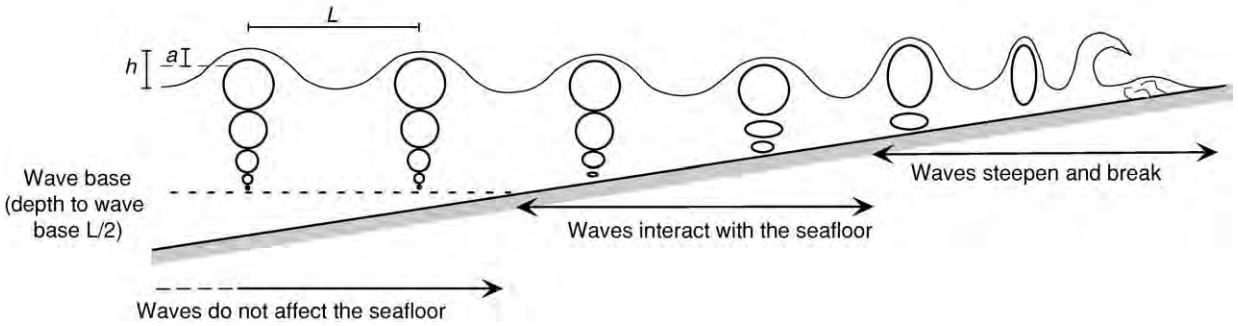


Figure 2 Key features related to water waves approaching a smooth dipping coast; *L* is the wave length, *h* is the wave height, and *a* is the wave amplitude.

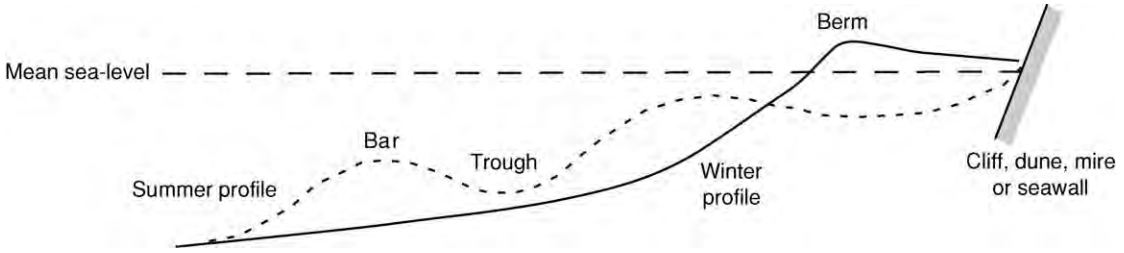


Figure 3 Typical summer and winter beach profiles. During the summer months, fair weather waves move sediment onshore and build up a berm. Winter storms erode the beach and deposit the sediment in the shoreface and offshore transition zone.

the water depth becomes less than the wave base and the waves interact with the seafloor. Waves hitting the coast may be classified as reflective, dissipative, or intermediate forms. Reflective waves strike very steep coastlines (with features such as cliffs or a seawall) and the majority of the energy is reflected directly back into the water body. Purely dissipative waves gradually lose their energy through frictional interaction with a very gently dipping seafloor and do not break at all. A majority of the waves that affect active shorelines are intermediate between these two forms. In intermediate cases, interaction with the seafloor slows the waves and pushes them up; the waves steepen until they eventually collapse or break. If the shoreline is steep, the waves will break as plunging forms, whereas if it inclines more gently, the waves will tend to surge in a landward direction. Steeper shorefaces with plunging waves tend to be simple systems in which the waves break directly on the beach. Most shorefaces dip gently and have a complex topography, including numerous bar forms over which the waves may break before reaching the shoreline, to reform and then break again. These systems include complex sets of plunging and surging waves.

During fair weather periods, waves tend to move sediment in an onshore direction because water loss due to percolation into permeable beach sediments reduces the efficiency of the backflow. Large storms tend to erode the beach and to deposit sediment offshore, or in the case of barrier systems, in the lagoon behind the beach. During the summer months, when few storms occur, there is a net onshore migration of sediment; during the winter storm period, the beach is frequently eroded and does not have time to be repaired. Consequently, typical summer and winter profiles for beaches may be defined (Figure 3). The preserved rock record is a complex mixture of the depositional products of the repeated seasonal changes.

If waves approach the shoreline perpendicularly, there is no net movement of sediment along the coast. However, in many cases, waves from the open ocean approach the shoreline obliquely. This results in an oblique onshore movement of sediment with the incoming wave. The returning swash tends to be more perpendicular to the shore, which in turn leads to migration of sediment parallel to the shore; this is termed 'longshore sediment transport', or longshore drift (Figure 4). Transport of sediment by longshore drift is a key factor in moving volumes of sediment and in controlling the morphology of wave-dominated shoreline systems.

The sedimentary structures produced by waves show an increase in bedform asymmetry towards the shoreline, passing from oscillatory, symmetrical wave ripples near the wave base into combined-flow wave ripples as the water becomes shallower (Figure 5). In the shallowest part of the system, shear between the water mass and the seabed can

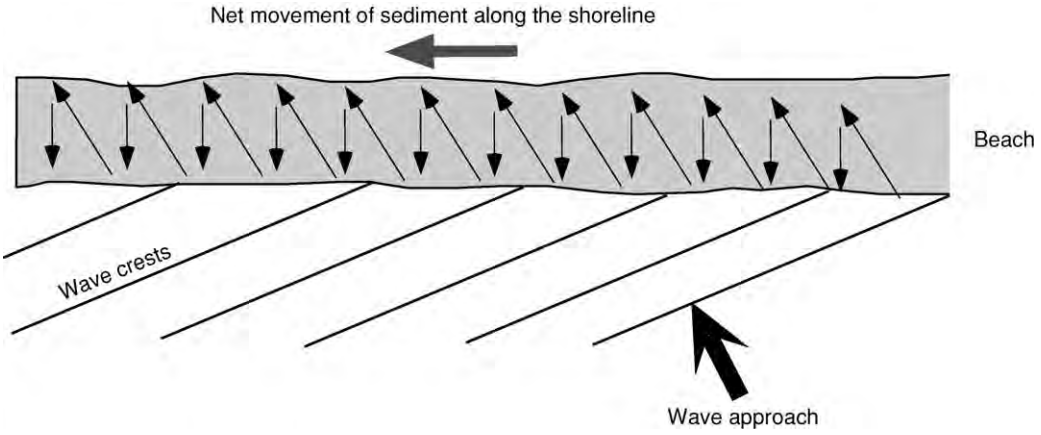


Figure 4 Longshore sediment transport (longshore drift). Waves approach the shoreline at an oblique angle. Sediment is moved up and across the beach by the incoming wave; the returning flow is weaker and tends to be straight back (i.e., shore normal). The net result is that sediment is moved along the coastline.

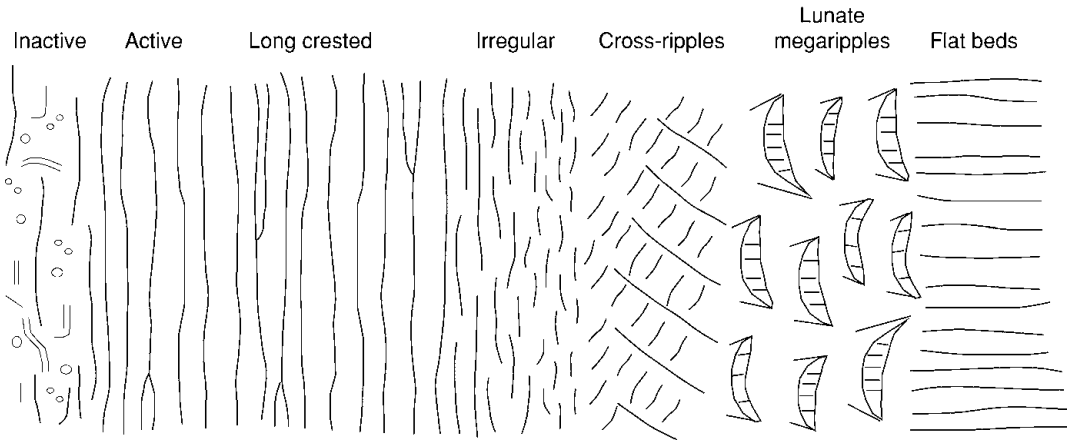


Figure 5 Plan view of the shoreface showing the distribution of bedforms resulting from the onshore approach of shoaling waves. Reproduced with permission from Clifton NE, Hunter RE, and Phillips RL (1971) Depositional structures and processes in the non-barred high-energy nearshore. *Journal of Sedimentary Petrology* 41: 651–670.

produce a variety of current-generated structures, including ripples, plane beds, and dunes, producing trough and tabular cross-bedding. Breaking waves generally produce upper-phase plane beds, although the returning swash current may also generate current ripples. In the deeper part of the system, combined flow associated with very large storm-related waves produces compound, three-dimensional low-angle bedforms called hummocks, which in turn produce hummocky cross-stratification (*see* **Sedimentary Environments:** Storms and Storm Deposits).

Classification and Geomorphology of Shoreline Systems

Geomorphologically, coastal systems can be characterized by the medium- to long-term movement of the shoreline (Figure 1B). In transgressive systems, the shoreline moves towards the hinterland and land is lost to the sea, whereas in regressive systems, the shoreline moves basinward and new land is created on the coastal plain. These movements are controlled by changes in sea-level and by the amount of sediment that is supplied to the system. Sea-level change may occur due to global changes in sea-level (eustasy) or may be driven by local uplift or subsidence within the basin. Commonly, a combination of the two forces is in play; this is termed 'relative sea-level change'. If relative sea-level is falling, the shoreline will move basinward (termed a 'forced' regression). Sea-level rise does not always lead to transgression, because the movement of the shoreline also depends on whether enough sediment is supplied to the system to fill the new space that is created. If more sediment is supplied than space is created, then the shoreline will prograde (i.e., will be regressive). If, however, the

rate of space creation is greater than the rate of sediment supply, then the shoreline will move landward (i.e., will be transgressive).

Transgressive shorelines may be further subdivided into low- and high-relief forms. Low-relief transgressive shorelines generally occur when relative sea-level rise floods a former delta or coastal plain. In such cases, systems may contribute to the geological rock record and their deposits will either be barrier islands in systems with a moderate to low tidal range, or tidal flats in systems with a higher tidal range. High-relief transgressive shorelines occur where the sea-level rise is flooding bedrock. In such cases, the systems are erosional. High-relief transgressive shorelines do not contribute to the geological rock record, although they are an important part of modern systems (see later).

Progradational Wave-Dominated Shoreline Systems

Progradational shoreline systems, which contribute significantly to the geological rock record, are classified by the dominant depositional mechanism: wave, fluvial (*see* **Sedimentary Processes:** Fluvial Geomorphology), or tidal. Progradational wave-dominated systems include shorefaces and wave-dominated deltas. In a shoreface, the sediment is introduced by longshore drift; in a wave-dominated delta, the sediment is originally introduce by a fluvial system and is then reworked by wave processes. In reality, especially in the rock record, it can be very difficult to distinguish the two, and because the depositional processes are the same, it is useful to consider them as largely synonymous. In both cases, the system includes a beach that lies between the low tidemark and the next significant feature in a landward direction, which may be dunes, a coastal forest, a mire, a cliff, or in many modern systems a seawall. Between the low tidemark and the FWWB is the shoreface, which includes the sediment that is being reworked by fair weather wave action. Seaward of the FWWB is the offshore transition zone where, for the majority of time, deposition consists of very fine-grained suspended sediment, with brief influxes of storm-derived coarser deposits. Seaward of the SWB is the offshore shelf where waves no longer play a role in the introduction or deposition of sediment (Figure 6).

Within a wave-dominated system, sediment is introduced by longshore drift, either from the reworking of a fluvial system (wave-dominated delta) or from wave erosion of the bedrock or older shoreline deposits. In plan view, a variety of features may exist within the depositional system. Offshore, seaward of the point at which the storm wave base interacts with the seafloor, deposition is restricted to mud deposited from suspension or to coarser sediments introduced by sediment gravity flows. In the zone between the SWB and FWWB, storms deposit hummocky cross-stratified

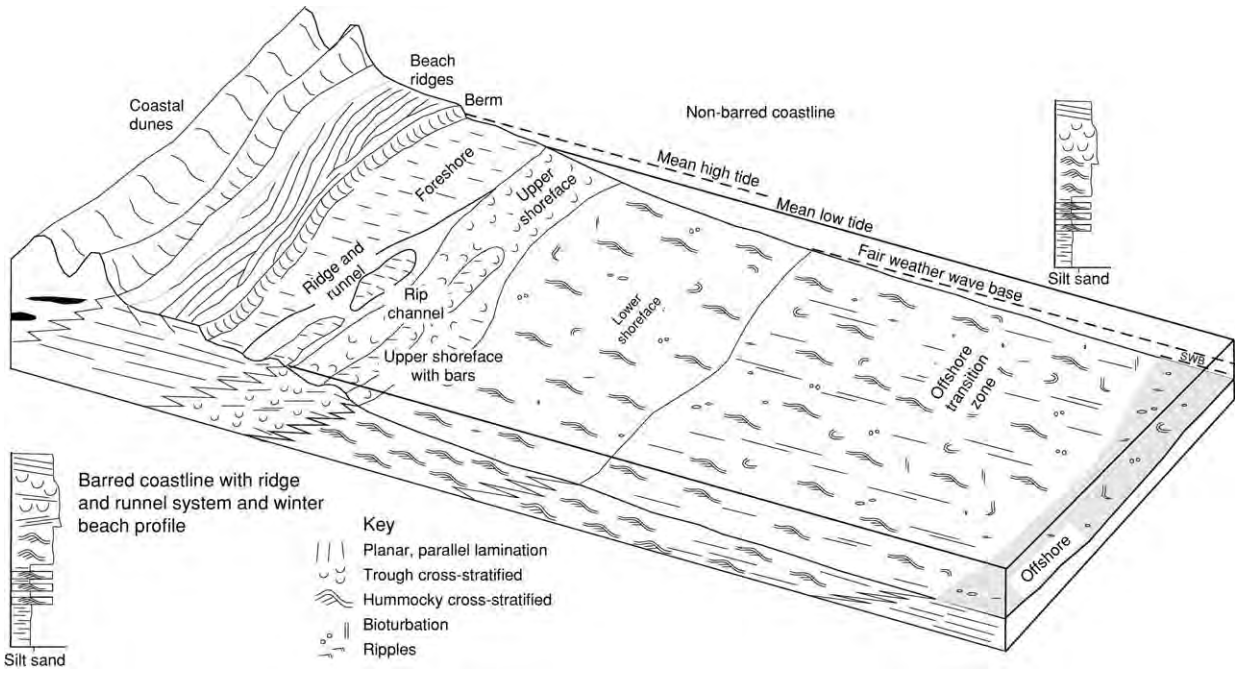


Figure 6 Facies model for a progradational wave-dominated shoreline system. View shows both barred and non-barred coastlines. Sediment is introduced by longshore drift. No scale is implied because the depths to the fair weather wave base and the storm wave base (SWB) are controlled by the size of the basin and the local climate.

(HCS) sheets of sandstone, typically less than 0.5 m thick; thin layers of coarser material (up to granule grade) may also be deposited if such material is available on the beach. These sheets have erosive bases whereas the top surfaces commonly exhibit wave ripples formed during the waning part of the storm. Bioturbation of the storm bed by opportunistic organisms is common. Between storm events, mud is deposited from suspension.

The shoreface is the interval that lies landward of the point at which the fair weather wave base interacts with the seafloor. Within the lower part of the shoreface, sediment is deposited as sheets of HCS sand during storms, and fair weather wave action prevents the deposition of mud between the storm events, resulting in a sand-rich depositional system. The point at which the fair weather waves actively move sediment in a landward direction marks the base of the upper shoreface. The upper shoreface may be a simple seaward-dipping surface or it may contain one or even several large bar forms. Simple, 'non-barred' shorelines are generally steeply dipping and coarse grained. They are covered with small dune forms that migrate in a landward and longshore direction; plunging waves generally produce seaward-dipping planar laminations on a relatively simple beach (**Figure 7**). Non-barred shorelines prograde through the steady addition of sediment to the beach and shoreface. Modern examples include the Pacific coast of North America. The Book Cliffs of Utah display superb Cretaceous examples of non-barred shoreline deposits (**Figure 7D, E, H, and J**).

Barred shorelines are more common than non-barred systems and contain large (up to 50 m wide and hundreds of metres long) shore-parallel barforms separated by narrow rip-channels that run perpendicular to the shoreline. Water and sediment are carried in an onshore direction by incoming waves that break over the bars and in an offshore direction by currents running through the rip-channels. The combination of these processes generates circulatory cells (termed 'longshore cells') within the shoreface. Barred shorelines are extremely dynamic and change markedly on a seasonal basis. Bars are generally built during the winter months as storms move sediment offshore from the beach. Bars can be symmetrical or asymmetrical and contain a series of seaward-dipping planer-beds deposited by storms on the seaward side and trough cross-bedding and landward-dipping planar-beds deposited by fair weather wave reworking. Rip-channels are either erosional or contain seaward-directed trough cross-bedding. During the summer months, the bars migrate landward and may emerge above the low tidemark, where they become part of the beach.

Within barred shorelines, the beach is more complex, compared to the simple seaward-dipping planar surface that exists in non-barred systems. Beaches within barred coastlines comprise a series of subzones, not all of which are always present. The lower intertidal portion includes a series of low, shore-parallel barforms, or 'ridges', separated from one another by hollows, or 'runnels' (**Figure 6**). Sedimentary features within this ridge and runnel system include low-angle dipping planar laminations on the ridges and shore-parallel current ripples formed by the draining of the runnels as the tide recedes. These ridge and runnel systems generally develop during the winter months when storms occur most frequently, but may also be related to the emergence of landward-migrating bars during the summer period.

The main part of the intertidal zone is termed the 'foreshore' and comprises a series of seaward-dipping planar surfaces formed by the breaking waves. The mean grain size of the sediment will strongly control the dip of the foreshore. If the grain size is coarse, the water brought onshore by the breaking waves will rapidly percolate into the sediment, reducing the backwash and producing a steep beach. Conversely, if the sediment is fine and less permeable, the beach will be gently dipping. The top of the foreshore is marked by a ridge called a 'berm'. The berm builds during the summer months as the fair weather waves push sediment onto the beach. In the absence of large storms, beaches may develop a series of parallel berms.

Beyond the berm lies the backshore. This is the area above the mean high tidemark, where sediment is introduced only during extremely high tides or by storms. If a lot of coarse (cobble to pebble grade) material is available, a storm ridge may be formed. This is typically composed of well-rounded, well-sorted, discoid clasts with a seaward imbrication that are thrown up during storms and are coarse enough to resist fair weather reworking. In the absence of a storm ridge, the backshore may comprise partially vegetated sand flats in which aeolian processes rework the sediment that is infrequently washed in.

Vertical Succession through a Wave-Dominated Shoreface Succession

The vertical succession deposited during the progradation of a wave-dominated shoreline succession reflects the lateral facies relationships as described. The onshore coarsening of grain size and increasing bedform asymmetry are recorded vertically in the rock record (**Figure 6**). A typical vertical succession from base to top will exhibit mudstones and siltstones deposited in the offshore setting at the base. Moving upward, the appearance of the first significant

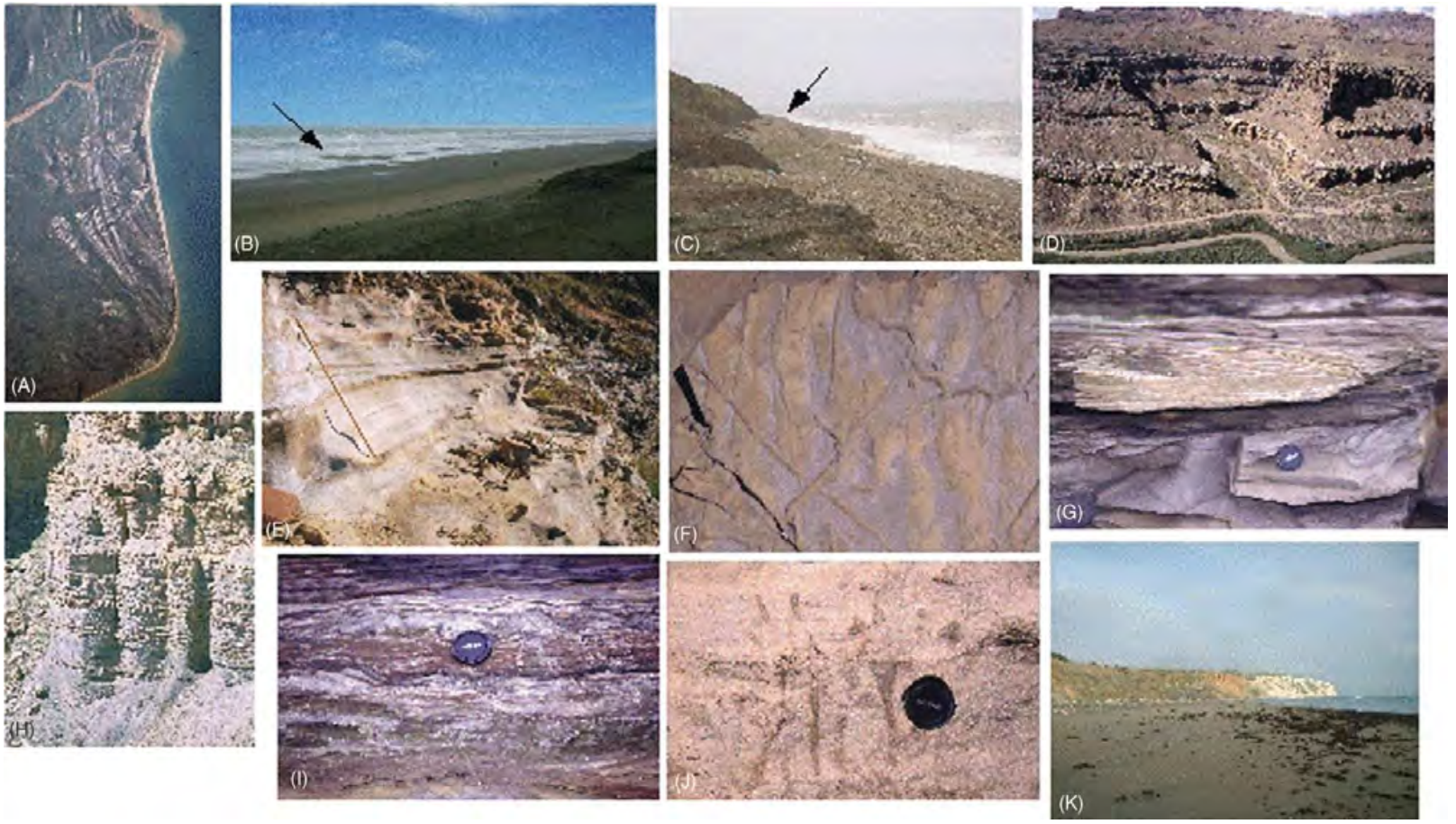


Figure 7 Modern and ancient wave-dominated coastlines. (A) Space shuttle photo showing the linear shoreline and well-developed beach ridges in the wave-dominated Paraiba Delta, Brazil. Image courtesy of Earth Sciences and Image Analysis Laboratory, National Aeronautics and Space Administration Johnson Space Centre (see website at http://eol.jsc.nasa.gov). (B) Modern wave-dominated beach (Dutch coast); note the low relief of the foreshore, the dunes on the left-hand side, and the ridge and runnel system (arrow) at the lower part of the beach. Photograph by Katharine Mathers; reproduced with permission. (C) Modern coarse-grained, non-barred wave-dominated shoreline (Isle of Wight, UK); note the well-developed berm (arrow) and the steep dip of the foreshore. (D) Book Cliffs of eastern Utah, USA, showing world-class outcrops of wave-dominated shoreface deposits. Laterally continuous, white sandstones are foreshore and upper shoreface deposits of successive progradational episodes. (E) Upper shoreface and foreshore deposits from the Book Cliffs of eastern Utah. The low-angle dipping planar laminations were deposited by waves breaking on a beach, and the trough cross-stratified sandstones (base of picture) were deposited by shoaling waves in the upper shoreface (pole = 1.2 m). (F) Bioturbated and rippled lower shoreface sandstones (Wasatch Plateau, Utah). (G) Hummocky cross-stratified sandstones from the offshore transition zone. (Carboniferous; Berwick, UK). (H) Offshore transition zone deposits comprising interbedded hummocky cross-stratified sandstones (brown) and siltstones (grey), from the Book Cliffs of eastern Utah. (I) Bioturbated, wave-rippled heteroliths from the offshore transition zone (Berwick, UK). (J) Bioturbated shoreface deposits. Low-angle, swaley cross-stratification cut by large, vertical *Rosselia* burrows (Wasatch Plateau, Utah). (K) Modern erosive beach from the Isle of Wight. Note the landslip deposits in the midground and the steep, retreating cliffs of chalk in the background. The foreshore is covered in a thin veneer of sand reworked from the cliffs.

hummocky cross-stratified storm beds represents the SWB and marks the base of the offshore transition zone (OTZ), which is characterized by interbedded sandstones with HCS and bioturbated mudstones. There is an upward increase in the thickness and the proportion of sandstone beds within the OTZ to the point at which they become amalgamated and no significant mudstones interbeds are observed. This point in the succession represents the position of the FWWB. The overlying amalgamated storm deposits are deposits of the lower shoreface. The lower shoreface also locally includes intervals that are heavily bioturbated, and top-truncated hummocks sometimes termed 'swaley cross-stratification' (SCS).

As previously noted, the upper part of the shoreface in a non-barred coastline will comprise simple landward and longshore migrating dunes that produce trough cross-bedding; this cross-bedding passes up into the seaward-inclined planar cross-bedding that represents the foreshore. In a barred coastline system, the vertical succession will be more complex, with a series of landward- and seaward-dipping planar stratified beds, interbedded with trough and ripple cross-stratified units and frequent erosion surfaces. In such cases, it may be difficult to distinguish the upper shoreface from the beach, except by the vertical transition to the overlying backshore deposits.

Muddy Coastlines and Cheniers

The majority of depositional wave-dominated shorelines are sand rich because waves will rework mud-grade material, moving it offshore in suspension. However, if mud is the only material present and is available in large enough quantities, muddy wave-dominated coasts can occur. Such systems build very large, low-angle dipping muddy barforms that are held together by the cohesiveness of the sediment. The very low angle dissipates all wave action and prevents wave erosion. Periodic influxes of sand and/or concentrations of shell debris can produce seaward-dipping beds associated with prominent ridges on the beach. These are termed 'cheniers'. Such systems are seen on the Surinam coast of South America and in areas adjacent to the Mississippi delta.

Low-Relief Transgressive Coastlines

Low-relief transgressive shorelines occur when the rate of sediment supply is less than the rate of relative sea-level rise and the coastal zone is a low-lying feature, typically a former depositional surface such as an abandoned delta top or coastal plain. In systems with a moderate to low tidal range, the shoreline will generally develop barrier islands. Barrier islands are linear, shore-parallel strips of land that separate the open ocean from a sheltered lagoon or intertidal flat area; the intertidal area may be connected to the marine realm through tidal inlets. There are extensively developed barrier islands on the Texas gulf coast (Figure 8). Many of the processes that are active on the seaward side of the barrier are identical to those described for the shoreface and beach systems in the previous section. The key difference is that

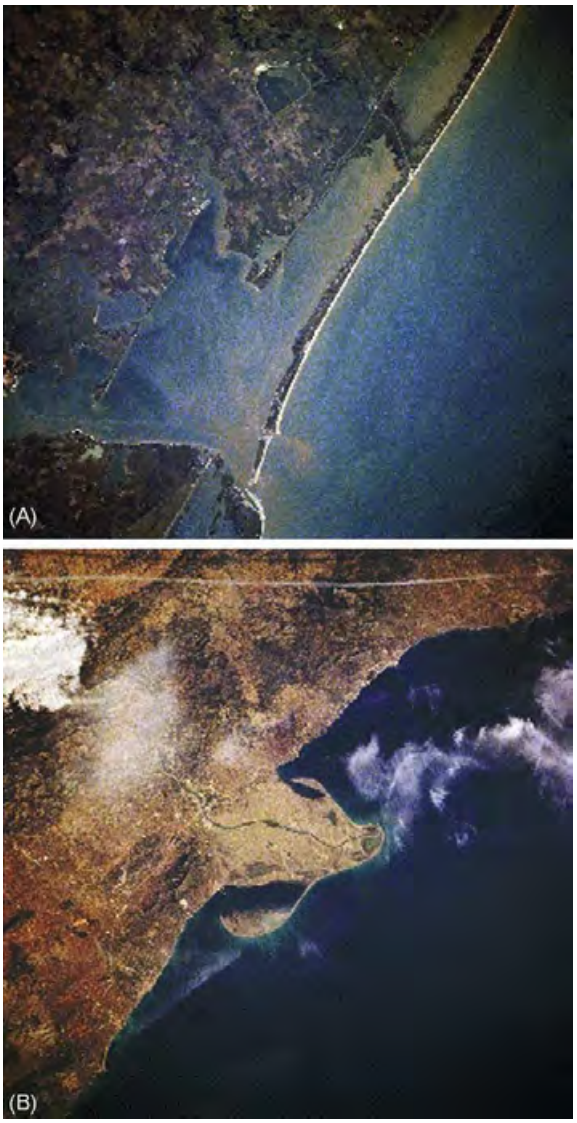


Figure 8 Space shuttle photographs of transgressive wave-dominated shorelines. (A) Matagorda Bay, Texas, USA. A narrow barrier island system shelters the lagoon from the open ocean. The irregular left-hand side of the barrier island is due to washover fans. Note the small tidal inlets towards the base of the photo. (B) Ebro Delta, Spain. Thirty-five dams along the river have significantly reduced the sediment supply to the delta, which has subsequently been transgressed due to relative sea-level rise caused by compactional driven subsidence. Note the development of spits and barriers on the wings of the delta. Both images courtesy of Earth Sciences and Image Analysis Laboratory, National Aeronautics and Space Administration Johnson Space Centre.

during large storms, some of the sediment is washed into the lagoons and is deposited as sheets of planar bedded and rippled deposits; these are termed 'washover fans' (Figure 9). Because sediment is removed from the seaward side of the barrier and is deposited on the landward side, the shoreline moves landward. This process, termed barrier 'rollover', can be relatively rapid. Parts of the modern-day Galveston Island in Texas have migrated 80 m in the past 100 years. Generally, the migrating barrier system leaves very little sediment behind and the geological record of the transgression is an erosive surface, often associated with a lag of pebble or shell debris, concentrating the coarsest material from the underlying unit. The barrier island will be preserved at the point at which it stops migrating landward and the system starts to prograde. If the sea-level rise is very rapid, the entire barrier complex may be preserved by drowning.

Barrier island systems that occur in mesotidal settings are cut by tidal inlets through which tidal flows connect the lagoon with the open marine realm. In such cases, the barrier is also locally reworked by tidal currents passing through inlets. Deposition associated with the tidal inlet will include small subaqueous deltas on the landward and seaward sides, termed 'flood' and 'ebb' tidal deltas, respectively. These deltas are composed of rippled and planar cross-stratified sandstones that interfinger with the shoreface deposits on the seaward side and with the lagoonal muds and washover deposits on the landward side. Deposition of trough and tabular cross-bedding, often showing tidal draping, may also occur within the inlet channel. This will frequently be reworked but can be preserved if the position of the inlet migrates laterally.

Although numerous authors have discussed progradational barrier systems, especially in modern settings, geologically such features are very short lived; they represent the transition from transgressive barrier systems to land-attached shorelines because, if the barrier progrades, the lagoon will rapidly (over geological time) be infilled. Progradational systems with a strong component of longshore sediment transport can also develop beach-attached spits and even barrier systems that also shelter a lagoon on their landward side; however, these lagoons may be quickly filled by sediment that is introduced from the landward side by runoff and is washed over by storms from the seaward side.

High-Relief Transgressive Coastlines

High-relief coastlines may occur when the system is in net erosion and the wave action produces a cliff. The dominant processes are wave action and mass wasting. Fluvial and tidal processes are negligible. Wave action undercuts the cliff, which eventually collapses onto the foreshore; the cliff sediment is then reworked by wave processes and is transported either offshore or along the coastline by longshore drift. In its simplest form, a rocky coastline typically comprises a cliff and a wave-cut platform (Figure 10). The nature of the bedrock geology, rates of relative sea-level fall, and the wave climate will characterize any particular example. Variations in the bedrock geology due to lithological variations and structural geology (presence of faults and fractures) will result in differential erosion, leading to a variety of landforms, including stumps, stacks, arches, and caves

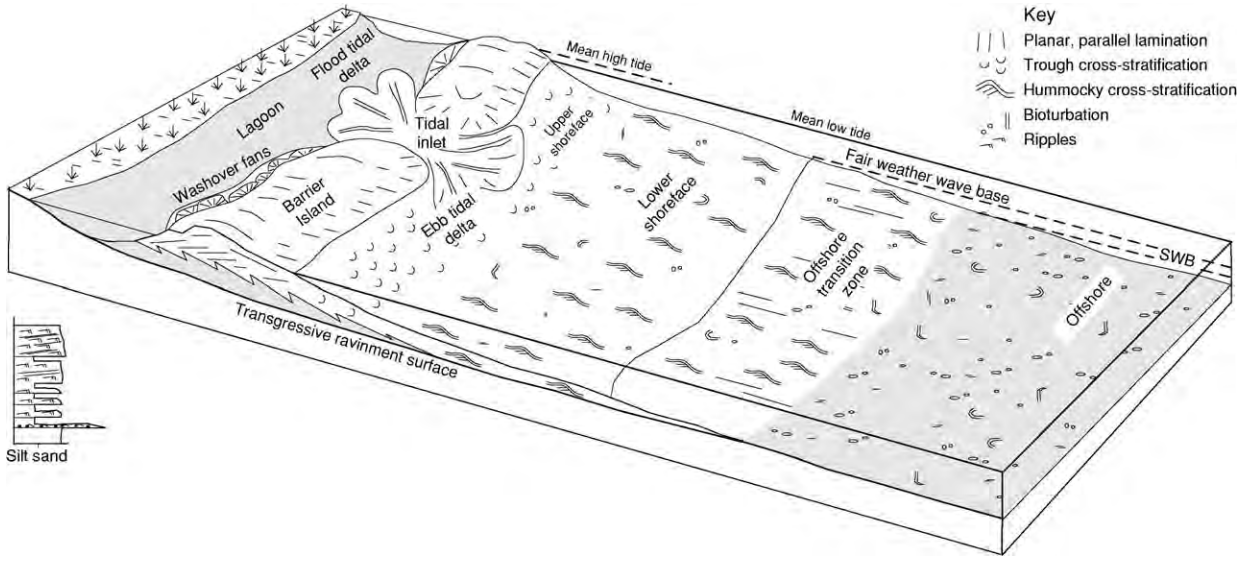


Figure 9 Facies model for a transgressive wave-dominated shoreline system. The barrier island migrates landward through erosion on the seaward side and by deposition in washover fans and flood tidal deltas. SWB, Storm wave base.

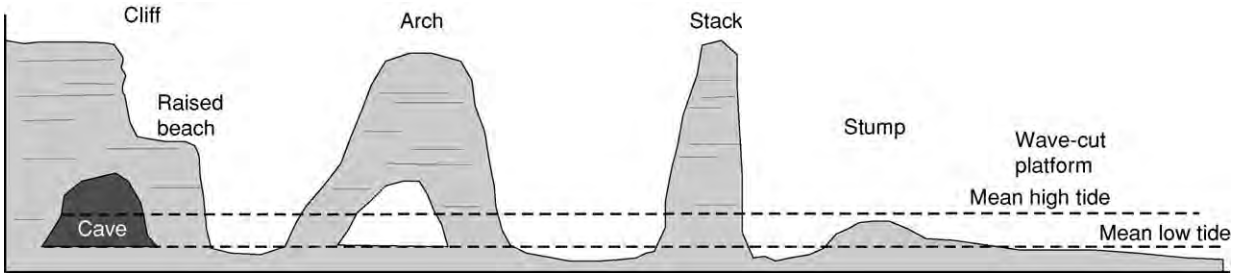


Figure 10 Geographical features of a high-relief erosional coastline.

(Figure 10). Changes in relative sea-level may lead to the formation of wave-cut terraces or marked wave-cut notches.

The geomorphological character of a cliff is a product of the bedrock geology and the climate. Hard, resistant bedrock will result in high, steep cliffs that retreat slowly and show pronounced wave-cut notches at the base; softer deposits will result in lower, less steep, rapidly retreating clifflines. The process of cliff retreat is enhanced in a wetter climate and is retarded in more arid regions. The wave-cut platform is a horizontal to low-angle seaward-dipping surface that occurs between mean high watermark and wave base. The wave-cut platform may be covered in a thin veneer of reworked sediment or may be bare. Where present, sediment may range in grain size from sand particles to boulders. Clasts are generally well rounded due to frequent reworking by waves. Wave movement of the retreating cliffline-derived sediment is a key process in the scouring of the platform.

As erosional features, rocky coastlines have little preservation potential in the geological rock record, although they may account for significant amounts of geological time. Preservation will be limited to an erosional surface that records the final position of the shoreline prior to a change in conditions that led to the onset of deposition. Immediately above this surface a thin veneer of sediment may be preserved.

See Also

Sedimentary Environments: Carbonate Shorelines and Shelves; Storms and Storm Deposits. **Sedimentary Processes:** Fluvial Geomorphology.

Further Reading

Boyd R, Dalrymple RW, and Zaitlin BA (1992) Classification of clastic coastal depositional environments. *Sedimentary Geology* 80: 139–150.

Coe A (ed.) *The Sedimentary Record of Sea-Level Change*. Cambridge: Cambridge University Press.

Clifton HE, Hunter RE, and Phillips RL (1971) Depositional structures and processes in the non-barred high-energy nearshore. *Journal of Sedimentary Petrology* 41(3): 651–670.

Davidson-Arnott RGD and Greenwood B (1976) Facies relationships on a barred coast, Kouchibouguac Bay, New Brunswick, Canada. *Beach and Nearshore Sedimentation. SEPM Special Publication No. 24*, pp. 149–168. Tulsa, OK: Society for Sedimentary Geology.

Davis RA and Fitzgerald DM (2004) *Beaches and Coasts*. London: Blackwell Science.

Davis RA, Fox WT, Hayes MO, and Boothroyd JC (1972) Comparison of ridge and runnel systems in tidal and non-tidal environments. *Journal of Sedimentary Petrology* 40(2): 413–421.

Dominguez JML, Bittencourt ACSP, and Martin L (1992) Controls on Quaternary coastal evolution of the east-northeastern coast of Brazil: roles of sea-level history, trade winds and climate. *Sedimentary Geology* 80: 213–232.

Elliott T (1986) Clastic shorelines. In: Reading HG (ed.) *Sedimentary Environments and Facies,* pp. 143–177. London: Blackwell Scientific Publications.

Heward AP (1981) A review of wave-dominated clastic shoreline deposits. *Earth Science Reviews* 17: 223–276.

Walker RG and Plint GA (1992) Wave and storm dominated shallow marine systems. In: Walker RG and James NP (eds.) *Facies Models: Response to Sea-Level Change,* pp. 219–238. Ottawa: Geological Society of Canada.

Storms and Storm Deposits

P Myrow, Colorado College, Colorado Springs, CO, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Numerous studies of modern shelves indicate that storms have wide-ranging effects in a spectrum of environments from outer shelves to coastal plains. Early oceanographic studies led to interest in finding analogues in the rock record. Resulting facies models for storm-influenced depositional systems were not published until the late 1970s and early 1980s, later than for many other depositional systems (e.g. fluvial, tidal-influenced shoreline, submarine fan). There are a number of reasons for this. First, the oceanographical study of modern storms is extremely difficult. Direct observations are challenging and dangerous. Quantitative analysis also required the development of current meters and other devices that could survive the effects of strong storm-generated currents and waves. Second, storms are extremely complex systems that include strong currents and waves, both of which vary both temporally and spatially during an individual storm. Third, although tempestites, individual storm-generated sandstone beds, were attributed to deposition from storm-generated currents fairly regularly by the mid-1970s, features considered to be diagnostic of deposition by storm processes, such as hummocky cross-stratification (HCS), were not identified early in the study of sedimentary structures. Many case studies of ancient storm-influenced deposits and modern settings indicate that storms have wide-ranging effects in a spectrum of environments from outer shelves to coastal plains. The most common storm deposits in the rock record are interbedded sandstone and bioturbated shale (Figure 1). The sandstone beds were deposited as a result of powerful near-bottom water motions produced by waves and currents, and these are the main focus of the following discussion. Additional storm-generated features include gutter casts, pot casts, gravel and conglomerate accumulations, and reworked or transported shell beds.

Figure 1 The deposits of storm-influenced shelves consist of tempestites or storm-generated sandstone beds and shale beds deposited between storm events. Note how the sandstone beds tend to pinch and swell in thickness, a common feature of tempestites. These strata, exposed in the central Transantarctic mountains, were deposited along the East Antarctic continental margin during the Cambrian. Hammer for scale.

Early Facies Models

Early studies established the basic idea that storms result in the transport of sediment from generally well-sorted shoreline sources onto shelves, as deduced from lithofacies patterns of decreasing bed thickness and evidence for less powerful currents from onshore to offshore. Despite more than 30 years of work on storm-influenced shorelines and ancient deposits, there remains considerable debate concerning the processes by which sediment is transported seawards during storms. Early discussions of storm processes produced two radically different perspectives from oceanographers and from geologists who worked on ancient deposits. Stratigraphical and sedimentological data indicated a potential role of density-induced forces, in part because many tempestites closely resemble turbidites. The first facies model of storm-influenced shelves invoked the mobilization of sand at the shoreline, the creation of dense dispersions, and subsequent gravity-driven offshore transport. Data from many ancient deposits indicate transport nearly perpendicular to the shore for long distances. In this view, storm-deposited sandstone beds are indicative of nonuniformitarian processes, and the very rare storm events responsible for these beds (with recurrence intervals of tens of thousands of years) are presumably significantly different (quantitatively if not qualitatively) from those studied by oceanographers.

Oceanographic Studies

Although the results of major storms on modern shorelines were first described in the late 1960s, the dynamics of modern storm-generated flows were not outlined until the 1980s by oceanographers who described the effects of storms on shelf and nearshore circulation. Winds generated by storms move surface water, which is in turn deflected by Coriolis

acceleration. This deflection affects much of the water column and results in net transport at 90° to the wind direction (to the right in the northern hemisphere). One common consequence of such transport is that under certain wind conditions surface waters may move towards, and bank up against, the shoreline. The resulting sloping sea surface causes an onshore–offshore pressure gradient, which tends to drive flow offshore along the sea bottom. Coriolis force affects water particles moving in this flow at right angles to the flow direction, and an equilibrium is eventually established between the offshore-directed pressure gradient and the Coriolis force, such that the majority of storm-generated flows are alongshore. The surface flows are more landward oriented, and the bottom flows, which are affected by bottom friction, tend to be oriented at a slightly higher angle offshore.

The conflict between the oceanographical observations of shore-parallel geostrophic storm flows and the geological data that indicates both cross-shelf transport and the deposition of turbidite-like beds has never been fully resolved. Theoretical work and numerical simulations of combined-flow boundary layers (Figure 2) are important in addressing this problem and have been used by some workers to suggest that offshore transport of sand could take place during storms under combined flow. Since geostrophic currents are oriented slightly oblique to shore, the offshore component of the flow is added to the seawards strokes of storm waves and acts against the shorewards strokes, resulting in an asymmetry of shear stress. It was thus postulated that this asymmetry was responsible for the offshore sediment transport of sand and for the unimodal offshore-directed sole-mark orientations on the bases of tempestites. Studies in the ancient sedimentary record have affirmed the importance of geostrophic combined flows in some shoreface and nearshore storm deposits, but it remains unclear to what degree such flows are responsible for cross-shelf transport in modern systems or to what degree they were responsible for the deposition of ancient tempestites.

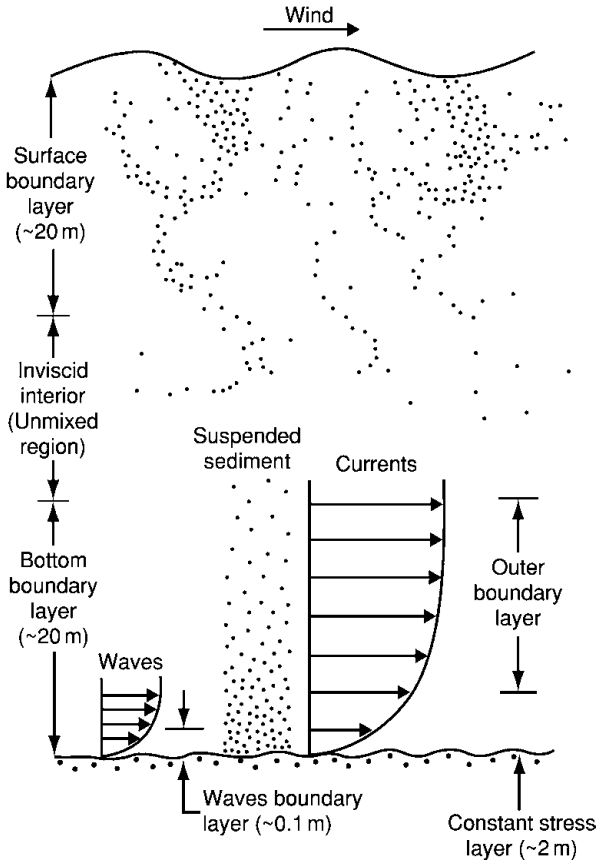


Figure 2 Subdivision of the water column into three layers during storms: a highly turbulent surface boundary layer, an inviscid interior or core flow layer, and a bottom boundary layer. The bottom boundary layer consists of a thin wave boundary layer and a thicker outer boundary layer produced by currents. Modified from Trowbridge and Nowell (1994).

Interpreting Ancient Tempestites

The large number of case studies of tempestites has led in recent years to a greater appreciation of the potential variety of storm processes. A wide range of storm deposits result from spatial and temporal variability in the degree to which three fundamental processes operate: geostrophic currents, storm-generated waves, and excess-weight forces (Figure 3). For instance, some studies have shown that geostrophic flows are in fact agents of transport and deposition in tempestites, as determined in part by shore-parallel palaeocurrents. Such flows cannot explain successions with offshore-directed palaeocurrents or sole markings, such as 'classic' flute marks, that are inconsistent with formation in the presence of waves. In fact, the wide range of sole-mark geometries alone supports genesis from various types of flows. Bidirectional and multidirectional orientations of prod marks and groove marks indicate strong wave activity. Flutes, unimodal prods, and well-oriented grooves probably reflect erosion by powerful unidirectional flow. Finally, a range of sole marks including curvilinear and quasi-sinusoidal forms record erosion by a variety of combined flows.

The internal structures of tempestites are equally diverse and relate to the temporal and spatial histories of the three basic components of storm-related flow described earlier. Most tempestites contain evidence of flow deceleration in the form of either normal grading or particular successions of stratification. They also contain stratification representative of a wide range of bedforms, from purely unidirectional

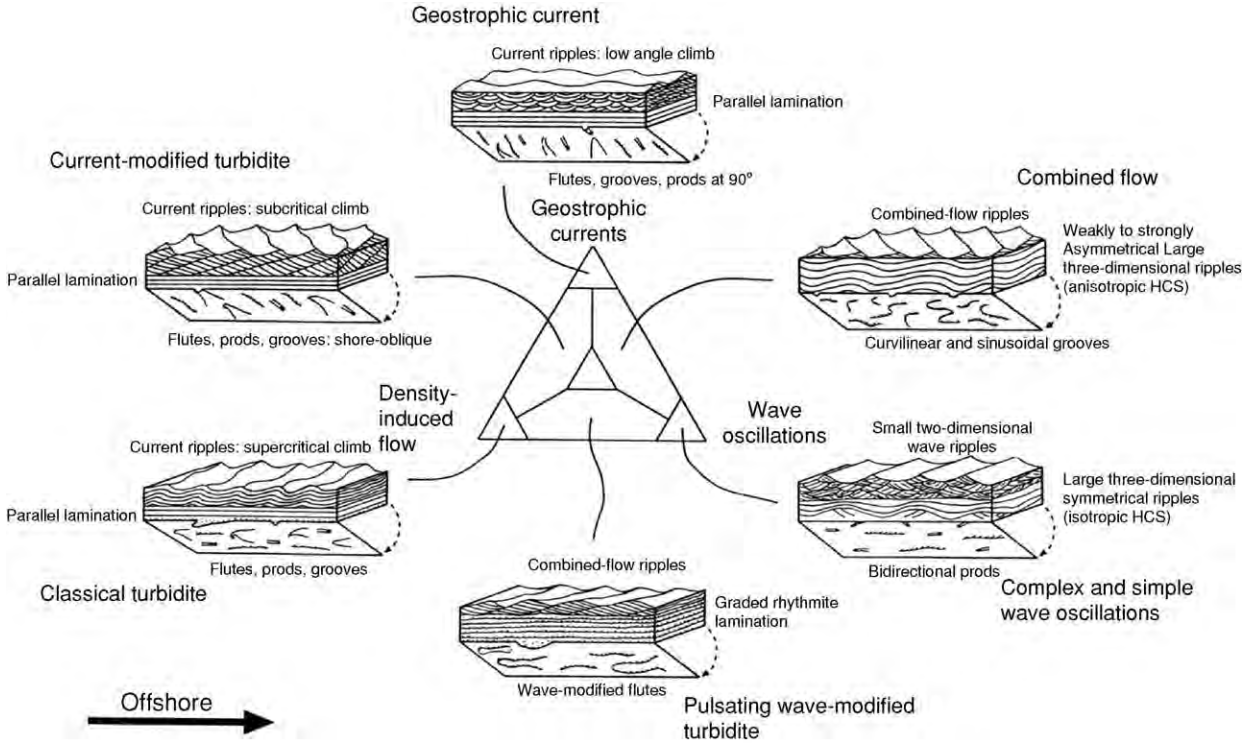


Figure 3 Heuristic model for the origin of stratification and sole markings for various possible combinations of storm processes. The diagram has three corners, representing density-induced flow due to excess-weight forces, geostrophic currents, and wave oscillations. The potential history of a storm event, in this three-part system, is extremely variable. The deposits of flows near the corners are well understood, but the central triangle and intermediate fields are poorly understood since so little is known experimentally about these fields, except for some information on simple rectilinear combined flows. HCS, hummocky cross-stratification. Reproduced from Myrow PM and Southard JB (1996) Tempestite deposition. *Journal of Sedimentary Research* 66: 875–887.

to purely oscillatory flow, including a spectrum of combined flows (Figure 4). The most diagnostic feature of storm deposition is HCS at various scales (Figures 5 and 6). The characteristics of HCS are well described and have been studied experimentally and with grain fabric analysis. There does not appear to be a size break in the bedforms that create HCS, namely three-dimensional, symmetrical to weakly asymmetrical, subaqueous dunes. Bedforms on modern shelves with HCS-like lamination have been attributed to combined flows, and experimental work has shown that hummocky bedforms are stable in complex oscillatory flows and wave-dominated combined flows (i.e. low velocities of current component).

Details of the geometry and internal structure of ancient tempestites indicate that beds range from those deposited in single events to those representing amalgamated accumulations produced by multiple events. The latter include thick beds that grade laterally into thinner beds separated by thin shale beds. These suggest that transport of sand onto shelves is in some cases incremental and involves a succession of storms. Some beds that appear to result from single events are potentially remobilized sand that simply records the last stage of transport and deposition.

Other tempestite beds show little or no evidence of amalgamation and are suggestive of powerful offshore-directed unidirectional flow during the erosional and initial depositional conditions of the storm. The nature of such flows is difficult to determine from the ancient. In many cases they may have been driven, at least in part, by excess-weight forces, caused by high near-bed suspended-sediment concentrations (SSCs). 'Sediment equilibrium' conditions exist when SSCs are caused by resuspension of sediment by storm-generated currents and waves. Very high SSCs (up to $4\,g\,l^{-1}$) and resulting powerful seaward-directed sediment-laden flows have been documented from shorefaces and inner-shelf settings. It is less clear to what distances such flows can carry sand offshore on the relatively low gradients common to modern continental shelves.

Excess-weight forces can also be generated extrinsically ('sediment disequilibrium' conditions), for example when rivers discharge sediment-laden floodwater into standing bodies of water and thus produce hyperpycnal (bottom-hugging) flows termed oceanic floods. Oceanic floods on the continental shelf of northern California adjacent to the Eel River have high SSCs (up to $2.5\,g\,l^{-1}$) of silt and

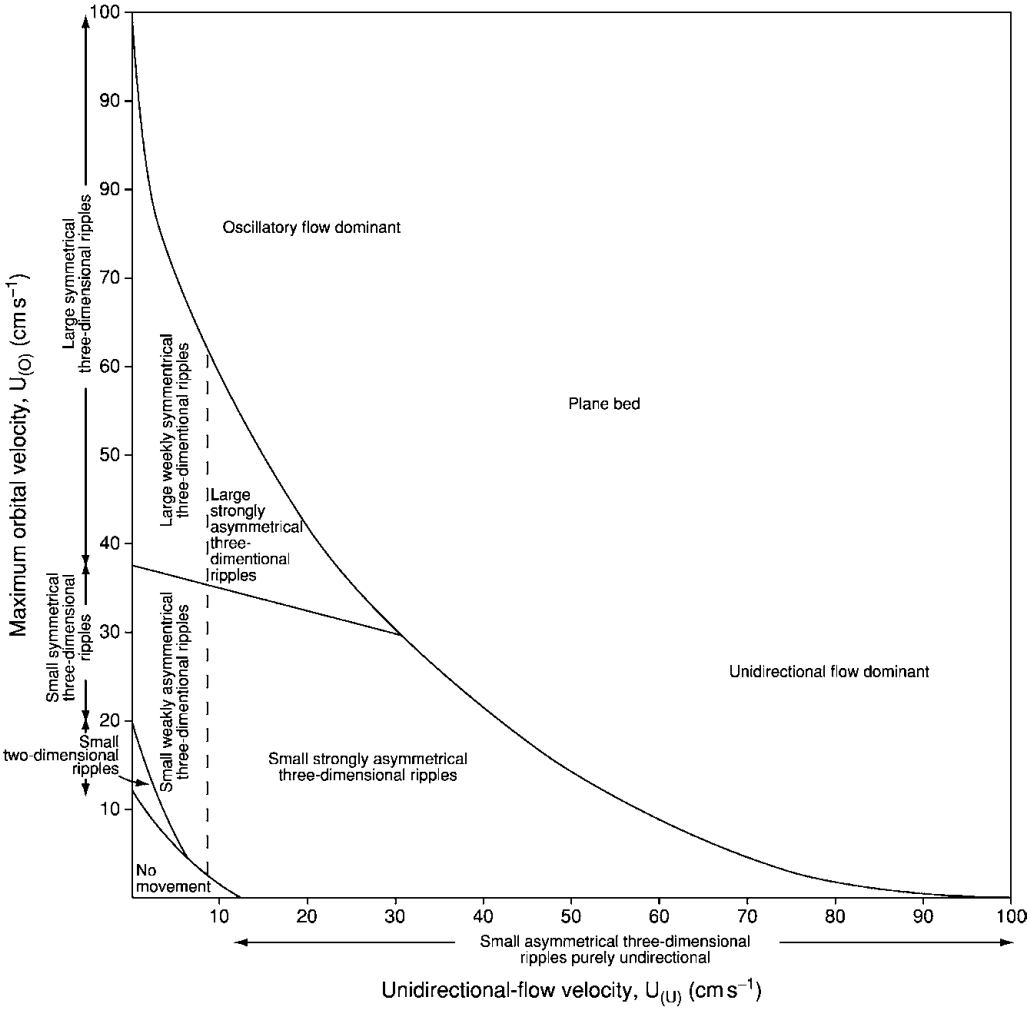


Figure 4 Bed-configuration stability diagram for combined flows in fine sand. Based on published data and on data from Arnott RW and Southard JB (1990) Exploratory flow-duct experiments on combined-flow bed configurations, and some implications for interpreting storm-event stratification. *Journal of Sedimentary Petrology* 60: 211–219 and Southard *et al.* (1990). The left side of the diagram (up to U_U of 25 cm s^{-1}) is drawn from the data in Arnott and Southard (1990), and the right side is extrapolated.

clay even at mid-shelf depths (50–70 m). Such floods are potentially a geologically important process for cross-shelf sediment transport, although it is unknown whether they could transport very fine and fine sand, the grain sizes that dominate most ancient tempestites. Facies models for some ancient sandstone tempestites cite nearshore sediment bypass (Figure 7), which is consistent with powerful unidirectional flows in very proximal positions, possibly associated with hyperpycnal oceanic floods.

In both sediment equilibrium and sediment disequilibrium conditions sediment may be transported in combined flows produced by a combination of excess-weight forces and storm waves. Such wave-modified turbidity currents differ from classic turbidity currents in that storm waves enhance boundary-layer shear stress and eddy viscosity, which reduces deposition from suspension and thus enhances gravity-driven transport. As long as turbulent diffusion by waves does not destroy the density stratification, it provides the energy necessary to maintain gravity-driven flow on relatively low gradients. This process has been documented in the cross-shelf movement of modern fluid mud layers, but not for sand-sized sediment. Ancient examples of wave-modified turbidites have internal sedimentary structures, including combined-flow ripples, that reflect the influence of both waves and gravity-driven flow (Figure 8).

Whereas the preceding discussion has emphasized the range of possible processes responsible for tempestite deposition and thus has explained the wide variety of sole marks and internal sedimentary structures in the ancient, there are two important aspects of tempestites that remain controversial. First, beds in the ancient that are interbedded with shale and were therefore deposited below the fair-weather wave base

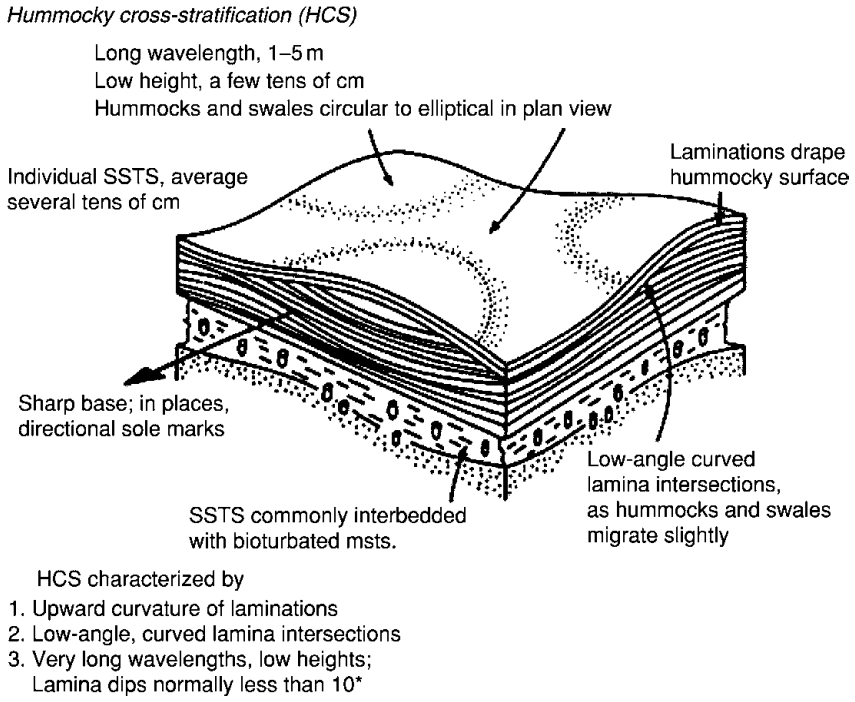


Figure 5 Block diagram illustrating the diagnostic features of hummocky cross-stratification (HCS). Modified from Walker RG (1984) Shelf and shallow marine sands. In: Walker RG (ed.) *Facies Models*, 2nd edn., pp. 141–170. Reprint Series 1. Toronto: Geoscience Canada.

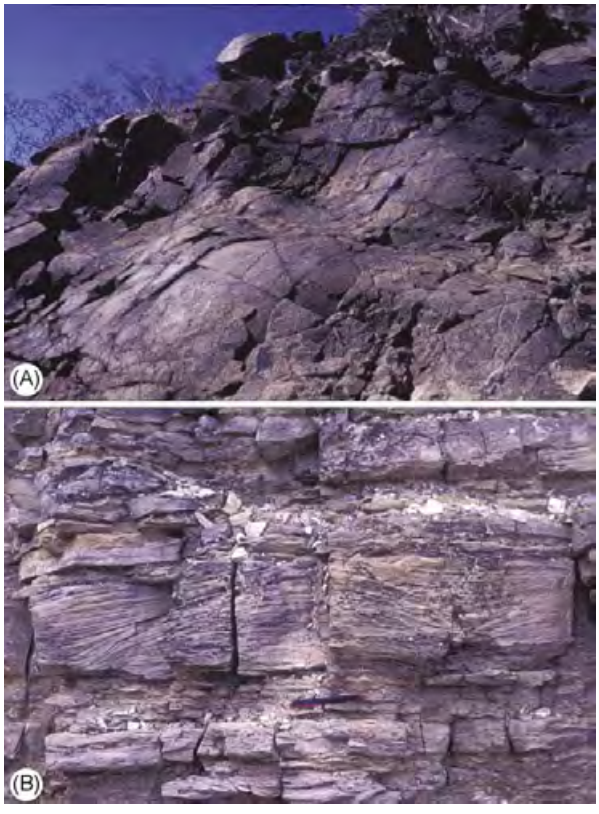


Figure 6 Hummocky cross-stratification (HCS). (A) Bedding plane view of exhumed HCS bedforms. Note the rounded symmetrical hummocks and swales. From Neoproterozoic Johnny Formation, Death Valley, California, USA. Pencil for scale. (B) Bed showing classic HCS features including upward-doming laminae, fanning of laminae, low-angle lamination, and low-angle curved erosion surfaces. From the Triassic Moenkopi Formation, central Utah, USA. Pencil for scale.

are considerably thicker than sand beds deposited on the shelf as a result of modern storms and appear to have a greater cross-shelf distribution than their modern counterparts. In unusual settings, such as the shallow (less than 20 m), flat, epicontinental shelf off the Yukon Delta in the Bering Sea, sand is carried offshore up to 100 km, although the beds range in (precompaction) thickness from 10–20 cm in the nearshore to less than 5 cm offshore. In this case, strong friction forces may have aided offshore transport, and thus such a setting may be a good analogue of epicontinental deposits, including those of the Cretaceous interior seaway of North America. In many modern settings sand beds of any appreciable thickness do not extend much below 5–10 m water depth. The thicknesses of modern storm beds fall in the small-to-average range for ancient tempestites, which commonly reach (compacted) thicknesses of many tens of centimetres and locally up to a metre or more. It is unclear to what degree thinner modern tempestites may simply reflect insufficient study of the modern, particularly since their record is commonly rapidly lost offshore owing to bioturbation. It

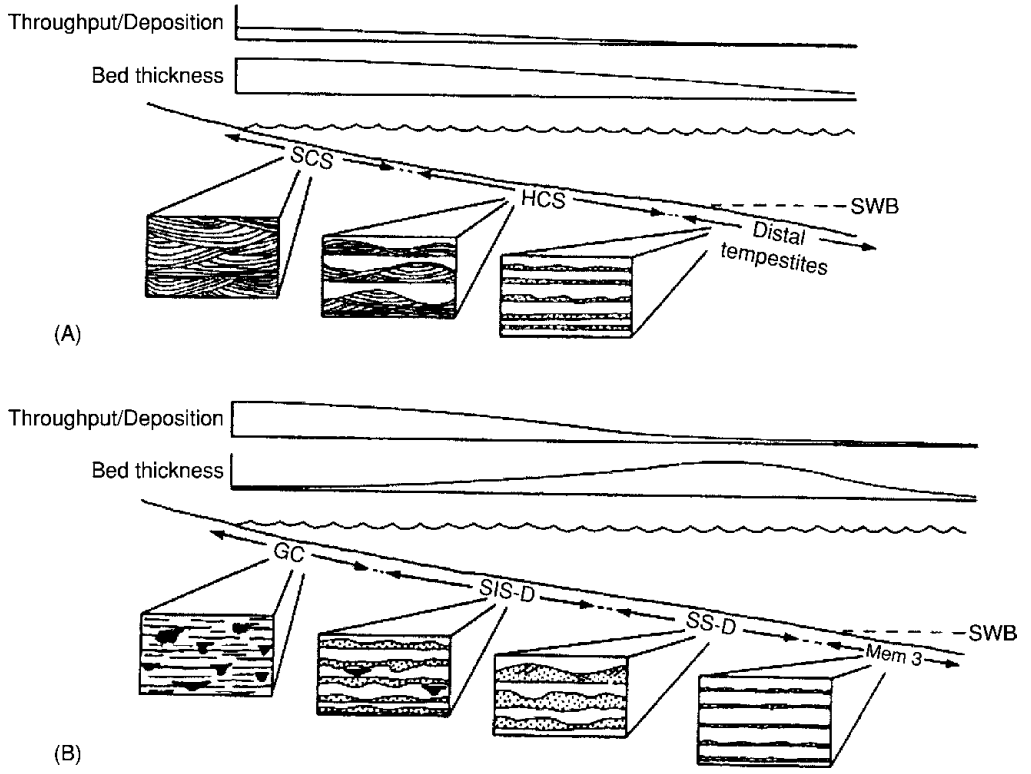


Figure 7 Comparison of (A) standard tempestite model and (B) bypass model (Myrow 1992). The spatial distribution of facies in each model is illustrated with onshore–offshore profiles and associated qualitative changes in bed thickness. The degree of sediment bypass (throughput to deposition ratio) is plotted above each profile. (A) Proximality trends in the standard model include constantly decreasing bed thickness away from the shoreline with a shift from swaly cross-stratified sandstone to hummocky cross-stratified (HCS) sandstone to distal turbidite-like beds. (B) The bypass model shows a bed-thickness trend that first increases then decreases away from the shoreline. The proximal setting is one of bypass, with erosion being the dominant process during storms (many gutter casts). Passing seaward, gutter casts die out and bed thickness increases. HCS is formed in the thicker sandstone beds farther out on the shelf. Below the storm wave base (SWB) distal tempestites are turbidite-like in character. The shelf gradient for the two models is shown to be the same simply for convenience.

is also difficult to establish accurate water depths of deposition from ancient shelf deposits, so the degree to which the ancient is nonuniformitarian in terms of bed thickness and cross-shelf transport is still unresolved. Finally, comparison of ancient shelf deposits with modern shelves is hampered by the fact that the latter may be strongly out of equilibrium owing to Holocene sea-level rise.

A second problem, seemingly forgotten in the literature, is that the time between the deposition of successive tempestite beds in many deposits is seemingly many orders of magnitude too large to explain simply as a series of storm events. Data tabulation suggests that preservation recurrence intervals (time between deposition of successive tempestite beds) in a small sample of ancient successions are of the order of 400–15 000 years. High-precision geochronological studies have never been done on tempestite successions, but they are necessary to establish a reasonably accurate picture of the recurrence interval of tempestite beds in the ancient. If one assumes that the recurrence intervals calculated for the ancient are reasonably accurate, then two possibilities arise for their explanation. The simple explanation would be that only extremely large and rare storm events are preserved in the rock record and that, given the recurrence intervals mentioned above, such events would potentially be qualitatively or quantitatively different from any documented in recent deposits. Such a stance might be difficult to justify on the basis of the physics of storms or on basic oceanographical principles, although changes in the Earth's climate, ocean circulation, and palaeogeography could certainly lead to nonuniformitarian atmospheric circulations including extremely large storms.

An alternative explanation is that the apparent recurrence intervals are artefacts of processes that bias the stratigraphical record. Given the range of recurrence intervals, there exists the possibility that relative sea-level changes may have played a role by shifting the source of sandy sediment seaward or landward. In other words, shoreline migration may

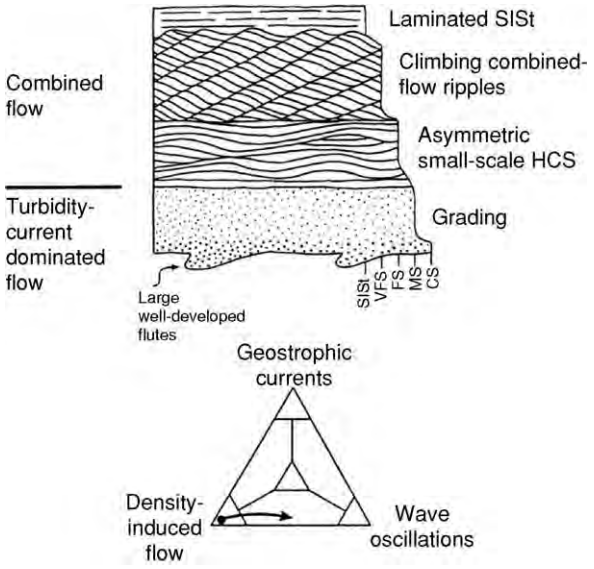


Figure 8 Generalized wave-modified turbidite bed. Temporal history of generalized bed is shown in the triangular diagram of storm-related effects (see also Myrow PM and Southard JB (1996) Tempestite deposition. *Journal of Sedimentary Research* 66: 875–887). HCS, hummocky cross-stratification.

have enhanced apparent cross-shelf transport. Sedimentological data must therefore be combined with sequence-stratigraphical data to understand better the bed geometries and transport distances of ancient tempestites. The stratigraphical distribution of tempestites may also be understood within a sequence-stratigraphical framework, particularly as sea-level has a strong control on shelf profile. For instance, geostrophic flow may be more common during transgressive to early highstand conditions, when accommodation space is higher and there is greater potential for Coriolis deflection, whereas more offshore-directed flow may take place during late highstand conditions characterized by greater bottom friction. Tempestites are probably more abundant during late highstand to early lowstand systems tracts of third- or higher-order sequences, when sand is mobilized from the coast during relative sea-level fall.

Carbonate Systems

Although the previous discussion applies to both carbonate and siliciclastic systems, there are also depositional features that are specific to, or most common in, carbonate systems. The most abundant such features include flat-pebble conglomerate beds and shell-bed accumulations. Flat-pebble beds are an important component of carbonate deposits, particularly in Cambrian and Lower Ordovician strata, and are generally considered to be accumulations of intraclasts produced by the storm reworking of early-cemented carbonate in shoreline and shallow subtidal environments. Detailed sedimentological analyses indicate that some flat-pebble beds were deposited from storm-generated combined flows, although it is likely that there are many other processes that could produce such beds.

Shell-rich grainstone tempestites accumulate as a result of winnowing by storm waves, commonly in shoreline and nearshore settings, and they have characteristics that are different from those of tempestites produced in condensed sections within maximum flooding intervals. In shallow epicratonic seaways, winnowing and concentration of shell debris occurs during lowstand conditions, and further reworking and basinward transport occurs during early transgression. Pronounced trends in the thickness, number, and character of shell beds occurred over the Phanerozoic, and these relate to evolutionary trends and complex taphonomic variables.

See Also

Sedimentary Environments: Depositional Systems and Facies; Carbonate Shorelines and Shelves. **Sedimentary Processes:** Depositional Sedimentary Structures; Particle-Driven Subaqueous Gravity Processes. **Sedimentary Rocks:** Mineralogy and Classification; Sandstones, Diagenesis and Porosity Evolution. **Unidirectional Aqueous Flow**.

Further Reading

Arnott RW and Southard JB (1990) Exploratory flow-duct experiments on combined-flow bed configurations, and some implications for interpreting storm-event stratification. *Journal of Sedimentary Petrology* 60: 211–219.

Dott RH Jr and Bourgeois J (1982) Hummocky stratification: significance of its variable bedding sequences. *Geological Society of America Bulletin* 93: 663–680.

Duke WL (1990) Geostrophic circulation or shallow marine turbidity currents? The dilemma of paleoflow patterns in storm-influenced prograding shoreline systems. *Journal of Sedimentary Petrology* 60: 870–883.

Einsele G (1996) Event deposits: the role of sediment supply and relative sea-level changes–overview. *Sedimentary Geology* 104: 11–37.

Einsele G and Seilacher A (1982) *Cyclic and Event Stratification*. Berlin: Springer-Verlag.

Kidwell SM (1991) The stratigraphy of shell concentrations. In: Allison PA and Briggs EG (eds.) *Taphonomy: Releasing the Data Locked in the Fossil Record*, pp. 211–290. Topics in Geobiology, 9. New York: Plenum Press.

Kreisa RD (1981) Storm-generated sedimentary structures in subtidal marine facies with examples from Middle and Upper Ordovician of southwestern Virginia. *Journal of Sedimentary Petrology* 51: 823–848.

Leckie DA and Krystinik L (1989) Is there evidence for geostrophic currents preserved in the sedimentary record of inner to middle-shelf deposits? *Journal of Sedimentary Petrology* 59: 862–870.

Myrow PM (1992) Bypass-zone tempestite facies model and proximality trends for an ancient muddy shoreline and shelf. *Journal of Sedimentary Petrology* 62: 99–115.

Myrow PM and Southard JB (1991) Combined-flow model for vertical stratification sequences in shallow marine storm-deposited beds. *Journal of Sedimentary Petrology* 61: 202–210.

Myrow PM and Southard JB (1996) Tempestite deposition. *Journal of Sedimentary Research* 66: 875–887.

Snedden JW, Nummedal D, and Amos AF (1988) Storm- and fair-weather combined flow on the Central Texas continental shelf. *Journal of Sedimentary Petrology* 58: 580–595.

Swift DJP and Niedoroda AW (1985) Fluid and sediment dynamics on continental shelves. In: Tillman R, Swift DJP, and Walker RG (eds.) *Shelf Sands and Sandstone Reservoirs,* pp. 47–133. Society of Economic Paleontologists Mineralogists Short Course Notes 13. Albuquerque: Society of Economic Paleontologists Mineralogists.

Walker RG (1984) Shelf and shallow marine sands. In: Walker RG (ed.) *Facies Models,* 2nd edn., pp. 141–170. Reprint Series 1. Toronto: Geoscience Canada.

Walker RG (1985) Fluid and sediment dynamics on continental shelves. In: Tillman RW, Swift DJP, and Walker RG (eds.) *Shelf Sands and Sandstone Reservoirs,* pp. 243–295. Society of Economic Paleontologists Mineralogists Short Course Notes 13. Albuquerque: Society of Economic Paleontologists Mineralogists.

SEDIMENTARY PROCESSES

Contents

Erosional Sedimentary Structures
Depositional Sedimentary Structures
Post-Depositional Sedimentary Structures
Aeolian Processes
Catastrophic Floods
Deep Water Processes and Deposits
Fluvial Geomorphology
Glaciers
Karst and Palaeokarst
Landslides
Particle-Driven Subaqueous Gravity Processes
Deposition from Suspension
Fluxes and Budgets

Erosional Sedimentary Structures

J Collinson, John Collinson Consulting, Beech, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

The accumulation of sediments into the rock record is not a process of continuous deposition, but rather the net result of interaction between episodes of deposition, non-deposition, and erosion. The erosional episodes are commonly recorded by distinctive sedimentary structures whose recognition and interpretation enhance the overall understanding of the sedimentary record. This article deals with some of the processes involved in sediment erosion and also reviews the range of distinctive structures that are produced both on present-day sediment surfaces and in the rock record. It deals solely with features that are present where overall accumulation has occurred or could occur. It does not deal with erosional features that form in subaerial landscapes undergoing long-term erosion.

Erosive Processes

Critical Erosion Velocity

When water or air flows over a bed of loose particles, the boundary shear stress caused by the moving fluid tends to initiate particle movement. There is a critical boundary shear stress (sometimes expressed as a critical erosion velocity) above which particle movement occurs. For sediment of rather uniform grain size, the critical shear stress will increase as the grain size increases. However, for small grain sizes (silt and mud), the situation is less straightforward, as the combined effect of increased cohesive strength and lower surface roughness means that higher velocities are needed to initiate movement. This relationship is illustrated for water in the Hjulström–Sundborg diagram (Figure 1), which operates ideally for well-sorted sediments. A poorly sorted substrate may be winnowed as the critical shear stress of the finer particles is exceeded, leading to the development of an armoured lag on the surface that inhibits further erosion.

Sediment Cohesion

The cohesive strength of damp or wet, fine-grained sediment results from the increasing importance of intergranular electrochemical forces as the surface area to volume ratio of the grains increases and as clay minerals make up an increasing proportion of the sediment. The cohesive strength of fine-grained sediments increases with the time following deposition and is augmented by shallow burial. As well as influencing the critical shear stress of fine-grained sediment, the cohesive strength also complicates the way in which fine-grained sediments are eroded. The particle-by-particle erosion that characterizes non-cohesive sediment is replaced by the erosion of grain aggregates of cohesive sediment. These range from small rip-up clasts to larger blocks of eroded mud (Figure 2). A cohesive substrate is also able to retain detailed erosional relief and to stand at steep, even overhanging, angles. This is critical to the preservation of the many erosional structures discussed below.

Role of Abrasion

The relationships between the critical boundary shear stress and particle size, as expressed in the Hjulström–Sundborg diagram, are complicated further by the fact that natural currents capable of erosion are likely to be carrying material in suspension. Sand and silt particles within this suspended load will enhance the erosive capability of the current by providing an abrasive supplement to the boundary shear stress.

Erosional Sole Marks

Erosional sole marks are relatively small-scale structures preserved as casts on the bases ('soles') of sandstone or, more rarely, limestone beds usually in interbedded sandstone/mudstone sequences. They vary in style and are valuable indicators of erosional and depositional processes. They also provide palaeocurrent information and act as 'way-up' indicators in deformed successions. Their common occurrence on the bases of sandstone beds in interbedded sequences suggests that the sands were deposited as episodic, high-energy events in settings in which quiet conditions normally prevailed, with background sedimentation being from suspension. The fact that erosional sole marks occur on the bases of the sandstones suggests that the high-energy events were erosional in their initial stages, prior to the onset of sand deposition. Such surges, which decelerated through time, typically occur as turbidity currents in deep water, as

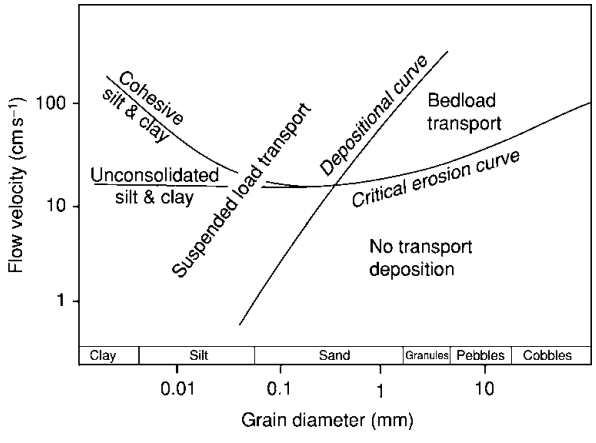


Figure 1 Plot of grain size against stream power showing the velocity at which sediment is set in motion or eroded. Note the different behaviour of fine-grained sediment depending on whether it has had time to develop cohesive strength.

Figure 2 Blocks of cohesive mud occurring as clasts in the base of a channel. Recent terrace deposits, Tana Valley, Norway.

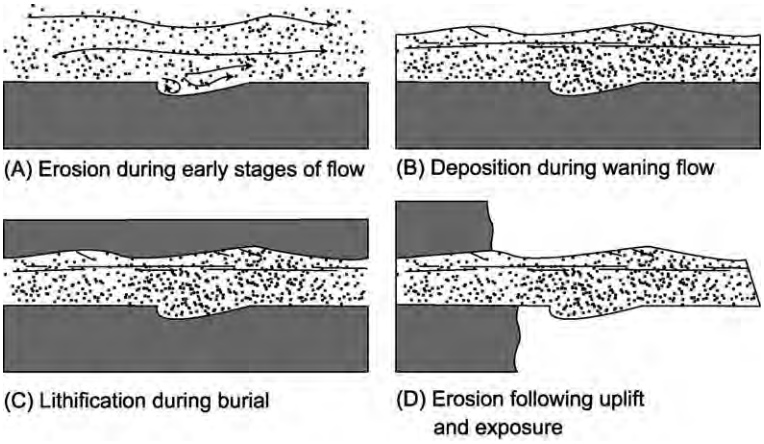


Figure 3 Schematic diagram to show the main stages in the development and preservation of erosional sole marks in an interbedded succession of cohesive sediment and coarser, high-energy event beds.

storm surges in shelf settings, and as flood events in deltaic and alluvial settings.

The erosive initial stages of such flows are likely to be highly charged with suspended sediment, which may include coarse particles and, consequently, erosion may result from a combination of both fluid shear and grain abrasion. If the transported clasts are sand and silt, abrasion will augment and accelerate fluid scour. If, on the other hand, the clasts are quite large, their interaction with the bed may create distinctive structures that relate directly to that interaction. Such structures are 'tool marks'. Erosional forms created by both types of process occur initially as negative relief on the bed, and their preservation and later recognition depend on that surface being rapidly overlain by coarser sediment, giving a positive cast of the relief (Figure 3). The survival of the surface relief through the transition from erosion to deposition depends on the cohesive strength of the substrate, which typically preserves steep and detailed features.

Sole Marks due to Fluid Turbulence

These structures relate to the pattern of fluid turbulence close to the bed and are essentially smooth features related to fluid scour, probably augmented by abrasion. They may be classified into three main groups: obstacle scours, flute marks, and longitudinal furrows.

'Obstacle scours' occur around an object lodged on the bed. This is most often a pebble or a shell fragment and the scour is preserved as a horseshoe-shaped ridge that wraps around the upstream side of the obstacle (Figure 4). The ridge was a trough during erosion and its form results from the locally enhanced boundary shear stress as flow was forced to accelerate around the object. Spiral eddies which are shed off either side of the obstacle give the tails that gradually die out downstream.

Figure 4 Obstacle scour around a pebble on the base of a sandstone bed. The horseshoe-shaped ridge is the cast of a scour trough eroded around the upstream side of the pebble as the flow accelerated past the obstacle. Eocene, Ainsa, Spanish Pyrenees.

'Flute marks' occur in isolation and in groups, sometimes in linear arrays parallel to flow. They are characterized by a rounded nose at the deepest part of the scour from which the structure flares away in a V-shaped form (Figure 5). The nose points upstream and the shape of the flute reflects the action of a local eddy at the bed, enhanced by the growing flute. Their shape in plan ranges from very sharp V-forms to features with a straighter transverse upstream edge. Flutes may have been initiated at some initial irregularity on the bed surface that caused flow separation and heightened bed shear stress. The irregularity responsible may sometimes be apparent, but in most cases is not. The separation eddy, in turn, led to local scour, enhancing the relief, so that, once initiated, a positive feedback loop was set up between the turbulence and the bed topography. The downstream flaring of the flute away from the nose reflects the mixing of the eddy back into the body of the flow. At the head

Figure 5 Flute casts on the base of a turbidite sandstone bed showing the noses pointing in the upstream direction. Eocene, San Sebastian, Spain.

of the flute, the axis of the separation eddy would have been transverse to flow and the axes would have rotated parallel to flow along the downstream flaring margins.

'Longitudinal furrows' occur on the bases of sandstone beds as patterns of broadly parallel ridges that have a rounded profile and are separated by narrower grooves. This reflects an erosional morphology of rounded parallel furrows separated by cuspate ridges. Along the furrows, it is quite common to find rounded noses that all point in the same direction, similar to small flutes. Indeed, there is a continuous series of erosional forms between clear flutes at one end and smooth parallel furrows at the other. The furrows are thought to result from the erosive effects of paired spiral eddies with their axes parallel to flow and with opposed senses of rotation. The furrows correspond to divergence zones where eddies direct flow down on to the bed, whilst the cuspate ridges lie beneath the convergence zones where flow components lift off from the bed.

'Gutter casts' are features that seem to occur preferentially in shallow marine settings where storm currents are typically the agent responsible. They differ from longitudinal furrows in being larger and separated from one another by areas of flat bedding surface rather than cuspate ridges (Figure 6). They are typically many centimetres wide and a few centimetres deep and they may be separated by flat sectors up to metres in width. Their fills sometimes carry a coarse lag in the base, often made up of fossil fragments. In plan view, they tend to be rather continuous, may be somewhat sinuous, and may split and rejoin, suggesting an anastomosing pattern. They probably result from rather sustained currents, possibly driven by wind or storm effects.

All the scours related to fluid scour have a strong orientation parallel to the current, which can be used to infer palaeoflow direction. With longitudinal scours without clear noses, it may be impossible to infer the sense of movement.

Figure 6 Lower bedding surface of a limestone bed from within a shallow marine interbedded limestone–mudstone succession showing gutter casts. Ordovician, north-west Greenland.

Tool Marks

These constitute a diverse suite of structures that are produced by the interaction of objects ('tools') carried by the flow with a soft cohesive substrate. The shape and scale of such structures vary with the nature of the tools and the way in which they impinge on the bed (Figure 7). Tools marks can generally be distinguished from structures produced by fluid scour by the fact that they tend to have sharp, often angular relief, reflecting the shape of the tool, compared with the much smoother shapes of fluid scour marks. The tools themselves are seldom seen associated with the tool marks that they created, and one can only imagine or infer what they may have been. They are most likely to have been pebbles, mud rip-up clasts, fossils, or wood fragments. In some cases, delicate relief on the marks can be related to features of the tool, such as ribbing on a shell.

Tools may interact with the bed in a variety of ways, and a loose terminology is applied to the results. Where tools remain in contact with the bed for some distance, they gouge 'grooves', usually with sharp margins and sometimes changing in form as the tool rotates slowly. Where the tools are in intermittent contact with the bed, a string of similar impact marks

Figure 7 A variety of tool marks on the base of a thin turbidite sandstone bed showing prod marks and bounce marks. The series of distinctive bounce marks towards the top of the slab is thought to have been made by a fish vertebra, unlike most tool marks where the nature of the tool is unknown. Oligocene, Polish Carpathians.

Figure 8 Chevron cast on the base of a turbidite sandstone. Ordovician, Kirkudbrightshire, Scotland.

with more or less regular spacing may form 'bounce marks'. These may be symmetrical or may have an asymmetry reflecting the differing angles of impact and rebound. Where tools make isolated impact with the bed, single impressions occur. These 'prod marks' commonly have a clear asymmetry, gently inclined on the upstream side, recording the low-angle approach trajectory, and steeper on the downstream end where the tool was lifted back into the flow.

Less common are 'chevron marks' where the tool appears to have been dragged across a rather viscous muddy substrate so that the sediment has been rucked into a series of chevron-like folds that face in a downstream direction (Figure 8).

Most tool marks are good palaeocurrent indicators and, in most cases, both direction and sense of movement can be deduced. Highly elongate forms, such as grooves and bounce mark sequences, give the best directional measure, whilst the asymmetry of prod marks provides the most reliable indication of sense of movement. It is often the case that a variety of forms occur together on the same surface and it is not uncommon for a range of directions to be present, suggesting that the current responsible changed direction whilst in the erosive mode.

Erosional Surface Forms

On present-day subaerial sediment surfaces, e.g., tidal flats, beaches, river beds, and desert flats, it is possible to find structures that record erosion by recent currents, both water and wind. Examples of these features also occur in the rock record, but they are less common than erosional sole marks because many are developed in sand, which is likely to be reworked by subsequent currents. Erosional forms are also detected and mapped on the seafloor by remote sensing techniques on both sandy and muddy substrates. Ancient small examples are seen in the rock record at outcrop, but some larger examples probably go undetected.

'Obstacle scours' occur around pebbles and shells on sandy river beds and on beaches. They are curved troughs that wrap around the upstream side of the obstacle and flare away downstream. They relate to the local acceleration of the current around the obstacle and the generation of a local eddy system. On stream beds, the sand may be rippled and the obstacle scour may, in effect, be a distortion of the ripple pattern. On beaches, the current responsible is the backwash of the waves and often they occur on an otherwise flat sediment surface.

'Wind ridges' occur where strong winds blow across damp sand as on a beach. As the sand dries out differentially, dry sand is set in motion by the wind, whilst damper patches retain some coherence and resist erosion. These will typically become streamlined as erosion goes on around them, producing ridges with blunt noses at their upwind ends and tails that flare away downwind. Rare examples are known in some ancient lake deposits where the surface was inundated and draped with finer sediment.

'Seafloor flutes and lineations' have been detected by remote sensing techniques, particularly side-scan sonar. They occur on both sandy and muddy seafloors and can result from both tidal currents and deep-sea turbidity currents. Lineations eroded in muddy seafloors may be many hundreds of metres in length and form spaced arrays. Large-scale flutes, which are

known from the surfaces of some submarine fans, are tens to hundreds of metres in length and seem particularly to be features of areas where large flows have spilled over from fan channels.

The recognition of ancient examples of such features requires exceptional exposure. The largest forms would commonly be categorized as channels. With extensive bedding surface exposure, large, flute-like forms (megaflutes) occur on the upper surfaces of sand units in turbidite sequences and are typically overlain by mud drapes and fills. It may well be the case that, where erosion was followed by the deposition of sand, the features usually go undetected.

Channel Forms

Channels occur in a wide range of settings as an integral part of the depositional process. In some cases, channels and their fills constitute only a small part of the succession; in other settings, the successions may be dominated by channel fills. Given good outcrop, the recognition of channels is often very straightforward, as an erosional surface, with significant relief, separating two discrete bodies of sediment, the overlying unit commonly being coarser grained than that below (Figure 9). The shapes of channel erosion surfaces are quite varied, ranging from strongly curved, concave-upwards forms to horizontal surfaces. The observed shape records the interaction of the active channel and its behaviour during its active life. Channels that maintain a stable position generate a cross-section that is very close to that of the active channel and the resultant sediment body is lenticular. Actively migrating channels generate an erosion surface with an extensive flat sector and steeper margins, so that the channel sediment body has a tabular shape. Channel stability reflects the energy of the flow within the channel and the stability of the material in the channel banks. With cohesive bank material, channels are less likely to migrate. For non-cohesive bank material, the greater tendency for migration means that channel margins are less likely to be seen in the rock record, and erosion surfaces have to be inferred from the concentration of larger clasts, both exotic and intraformational, in a lag conglomerate directly above the erosion surface. Where channel margins are cut into dominantly cohesive sediment, but with some interbedding of lithologies, it is quite common for the interbedding to be reflected in the morphology of the erosion surface. Stepped profiles and even overhangs may occur. These smaller features are often valuable for the measurement of channel trend directions, particularly where accompanied by sole marks, such as flutes.

Figure 9 Part of a channel margin within a turbidite succession. The stepped nature of the junction in part reflects the lithological differences in the underlying thin-bedded succession. The thick beds of sand that make up the channel fill reflect the stronger currents that flowed through it. Namurian, North Derbyshire, England.

Other Erosion Surfaces

Many bedding surfaces have an element of erosion in their genesis, but hardly qualify as 'sedimentary structures' and will not be dealt with in detail. Marine transgressions may often be erosive in nature, and the landward advance of a shoreface may produce a horizontal and relatively flat surface (ravinement surface) recognized by a sharp upwards change from more nearshore to more offshore facies, possibly associated with a lag conglomerate. In aeolian settings, major bounding surfaces (supersurfaces) may be associated with erosion down to the water table and are recognized in some cases only on the basis of their wide extent.

Channel-like features may result from processes other than fluid scour, especially in finer grained sediments and particularly those deposited on subaqueous slopes. Instability, often due to rapid deposition, may give rise to failure and the initiation of slumps and slides. The surfaces over which such movements take place may have concave-upwards forms, especially in the up-slope slump scar area, and these can superficially resemble channels. Generally, the context and the associated sediments allow the differences to be recognized.

Other erosional forms result from the dissolution of lithified sediments, typically limestones, so that irregular palaeokarstic surfaces may be incorporated into the rock record.

See Also

Sedimentary Processes: Depositional Sedimentary Structures; Post-Depositional Sedimentary Structures; Aeolian Processes; Karst and Palaeokarst; Particle-Driven Subaqueous Gravity Processes. **Sedimentary**

Rocks: Mineralogy and Classification. **Unidirectional Aqueous Flow**.

Further Reading

Allen JRL (1982) *Sedimentary Structures: Their Character and Physical Basis. Developments in Sedimentology 30A & B,* pp. 593–663. Amsterdam: Elsevier.
Allen JRL (1985) *Principles of Physical Sedimentology,* ch. 8. London: Allen & Unwin.
Collinson JD and Thompson DB (1989) *Sedimentary Structures,* 2nd edn., ch. 3–4. London: Chapman & Hall.
Dzulynski S and Walton EK (1965) *Sedimentary Features of Flysch and Greywackes.* Amsterdam: Elsevier.
Leeder MR (1982) *Sedimentology; Process and Product,* ch. 5, 6, 9. London: Allen & Unwin.
Leeder MR (1999) *Sedimentology and Sedimentary Basins,* ch. 5, 6, 10. Oxford: Blackwell Science.
Mutti E (1992) *Turbidite Sandstones.* Milan: Agip.
Pettijohn FJ and Potter PE (1964) *Atlas and Glossary of Primary Sedimentary Structures.* Berlin: Springer.
Ricci-Lucchi F (1970) *Sedimentografia.* Bologna: Zanichelli.
Selley RC (2000) *Applied Sedimentology,* 2nd edn., ch. 5. San Diego: Academic Press.

Depositional Sedimentary Structures

J Collinson, John Collinson Consulting, Beech, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Depositional sedimentary structures reflect closely the processes by which sediment was transported immediately prior to deposition, and therefore provide an important starting point for environmental interpretation through facies analysis (*see* **Sedimentary Environments:** Depositional Systems and Facies). Their interpretation requires an appreciation of the hydrodynamics or aerodynamics of fluid–sediment interactions.

Sediment Transport and Deposition

Excluding highly concentrated sediment–water mixtures, sediment is transported by fluid in two distinct ways. Finer grained material, commonly clay and silt, but sometimes including sand and coarser material, is carried along with the fluid in 'suspension', when particles are supported by the upwards component of turbulence. Coarser grained material tends to move in intermittent contact with the bed, once its critical boundary shear stress has been exceeded. This movement, referred to generally as 'bedload transport', takes place through rolling or, more commonly, bouncing of grains on the bed, a process called 'saltation'. At its most vigorous, saltation grades into suspension. Sand particles in saltation, on colliding with a bed of similar sand, set other grains in motion as their kinetic energy is dissipated, a process especially important in wind-blown transport.

Bedload Transport

Bedload transport in water results from unidirectional currents, wave action, and combinations of the two. Clearly, for waves to affect sediment movement, the bed has to be above wave base (i.e., within the depth range of the waves), and it is important in many settings to recognize the distinction between fair weather and storm wave base.

Suspension

Sediment is carried in suspension provided that the intensity of turbulence within the fluid remains high. A current that is underloaded with suspended sediment can take more material into suspension up to a certain critical capacity. Sediment is deposited when the level of turbulence can no longer support all the suspended grains. This usually occurs when the current decelerates, often as a result of flow expansion at a channel mouth, but also through the waning of a high-energy event, such as a storm or flood. Coarser particles, with higher settling velocities, fall from suspension first. Sedimentation is commonly accelerated in river mouth and estuarine settings where a freshwater suspension meets saline water and particles coagulate into larger flocs. Deposition from suspension occurs in many settings and produces a variety of types of bedding and lamination. 'Bedding' is used for units of centimetres or larger thickness, whilst 'lamination' is used for features at the scale of millimetres.

At coastal delta fronts, a floating plume of turbid freshwater floats over denser seawater and may extend many kilometres offshore before it disperses. In freshwater lakes, the reverse situation may apply, with denser turbid water hugging the lake floor as a density underflow. Such relatively sustained underflows contrast with more episodic underflow surges, which are features of deep marine settings. These turbidity currents carry a wide variety of grain sizes in suspension and the interaction of density contrast, down-slope movement, resultant turbulence, and

consequent sediment load leads to a phenomenon referred to as 'autosuspension'. Changes to the balance amongst the variables, mainly through changes in bottom gradient, lead to accelerations or decelerations and to phases of erosion and deposition, respectively.

Lamination in Fine-Grained Sediments

In settings in which low energy prevails, only the finest sediment, carried in suspension, settles as mud. If the bottom waters are well oxygenated, sediment will usually be disturbed by burrowing organisms and the depositional lamination destroyed. If the bottom waters are anaerobic, a fine depositional lamination may be preserved. In all cases, burial compaction flattens the original fabric and creates a foliation or lamination of later origin. Only where early cementation occurs, usually as concretions, is depositional and early postdepositional fabric and lamination preserved.

In areas of higher sediment supply, such as a coastal river mouth, lamination and bedding in sediments deposited from suspension often result from changes in river discharge. In marine settings, deposition from a floating turbid plume can give gradational thin bedding or lamination due to gradual changes in discharge or in the position of the plume. In lakes, where underflows are more common, bedding and lamination may be more sharply defined, particularly in proglacial settings where seasonal discharge variations are shown by varves (rhythmites; *see* **Sedimentary Environments:** Lake Processes and Deposits).

In certain tidal settings, particularly estuaries, where fine-grained suspended sediment occurs due to both river supply and erosion by waves and tidal currents, the deposition of mud is particularly associated with slack water periods between tidal flows. These occur four times a day in subtidal settings and twice a day in intertidal areas. High concentrations of suspended mud and the effects of flocculation cause the deposition of discrete 'mud drapes' within the space of a few hours. When associated with sand deposition, reflecting higher energy tidal flows, these help to create 'heterolithic facies', which are discussed below.

Aqueous Sandy Bedforms and their Internal Structures

Once sand or coarse silt begins to move under a unidirectional water current, the sediment surface takes on a succession of dynamic morphologies in response to changing current strength. These morphologies are termed 'bedforms', and their hydrodynamic significance and relationships to internal lamination in sands and sandstones are key elements in interpreting processes and environments of deposition.

Current Ripples and Ripple Lamination

When sand starts to move, the surface becomes covered by 'current ripples', small-scale, repetitive, asymmetrical bedforms, a few centimetres high and a few tens of centimetres in wavelength (Figure 1). Their dimensions scale with grain size rather than with the properties of the flow. Their asymmetrical profile parallel to flow has a steep 'lee side' facing downstream and a gently inclined 'stoss side' facing upstream. Soon after formation, current ripples have rather straight crest lines, orientated transverse to flow, but this pattern soon breaks down into more complex three-dimensional 'linguoid' geometries, whereby the ripples have strongly curved crest lines, usually convex downstream and with scour pits in front of gaps between adjacent ripples. These shapes are intimately related to the patterns of fluid turbulence over the ripples, with flow separation at the crests and downward-directed eddies converging to create lee-side scour pits. As an array of ripples moves downstream, sand is swept from the scour pits over downstream stoss sides and deposited on lee sides. Ripples and dunes (dealt with below) are together referred to as the 'lower flow regime' of bedload transport (Figure 2).

The deposition of sand on the lee sides of ripples leads to inclined lamination that records successive positions of the lee side. This cross-lamination is commonly defined by micas and other platy grains. It is preserved only when ripple migration takes place during vertical aggradation of the bed (Figure 3).

Figure 1 Small-scale asymmetrical current ripples formed by a unidirectional flow from left to right. Gently inclined stoss sides are inclined upstream and steeper lee sides face downstream. Note the variability in the curvature of the crest lines. Tana River, Norway.

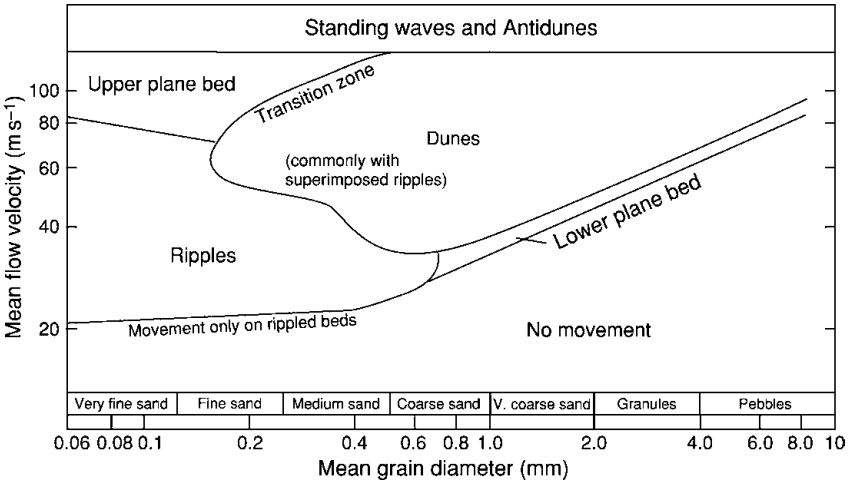


Figure 2 A plot of stream power vs. grain size showing the fields of occurrence of the main classes of bedform produced by unidirectional water flows.

Figure 3 Climbing ripple cross-lamination in fine sand. The trajectories of the ripple crests reflect the balance between the rate of bed accretion and rate of ripple migration. Where the angle of climb exceeds that of the ripple stoss sides, stoss-side lamination is preserved. Holocene terrace deposits, Tana Valley, Norway.

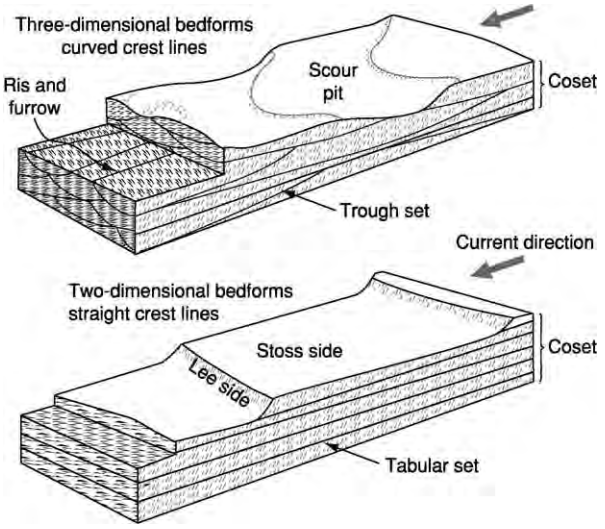


Figure 4 Block diagram to show the main types of three-dimensional geometries of cross-bedding and cross-lamination and the terminologies used.

Where migration occurs without aggradation, lee-side scour removes the cross-laminated sand deposited by the preceding ripple. When aggradation occurs, the trajectory of the scour zone climbs downstream, allowing the partial preservation of cross-laminated sand. The angle of climb determines the extent of preservation, but, in many cases, the angle is not apparent because of the random component of erosion created by migrating scour pits. Where climb angles are high, due to high aggradation rates, the inclination of erosion surfaces between cross-laminated sets may be apparent, and the style is referred to as 'ripple drift' or 'climbing ripple lamination' (Figure 3). Where climb angles exceed the inclination of the stoss sides, stoss-side laminae are preserved with no erosion between sets.

In three dimensions, the geometry of ripple cross-lamination most commonly occurs as a series of intersecting troughs, each filled with curved cross-laminae that are concave both upwards and downstream (Figure 4). The troughs are the curved erosion surfaces cut by the migration of the scour pits and the laminae record their filling by the migration of the scour pits. Usually, only the lower parts of lee-side laminae are preserved. In plan view, trough cross-lamination appears as a pattern of flow-parallel zones of curved laminae, known as 'rib and furrow', which is an important palaeocurrent indicator. Palaeocurrents are most accurately measured from the axes of the troughs. Rib and furrow is a more

common record of current ripples than the preservation of ripple forms on bedding surfaces. Where ripple forms are preserved, they are usually overlain by a drape of finer grained sediment due to a rapid reduction in current strength.

Dunes, Sand Waves, and Cross-Bedding

As current velocities increase above those appropriate for ripples, sand beds deform into larger bedforms, termed 'dunes' or 'sand waves', provided that the water is deep enough (Figure 2). These have a similar asymmetry to current ripples, although the wavelengths are greater, seldom less than 1 m and extending to many tens of metres. Heights are in tens of centimetres or even metres. There is a clear hydrodynamic distinction between ripples and dunes, illustrated by the superimposition of ripples on the larger forms under equilibrium conditions. Dunes, therefore, are not just large ripples. The orientations of superimposed ripples reflect the direction of flow close to the bed over and around the dunes, recording patterns of eddy separation and reattachment, especially those associated with scour pits.

Dunes and sand waves occur in a wide range of plan forms, from straight crested through highly three dimensional, and also span a wide range of height to length ratios (Figures 5 and 6). Steep, strongly three-dimensional forms are mainly a response to strong and relatively deep flows, whilst straight-crested or low-relief forms, which tend to lack scour pits, are responses to less strong or more shallow flows. The dimensions of dunes and sand waves reflect the properties of the flow, rather than the sediment. They seem particularly susceptible to changes in flow depth. However, it is quite difficult to interpret large bedforms in terms of flow parameters as they are commonly out of equilibrium with the prevailing current. Large volumes of sand must be moved to remould large forms and, in rivers where discharge varies quickly or in tidal settings, large-scale bedforms are often in a state of disequilibrium. In intertidal settings or on emergent river beds, larger bedforms are commonly modified by the process of emergence. Changing flow patterns over and around the forms and the action of waves both rework ripple patterns and reduce the angle of the lee-side slipfaces. In subtidal settings, sand waves often are present on the seabed where they are susceptible to semi-diurnal reversals of flow. Asymmetrical profiles may reflect quite a small asymmetry in strengths of ebb and flood tides.

Figure 5 Small, three-dimensional dunes with well-developed scour pits in their lee-side areas. Note how the superimposed ripples fan out to reflect the pattern of eddying over the dunes close to the bed. Note also the difference in scale of the dunes and the ripples. Tana River Delta, Norway.

Figure 6 A wide sandy braided bed of the Tana River, Norway, showing large, low-relief sand wave bedforms. These examples have curved crest lines (convex downstream), superimposed ripples, and, locally, smaller dunes on their upper surfaces. Whilst their spacing is in hundreds of metres, the heights of their lee faces seldom exceed 1 m. The larger areas of sand, which separate the river into discrete channels, are compound bedforms or bars. Tana River Delta, Norway.

As with ripples, internal lamination produced by dune-scale bedforms is dominated by inclined lamination, generally referred to as 'cross-bedding' (Figure 7). The individual dipping laminae are termed 'foresets', whilst packages of foresets produced by the migration of individual dunes are 'sets'. Sets range in thickness from around 10 cm up to several metres, and foresets are generally defined by slight differences in grain size, related to avalanches of grains (grain flows) on the lee side and to fluctuating turbulence in that area. In vertical section, some foresets are straight with an angular basal contact (termed 'planar'), whilst others have a tangential or asymptotic contact. The sets fall into one of two classes depending on the shape of the set boundaries. Where these are broadly subparallel and flat-lying, the sets are described as 'tabular'. Where set boundaries are concave-upwards erosion surfaces, they compare, except in scale, with ripple cross-lamination, and the term 'trough cross-bedding' is used. The foresets of trough sets are predominantly concave upwards and tangential, whilst the foresets of

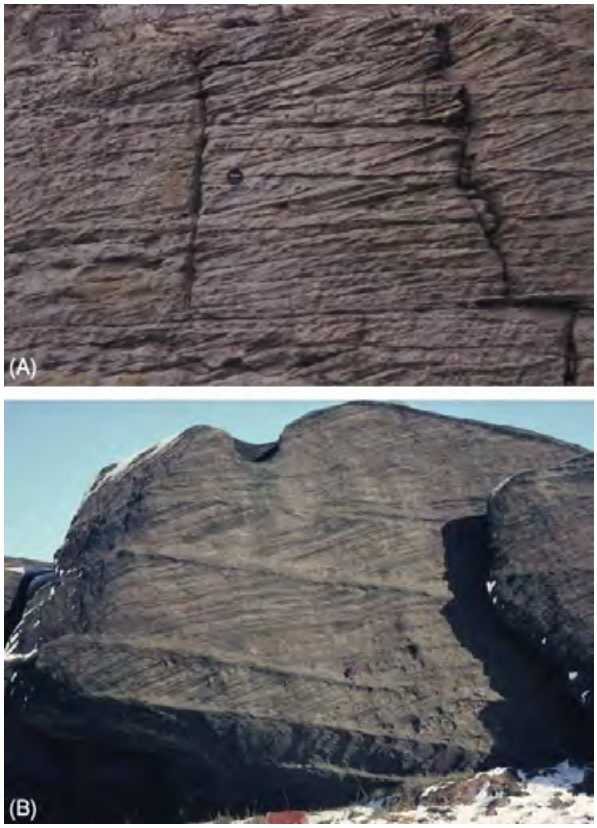


Figure 7 Examples of cross-bedding in sandstones. (A) Coset of small sets with slightly asymptotic foresets. The bounding surfaces between sets are themselves inclined downstream, suggesting deposition on the downstream side of a larger compound bar form. Westphalian, Northumberland, England. (B) Tabular sets with thin units of ripple lamination between sets. The foresets are essentially planar. Namurian, Staffordshire, England.

tabular sets vary in shape. This reflects the association between trough sets and lee-side scour pits and the fact that steep, three-dimensional dunes, which commonly have such pits, are associated with stronger currents. Tabular sets record the migration of dunes or sand waves without scour pits, probably with straight or only gently curved crest lines.

Changes in the shape of foresets along a set record changes in flow strength during the life of the bedform. They are a particular feature of tidal sediments and the deposits of rivers with large discharge variations. Wave action on emergence during the falling stage may create low-angle erosion surfaces, truncating the foresets. Succeeding foresets re-establish the higher angle dips above these 'reactivation surfaces' (Figure 8). Flow reversals in subtidal settings produce similar structures and tidal slack water can cause mud drapes to be interbedded with the foresets. In exceptional cases, the spacing of drapes can be interpreted in terms of spring-neap tidal cycles ('tidal bundles').

Figure 8 Reactivation surface within a tabular cross-bedded set produced by the migration of a low-relief sand wave. The discontinuity reflects modification of the lee side of the bedform during emergence during falling water discharge. Tana River, Norway.

Compound Bedforms and Bars

In many wide, sandy rivers, the large repetitive bedforms described above are organized into compound forms, which commonly appear to scale with the channel itself. Some of these features are attached to channel banks, whilst others are located in mid-channel. Some examples have a clear asymmetry to their stream-parallel profile with a slipface at the downstream end. In other cases, dunes descend a more gently inclined lee side. These compound features are commonly referred to as 'bars', of which there are many types. The important point, in terms of internal structure and its interpretation, is the relative movement directions of dunes and of the accretion of the bar. Where the bar advances downstream, in the same direction as the dunes, either a large, cross-bedded set will be present beneath smaller cross-beds, or the bounding surfaces between sets will be inclined downstream. Where bars move laterally as a result of dunes migrating along their flanks, bounding surfaces dip in a direction highly divergent from that of the cross-bedding (lateral accretion surfaces).

Upper Flow Regime Bedforms and Lamination

In situations in which water flow is very rapid and/or very shallow, saltating sand close to the bed moves continuously over an essentially flat surface. This phase of movement is known as 'plane bed transport' with the moving grain layer termed a 'traction carpet'. When the flow is particularly strong, the bed develops gentle undulations, forming ridges transverse to flow or three-dimensional domes. When this happens, the water surface also takes on a wave form which parallels that on the sediment surface. These undulations are 'standing waves' (Figure 9). Where the standing waves on the water

surface grow beyond a critical height, they move rapidly upstream and break, a case described as 'antidunes'. Breaking antidunes temporarily wipe out the undulations and the cycle of growth and breaking is repeated. This whole family of transport mechanisms is referred to as 'upper flow regime' transport.

Standing waves and antidunes are highly ephemeral features and preservation of their deposits is rare. However, where plane bed conditions coincide with bed aggradation, distinctive parallel and horizontal lamination forms. This may be accentuated by mica and gives rise to flaggy sandstones that split easily parallel to bedding. The bedding surfaces themselves show a distinctive lineation ('parting lineation') (Figure 10) parallel to the current, which is a useful palaeocurrent indicator. Rarely, parting lineation occurs on the bedding surfaces of undulatory lamination, indicating the chance preservation of standing waves.

Figure 9 Standing waves on the surface of a shallow, rapidly flowing stream. In-phase undulations on the underlying, rapidly moving sand bed may produce undulating lamination with low preservation potential.

Figure 10 Upper bedding surface of parallel laminated sandstone showing a clear parting lineation or primary current lineation. This lies parallel to the direction of the shallow, rapid flow that produced it. Middle Jurassic, Yorkshire Coast, England.

Wave Ripples and their Lamination

The to-and-fro movement of water associated with waves is able to move sediment, both on its own and in conjunction with unidirectional currents. Sand moved by waves is readily moulded into ripples, which are distinguished in several ways from current ripples. Wave ripples, most distinctively, have straight crests, which may be rounded or quite cuspate in profile (Figure 11). Their profile is also commonly symmetrical, although not always so. Where the strength of the wave surge is stronger in one direction than in the other, or where waves coexist with currents, asymmetry results. Wave ripples vary quite widely in spacing, depending on the grain size, the strength of the waves, and the water depth. Less common are ripples produced by coexisting wave trains with differing propagation directions. These show interference patterns, reflecting the wave sets responsible.

As well as producing ripples in their own right, waves also modify pre-existing bedforms, particularly under emergent or near-emergent conditions. Wave-modified current ripples are particularly common on tidal flats.

Internally, wave ripples produce small-scale cross-lamination which is not always easy to distinguish from that of current ripples. Interfingering bundles of laminae with opposed dips, draping laminae over symmetrical ripple crests, and discordances between internal lamination and ripple forms all suggest wave influence, but many cases are difficult to fully diagnose (Figure 12).

As well as producing ripples, waves also create plane bed conditions. This occurs most readily in the swash zone of beaches where the run-up and backwash of waves produce the shallow, rapid flows

Figure 11 Small, straight-crested wave ripples with symmetrical, rather rounded crest lines. The straightness of the crest lines and the tuning-fork junctions along the crests are most characteristic of wave influence. Tana River Delta, Norway.

Figure 12 A wave-ripple bedding surface with clear, symmetrical ripple profiles, underlain by a unit of ripple laminated sand showing bundled laminae and trough forms in a view normal to the crest lines. Namurian, County Clare, Ireland.

Figure 14 Heterolithic interlaminated sand and mud typical of many tidal settings where energy levels fluctuate between strong currents capable of transporting sand and still conditions where mud falls from suspension to drape the bed morphology. Lower Cretaceous, Isle of Wight, England.

Figure 13 Sandstone bed with undulating lamination with convex-upwards sectors and undulatory erosion surfaces. This assemblage of hummocky and swaley cross-stratification is typical of deposition from storms under conditions of coexisting, intense current and wave activity. Upper Jurassic, Dorset, England.

needed. Parallel lamination and parting lineations result if the sand is preserved.

Wave–Current Interaction

During storms, shallow-water shelf or nearshore areas are commonly subjected to energy regimes involving a combination of waves and currents. Under such conditions, sand falling from suspension may be reworked into low-relief, rounded forms and may be re-eroded by higher energy surges. The result is an undulatory lamination, characterized by gentle convex-upwards patterns and gentle scours. Where the convex-upwards laminae dominate, the structure is called 'hummocky cross-stratification', and where the scours dominate, it is termed 'swaley cross-stratification' (Figure 13).

Heterolithic Lamination

In settings in which energy levels fluctuate and in which there is a significant amount of sediment in suspension, it is common for mixed sand–mud sediments to be deposited. These record the alternation of suspended and bedload sediment deposition and are a particular, although not exclusive, feature of tidal settings. The sandy components are often ripple laminated with ripple profiles commonly preserved, so that current and wave forms can be distinguished. Collectively, these interlaminated sediments are termed 'heterolithic' (Figure 14) and are subdivided using the proportions of sandy and muddy components. Where mud dominates and sand forms isolated ripple lenses, the term 'lenticular bedding' is applied, whilst ripple-laminated sand with discrete mud drapes is called 'flaser bedding'. More or less equal components are referred to as 'wavy bedding'. Some confusion surrounds the application of these terms. Laminae rich in mica and carbonaceous debris in normal ripple cross-laminated sand can look superficially like flaser bedding, but the term is best reserved for cases with a clear separation of sand and mud.

Aeolian Bedforms and Internal Bedding

Wind, blowing over dry sand, initiates movement in much the same way as water. Above a certain critical velocity or shear stress, grains begin to move as bedload. However, in comparison with water, wind moves a much narrower range of grain sizes in this way. The buoyant effect of water is much greater than that of air and, in addition, the viscosity of water dampens collision impacts between grains in saltation. The result is that sands moved by air tend to be very well sorted and also well rounded. The first bedforms to develop on dry sand are ripples. These have low relief compared with aqueous current

ripples and are less clearly asymmetrical. They are commonly straight crested with the bifurcation of crest lines.

The vigorous saltation leads to grains following asymmetrical trajectories: a steep uplift from the bed, and a lower angle approach and impact. The impact throws other grains into motion, but the extent of this depends on the inclination of the bed at the point of impact. If the bed slopes upwind, the impact is steep and energy is dissipated strongly into grains near the impact point, throwing some into saltation and advancing others by a process of creep. If the bed slopes downwind, energy is dissipated more gently and grains accumulate. Wind ripples reflect this behaviour and their wavelength scales to the saltation path length and, hence, to the wind strength. Where wind blows dry sand across a surface that is immobile, such as a rock surface, a gravel bed, or a damp surface, wind ripples may be widely spaced and may begin to cluster together into larger forms that are incipient dunes. Where the surface is damp, saltating grains may adhere on impact through surface tension and patterns of wind 'adhesion ripples' or 'adhesion warts' may form. These irregular warty forms have steeper sides pointing upwind. All wind ripples have quite a low preservation potential, but examples do occur in the rock record. Wind ripples commonly produce subparallel lamination, 'pinstripe' lamination with inverse grading, and rare cross-lamination. It is an important component in the cross-bedding produced by aeolian dunes.

Vigorous wind transport commonly leads to the development of larger bedforms, 'aeolian dunes' (*see* **Sedimentary Processes:** Aeolian Processes), which occur in isolation on hard substrates or as parts of larger accumulations of sand. Dunes vary greatly in both scale and shape, and it is quite common for dunes of several scales to occur superimposed upon one another. An important distinction is between dunes with their own slipfaces and those without such surfaces. Slipfaces occur downwind of crest lines where sand accumulates through grain flow avalanches. Such processes commonly occur in conjunction with grainfall, whereby grains are thrown over the dune crest and accumulate on the lee side. Where grainfall dominates, the inclination of the surfaces may be lower and reworking by wind ripples may occur, particularly in the lower, more gently inclined parts of the dunes ('plinths').

Where sand supply is insufficient for the substrate to be fully covered, 'barchan dunes' or, more rarely, 'linear (seif) dunes' occur. In sand seas, a variety of more complex forms occurs. These include transverse dunes, barchanoid forms, and star-shaped dunes. The largest forms, termed 'draa', commonly have superimposed smaller dunes, which leads to complex morphologies. The size of the largest forms means that they change shape only over long periods of time and thus reflect a wind regime rather than a particular wind episode. Some are out of phase with the current regime as a result of lag effects.

Internally, dune sands show complex cross-bedding at scales up to many metres thick. The inclined laminae are commonly well defined and it may be possible to differentiate grainfall, grain-flow and wind-ripple laminae. Wind-ripple laminae generally occur in more gently inclined intervals, whilst grain flow is commonly close to the angle of rest (*ca*. 30°). Aeolian dune cross-bedding is characterized by discordances or bounding surfaces at several scales. Individual cross-bedded sets are separated from one another by erosion surfaces, and internally they may also be punctuated by low-angle erosion surfaces similar to the reactivation surfaces of aqueous cross-beds. These record the complexity of the wind regime whereby the lee sides of dunes become sites of erosion during particular wind episodes. They may also result from the migration of dunes over slipfaceless draas or the oblique migration of scour pits along the flanks of dunes.

Structureless Sand and Sandstone

Not all sand and sandstone has visible internal lamination. This happens for several reasons. First, it may have been deposited in that way. The dumping of sediment from suspension from, for example, a decelerating current, may have been so rapid that there was no time for the sand to be reworked into bedforms and laminae. On the other hand, sand that was initially laminated may have lost its lamination through later remobilization, perhaps due to liquefaction. Finally, lamination may be lost through the activities of burrowing organisms, which have the ability to totally homogenize laminated sediments (*see* **Trace Fossils**).

Decelerating Flows and the Bouma Sequence

Many sandstone beds in interbedded sandstone–mudstone successions show an internal vertical sequence of lamination types that are diagnostic of decelerating flows. The sandstones are sharp based and may have erosional sole marks (*see* **Sedimentary Processes:** Erosional Sedimentary Structures). Above their bases, they show one or more intervals of five different lamination types that occur in a constant relative order even though they are rarely all present in the same bed. These five types are: A, structureless

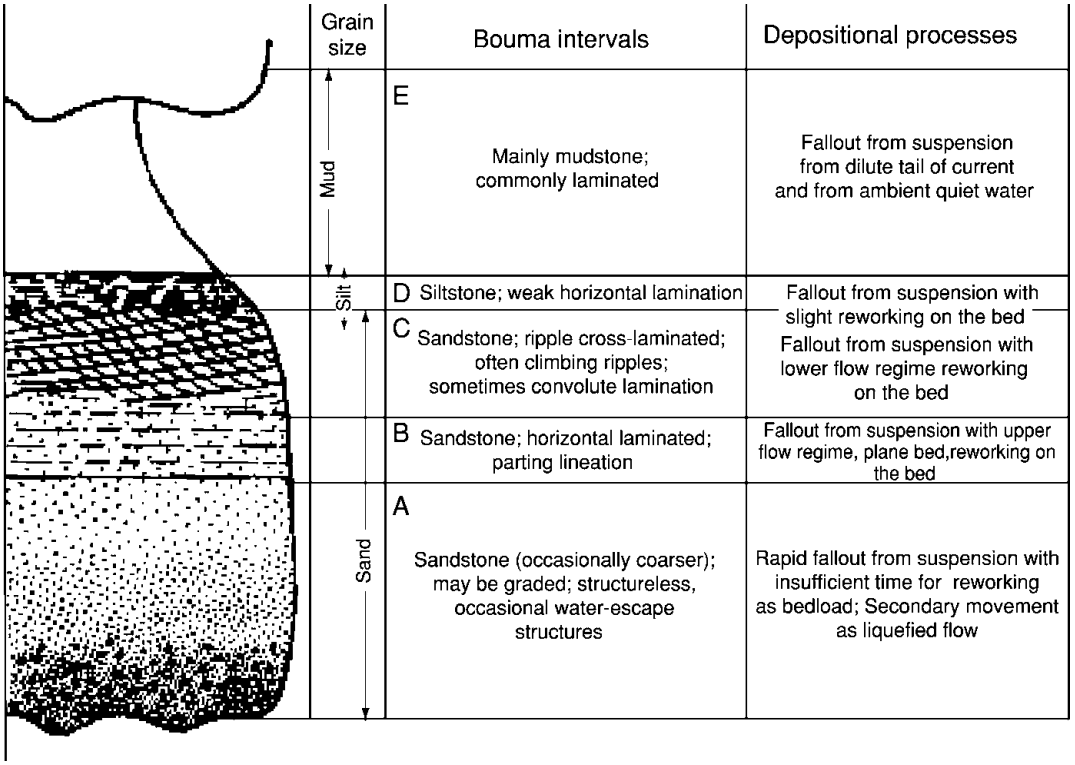


Figure 15 A schematic profile through a turbidite sandstone bed showing a complete sequence of Bouma intervals and their hydrodynamic interpretation.

(commonly graded) sandstone; B, parallel laminated sandstone; C, ripple cross-laminated sandstone; D, horizontally laminated (commonly silty) sandstone; E, mudstone. The complete succession of lamination types is known as the 'Bouma sequence' (Figure 15). The A interval records rapid deposition from suspension, with graded bedding reflecting sorting within the flow of a mixed grain size population. The B interval records upper flow regime plane bed transport, whilst the C interval records ripples moving in the lower flow regime. Climbing ripple lamination is quite common, recording rapid rates of bed aggradation, and convolute lamination may also result from rapid deposition. The D interval, often poorly displayed in weathered exposures, records fallout from suspension with weak reworking on the bed, whilst the E interval grades into the background sedimentation from suspension.

The overall succession, however complete or incomplete, records an episodic current, possibly erosional in its early stages, and progressively decelerating thereafter. Most commonly, the Bouma sequence is associated with deep-water settings where the episodic flows are turbidity currents (*see* **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes), but similar decelerating flows occur in shallow-water settings of floodplains, shallow lagoons, and lakes due to crevassing of river banks during floods. Similar sands also occur in shallow marine settings as a result of storms, although the Bouma sequence may then be modified to include hummocky cross-stratification and wave ripples.

Figure 16 Coarse fluvial gravels showing clear pebble imbrication. The flat surfaces of the clasts dip upstream, indicating that flow, in this case, was from left to right. Holocene, Pyrenees, Spain.

Gravel Forms and Fabric

Whilst the preceding section deals mainly with sands and finer grained sediments, powerful water currents also transport gravel, and its deposition can lead to distinctive structures and fabrics (*see* **Sedimentary**

Rocks: Rudaceous Rocks). Where gravel moves as a traction load, on a stream bed or on a beach, it may become lodged on the bed, particularly where clasts land amongst similar sized clasts. If the clasts are flattened, they will tend to be deposited in an imbricated fabric with an upstream dip. This is the most stable position beneath the prevailing flow and, when seen in the rock record, 'imbrication' (Figure 16) is a valuable palaeocurrent indicator.

Gravel is also deposited in bars on river beds and these commonly have slipfaces at their downstream ends. Cross-bedding results from the avalanching that takes place there. Depending on the overall sediment population, sandy and gravel-rich foresets may be intermixed.

See Also

Sedimentary Environments: Depositional Systems and Facies; Lake Processes and Deposits. **Sedimentary Processes:** Erosional Sedimentary Structures; Post-Depositional Sedimentary Structures; Aeolian Processes; Particle-Driven Subaqueous Gravity Processes. **Sedimentary Rocks:** Rudaceous Rocks. **Trace Fossils**. **Unidirectional Aqueous Flow**.

Further Reading

Allen JRL (1982) *Sedimentary Structures: Their Character and Physical Basis. Developments in Sedimentology 30A & B*, pp. 593, 663. Amsterdam: Elsevier.

Allen JRL (1985) *Principles of Physical Sedimentology.* ch. 4. London: Allen & Unwin.

Brookfield ME and Alhbrandt TS (eds.) (1983) *Eolian Sediments and Processes. Developments in Sedimentology 38*, p. 660. Amsterdam: Elsevier.

Collinson JD and Thompson DB (1989) *Sedimentary Structures*, 2nd edn., ch. 5–7. London: Chapman & Hall.

Leeder MR (1982) *Sedimentology; Process and Product*, ch. 6–8. London: Allen & Unwin.

Leeder MR (1999) *Sedimentology and Sedimentary Basins*, ch. 4–12. Oxford: Blackwell Science.

Mutti E (1992) *Turbidite Sandstones*. Milan: Agip.

Pettijohn FJ and Potter PE (1964) *Atlas and Glossary of Primary Sedimentary Structures*. Berlin: Springer.

Pye K and Lancaster N (eds.) (1993) *Aeolian Sediments Ancient and Modern. Special Publication of the International Association of Sedimentologists 16*, p. 167. Oxford: Blackwell Scientific Publications.

Reineck HE and Singh IB (1980) *Depositional Sedimentary Environments*, 2nd edn. Berlin: Springer.

Selley RC (2000) *Applied Sedimentology*, 2nd edn., Ch. 5. San Diego: Academic press.

Post-Depositional Sedimentary Structures

J Collinson, John Collinson Consulting, Beech, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Primary sedimentary structures and lamination in clastic sediments result from deposition under different hydrodynamic regimes. However, there are other secondary structures, which originate during or very soon after deposition as a result of sediment instability, extreme climatic effects, or chemical changes during burial. The last group involves the precipitation of mineral cements within the sediment. Some of the early physical structures deform primary depositional structures and lamination, whilst others originate entirely as a result of the deformation processes. Those associated with cementation often help to preserve original lamination and other delicate features, particularly in fine-grained sediment, that would otherwise have been destroyed by compaction.

Soft-Sediment Deformation Processes

Early formed soft-sediment deformation structures are typically features of sandstones and finer sediment. Deformation in gravels and conglomerates is rare, except in glacial and periglacial settings, or where there is abundant muddy matrix. The structures have an entirely physical origin and must be distinguished from organic deformation structures, such as burrows and trails of animals and disturbance by plant roots (*see* **Trace Fossils**). In addition, tectonic structures are often geometrically identical with syndepositional structures, and the boundaries between widespread tectonic deformation and more local, sediment-induced deformation may be blurred.

Deformation structures are quite difficult to classify as they result from rather diverse processes. However, all attempts at interpreting physical soft-sediment deformation structures must consider two discrete questions. First, why did the sediment lose its strength and become susceptible to deformation? Second, what forces operated on the weakened sediment to cause the deformation?

In most cases, the sediment loses frictional shear strength and behaves, briefly, as a plastic or a fluid. In other cases, a loss of tensile strength can lead to brittle deformation as soft-sediment faults. Loss of frictional shear strength may result in total remobilization of the sediment and transformation into a gravitational

mass movement, leading to a phase of resedimentation. However, this article deals only with situations in which the loss of strength was short-lived and the sediment regained its strength before it moved downslope, thus preserving structures that record more or less *in situ* deformation.

Deformation occurs within sediment when intergranular forces are unable to resist applied stresses, which are usually, but not exclusively, gravitational. The shear strength of sediment is normally expressed by the equation:

$$\tau = C + (\sigma - p)\tan\varphi$$

where τ is the shear strength, C is the grain cohesion, σ is the pressure normal to shear, p is the excess pore fluid pressure, and φ is the angle of internal friction. This means that sediment will fail when the applied stress exceeds τ; such a condition will be favoured by a lower cohesive strength C, by changes in grain packing to reduce $\tan\varphi$, or by an increase in pore fluid pressure p beyond a certain critical value. Such conditions can occur for several reasons, discussed below. Strain, when it occurs, may be isotropic, dispersed throughout the body of sediment, or concentrated on discrete slip surfaces.

Loss of Strength

In the shear strength equation set out above, the presence of fine-grained, especially muddy sediment tends to increase the role of the cohesive strength, which is largely a function of grain size, whilst in better sorted, mud-free sediment, frictional forces dominate. In both cohesive and non-cohesive sediments, high pore water pressure commonly leads to a loss of shear strength, but this condition can have different causes and durations in different sediments.

In fine-grained, muddy sediments, excess pore fluid pressure is often caused by relatively rapid deposition. Low permeability inhibits the escape of pore water and hence retards sediment compaction, giving overpressured conditions. In addition, the decay of organic matter within the sediment can generate gases, which, if unable to escape, also contribute to pore pressure. Such overpressured conditions occur close to the sediment surface where highly mobile mud may be susceptible to deformation or flowage. They also develop at a larger scale, and in a longer time frame, within thick piles of sediment, such as those deposited by prograding deltas or deltaic margins (e.g., Mississippi, Niger). In such cases, increasing pore water pressure with burial may eventually exceed lithostatic pressure, leading to failure. This results in plastic flowage of a thick, overpressured layer. These movements set up extensional stresses in the overburden, which can lead to the development of discrete slip surfaces that cut up-section as listric faults when the tensile strength of the sediments is exceeded. Such surfaces may also become the bounding surfaces of sediment gravity slides. These conditions, and the resultant movements, develop and persist over long intervals of time; in the case of the largest deltaic margins, over millions of years.

Rapidly deposited sands, lacking significant fine-grained sediment, commonly have rather loose grain packing. They are particularly susceptible to shock, such as by an earthquake, heavy wave action, further rapid deposition, or a sudden rise in water level. Shock breaks grain contacts and induces tighter packing and, as a result, excess pore water is present, raising the pore water pressure and causing temporary liquefaction, which will last until the extra pore water escapes. In the case of sands, the high permeability usually means that liquefaction is short-lived. Whilst liquefied, the sediment–water mixture deforms readily as a Newtonian fluid in response to gravitational and other applied stresses. As pore fluid escapes, usually upwards to the sediment surface, grain contacts are re-established and a rising front of reconsolidation 'freezes' the deformed sediment. Deformation is usually identified through the distortion of depositional lamination but, with extreme or sustained liquefaction and deformation, depositional lamination is totally obliterated and structureless sand or slurried textures are produced. The upwards movement of escaping pore water may also create deformation, both by distorting depositional laminae and by creating new vertical structures, such as dishes, pipes, and sheets, within which sediment–water mixtures move as fluidized flows.

Deforming Forces

Virtually all the forces that act upon sediment weakened by the above processes are gravitational. These can be divided into two major classes: a down-slope component of gravity and the action of gravity on inverted density gradients. In addition, shear at the sediment surface due to flowing water or ice, or to sediment gravity flows, may also deform the sediment.

The down-slope gravitational component may be sufficient to overcome the cohesive strength of muddy sediment on a slope. This can lead to the detachment of a slab of sediment on a basal shear surface and to internal deformation within the moving layer. The style of deformation varies depending upon the position within the moving sheet. Stretching, particularly at the up-slope end, may lead to tensile failure and to the development of brittle extensional structures, whilst compression, commonly in a down-slope setting, causes plastic folds. These styles of deformation are associated with a family of mass

movements dealt with in detail elsewhere (*see* **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes). Also partly driven by down-slope gravitational components are sedimentary growth faults. These are related to failures in deeply buried overpressured muds. Extension on the faults is also driven by lateral flowage of these muds due to vertical loading by the overlying sediment pile.

Density inversions occur as a result of both depositional processes and the early loss of pore water. A sand layer, deposited on mud, will commonly have a higher density than the mud and, if the mud is weak, the instability may lead to loading of sand into the mud. Where the sand is of uneven thickness, as with ripple lenses or erosional sole marks, loading may be concentrated where the sand is thickest. Extreme cases of loading show detached balls of sand, usually encased in slurried mud or silt: so-called ball and pillow structures.

Within liquefied sand layers, which commonly result from shock, patterns of density inversion may be more subtle. Internal depositional lamination may be contorted in convolute lamination, some of which reflects the upward escape of water or low-density sediment–water mixtures. Very rapidly deposited sands may lack depositional lamination and experience liquefaction during the depositional process. The dewatering of such sands and the elutriation of fines by escaping water can produce subtle features, such as dish structures and water-escape pillars and pipes.

Deeper burial and associated differential compaction lead to the deformation of larger scale depositional geometries. The cross-sections of channel sand bodies, in particular, are susceptible to change from planar-convex forms to concavo-convex patterns where encased in muds. Where channel sands are encased in mud, as in some deep-water depositional settings, dewatering of the muds during burial can lead to high pore water pressures in the sands. This can eventually lead to rupture of the mud seal and to the injection of liquefied sands into the muds as sills and dykes. When such processes occur near the sediment surface, liquefied sand may be extruded as slurry flows or as sand volcanoes.

Where sands are deposited in large bedforms, such as dunes and sand waves, they may be liquefied through rapid deposition on the lee slope. If this happens whilst the current is still active, the cross-bedded sands, developed by lee-side deposition, may be sheared and the foresets dragged into folds, giving overturned cross-bedding.

The movement of ice over unconsolidated sediment also generates large-scale and intense deformation. This occurs near the front of a glacier, where glacial advance operates as a bulldozer to fold and thrust proglacial deposits. In addition, the basal layers of a mobile glacier may be rich in sediment. Shearing across the basal layers can lead to intense deformation, although the rheology mainly involves plastic deformation of ice rather than sediment liquefaction or cohesion.

Soft-Sediment Deformation Structures

This section deals with the descriptive aspects of soft-sediment deformation and relates the structures to the deforming processes. The structures themselves do not fall naturally into well-defined classes, but Figure 1 shows a scheme which relates all soft-sediment structures to processes. This diagram illustrates the possible interactions between different causes of a loss of strength and deforming forces, and shows that processes are gradational with tectonic deformation and with resedimentation. It also shows the wide range of scales across which deformation occurs, from microloading to growth faults at the scale of continental slopes, and highlights the differences in timing and duration of deformation. Structures formed during deposition, such as overturned cross-bedding, contrast with structures that require significant later burial, such as growth faults. Commonly occurring deformation structures are described below, in order of increasing scale, but moderated by considerations of process.

Load Casts and Pseudonodules

These usually occur at the interface between a mud layer and an overlying sand layer. Downward-facing bulbous load casts of sand are separated by upward-pointing flame structures of mud (Figure 2). Some occur on otherwise flat bedding surfaces, whilst others accentuate erosional or depositional irregularities. In extreme cases, sand may be detached from its source bed and occur as isolated load balls, or pseudonodules, within highly disturbed muds or silts, a features sometimes referred to as a 'ball and pillow' structure. Internal lamination within both sand and mud is highly contorted. Loading is caused by gravity acting on unstable density stratification where the sand layer is denser than the muds and where both layers are weakened by excess pore fluid.

Convolute Lamination

This commonly occurs within sands or silts and involves the distortion of depositional lamination by folding of variable intensity (Figure 3). In some cases, the folding is chaotic, whilst in others upward movement of escaping water has dragged lamination into upright folds. Cuspate folds with sharp anticlines and

<table>
<tr><th colspan="2" rowspan="2">Nature of deforming force \ Loss of strength</th><th colspan="3">Exceed strength of sediment</th><th colspan="2">Liquidize</th></tr>
<tr><th>Internal tensile (Brittle)</th><th>Internal cohesive (Plastic)</th><th>External surface cohesive (Plastic)</th><th>Liquefied</th><th>Fluidized</th></tr>
<tr><td colspan="2">Gravitational body force on slope</td><td>Slides</td><td>Slumps</td><td>Slumps and slides</td><td colspan="2">Debris flows</td></tr>
<tr><td colspan="2" rowspan="2">Unequal confining load</td><td rowspan="2">Growth faults</td><td colspan="2" rowspan="2">Loaded ripples shale ridges and diapirs</td><td colspan="2">Loaded ripples and sole marks</td></tr>
<tr><td></td><td>Clastic dykes sand volcanoes</td></tr>
<tr><td rowspan="4">Gravitationally unstable density gradient (density inversion)</td><td>Continuous</td><td rowspan="5">Soft sediment faults</td><td colspan="2" rowspan="2"></td><td>Convolute lamination</td><td></td></tr>
<tr><td>Within a single layer</td><td>Dish structures</td><td>Water-escape pipes and pillars</td></tr>
<tr><td>Multiple layer, not pierced</td><td colspan="3"></td><td colspan="2">Bedding surface load casts</td></tr>
<tr><td>Multiple layer, pierced</td><td colspan="2">Shale ridges and mud diapirs</td><td colspan="2">Ball and pillow/pseudonodules Isolated load balls</td></tr>
<tr><td rowspan="2">Applied shear stress</td><td>Current drag</td><td colspan="2" rowspan="2"></td><td>Overturned cross-bedding</td><td></td></tr>
<tr><td>Vertical</td><td></td><td>Water-escape pipes and pillars</td></tr>
</table>

Figure 1 The relationship of various post-depositional structures to the nature of the loss of strength and the type of deforming force. Slumps and slides, which involve lateral resedimentation of material, are dealt with in detail in (*see* **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes). Modified from Collinson JD (1994) Sedimentary deformational structures. In: Maltman A (ed.) *The Geological Deformation of Sediments*, pp. 95–125. London: Chapman & Hall; and Collinson JD (2003) Deformation structures and growth faults. In: Middleton GV (ed.) *Encyclopedia of Sediments and Sedimentary Rocks*, pp. 193–195. Dordrecht, Kluwer: Academic Publishers.

Figure 2 Small-scale load casts within thin-bedded turbidites. The downward-sinking sand lobes are separated by upward-pointing flames of mud. Carboniferous, North Cornwall, England.

more rounded synclines are typical. In many cases, depositional lamination can still be identified, commonly cross-bedding or cross-lamination. The good sorting and high porosities of aeolian sands make them particularly susceptible to liquefaction when the water table rises rapidly through them, giving convolute lamination on a large scale. Convolute lamination is common in turbidite sandstones, often

Figure 3 Convolute bedding developed in coarse cross-bedded fluvial sandstones as a result of the sediment experiencing a short-lived period of liquefaction shortly after deposition. Namurian, Yorkshire, England.

associated with the ripple laminated (Bouma C) interval (*see* **Sedimentary Processes:** Depositional Sedimentary Structures; Deep Water Processes and Deposits).

All examples reflect liquefaction of the sand, which can happen for a variety of reasons, including seismic shock, rapid sedimentation, breaking waves, or a shift in water table. Where seismic shock caused liquefaction, deformed layers may be very widespread and have stratigraphical significance. Upward escape of excess pore water, internal density inversions, and down-slope components of gravity may all contribute to deformation. Dewatering leads to reconsolidation of the sediment from the bottom up, 'freezing' the deforming laminae. Protracted liquefaction may lead to total homogenization.

Overturned Cross-Bedding

This is a special case of convolute lamination in which liquefaction occurred in an actively migrating bedform (Figure 4). Current shear on the sediment–water interface dragged the liquefied sand in a down-current direction in an essentially laminar style. An upward-migrating front of reconsolidation allowed higher parts of the sediment to be sheared longer, giving down-current-facing recumbent folds.

Dish and Pillar Structures

These are most common in thick, otherwise massive sands and result from dewatering following rapid

Figure 4 Overturned cross-bedding in two small sets within a shallow marine sandstone. This was caused by ongoing bed shear by a current during a short-lived loss of strength. The deformed sand was progressively 'frozen' as a front of reconsolidation moved upwards through the sediment. Late Precambrian, north-east Greenland.

Figure 5 A thick bed of sandstone in a turbidite setting showing intense development of dish structures due to rapid escape of pore water shortly after rapid deposition. Eocene, Fuenterrabia, Spain.

deposition. Dish and pillar structures are direct products of dewatering and are not distortions of pre-existing lamination. Dish structures are thin concave-upward, subhorizontal zones of slight clay enrichment produced by local filtering out of the elutriated fines (Figure 5). Their upturned edges merge with pillar structures, which are vertical conduits of water escape and record fluidized or near-fluidized conditions.

Sand Injection Structures

Where a sand body is encased in mud and progressively buried, compaction and dewatering of the muds can lead to overpressure in the sands. If the fluid pressure exceeds the tensile strength of the overlying mud, rupturing occurs and complex networks of dykes and sills of liquefied sand are intruded into the muds. This appears to be quite common in subsurface channel sand bodies in deep-water and slope settings, although outcrop examples are rare.

Small-scale injection dykes occur in mudstones of interbedded sandstone and mudstone successions. These may be folded ptygmatically because of differential compaction. Sand within dykes is typically structureless, although some examples show foliation parallel with dyke margins. Remobilized sands are increasingly being recognized as important petroleum reservoirs in the North Sea and elsewhere in the world.

Sand Volcanoes and Extruded Sheets

Sometimes upward-intruding liquefied sand, both dykes and pipes, penetrates to the free sediment surface where the sand is extruded. Where the escape conduits are pipes, sand volcanoes form, at scales from a few centimetres to a few metres (Figure 6). If the intrusions are dykes, fissure extrusion occurs and liquefied sand may flow as sheets before dewatering and 'freezing'. Where sand dykes penetrate muddy debris flows, extruded sand may be interfolded with the upper levels of the debris flow.

Mud Diapirs

Overpressure develops where muds are buried quite rapidly, as in a prograding delta. The resultant loss of strength and the significant overburden lead to both vertical and horizontal flowage of the muds. Where the overlying sediment is denser than the mud, as in an upward-coarsening deltaic succession, mud may rise vertically as a diapir or mudlump, pushing aside or penetrating mouth bar sands. In the Mississippi delta, mudlumps rise to sea-level and create short-lived islands offshore from distributary mouths. Mud diapirism may lead to great variability in the thickness of mouth bar sands, which thicken into withdrawal synclines between diapirs (Figure 7).

Horizontal flowage of overpressured muds leads to the extrusion of muds at the base of the prograding slope as imbricate thrusts and folds. The same motion is partly responsible for extensional stresses that drive extensional sedimentary growth faults in deltaic successions. Mud diapirs are widespread in Tertiary deltas around the world and may help to trap petroleum.

Slumps and Slides

Gravitational body forces acting on sediments lying on a slope lead to slumps and slides when cohesive (plastic) and tensile (brittle) strengths are exceeded. In both types, shearing is concentrated on discrete basal surfaces. They differ principally in their internal deformation. Slides are largely undeformed, but may have internal faults, whilst slumps show plastic folding. The two styles may coexist in the same deformed unit. The fold orientation may indicate palaeoslope direction, but analysis must be carried out with caution. Gravitational mass movements are dealt with more fully in (*see* **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes).

Figure 6 Sand volcanoes on top of a slump sheet. Dewatering of the slump led to rapid upwards injection of a sand slurry which was extruded at a point vent to give the volcano. Namurian, County Clare, Ireland.

Figure 7 Mud diapir penetrating delta front mouth bar sands. Overpressuring of the buried muds led to a loss of strength which, combined with a density inversion, led to the upwards flow of mud. Namurian, County Clare, Ireland.

Sedimentary Growth Faults

These occur in deltaic successions. They are large-scale features and require exceptional exposures to be identified at outcrop (Figure 8). In small exposures, they may be confused with later tectonic faults. They are caused by shear failure at discrete dislocations within the sedimentary pile, and they commonly occur as listric surfaces that sole out on a basal '*décollement*'. In small examples, within a single progradational unit, the basal surface may be within particularly fine-grained mudstone directly above a basal flooding surface. In larger examples, penetrating several progradational units, the *décollement* may be within a thick interval of overpressured muds, whose lateral flowage towards the free surface slope combines with gravitational forces on the hanging wall block to drive that block down-slope. Growth faults are characterized by slow and continual movement

Figure 8 Small, synsedimentary growth fault within deltaic sediment. Note the thickening of sands into the hanging wall of the fault and the roll-over within these sands. The synsedimentary nature of the fault is shown by the fact that the uppermost beds pass over the fault without disturbance. Namurian, County Clare, Ireland.

and are aided by progressive sediment loading. Their hanging wall areas act as local depocentres in which thickened delta-front sediments are deposited and preserved. Continued progradation of individual deltas or of the continental slope causes the active faulting to shift progressively basinwards. Large growth faults are important traps for petroleum in deltas such as the Mississippi and the Niger.

Climatically Induced Structures

In addition to the soft-sediment structures described above, in which a temporary loss of strength causes the disturbance, there are some rather diverse structures which result from climatic effects.

Desiccation and Other Cracks

When surface layers of mud or silt dry out, they contract and create isotropic horizontal tension. If this exceeds the tensile strength of the mud, polygonal (commonly hexagonal) cracks develop which may be filled with wind-blown or water-lain sand (Figure 9). The dimensions of the polygons are a function of the mud layer thickness, and several scales may coexist. Subaqueous shrinkage (synaeresis) of surface mud layers also leads to cracks, although these often form less regular patterns. They are thought to result from volume changes of clays within the surface layer. Synaeresis cracks may commonly be differentiated from subaerial desiccation cracks because they are infilled with mud, not sand. Large-scale polygonal cracks also develop through protracted freeze–thaw processes in periglacial settings.

Figure 9 Desiccation cracks on an upper bedding surface in ephemeral lacustrine sediments. Carboniferous, New Brunswick, Canada.

Evaporite Pseudomorphs

Evaporation of bodies of saline water leads to saturated solutions and to the precipitation of minerals on the floor of the lake or lagoon. If saturated conditions persist, large, euhedral crystals may grow at or just below the muddy sediment surface. When a subsequent flood or storm introduces both fresher water and coarse sediment, the crystals may be dissolved and their forms buried and cast by the coarser material. The bases of sandstone interbeds in playa lake sequences are often decorated by cubic pseudomorphs of halite, commonly with hopper faces (Figure 10), whilst in marine or lagoonal settings, gypsum pseudomorphs may occur.

Figure 10 Halite pseudomorphs on the base of a fine-grained sandstone bed. Triassic, Cheshire, England.

Raindrop Impressions

These occur as small circular or elliptical depressions on the upper surfaces of beds of fine sediment. They record the impact of raindrops or hailstones and their preservation depends on the cohesive strength of the sediment. They are quite often associated with desiccation cracks.

Periglacial Deformation

This results mainly from the expansion of water on freezing and the contraction of frozen sediment during extreme freezing. This takes place on a seasonal basis in an active layer close to the surface and creates metre-scale patterns of polygonal cracks. Open cracks are filled with wind-blown sediment and material collapsing from the crack sides during periods of thaw. Expansion on freezing leads to the pushing up of ramparts on the edges of the cracks. In vertical section, frost polygons appear as sharply tapering wedges with complex, poly-phase fills. Convolute

Figure 11 Large, discoid, calcite-cemented concretions developed in mudstones. Lower Jurassic, Yorkshire, England.

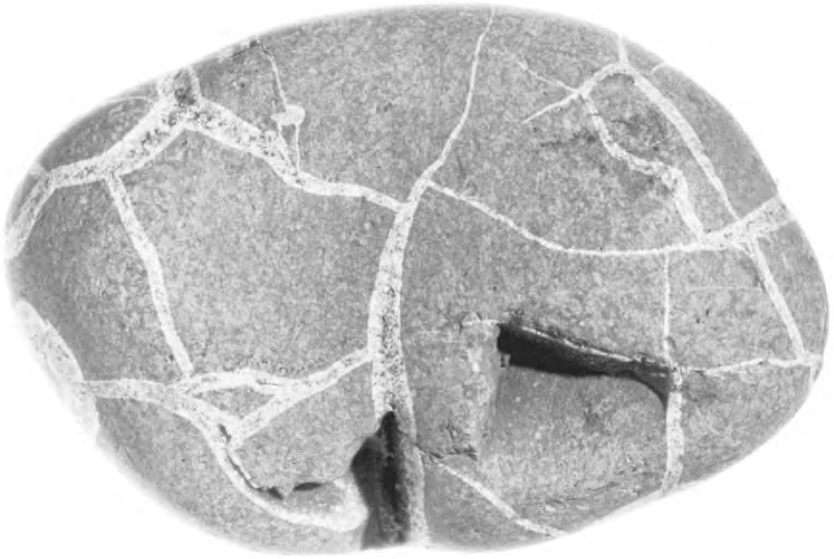


Figure 12 Septarian nodule developed in mudstone with calcite-filled cracks which demonstrate shrinkage of the concretion. Kimmeridge Clay, Dorset, England. Courtesy of R Selley.

bedding also occurs in periglacial sediment due to freeze–thaw action termed cryoturbation.

Concretions

During the burial of sediment, pore water may precipitate minerals in the pore spaces. In some cases, this is as pervasive cement, but, in other cases, it occurs in discrete patches around nuclei (Figure 11). In some cases, these cemented concretions or nodules occur apparently at random, whilst others follow bedding or are nucleated on obvious earlier inhomogeneities, such as fossils, burrows, or root traces. Some concretions, which form at shallow burial depths in fine-grained sediment, predate major compaction and, by resisting compaction, help to preserve lamination and delicate fossils that would otherwise be flattened beyond recognition. Other early concretions (e.g., calcrete nodules) are an integral part of soil textures and are discussed elsewhere (*see* **Soils:** Modern; Palaeosols). Most concretions occur during deeper burial and many have distinctive internal features. 'Septarian nodules' have internal cracks filled with calcite, reflecting shrinkage during their development (Figure 12). Other concretions have complex nested conical fractures, termed 'cone-in-cone structure'.

See Also

Sedimentary Processes: Depositional Sedimentary Structures; Deep Water Processes and Deposits; Particle-Driven Subaqueous Gravity Processes. **Soils:** Modern; Palaeosols. **Trace Fossils**.

Further Reading

Allen JRL (1982) *Sedimentary Structures: Their Character and Physical Basis. Developments in Sedimentology 30A & B*, pp. 593, 663. Amsterdam: Elsevier.

Allen JRL (1985) *Principles of Physical Sedimentology*, ch. 10. London: Allen & Unwin.

Collinson JD (2003) Deformation structures and growth faults. In: Middleton GV (ed.) *Encyclopedia of Sediments and Sedimentary Rocks*, pp. 193–195. Dordrecht, Kluwer: Academic Publishers.

Collinson JD (1994) Sedimentary deformational structures. In: Maltman A (ed.) *The Geological Deformation of Sediments*, pp. 95–125. London: Chapman & Hall.

Collinson JD and Thompson DB (1989) *Sedimentary Structures*, 2nd edn., ch. 9. London: Chapman & Hall.

Jones ME and Preston RMF (eds.) (1987) *Deformation of Sediments and Sedimentary Rocks. Special Publication of the Geological Society, London, 29*. London: Geological Society.

Leeder MR (1982) *Sedimentology; Process and Product*, ch. 11. London: Allen & Unwin.

Leeder MR (1999) *Sedimentology and Sedimentary Basins*, ch. 12. Oxford: Blackwell Science.

Mutti E (1992) *Turbidite Sandstones*. Milan: Agip.

Pettijohn FJ and Potter PE (1964) *Atlas and Glossary of Primary Sedimentary Structures*. Berlin: Springer.

Reineck HE and Singh IB (1980) *Depositional Sedimentary Environments*, 2nd edn. Berlin: Springer.

Selley RC (2000) *Applied Sedimentology*, 2nd edn., ch. 5. San Diego: Academic Press.

Aeolian Processes

N Lancaster, Desert Research Institute, Reno, NV, and United States Geological Survey, Reston, VA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Aeolian processes, involving erosion, transportation, and deposition of sediment by the wind, occur in a variety of environments, including beaches, semi-arid and arid regions (e.g., cold and hot deserts), agricultural fields, and some terrestrial planets, notably Mars and possibly Venus. Common features of these environments are a sparse or non-existent vegetation cover, a supply of fine sediment (clay, silt, and sand size), and strong winds. Aeolian processes are responsible for the emission and/or mobilization of dust and formation of areas of sand dunes. They largely depend on other geological agents (e.g., rivers, waves) to supply sediment for transport. Despite assertions by many early workers, erosion of bedrock by the wind is not a significant process in most regions.

Sediment Movement by the Wind

Most naturally occurring airflow is turbulent. Mixing of turbulent eddies, each with a different velocity and direction, together with the effect of friction at the surface, results in a wind profile with a logarithmic increase in velocity with height above the aerodynamic roughness length (z_o) (Figure 1). The aerodynamic roughness length is determined by the particle size and roughness of the surface. For sand surface, z_o is typically 1/30th of the mean particle size. For surfaces with scattered isolated roughness elements such as plants or rocks, aerodynamic roughness scales with the roughness density (the volume of roughness elements per unit area). The existence of roughness elements plays an important role in the interaction of the wind with the surface because the available wind shear stress is partitioned between the roughness elements and the intervening surface. A high density of roughness elements will therefore reduce the wind shear stress on the surface so that sediment mobilization and transport may be reduced or eliminated.

Particle movement is achieved via a combination of direct wind shear stress on the surface and atmospheric turbulence. There are three modes of sediment transport by wind: creep or reptation, saltation, and suspension (Figure 2). The mode of transport depends primarily on the ratio between particle settling velocity, and hence particle size, and wind shear stress and turbulence intensity. Very small particles (<20 μm) are transported in suspension (tens of kilometres or greater) and are kept aloft by turbulent eddies in the wind. True suspension occurs when the particle settling velocity is very small compared to the turbulence intensity of the wind. Larger particles (20–63 μm) undergo short-term suspension for distances of tens to hundreds of metres; material of sand size (63–1000 μm) is transported mainly in a series of short hops (saltation), in which the vertical component of wind velocity (turbulence) has no effect on particle trajectories. Material coarser than 500 μm diameter is transported in contact with the surface by reptation and creep. The modes of transport are interdependent: saltating sand particles eject silt- and

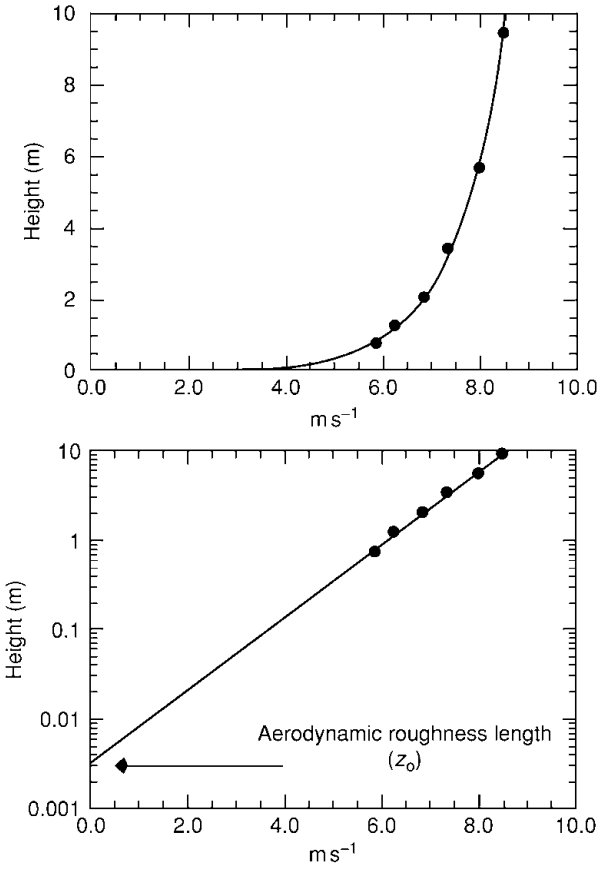


Figure 1 Development of wind profiles. The wind profile equation is $u_* = u/\kappa \ln(z/z_o)$, where u_* is wind shear velocity, u_* is wind speed at height z, and κ is the von Karman constant (~0.4). Field data from Lancaster N, Rasmussen KR, and Greeley R (1991) Interactions between unvegetated desert surfaces and the atmospheric boundary layer: a preliminary assessment. *Acta Mechanica, Suppl.* 2: 89–102.

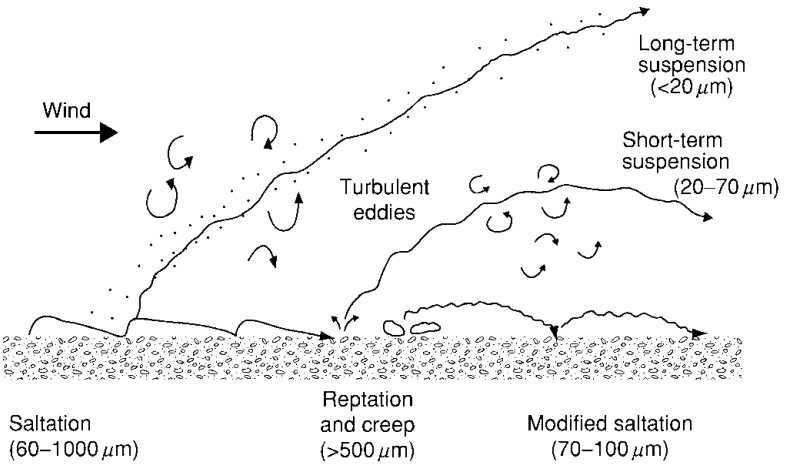


Figure 2 Modes of sediment transport by the wind. Reproduced with permission from Pye K (1987) *Aeolian Dust and Dust Deposits*. London: Academic Press.

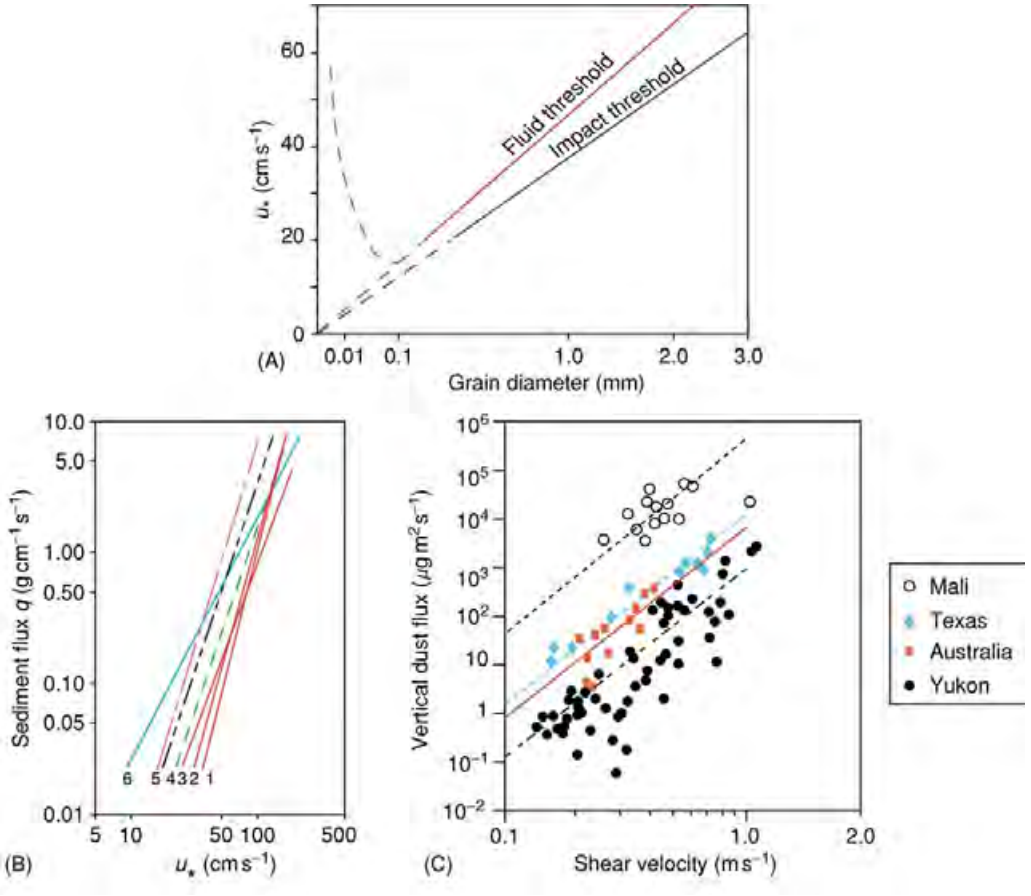


Figure 3 Transport of sediment by the wind. (A) Relation between threshold wind shear velocity and particle size (from Bagnold RA (1941) *The Physics of Blown Sand and Desert Dunes*, p. 265. London: Chapman and Hall). (B) Mass flux of sand as a function of wind shear velocity. Data from laboratory wind tunnel experiments (from Lancaster, 1995). (C) Relations between horizontal flux of sand-sized particles and vertical flux of dust (from Nickling WG, McTainsh GH, and Leys JF (1999) Dust emissions from the Channel Country of western Queensland, Australia. *Zeistchrift für Geomorphologie, Supplement-Bd.* 116: 1–17.) Data from field experiments.

clay-size particles into the wind and impact coarse grains that are rolled along the bed.

Grains begin to move and sediment is entrained by the wind when fluid forces (lift, drag, and moment) exceed the effects of the weight of the particle and any cohesion between adjacent particles as a result of moisture, salts, or soil crusts. There is a cascading effect as the first particles to move dislodge or impact other grains, so that the number of particles in transport increases exponentially. These particles extract momentum from the wind, reducing its near-bed velocity, so that the transport rate reaches a dynamic equilibrium state over a period of 1 to 2 s. The threshold wind speed at which grains begin to move is strongly dependent on particle size (Figure 3A). As in water, there is a particle size at which the transport threshold is at a minimum. For quartz sand in air, the minimum threshold velocity is associated with fine sand (approximately 100 μm diameter).

Most sand particles are moved by saltation in a layer close to the bed, with an exponential decay in the number of particles moved with height above the bed. Particles in saltation move in a parabolic path with a steep ejection limb and a gradual return to the bed. The impacting grains may rebound directly (successive saltation), deform the bed, eject fine particles, or move coarse grains a short distance by reptation or surface creep. The mass flux of sand has been determined by numerous laboratory wind tunnel and field studies to be proportional to the cube of wind shear velocity above a threshold value (Figure 3B). For any wind shear velocity, there is a potential rate of sand transport or transport capacity; this is reached only when the availability of sediment is unrestricted (e.g., most loose sand surfaces). In these conditions, the wind is saturated with respect to transport capacity. If the actual transport rate is less than the potential rate, then the wind is undersaturated. Very fine grains (silt and clay size) are inherently resistant to entrainment, yet are readily transported by the wind. Studies have shown the critical role of impacting sand grains in the mobilization of silt- and clay-size particles and have demonstrated the close relations between the horizontal flux of sand-size particles and the vertical flux of fine particles. In these situations, the horizontal mass transport rate is directly related to shear velocity (Figure 3B), so dust emissions scale to the fourth power of wind shear velocity (u_*^4; Figure 3C). Where there is a limited supply of particles able to abrade soil clods or playa crusts, dust emissions are limited by the supply of particles rather than by the wind shear velocity, and the vertical flux of dust is almost independent of wind shear velocity.

Based on the principle of sediment continuity, winds are erosional if transport rate (or wind shear velocity) increases downwind; deposition occurs when transport rates decrease in the direction of transport, and sediment bypass occurs when there is no

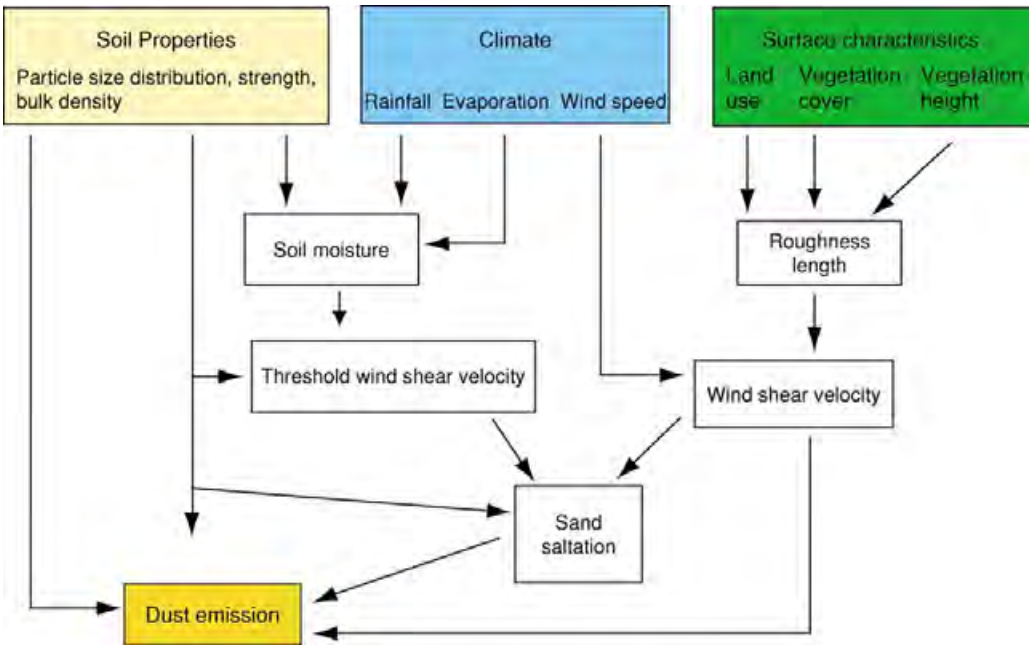


Figure 4 Controls of wind erosion processes on agricultural lands. Reproduced with permission from Leys JF (1999) Wind erosion on agricultural land. In: Goudie AS, Livingstone I, and Stokes S (eds.) *Aeolian Environments, Sediments, and Landforms*, pp. 143–166. Chichester and New York: Wiley.

change in transport rates. These principles are of vital importance to understanding of the dynamics of dunes and sand seas.

Wind Erosion

Erosion by wind involves two linked processes: abrasion (mechanical wearing of coherent materials, including playa crusts and clods created by tillage) and deflation (removal of loose material). Considerable attention has been devoted to the processes and rates of wind erosion because of their impact on agriculture, especially in semi-arid regions, and the implications of dust emissions for air quality. Wind erosion abrades crops, removes organic matter, nutrients, and fertilizer, and changes soil texture. The products of wind erosion (especially dust particles) impact air quality, atmospheric radiative properties, and human health, causing respiratory illnesses and transporting pathogens. Rates of wind erosion vary widely and for a given wind shear velocity are dependent on soil or sediment texture and the degree of crusting and cohesion. The highest emission rates for fine-grained sediment are associated with soils of loamy texture, especially those that have been disturbed by vehicular traffic and/or animals.

Empirical studies of wind erosion rates in relation to environmental conditions and agricultural practices have resulted in the development of predictive models, including the widely used wind erosion equation (WEQ) and the new Wind Erosion Prediction System (WEPS) developed by the United States Department of Agriculture (USDA). The considerable uncertainties associated with such empirically derived models have stimulated development of physically based models of wind erosion in which relations between the physical parameters of the wind and the surface are used to predict dust emissions (Figure 4).

Wind Erosion Landforms

Wind erosion landforms have not received the same level of research attention as have those associated with aeolian deposition. Landforms created by wind erosion can occur on several scales. Large-scale landforms include yardangs (Figure 5A), some of which may occur in systems of ridges and swales aligned with the wind. Yardangs are streamlined forms, with a blunt upwind face and a tapering leeward projection that appears to have evolved to minimize drag. Many yardangs are formed in Quaternary lacustrine deposits; others occur in bedrock, including Precambrian dolomite, Cambrian sandstone, and Mesozoic limestone. Large-scale systems of yardangs are also observed in several areas of Mars. Formation of yardangs appears to involve winds from a very limited range of directions and, at least for bedrock examples, long-term erosion. In some areas (e.g., Kalahari, West Texas), wind erosion has created enclosed depressions with their depth controlled by the water table or resistant strata (Figure 5B). Small-scale aeolian erosional landforms involve wind abrasion of clasts

Figure 5 Wind erosion landforms. (A) Yardangs in Dahkla Oasis, Egypt; (B) Kalahari pans; (C) Ventifacts in the Namib Desert.

and bedrock outcrops to form ventifacts (Figure 5C). Smoothing and polishing, faceting, fluting, and grooving of rock surfaces occur in areas of high-intensity sand transport and affect a variety of lithologies, including granite, marble, and basalt.

Dust Storms

Dust storms are defined as events in which visibility is reduced to 1 km or less as a result of blowing dust (Figure 6). Interest in air quality and in the impact of aerosols on atmospheric properties has promoted intensive study of dust storms and their occurrence in space using meteorological records and satellite images. The Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite images and the Total Ozone Monitoring Sensor(TOMS) have been used to identify major dust source areas and to track the dispersion of particulate matter (Figure 7). Major dust source areas include the Bodelé Depression in Tchad, the Aral Sea area, south-east Iran, and the loess plateau of China. These areas are characterized by seasonally strong winds and areas of fine-grained sediments (including those of playa lakes or previously deposited aeolian materials). The interannual frequency and magnitude of dust storms are strongly linked to climatic variability, so that in many areas dust emissions are inversely correlated to rainfall. Although dust storm intensity in some areas is affected by human impacts, including water diversion for irrigated agriculture and poor land use practices, the largest sources of dust emissions in the Sahara are in uninhabited regions.

Aeolian Deposits

Aeolian deposits include sand seas and dune fields, deposits of silt (loess), and fine-grained material that forms a significant component of desert margin and other soils. Many marine deposits, especially those downwind of major desert areas, also include significant proportions of wind-transported material.

Silt and Clay Size

Terrestrial deposits of wind-transported silt-sized quartz particles, termed loess, cover as much as 10% of Earth's land surface. Loess deposits are widespread in areas of northern China, central Russia, the Ukraine, Kazakhstan, central Europe, Argentina, Alaska, and the central United States. Much of the material was thought to be derived from silt particles produced by glacial grinding and supplied to aeolian processes by glacial outwash ('glacial loess'), but aeolian abrasion and other processes, including frost shattering, salt weathering, and fluvial comminution, are important, especially in the formation of 'desert loess'. At least some loess (e.g., in Colorado and Nebraska) has its origins in weathered bedrock (e.g., siltstones, volcaniclastic deposits). Loess deposits and their interstratified buried palaeosols provide a long record of Quaternary climate change (Figure 8). In these sequences, the rate of loess deposition varies

Figure 6 Reduced visibility from major dust storm in West Texas, 15 December 2003. Compare with clear day (inset). Photographs by J Lee, Texas Tech University.

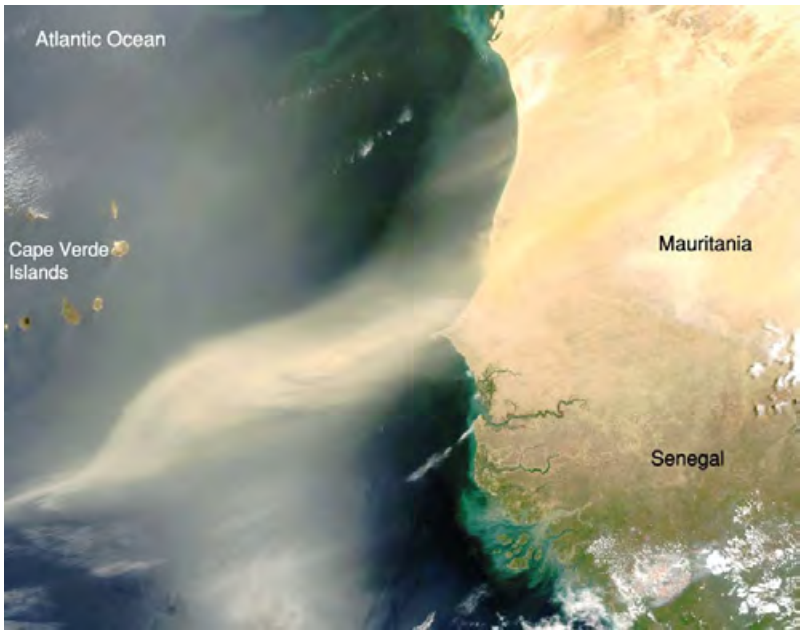


Figure 7 Dust plumes extend up to 200 km offshore from the western Sahara in Senegal and Mauritania, 30 April 2003. This was the fourth day of the storm. Details and many other dust storm images are available from the National Aeronautics and Space Administration website at http://www.visibleearth.nasa.gov.

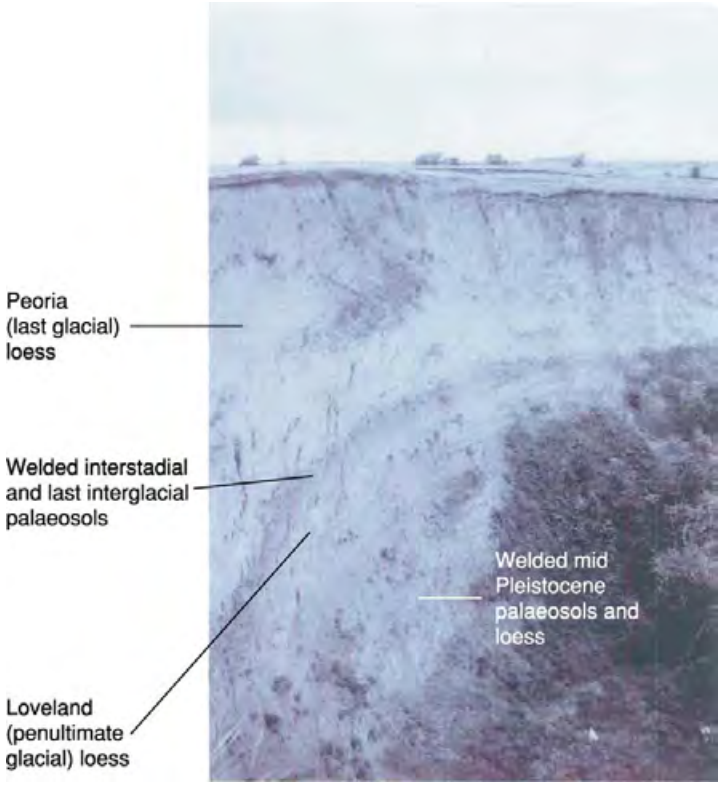


Figure 8 Loess deposits in Nebraska. Photograph and annotations by D Muhs, USGS.

over time, being greater in glacial periods (more arid, windier, less vegetation) compared to interglacial periods (less arid and windy, more vegetation), when the influx of particles was low enough to allow soils to form.

Silt- and clay-sized material of aeolian origin is also an important component of many desert margin soils. Deposition of silt plays a role in the formation of many stone pavement surfaces in desert regions (desert pavement, or 'reg'). These surfaces are characterized by a surface layer of gravel or larger clasts that overlie fine-grained materials. Older work argued that desert pavement surfaces were deflationary and that the mantle of coarse clasts resulted from winnowing and removal of fine material. Detailed studies of these surfaces now show, however, that the surface layer of gravel-sized clasts rests on an accretionary mantle of soil-modified dust that may be a metre or more thick. The dust is trapped by the clasts and deposited between them. The fine material is incorporated into the mantle by pedogenic processes (e.g., shrinking and swelling of clay minerals), so that the clasts remain at the surface as it inflates over periods of thousands of years. Thus desert pavement surfaces developed on bedrock and alluvial surfaces are actually depositional features.

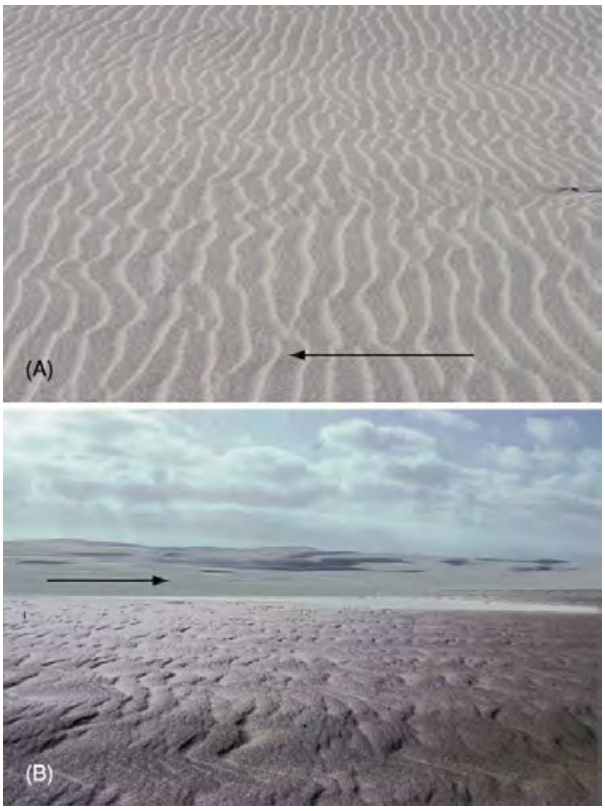


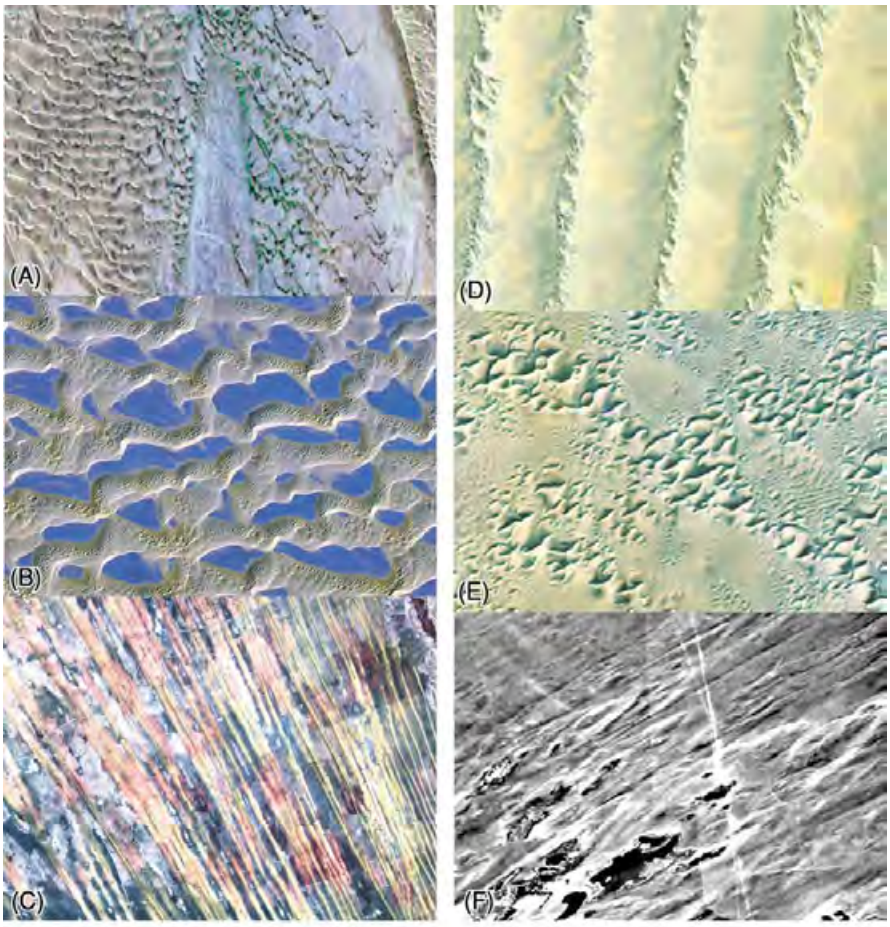
Figure 9 (A) Wind ripples in the Victoria Valley, Antarctica; (B) granule ripples (megaripples) at Skeleton Coast, Namibia. Arrows indicate formative wind direction.

Sand Dunes

Aeolian dunes form part of a hierarchical system of self-organizing bedforms developed in wind-transported sand; this system comprises wind ripples (spacing 0.05–2 m), individual simple dunes or superimposed compound and complex dunes (spacing 50–500 m), and compound and complex megadunes, or draa (spacing >500 m). Dunes occur wherever there is a sufficient supply of sand-sized sediment, winds to transport that sediment, and conditions that promote deposition of the transported sediment. These requirements are satisfied in two main environments: coastal areas with sandy beaches and onshore winds, and desert areas. Most dunes occur in contiguous areas of aeolian deposits termed 'sand seas' (>100 km^2) or 'dune fields'.

Wind ripples (Figure 9) typically have a wavelength of 0.05–0.2 m and an amplitude of 0.005–0.010 m. They are ubiquitous on sand surfaces except for those undergoing very rapid erosion or deposition, and form because a flat sand surface over which sand transport by saltation and reptation occurs is dynamically unstable. Field and laboratory wind tunnel experiments and numerical modelling suggest that ripple wavelength is a function of grain transport distance and that ripples develop as a result of stochastic variations in the flux of reptating coarser grains driven by the impacts of finer saltating grains. Once the initial variation in bed topography develops, variations in microtopography influence the intensity of saltation impacts and therefore the local mass flux of reptating grains, and the pattern becomes self-organizing by the merging and growth of a characteristic wind ripple wavelength and amplitude.

Aeolian dunes occur in a self-organized pattern that develops as sand surfaces respond to the wind regime (especially its directional variability) and the supply of sand. Sand dunes occur in four main morphological types (Figure 10): crescentic (transverse), linear, star, and parabolic. The simplest dune types and patterns form in areas characterized by a narrow range of wind directions (unidirectional wind regime). In the absence of vegetation, crescentic dunes will be the dominant form. Isolated crescentic dunes, or barchans, occur in areas of limited sand supply; they coalesce laterally to form crescentic (barchanoid) ridges as sand supply increases (Figure 10A and B). Linear dunes, which form in areas of bimodal or wide unimodal wind regimes, are characterized by their length (often more than 20 km), sinuous crestline, parallelism, and regular spacing (Figure 10C and D). The essential mechanism for linear dune formation is



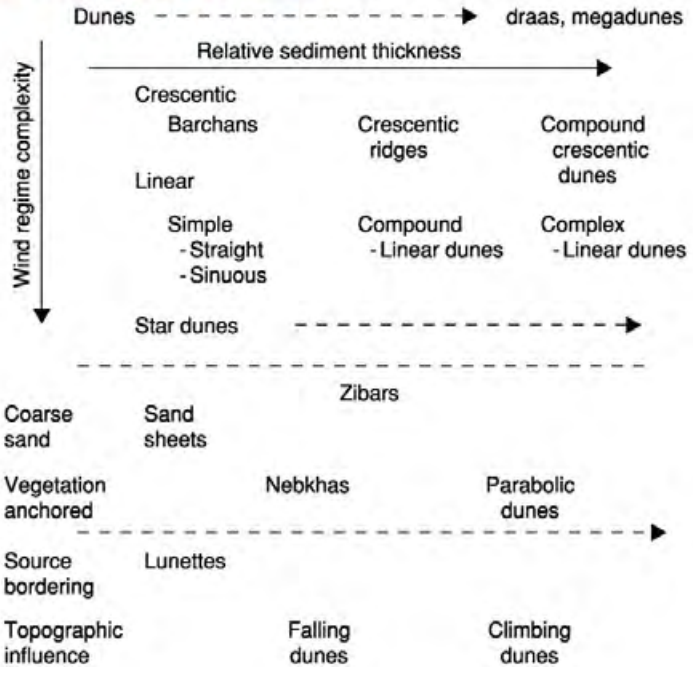


Figure 10 Morphological classification of aeolian sand dunes, and satellite images and aerial photographs of major dune types: (A) barchans and crescentic dunes, Namib Sand Sea; (B) compound crescentic dunes, Liwa, United Arab Emirates; (C) simple linear dunes, Kalahari Desert; (D) complex linear dunes, Namib Sand Sea; (E) star dunes, Gran Desierto, Mexico; (F) parabolic dunes, Casper, Wyoming.

the deflection of winds that approach the crest at an oblique angle, then parallel wind flow to the lee side and transport of sand along the dune, so elongating it. Star dunes have a pyramidal shape, with three or four sinuous sharp-crested arms radiating from a central peak and multiple avalanche faces (Figure 10E). Star dunes occur in multidirectional or complex wind regimes and are the largest dunes in many sand seas, reaching heights of more than 300 m. The development of star dunes is strongly influenced by the high degree of form–flow interaction, which occurs as a result of seasonal changes in wind direction, and the existence of a major lee-side secondary circulation. Parabolic dunes (Figure 10F) are characterized by a U or V shape with a 'nose' of active sand and two partly vegetated arms that trail upwind. They are common in many coastal dunefields and semi-arid inland areas and often develop from localized blowouts in vegetated sand surfaces. Other important dune types include zibars (low-relief dunes lacking discernible slip faces; commonly occur on sand sheets), nebkhas (hummock dunes; anchored by vegetation; common in may coastal dunefields), lunettes (form downwind of small playas; often composed of sand-sized clay pellets), and a variety of topographically controlled dunes (climbing and falling dunes, echo dunes).

Relations between dune types and wind regimes indicate that the main control of dune type is the directional variability of the wind regime (Figure 11). Grain size, vegetation cover, and sediment supply play subordinate roles.

Dune Processes and Dynamics

The development and morphology of all dunes are determined by a complex series of interactions between dune morphology, airflow, vegetation cover, and sediment transport rates. As dunes grow, they project into the atmospheric boundary layer so that streamlines are compressed and winds are accelerated up the stoss (windward) slope by up to 2.5 times, depending on the size and steepness of the dune. Flow acceleration results in an exponential increase in sediment transport rates towards the dune crest, resulting in erosion of the stoss slope. In the lee of the crest of dunes, wind velocities and transport rates decrease rapidly as a result of flow expansion between the crest and brink of the lee or avalanche face and flow separation on the avalanche face. Flow separation causes fallout of previously saltating sand grains and development of avalanche (slip) faces. Secondary flows, including lee-side flow diversion, are also important when winds approach the dune obliquely, and are an important process on linear and many star dunes. Migration of crescentic dunes occurs as a result of erosion of the stoss slope and deposition on the lee. Their movement can be described by $c = Q/yh$, where c is the migration rate, Q is the bulk volumetric sand transport rate, y is the bulk density of sand, and h is dune height. Linear dunes

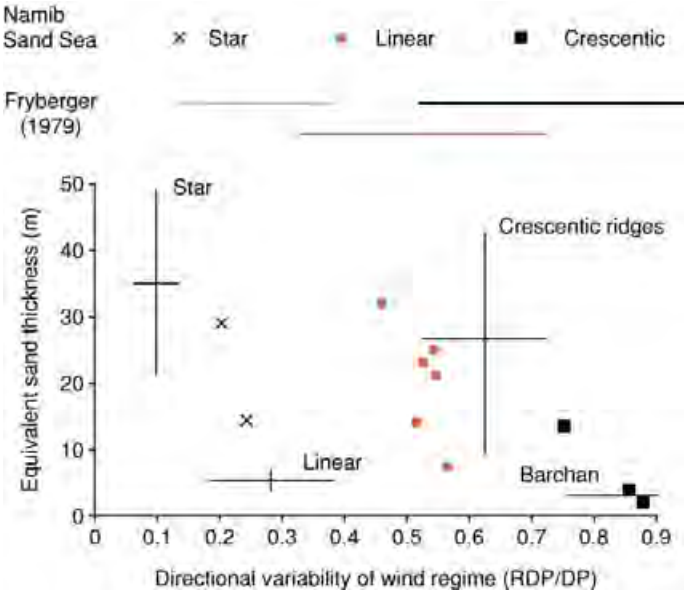


Figure 11 Relations between dune types and wind regimes. The directional wind variability index is the ratio of the resultant drift potential (RDP) to the drift potential (DP). Data from Fryberger SG (1979) Dune forms and wind regimes. In: McKee ED (ed.) *A Study of Global Sand Seas: United States Geological Survey, Professional Paper, 1052*, pp. 137–140. Reston, VA: USGS, Lancaster (1983), Wassan and Hyde (1983).

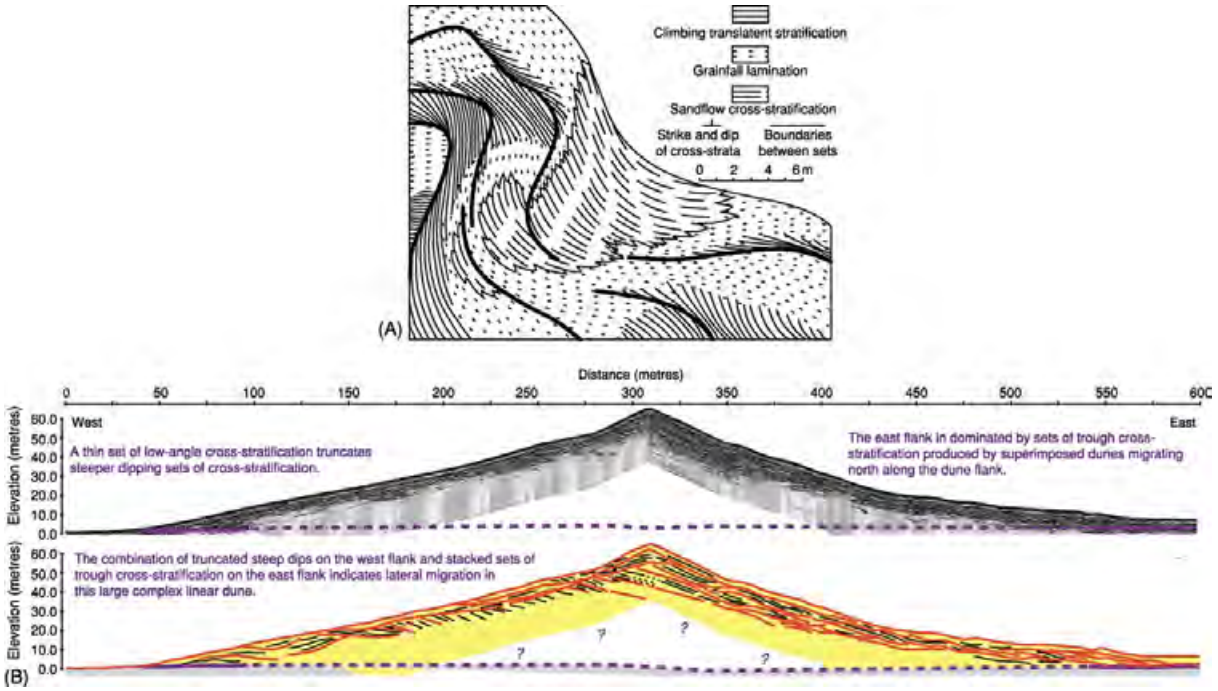


Figure 12 Dune sedimentary structures. (A) Crescentic dunes. Reproduced with permission from Hunter RE (1977) Basic types of stratification in small eolian dunes. *Sedimentology* 24: 361–388. (B) Linear dunes. Courtesy of CS Bristow, N Lancaster, and SD Bailey. Poster presented at British Sedimentological Research Group Annual Meeting, 1999.

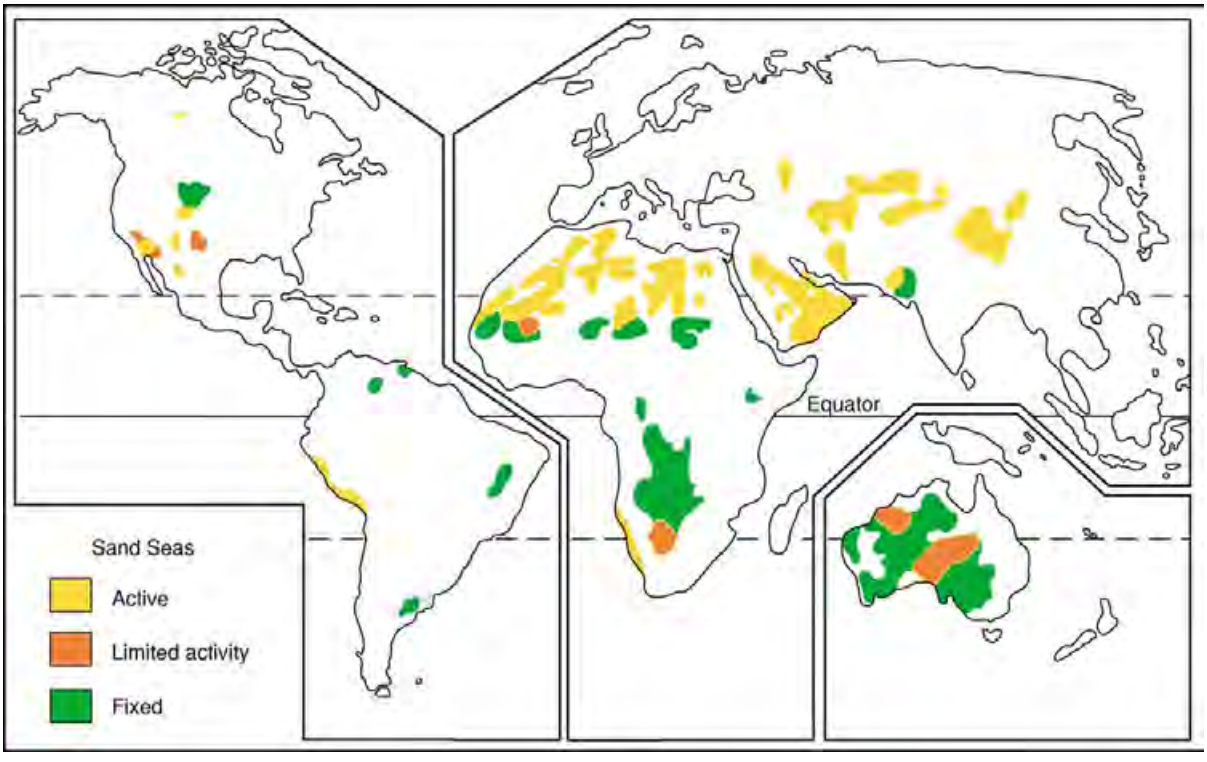


Figure 13 Distribution of major sand seas.

migrate laterally at a very slow rate and also extend downwind. Star dunes are essentially stationary forms.

Dune Sediments

Dunes are composed of moderately to well-sorted sands (63–1000 μm), with a mean grain size in the range 160–300 μm. Most dune sands are composed of quartz, but may include significant quantities of feldspar. Dunes composed of volcaniclastic materials occur in some areas. Carbonate-rich dune sands are found adjacent to many subtropical coastlines and gypsum dunes occur close to some playas.

Three primary modes of deposition occur on dunes: migration of wind ripples, fallout from temporary suspension of grains in the flow separation zone (grainfall), and avalanching of grains initially deposited by grainfall (grainflow). These deposition modes create three basic types of aeolian sedimentary structures: wind ripple laminae, grainfall laminae, and grainflow cross-strata. The packages of laminae may be separated by a hierarchy of bounding surfaces representing episodes of reactivation, dune growth, and migration of dunes in conditions of bedform climbing.

Crescentic dunes are dominated by grainfall and grainflow cross-strata (Figure 12A). Linear dune structures imaged using ground-penetrating radar show a change in the dominant type of sedimentary structures as the dune develops. Small dunes are dominated by stacked sets of cross-strata that dip in opposite directions and are formed by the migration of the elements of a sinuous crestline up the dune. The structures of larger dunes are dominated by trough cross-strata created by the migration of superimposed bedforms (Figure 12B). Sedimentary structures of star dunes are not well known, but are likely complex.

Sand Seas

Desert sand seas are major depositional landforms that comprise dunes of varying size, spacing, and morphological type; sand seas also have areas of sand sheets as well as interdune deposits (e.g., playas, lacustrine deposits) and extradune fluvial, lacustrine, and marine sediments. Major areas of sand seas lie in

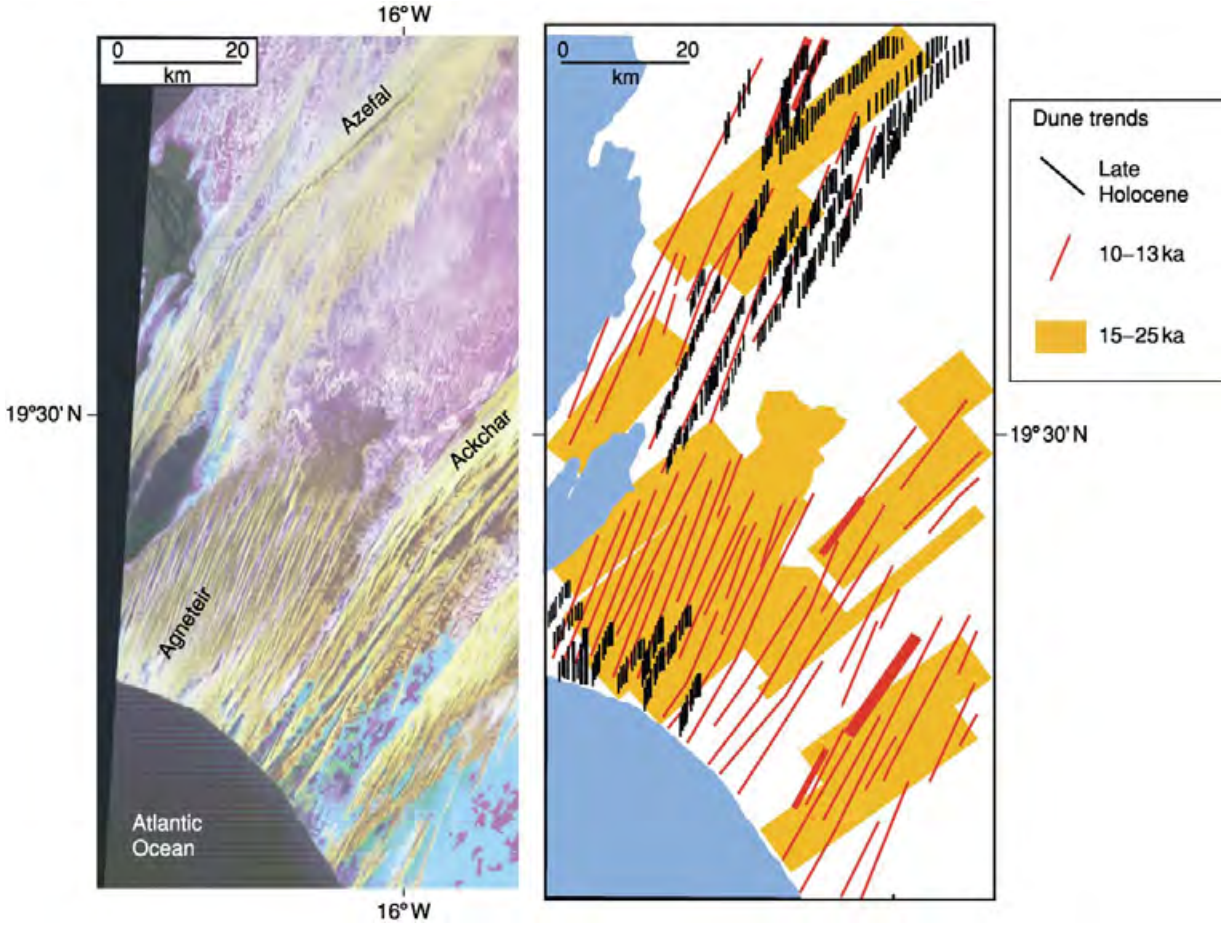


Figure 14 Linear dune generations in western Mauritania; LANDSAT image (left) and interpreted dune trends (right).

the tectonically stable desert regions of the Sahara, Arabian Peninsula, Australia, and southern Africa and in the enclosed basins of central Asia (Figure 13). Sand seas and dune fields form part of well-defined regional- and local-scale sediment transport systems in which sand is moved by wind from source areas (e.g., distal fluvial deposits, sandy beaches) via transport pathways to depositional sinks. Sand seas accumulate downwind of source zones wherever sand transport rates decrease as a result of regional and/or local changes in wind speed and directional variability, so that the influx of sand exceeds outflux.

Most sand seas exhibit significant spatial variations in dune type and sediment characteristics. These may be the result of regional changes in wind regime, but can also be the product of multiple periods of sand influx and stabilization, controlled by changes in climate and sea-level. Areas of different dune types therefore often represent distinct generations of dunes. Advances in dating techniques (e.g., optically stimulated luminescence) are beginning to provide absolute chronologies of dune generations in sand seas (Figure 14).

Preservation of Aeolian Deposits in the Rock Record

Formation of an aeolian rock record requires three sets of processes: sand sea construction, accumulation of aeolian strata, and preservation of the

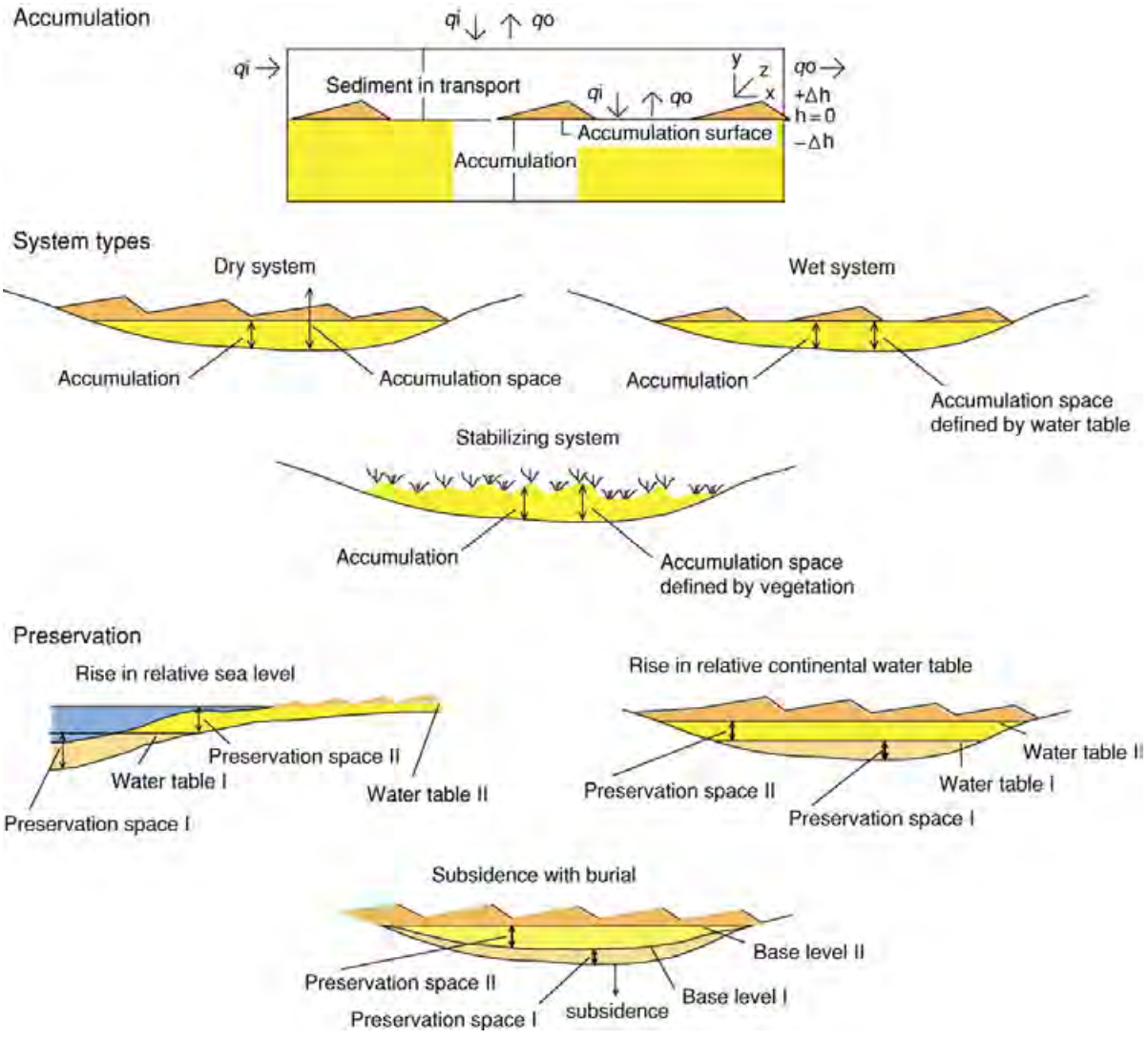


Figure 15 Accumulation and preservation of aeolian sands. Accumulation depends on the transport rate *qi* into and *qo* out of the system. Reproduced with permission from Kocurek G (1998) Aeolian system response to external forcing factors – a sequence stratigraphic view of the Saharan region. In: Alsharan AS, Glennie KW, Whittle GL, and Kendall CGSC (eds.) *Quaternary Deserts and Climatic Change*, pp. 327–338. Rotterdam/Brookfield: Balkema.

accumulation. Accumulation of aeolian deposits occurs when that sediment is transferred from bedforms to the accumulation such that its surface rises over time (Figure 15). In dry aeolian systems, the water table is sufficiently far below the surface that it plays no role in the dynamics of the accumulation, which is instead dominated by the wind. In wet aeolian systems, the water table is at or near the surface and so controls accumulation dynamics. Where vegetation controls accumulation dynamics, the system is known as a stabilizing aeolian system. In all cases, accumulation occurs when sediment transport rates decrease downwind and the excess sediment is deposited; and/or when the concentration of sediment in transport (approximately indicated by dune size) decreases over time. In dry systems, spatial or temporal decreases in wind strength give rise to accumulation of sediment by bedform climbing; in wet systems, transport rates

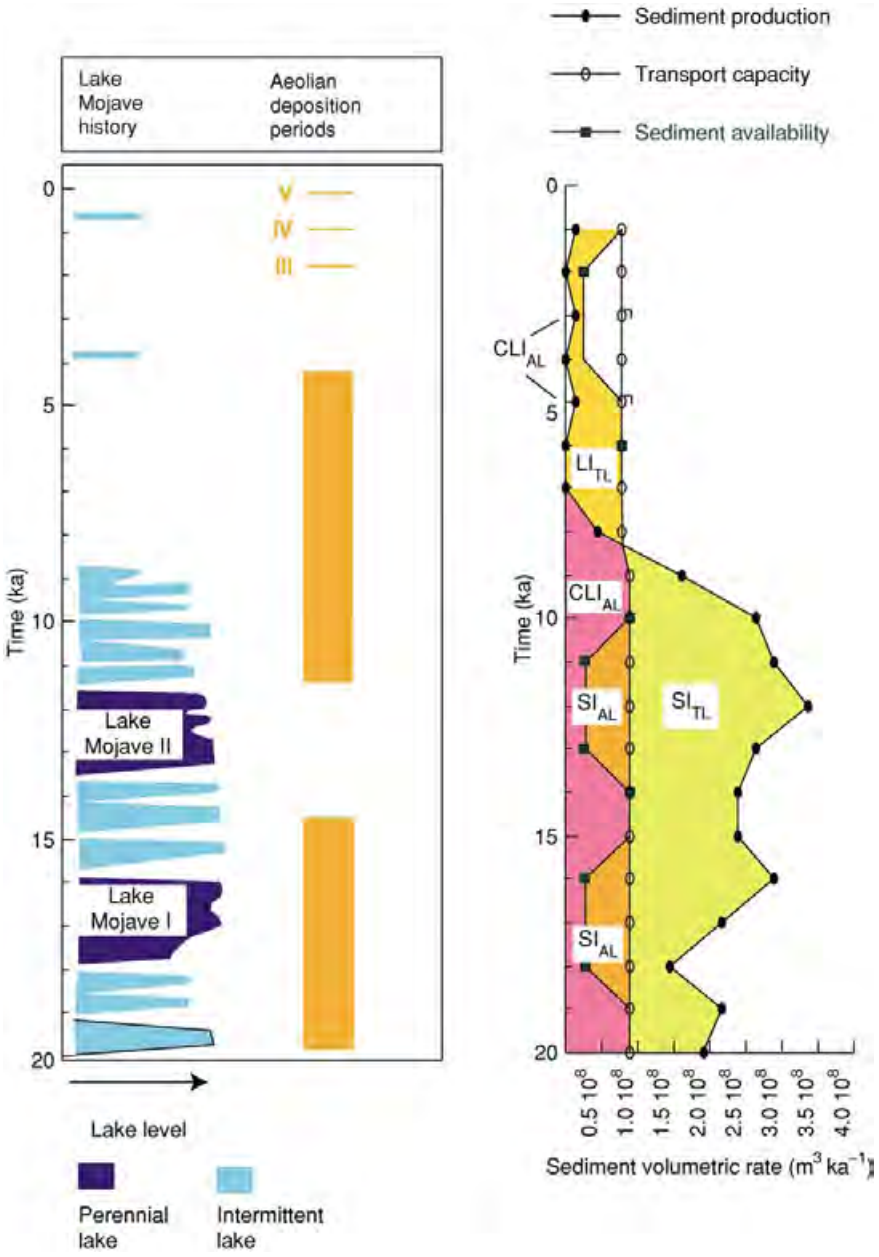


Figure 16 Example of controls on an aeolian system state in the Mojave Desert, both transport limited (TL) and availability limited (AL). CLI, Contemporaneous and lagged input; LI, lagged input; SI, stored input. Reproduced with permission from Kocurek G and Lancaster N (1999) Aeolian sediment states: theory and Mojave Desert Kelso Dunefield example. *Sedimentology* 46(3): 505–516.

decrease downwind because sediment is trapped by dunes migrating into a wet area, resulting in a decrease in dune size, or dunes shrink because the water table rises under the whole system.

For these accumulations to be preserved, there must be subsidence and burial and/or a rise of the water table through the accumulation. Because wind has the potential to erode any sand accumulation above the water table if it is undersaturated with respect to transport capacity, exhaustion of sediment supply may promote erosion unless the accumulation is buried by non-aeolian strata or the water table is at

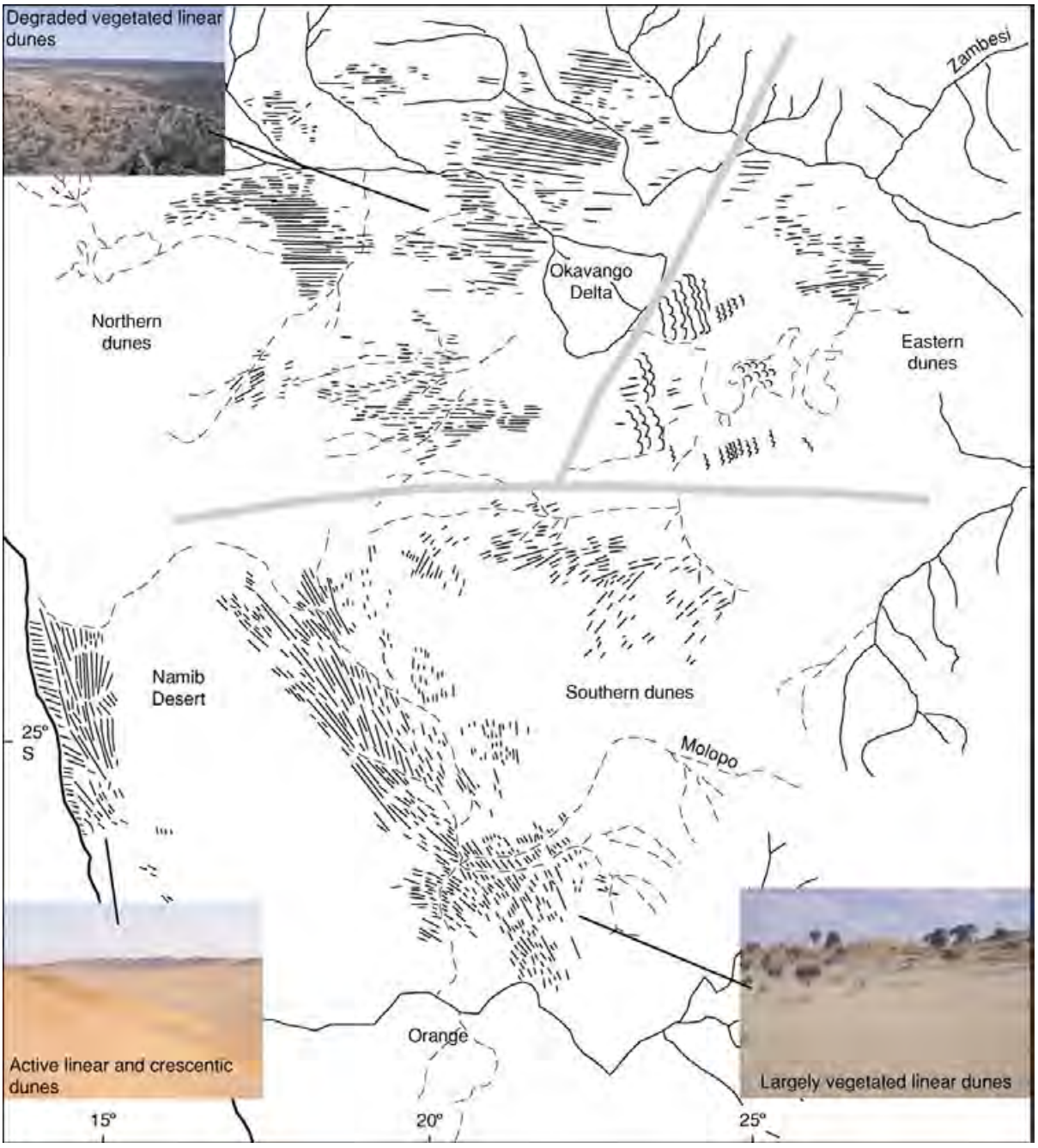


Figure 17 Relict dune systems in the Kalahari Desert of southern Africa. Dune patterns reproduced with permission from Lancaster N (2000) Eolian deposits. In: Partridge TC and Maud RR (eds.) The Cenozoic of Southern Africa, pp. 73–87. Cape Town, South Africa: Oxford University Press.

or near the surface. Preservation of aeolian accumulations is favoured in tectonic basins subject to a long history of sediment influx and subsidence, in coastal areas in which marine transgressions can bury aeolian accumulations or cause a rise in the regional water table, and in continental areas in which the water table rises permanently because of tectonism or climate. These conditions were satisfied during the formation of the extensive Jurassic aeolian sandstones of the western United States, which were deposited in a subsiding foreland basin supplied with large amounts of sediment from the North American Craton. Marine transgressions from the west preserved parts of the record via sabkha and marine strata. By contrast, most aeolian accumulations in the Sahara and many other deserts are unfavorably located with respect to preservation. One exception is that along the coast of Mauritania and its continental shelf, where aeolian sands sourced from inland craton areas are preserved by the postglacial rise in sea-level and the associated rise in water tables.

Effects of Climate and Sea-Level Change on Aeolian Systems

Many aeolian deposits and landforms formed during the Quaternary Era, thus they have been affected by climatic and sea-level changes associated with glacial–interglacial cycles. The effects of changes in external forcing factors are manifested by changes in the supply of sediment of a size suitable for transport by the wind, the availability of this sediment for transport (determined by vegetation cover and soil moisture), and the mobility of this sediment (controlled by wind strength). The interactions between these variables can be evaluated in terms of the state of the aeolian system and the limiting factors

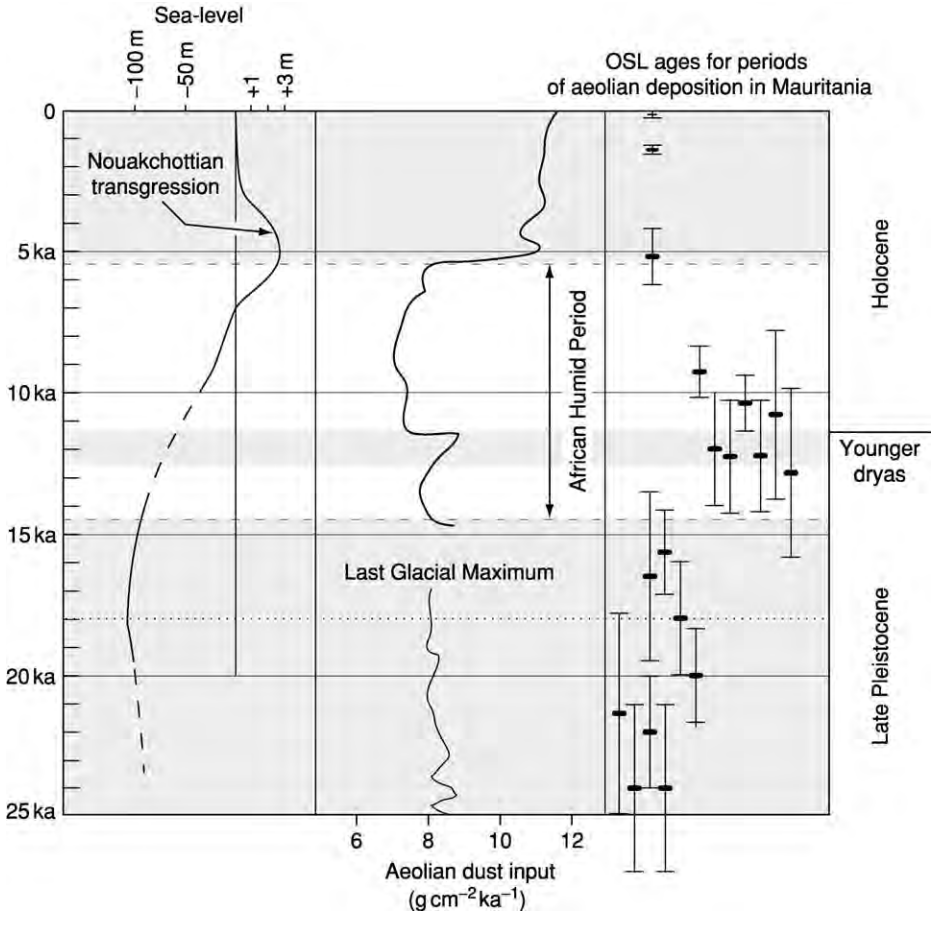


Figure 18 Relations between periods of dune construction, sea-level, and dust flux to the Atlantic Ocean (dust flux after deMenocal P, *et al.* (2000) Abrupt onset and termination of the African Humid Period: rapid climate responses to gradual insolation forcing. *Quaternary Science Reviews* 19(1–5): 347–361). OSL ages for aeolian deposition in Mauritania from Lancaster N, *et al.* (2002) Late Pleistocene and Holocene dune activity and wind regimes in the western Sahara of Mauritania. *Geology* 30(11): 991–994.

identified (Figure 16). The changes in system state are manifested as episodic accumulation of aeolian deposits, resulting, for example, in sequences of loess accumulation interspersed by palaeosols (Figure 8), or formation of different dune generations in aeolian sand seas and dune fields (Figure 14). Consideration of variations in the supply of sediment for dune construction or loess accumulation, together with changes in the availability and mobility of this sediment, has prompted new explanations of regional variations in loess stratigraphy in North America and periods of aeolian accumulation in the Mojave and Arabian deserts. For example, loess accumulation was low during the last glacial maximum in Alaska, despite high rates of sediment supply and mobility, because tundra vegetation is ineffective at trapping sediment. This is in contrast to areas in the Great Plains and Central Lowlands, where accumulations of thick sequences of loess during the last glacial period were favoured by high rates of sediment supply from glacial outwash and high rates of accumulation in the boreal forest that then covered these areas. In the Mojave Desert, high rates of aeolian accumulation and dune construction took place during times when palaeolake systems were drying out seasonally or annually. Sediment deposited in these basins during wetter periods was mobilized during periods of drier climates. Aeolian construction ceased when these supplies of sediment were exhausted (Figure 16).

In addition to being influenced by climate and sea-level change, aeolian sediments and landforms may also provide evidence for past climate changes. Examples include loess–palaeosol sequences, as well as dune systems, now vegetated and inactive, indicating the former extent of arid conditions and past wind regimes in what are today areas of semi-arid or subhumid climates (Figure 17), and increased rates of aeolian dust deposition in ocean sediments and ice cores (Figure 18).

See Also

Sedimentary Environments: Deserts. **Sedimentary Processes:** Erosional Sedimentary Structures; Depositional Sedimentary Structures; Deposition from Suspension. **Soils:** Palaeosols.

Further Reading

Bagnold RA (1941) *The Physics of Blown Sand and Desert Dunes*. London: Chapman and Hall.

Bristow CS, Bailey SD, and Lancaster N (2000) Sedimentary structure of linear sand dunes. *Nature* 406: 56–59.

Fryberger SG (1979) Dune forms and wind regimes. In: McKee ED (ed.) *A Study of Global Sand Seas: United States Geological Survey, Professional Paper, 1052*, pp. 137–140. Reston, VA: USGS.

Goudie AS, Livingstone I, and Stokes S (eds.) (1999) *Aeolian Environments, Sediments, and Landforms*. New York: Wiley.

Kocurek G (1998) Aeolian system response to external forcing factors – a sequence stratigraphic view of the Saharan region. In: Alsharan AS, Glennie KW, Whittle GL, and Kendall CGSC (eds.) *Quaternary Deserts and Climatic Change*, pp. 327–338. Rotterdam/Brookfield: Balkema.

Kocurek G and Lancaster N (1999) Aeolian sediment states: theory and Mojave Desert Kelso Dunefield example. *Sedimentology* 46(3): 505–516.

Laity JE (1994) Landforms of aeolian erosion. In: Abrahams AD and Parsons AJ (eds.) *Geomorphology of Desert Environments*, pp. 506–535. London: Chapman and Hall.

Lancaster N (1983) Controls of dune morphology in the Namib sand sea. In: Ahlbrandt TS and Brookfield ME (eds.) *Eolian Sediments and Processes. Developments in Sedimentology*, pp. 261–289. Amsterdam: Elsevier.

Lancaster N (1995) *Geomorphology of Desert Dunes*. London: Routledge.

Lancaster N (2000) Eolian deposits. In: Partridge TC and Maud RR (eds.) *The Cenozoic of Southern Africa*, pp. 73–87. Cape Town, South Africa: Oxford University Press.

Livingstone I and Warren A (1996) *Aeolian Geomorphology: An Introduction*. Harlow: Addison Wesley Longman.

Muhs DR and Bettis EA III (2003) Quaternary loess–paleosol sequences as examples of climate-driven sedimentary events. In: Chan MA and Archer AW (eds.) *Extreme Depositional Environments, Mega End Members in Geologic Time*, pp. 53–74. Boulder, Colorado: Geological Society of America.

Pye K (1987) *Aeolian Dust and Dust Deposits*. London: Academic Press.

Pye K and Tsoar H (1990) *Aeolian Sand and Sand Dunes*. London: Unwin Hyman.

Wasson RJ and Hyde R (1983) Factors determining desert dune type. *Nature* 304: 337–339.

Wiggs GFS (2001) Desert dune processes and dynamics. *Progress in Physical Geography* 25(1): 53–79.

Catastrophic Floods

A J Russell, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

At present there is no widely recognized definition of a flood other than a 'notable flow of water'. Since the diluvial theory of the biblical flood, cataclysms and catastrophic flooding have drawn the attention of geologists throughout the development of the subject and to the present day. Catastrophic floods occur worldwide within a wide range of depositional environments and hydroclimatic settings. Catastrophic floods represent a major natural hazard within riverine environments, generating substantial loss of life and infrastructure. As a surface process, catastrophic floods are capable of generating immense landscape change and the transfer of huge sediment volumes. Large meltwater outbursts associated with continental deglaciation have recently been invoked as a trigger mechanism for changes in ocean circulation and consequent climate changes.

Catastrophic floods have been defined in a number of ways. From an applied perspective, catastrophic floods can be defined in terms of human and monetary loss. Floods can also be judged truly catastrophic, depending upon their impact on channel morphology and sedimentology or the work they do in transporting sediment. The extent to which a flood is catastrophic also depends upon the time-scale over which one wishes to consider its impacts. Preservation of erosional landforms and sedimentary successions associated with floods over geological time-scales are likely to reflect truly catastrophic events and the presence of appropriate depositional environments for preserving evidence of high-magnitude floods. Over shorter time-scales, a flood may be judged catastrophic if its impacts result in an irreversible change in fluvial geomorphology. High-energy flows are known to result in a suite of 'catastrophic processes', which do not operate under non-flood conditions and which achieve substantial and distinctive impacts. Catastrophic processes can, however, also occur under nonflood flow conditions within, for example, steep bedrock fluvial systems and subglacial drainage systems characterized by high stream powers.

Causes of Catastrophic Floods

Catastrophic floods require the sudden input of large volumes of water to a flood routeway. Flood generation mechanisms provide a major control on flood magnitude, frequency, and characteristics (Figure 1). Traditional classification schemes of flood causes examine the relationship between floods and climatological variables. More recently, attention has focussed on identifying the physical limits on the magnitude of floods generated by various mechanisms.

Rainfall is the most common cause of flooding in rivers around the globe. Rainwater is a major input to the fluvial system, often having travelled through hillslopes or more rapidly as overland flow. Rainfall intensities tend to be high over short time periods and on a relatively local scale. Meteorological and climatological controls on flood-producing precipitation events can be identified over a range of spatial and temporal scales (Figure 2). This hydroclimatological approach uses understanding of the controls on flood-generating precipitation events to characterize the distribution of floods in space and time. Other approaches have considered maximum flood discharges, which can be yielded from a certain catchment area. Empirically derived 'envelope curves' illustrate major differences in rainfall–runoff relationships between different climatic regions. Envelope curves do not, however, consider the relative role of specific drainage basin characteristics in generating flood flows (Figure 3).

Snowfall generally reduces the immediate potential for flooding by putting water into storage for more gradual release upon melting. However, the progressive or rapid melt of a snow-pack can generate substantial floods, especially as a 'nival' flood in springtime. Nival floods commonly affect rivers with large catchments, which allow synchronous water input from all parts of the catchment area. Snowmelt rates are controlled by the amount of available energy. Heavy rainfall onto a snow-pack will enhance peak flood discharges from a catchment, although modest precipitation onto a snow-pack may delay flows, resulting in lower peak discharges. Nival flows are also associated with secondary storage-release floods resulting from the failure of temporary snow dams or river ice jams.

Melt of glacier ice is a major control on discharge within proglacial river channels. Although melt-induced flows commonly exhibit discharge variations on diurnal and synoptic time-scales, they are not normally considered floods unless associated with unusually high ablation. This may occur, for example, during subglacial volcanic eruptions or where there are enhanced levels of geothermal activity. More commonly, however, glacial meltwater is stored

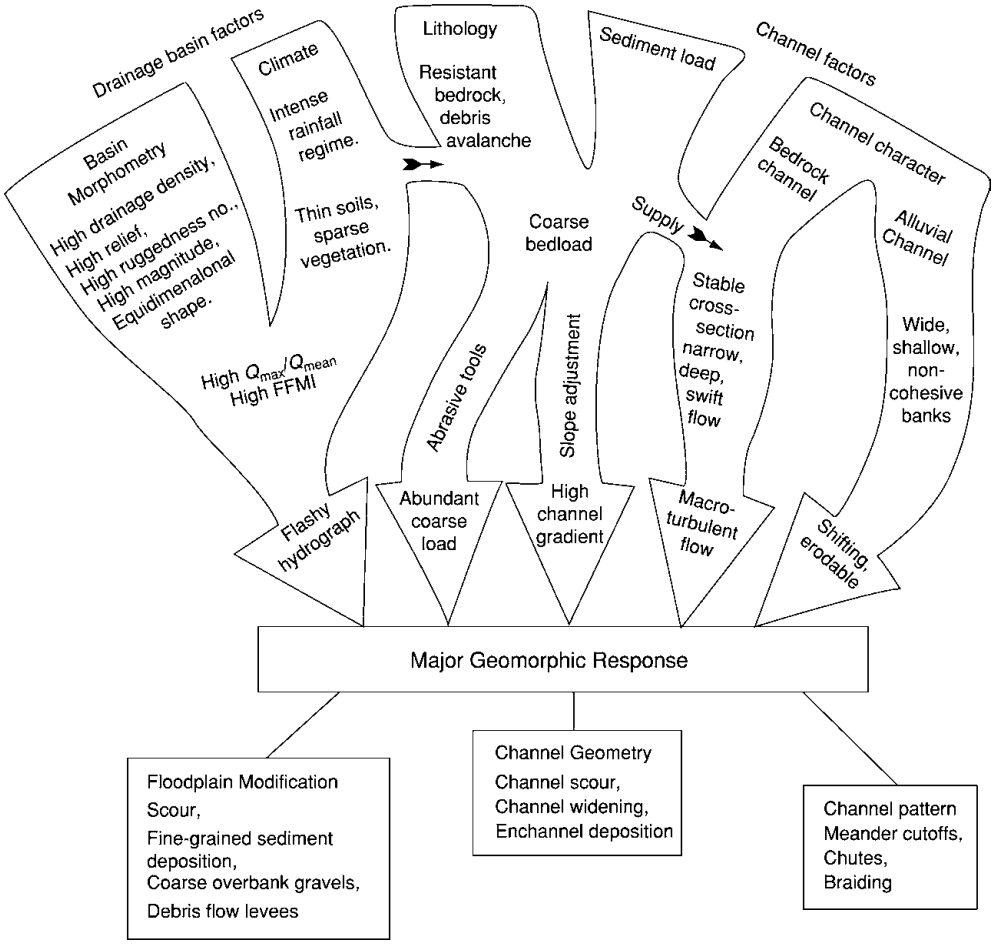


Figure 1 Controls on flood magnitude, frequency, and characteristics. After Kochel RC (1988) Geomorphic impact of large floods: review and new perspectives on magnitude and frequency. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 169–187. London: Wiley.

on top of, against, within, and underneath glaciers, providing the potential for storage-release events (Figure 4).

Natural or man-made dam failures can release outbursts of large volumes of water over relatively short time-spans. Natural dams commonly comprise vegetation debris, river ice, glacier ice, or sediment. Dam formation mechanism, composition, and geographic setting control dam longevity, failure mechanisms, and flood magnitude. The world's largest floods have resulted from the failure of glacier dams during the Late Quaternary. Man-made dams are also subject to failure, with the capability of releasing catastrophic floods many times greater in magnitude than could physically be generated by natural processes within the same catchment (Figure 5).

Accumulations of vegetation debris or coarse woody debris within river channels can obstruct flood flows, temporarily and partially damming floodwater. Infrequent high-magnitude flows within ephemeral rivers, for example, would allow sufficient time for coarse woody debris to accumulate within the channel between the highest magnitude events. Sudden dam failure commonly generates an enhanced flood peak as water released from storage is superimposed on the main flood hydrograph.

Although snow accumulation with river channels is a common occurrence in many parts of the world, there are relatively few reports of flooding produced by snow-dam failure. Outburst floods have been described from snow-dammed lakes in Antarctica and nival flood peaks enhanced by small snow-dammed lake outbursts in Canada. Snow damming mainly affects small low-order streams and can only pond water for very short time periods.

River ice break-up during the spring melt can lead to ice-jams at flow constrictions, resulting in considerable upstream inundation as well as downstream flooding on ice-jam failure. Unlike the snow-dammed lake bursts, ice-jam-related flooding occurs on major

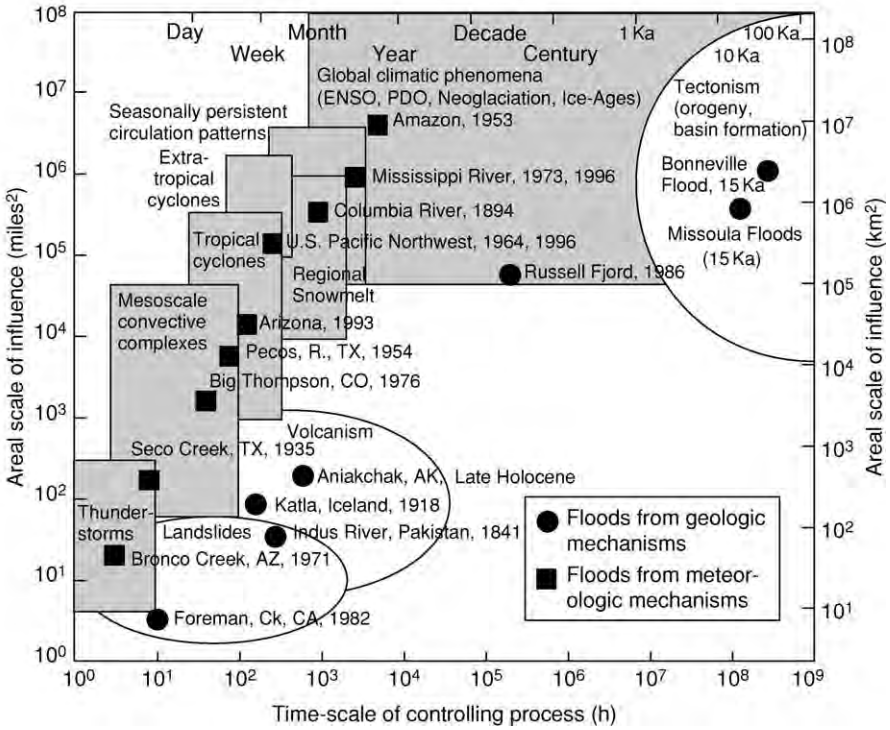


Figure 2 Space and time domain for flood-producing hydroclimatological conditions and flooding. After Hirschboeck (1988) Flood hydroclimatology. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 27–49, London: Wiley; and O'Connor *et al.* (2002) The geology and geography of floods. *Water Science and Application* 5: 359–385.

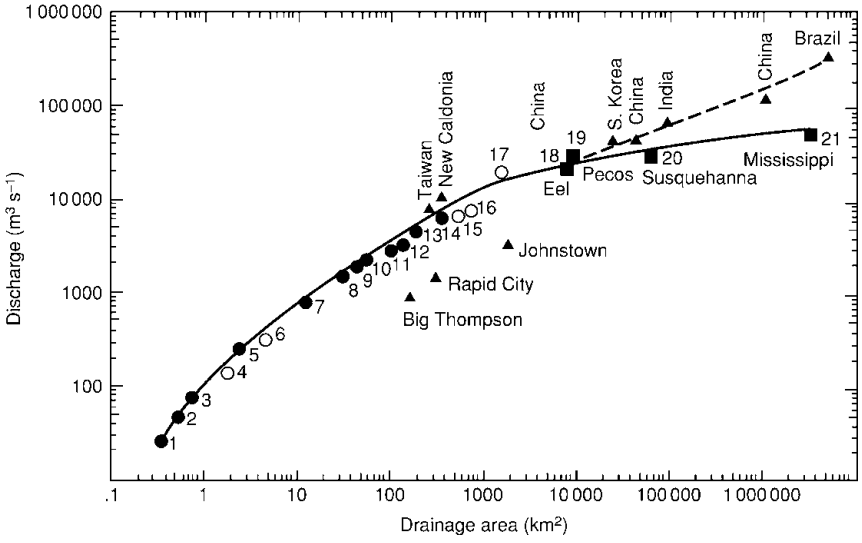


Figure 3 Envelope curve for maximum discharge generation in relation to catchment area. After Hirschboeck KK (1987) Catastrophic flooding and atmospheric circulation anomolies. In: Mayer LD and Nasch D (eds.) *Catastrophic Flooding* pp. 23–56. Allen & Unwin.

river systems. Within the northern hemisphere, northward draining rivers are particularly prone to ice-jams as the spring melt is initiated within the catchment headwaters before the lower reaches have thawed.

Within glaciated catchments, ice-cliff or subglacial tunnel collapse can temporarily dam meltwater rivers with a breccia of ice blocks. Damming commonly lasts only for periods of up to one hour, as meltwater flows are easily able to penetrate, float, and bulldoze

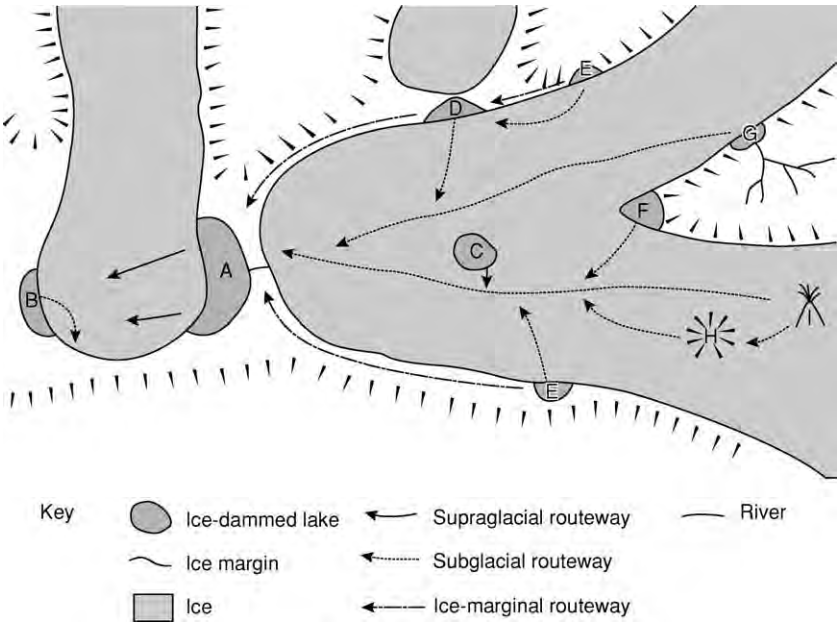


Figure 4 Ice-dammed lake storage locations. Letters A–I represent different meltwater storage locations, each associated with characteristic lake drainage triggers and jökulhlaup routing characterics. 'A' represents a tributary glacier which dams a trunk valley and which is subject to subaerial breach widenin. 'B' 'D' 'E' 'F' and 'G' represent proglacial and ice-marginal ice-dammed lakes in a variety of settings where drainage is triggered by mechanisms such as lake overspill, ice-dam flotation, syphoning and subglacial cavity formation. 'C' represents a supraglacial lake which taps through the entire thickness of the glacier. 'H' is a subglacial lake within either a suitable storage location such as a volcanic caldera, or as an up-turned bowl-shaped lake often associated with locations with high geothermal acitivity. Drainage triggers postulated for subglacial lakes include ice-dam flotation as well as Darcian flow through the glacier substrate. 'I' represents a subglacial eruption site, which may result in little or short-lived meltwater storage during the course of an eruption. High meltwater temperatures associated with the drainage of geothermal and volcanic events characteristically result in jökulhlaups with rapid rising stages. Jökulhlaup hydrographs have a longer rising stage and lower peak discharge the further the ice-dammed lake is from the glacier snout. After Tweed and Russell (1999) Controls on the formation and sudden drainage of glacier-impounded lakes: implications for jökulhlaup characteristics. *Progress in Physical Geography* 23: 79–110.

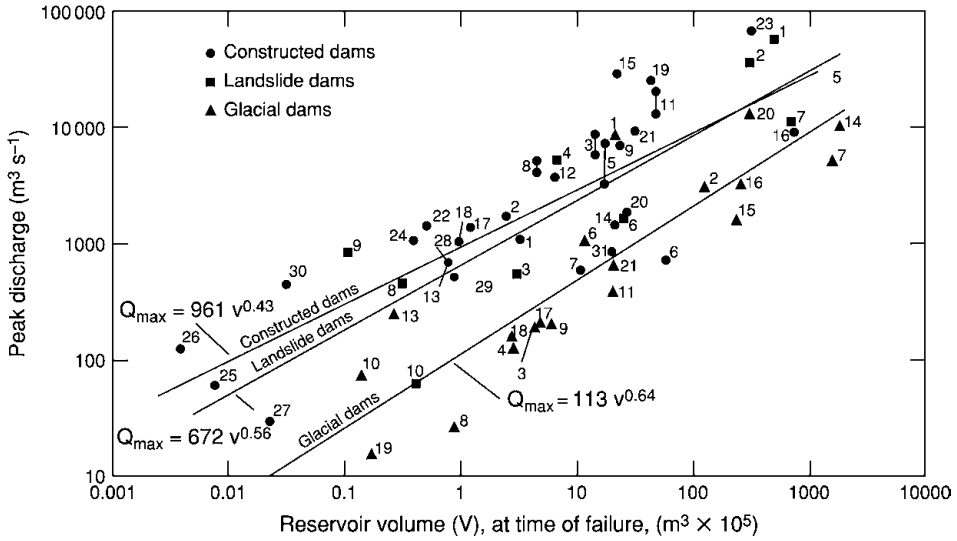


Figure 5 Relationships between lake volume and peak discharge for the failure of different dam types. After Costa (1988) Rheologic, geomorphic and sedimentological differentiation of water floods, hyperconcentrated flows, and debris flows. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 113–122. London: Wiley.

the breccia dam. Such events have been reported from locations such as the Alps, Greenland, and the Canadian Arctic and commonly result in floods with peak discharges up to an order of magnitude higher than 'normal' diurnal meltwater flows. These floods occur within either ice-marginal river systems or englacial tunnels, and they are more frequent earlier in the melt season when increasing meltwater discharges undermine and remove glacier ice from winter advance positions.

Glaciers are able to store water within ice-dammed lakes over time-periods of 10^1–10^2 years, generating glacier outburst floods or 'jökulhlaups' when they drain. Lakes can form in a number of locations: beneath, within, on top of, and at the edge of glaciers (Figure 4). Subglacial lakes can exist either within pre-existing topographic depressions, such as volcanic calderas and valleys, or within up-turned bowl-shaped lakes protruding upwards into the glacier. Such subglacial lakes are common in Iceland where high geothermal heat fluxes and frequent volcanic activity are common. In Iceland, subglacial Lake Grímsvötn drains every few years as meltwater takes time to accumulate, raising the lake to a level where drainage will occur. In 1996, rapid influx of meltwater from the nearby Gjálp subglacial eruption site resulted in the filling of Grímsvötn within one month. Approximately 3.4 km^3 of meltwater drained subglacially from the lake as a jökulhlaup, which lasted for three days.

Although Grímsvötn is one of the largest modern subglacial lakes known to drain, it is dwarfed by Late Quaternary jökulhlaups from glacial Lake Missoula, which drained over 2000 km^3 of water within just a few days. The release of water from sub- or englacial pockets has also been invoked as an explanation for sudden outburst floods or 'debacles' from Alpine glaciers. Large supraglacial lakes are also known from modern glaciers and ice-sheets and can involve the drainage of up to ~0.1 km^3 of meltwater subglacially over distances of tens of kilometres. Ice-marginal, ice-dammed lakes are relatively common at modern glaciers, and the presence of shorelines, deltas, and lacustrine deposits provide widespread evidence for the existence of large ice-marginal lakes associated with Palaeo ice-margins. Large proglacial lakes formed during the deglaciation of North America, and drained to the North Atlantic and Arctic oceans via a series of enormous spillway channels. For example, glacial Lake Agassiz is known to have drained 160 000 km^3 of meltwater within only a few years.

Although ice-dammed lakes occupy a variety of locations relative to the impounding glacier, the most significant control on jökulhlaup characteristics is whether the glacier enters a main valley as an outlet and has the potential to pond a lake within the main valley upstream of the glacier dam. In this case, the ice-dammed lake volume would be much larger than that of ice-marginal, ice-dammed lakes relative to the glacier dam. A number of mechanisms to account for the drainage of ice-dammed lakes have been proposed. In general terms, lake drainage will take place if there is a hydraulic gradient allowing lake water to flow into the glacier. However, a connection must also exist between the lake and the glacier for drainage to commence. In most cases ice-dammed lake drainage leaves the glacier intact, but complete ice-dam removal may take place where the lake volume is large relative to ice dam volume. The timing of ice-dammed lake drainage is frequently controlled by the retreat and advance of glaciers. Glacier thinning and retreat often allow the initiation of cycles of lake drainage (Figure 6). Jökulhlaup magnitude decreases during each cycle as impounded lake volume decreases during glacier retreat (Figure 6).

Glacier retreat within high mountain regions often creates highly unstable moraine-dammed lakes, which are prone to sudden failure and flood generation. Calving of glacier ice into these lakes can generate waves, which overtop and then incise the moraine dam, allowing the extremely rapid development of a breach, facilitating catastrophic evacuation of lake water. Melt out of buried ice within moraine dams may take decades, and it is common for moraine-dammed lakes to exist for considerable periods before drainage. Once a moraine dam has been breached, only a subsequent glacier advance can reinstate the dam.

Landslides can generate floods in two ways. Firstly, rapid input of landslide debris to lakes can generate displacement waves. The best-known example of such a flood was at the Vaiont reservoir in Italy, where a landslide generated a large displacement wave that washed up a hillside and over the dam. In Iceland, a rockfall generated a displacement wave known as the Steinholtshlaup, which constituted one of the most rapid flood-rising stages experienced in Iceland. Landslides frequently dam major valleys allowing the formation of large lakes. Dam failure may occur immediately after formation, or may be delayed, possibly for decades, until reservoir volume has built up to a sufficient level for drainage to occur.

Bedrock and sediment dams are also known to fail via catastrophic enlargement of their spillways, leading to large outbursts. The drainage of Lake Bonneville in the USA during the Late Quaternary generated an outburst with a peak discharge of approximately $1 \times 10^6\ m^3\ s^{-1}$, whilst draining 4800 km^3 of water over several months. In New Zealand the

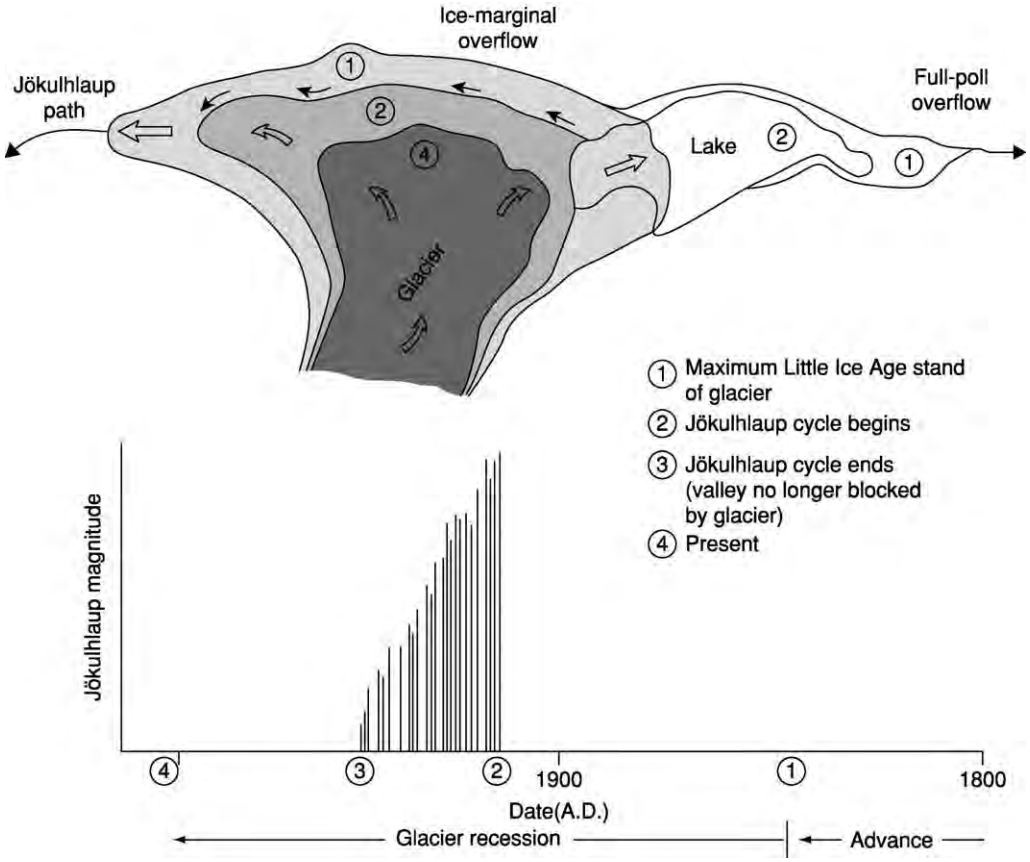


Figure 6 The jökulhlaup cycle. The upper diagram illustrates the advance and retreat of a glacier during the Little Ice Age. Ice positions 17–4 in the upper diagram correspond with the timescale illustrates in the lower diagram. The cycle of jökulhlaup drainage occurs as the glacier is thinning and retreating from its advance. Jökulhlaup magnitude within the cycle decreases as ice-dammed lake volumes decrease progressively during glacier retreat. After Evans SG and Clague JJ (1994) Recent climatic change and catastrophic geomorphic processes in mountain environments. *Geomorphology* 10: 107–128.

failure of caldera Lake Taupo generated a peak discharge of 1.7–3.5 $\times 10^5\,m^3\,s^{-1}$, draining a total of 20 km^3 of water. The failure in Alaska of Aniakchak caldera (3.7 km^3) generated a peak discharge of $1 \times 10^6\,m^3\,s^{-1}$. Deposition of alluvial fans at tributary junctions within ephemeral systems can block subsequent flood flows, resulting in enhanced flood peaks due to sudden dam failure.

Catastrophic Flood Characteristics

The discharge hydrograph is the most important characteristic of any flood as it defines flood duration and magnitude. Flood magnitude can be described by the peak discharge and total flood volume. Hydrograph shape reflects flood generation mechanisms and drainage basin characteristics. Different flood generation mechanisms result in different hydrograph types (Figure 7). Dam failure due to rapid breach development generates very rapid rises to peak discharge (Figure 7A). Volcanically generated (volcanoglacial) jökulhlaups also produce extremely rapid rises to peak discharge due to extremely high ice melt rates and glacier hydrofracturing. Rainfall-generated floods in ephemeral river-systems also display rapid rises to peak discharge followed by more sedate waning flow stages. Many jökulhlaup hydrographs exhibit gradual rising stages characterized by an exponential shape (Figure 7B). These jökulhlaup hydrographs are controlled by a positive feedback between meltwater flow and tunnel enlargement (Figure 7B and 7E). The rate of exponential discharge increase is controlled by ice-dammed lake water temperature, the length of the tunnel routeway, and height difference between the lake surface and the outlet. Flood hydrographs with multiple peaks can represent either nonsynchronous response of various subcatchments, complex channel networks to runoff, or temporary channel blockage and subsequent release. For volcanoglacial jökulhlaups, multiple peaks may be generated by peaks in volcanic activity. Similarly 'heartbeat' events associated with temporary tunnel blockage are characterized by a sudden fall and then rise in discharge (Figure 7C). Sudden onset

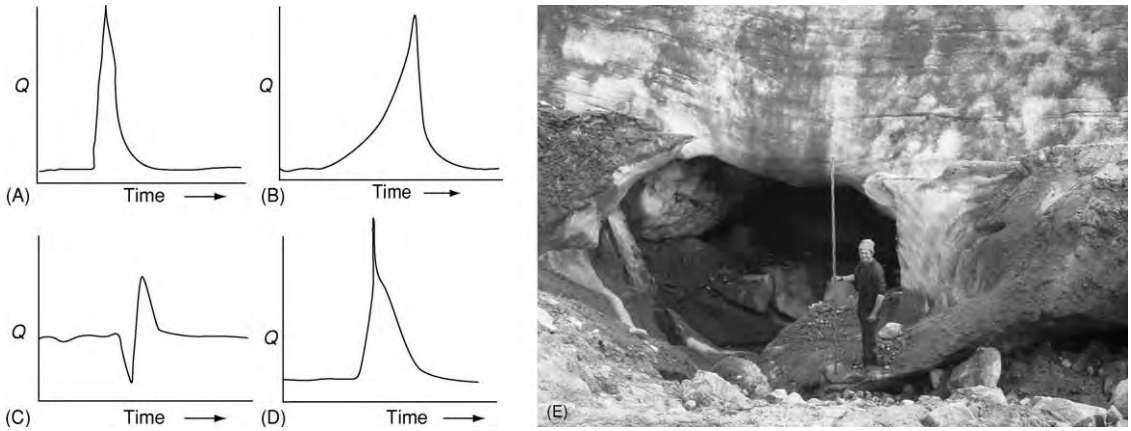


Figure 7 Flood hydrographs with contrasting shapes. (A) Rapid rise to peak discharge associated with sudden dam failure and glacier outburst floods associated with sheet flow and hydrofracturing. (B) Exponential rise to peak discharge is typical of floods where rising stage is controlled by progressive enlargement of ice-tunnels. (C) 'Heartbeat' event where temporary channel blockage results in a reduction in discharge followed immediately by a flood peak associated with dam failure. (D) Composite flood hydrograph characterized by the superimposition of one flood wave on top of another. (E) Jökulhlaup tunnel inlet generated by the drainage of a small ice-dammed lake in West Greenland.

jökulhlaups may fill ice-marginal basins within minutes, creating short-lived ice-dammed lakes, which in turn drain to produce a composite flood hydrograph where one flood peak is superimposed upon another (Figure 7D).

Catastrophic floods are often characterized by a dearth of direct process measurements. Most hydraulic variables are extrapolated from geomorphological and sedimentary evidence using hydraulic modelling techniques developed for lower magnitude flows. Calculated velocities are often extremely high ($<40\,m\,s^{-1}$), often an order of magnitude greater than that for nonflood flows. Flow depths in excess of 500 m have been reconstructed for the largest floods on earth. Flood flows commonly vary between sub- and supercritical states. Shear stress and unit stream power provide useful indicators of a flood's capacity to erode channel boundaries and to transport sediment. Flows are commonly macroturbulent, and involve considerable aeration of the flow (Figure 8A). Large-scale roughness generates tornado-like vortices known as 'kolks', which have the capability to pluck large fragments from the bedrock surface. High-flow velocities and violent flow separation can generate local instantaneous reductions in the vapour pressure of water, allowing the formation of water vapour bubbles within the flow. Violent collapse of the vapour bubbles releases large amounts of energy, enabling the pitting of bedrock surfaces. Flood flows commonly superelevate towards the outside of channel bends and exhibit cambered surface profiles (Figure 8B). In many cases it is very difficult to define the flood surface due to unsteady and macroturbulent flow conditions (Figure 8A and 8B). Flood peak discharges commonly attenuate downstream due to resistance effects and transmission losses within some systems. Floods and displacement waves with high peak discharges and relatively modest volumes show the highest rates of downstream attenuation.

Figure 8 (A) Macroturbulent flow during a jökulhlaup in West Greenland. (B) Superelevated flood flow during a jökulhlaup in West Greenland.

High flood powers and bed shear stresses allow grain sizes that would normally be transported as bed load to be transported as either suspended or

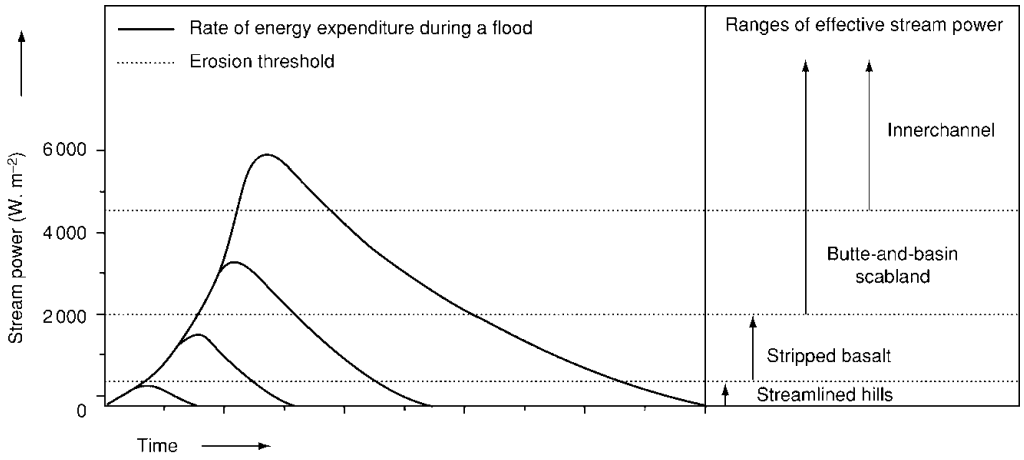


Figure 9 Progressive increase in stream power for various erosional channel morphologies in the Columbia River gorge. After Benito (1997) Energy expenditure and geomorphic work of the cataclysmic Missoula flooding in the Columbia River gorge, USA. *Earth Surface Processes and Landforms* 22: 457–472.

Figure 10 (A) Cataract erosion in the Jokulsá á Fjöllum system, northern Iceland. (B) Erosional bedforms sculpted by fluvially driven abrasion in high resistant bedrock, Kangerlussuaq, West Greenland.

wash load within the flow. A jökulhlaup in Greenland with a discharge of $1100\,m^3\,s^{-1}$ and powers of $3000\,W\,m^{-2}$ transported sediment of up to 10 cm in diameter in suspension, whilst the November 1996 jökulhlaup in Iceland with a peak discharge of $45–53\,000\,m^3\,s^{-1}$ and a power of $40\,000\,W\,m^{-2}$ was capable of moving 2-m-diameter boulders in suspension. These modern jökulhlaups only maintained such high stream powers locally and for a short duration. Giant Quaternary outbursts in the Altai Mountains of Siberia and from glacial Lake Missoula have generated flood powers of $10^5–10^6\,W\,m^{-2}$ over larger areas.

As erosional and sediment transport capacities are so high during floods, large amounts of sediment can alter flow conditions. Typical alterations include reduction of resistance to flow as sediment in transport blankets roughness elements, reductions of fall velocities of sediment in transport, and transition to upper flow regime bed conditions.

Sediment concentrations often exceed the values normally associated with fluidal flow conditions and become hyperconcentrated. Sediment bulking during floods may transform floods into debris flows, which have their own strength. Debris flow deposits directly reflect the entire flow thickness in contrast to water flood deposits, which only represent what was deposited at the bed of the flow. Sediment concentrations within catastrophic floods may also increase towards the base of the flow, leading to hyperconcentrated near-bed conditions, often referred to as a traction carpet. This may account for the deposition of units more typical of hyperconcentrated flow conditions by seemingly turbulent high-energy fluvial flows.

High sediment concentrations within floods can only be achieved if there is sufficient sediment readily available for entrainment. Sediment availability to catastrophic floods varies, depending upon flood generation mechanism, lithology, and weathering environment. Sediment bulking is usually most efficient within catchment headwaters, where precipitation inputs are high and stream and hillslope systems are

Figure 11 Subglacially eroded rip-up clast deposited by the November 1996 jökulhlaup near to the glacier margin.

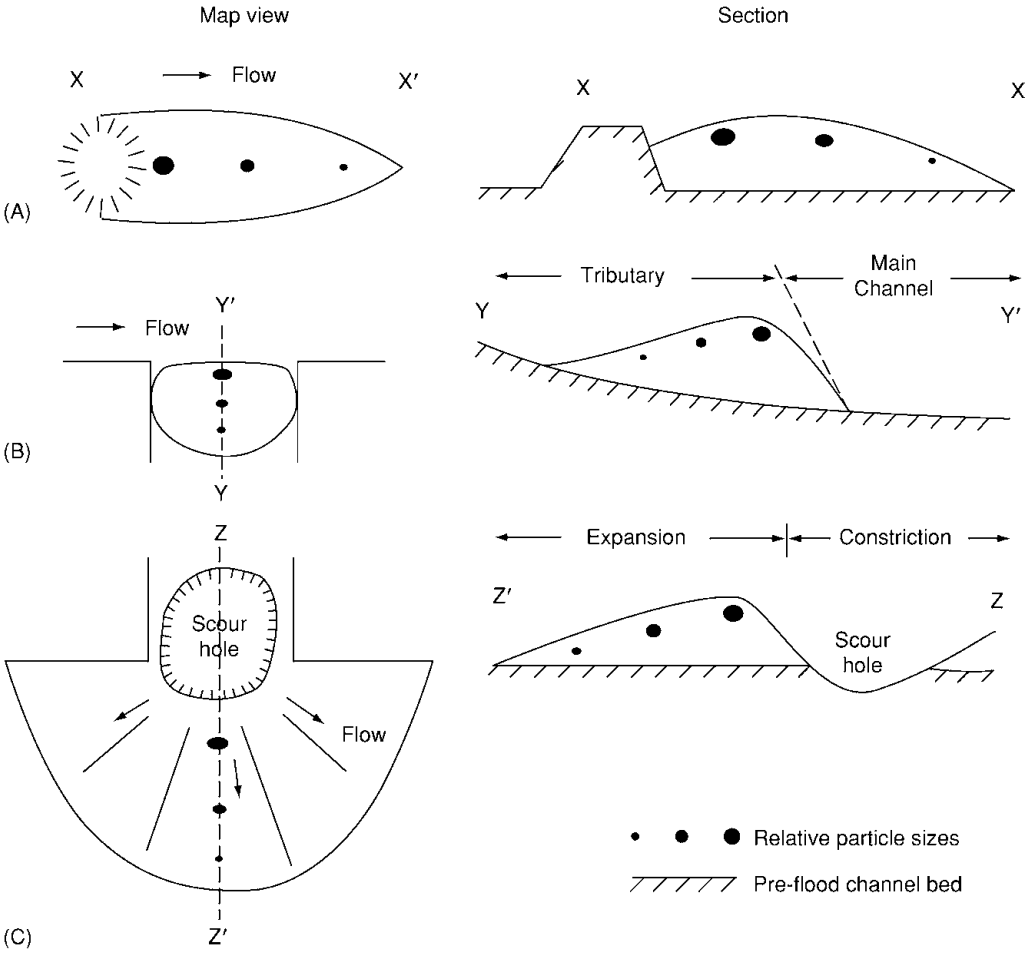


Figure 12 Role of channel geometry in controlling the morphology and sedimentology of flood deposits. Adapted from Baker (1984) Flood sediments in bedrock fluvial systems. In: Koster EH and Steel RJ (eds.) *Sedimentology of Gravels and Conglomerates*. Canadian Society of Petroleum Geologists Memoir 10, pp. 87–98.

coupled. In many cases, hillslope failure may transform directly into debris flows. Some floods may acquire large volumes of sediment at source. Jökulhlaups associated with subglacial volcanic eruptions contain large volumes of primary eruption products, and dam-failure outbursts are often heavily freighted with sediment eroded from the dam breach. Sediment concentration during nonvolcanic glacier outburst floods is usually supply limited, as large volumes of water dilute a finite volume of available sediment. Catastrophic floods can also bulk-up by eroding their channel boundaries.

Sediment transport during catastrophic floods exceeds nonflood transport rates by orders of magnitude. Such flows are able to access sediment sources beyond the reach of lower magnitude flows and are capable of exceeding thresholds for bedrock erosion. However, infrequent high-magnitude events may not be the dominant transport mechanism in catchments dominated by high productivity of fine-grained sediment.

Geomorphic and Sedimentary Impact

Flood impacts can be either erosional or depositional, with the occurrence, distribution, and character of specific impacts controlled by the spatial and temporal distribution of stream power. Erosional landforms are generated when flood power exceeds channel resistance to erosion. Flood flows not only have to exceed local erosional thresholds, but must persist long enough to generate distinctive suites of erosional landforms. Repeated Late Quaternary outburst floods from glacial Lake Missoula have created the world's most spectacular erosional landscape. Within the Columbia gorge system erosional landforms have been associated with various stream power ranges. Only the largest of the Missoula floods had sufficient power and duration to develop the entire suite of erosional features: streamlined hills, stripped basalt, butte and basin scabland, and inner channels, in order of ascending stream power (Figure 9). Although landscapes of flood erosion owe their detailed morphology to the interaction between hydraulic processes and local bedrock strength and structure, bedrock channels exhibit numerous common characteristics. Cataracts are common where flood flows excavate a more resistant cap-rock, allowing for rapid headward migration of a plunge pool (Figure 10A). Large-scale potholes are common where deep flood flows allow kolks to form. Where bedrock surfaces are stable, they can be abraded by sediment in transport to produce a range of erosional bedforms (Figure 10B). The widespread distribution of bedrock bedforms in parts of North America has been interpreted as the product of catastrophic subglacial sheet floods. Large channels and lake overspills containing large-scale and widespread erosional evidence are common in many

Figure 13 (A) Oblique aerial photograph of valley confined outwash plain subject to jökulhlaups showing proximal expansion bars. (B) Photograph of 1987 jökulhlaup flows across a outwash plain in West Greenland 1987. Note high-energy proximal flows expanding from a flow constriction and relatively tranquil flows in the distal area related to backwater ponding upstream of a flow constriction.

parts of the world subject to catastrophic floods. As well as eroding bedrock, floods have the potential to incise unconsolidated sediment. The November 1996 jökulhlaup in Iceland excavated large volumes of subglacial sediment, providing enormous numbers of rip-up clasts to the ice proximal outwash plain (Figures 11 and 15B). Large buried tunnel valleys in former glaciated regions have been attributed to Quaternary catastrophic floods. Large tracts of the channelled scablands inundated by the Late Quaternary Missoula outburst floods have eroded into loess, generating large-scale streamlined islands.

Catastrophic flood deposition is controlled by spatial and temporal variations in stream power. Downstream variations in flood channel morphology control stream power variations, resulting in zones of erosion, nondeposition, or deposition. Flood deposition within confined bedrock channels is limited either to the flood-waning stage or to specific locations such as the mouths of tributary channels, embayments, or the lee of obstacles to flow (Figure 12). Large bars radiate downstream from flow expansions and commonly surround a scour hollow or zone of nondeposition. Boulder ridges or 'berms' are also frequently deposited at sudden flow expansions at the boundary between the main and secondary flows (Figures 12 and 13A). Flow constrictions commonly generate considerable backwater effects, which build up progressively during the rising stage and subside during the waning stage (Figure 13B). Deposition takes place into the backwater zone under progressively lower energy flow conditions (Figure 13A and 13B). High-stage backwater deposits are reworked by higher energy waning stage flows. If during the flood the channel constriction is widened by erosion, the backwater effect will reduce substantially and subsequent backwater effects will be reduced. Unconfined flood flows tend to be shallower with high rates of flow dissipation consequently, resulting in thinner flood units.

Although catastrophic flood deposits are commonly identified by the presence of anomalously large clasts, indicative of high sediment transport capacities, this only applies where coarse-grained sediment is available for entrainment. However, for many fluvial systems coarse-grained sediment is supply limited. Other criteria are therefore needed to characterize flood deposits. Landforms and deposits may be interpreted to be of flood in origin if they are sufficiently large in scale relative to those associated with normal flows within the same fluvial system. Giant flood bars often display large-scale cross-stratification, which could easily be misinterpreted as delta foreset beds. Flood units are usually laterally extensive and contain a relatively uniform suite of clast lithologies compared with nonflood units. Flood deposits often contain rip-up clasts composed of blocks of sediment eroded from the river bed or banks (Figures 11 and 15B). Mud clasts are common within ephemeral river systems where silts and clays are ripped up during the flood-rising stage. Glacier proximal jökulhlaup deposits commonly contain rip-ups derived from subglacial fluvial excavation, which comprise sheared and folded subglacial diamicton or older outwash sediments (Figures 11 and 15B). Large-scale ice block obstacle marks and their associated secondary melt-out structures are diagnostic of high-magnitude glacier outburst floods (Figure 14). Simultaneous progradation, aggradation, and backfilling within the ice-marginal zone are

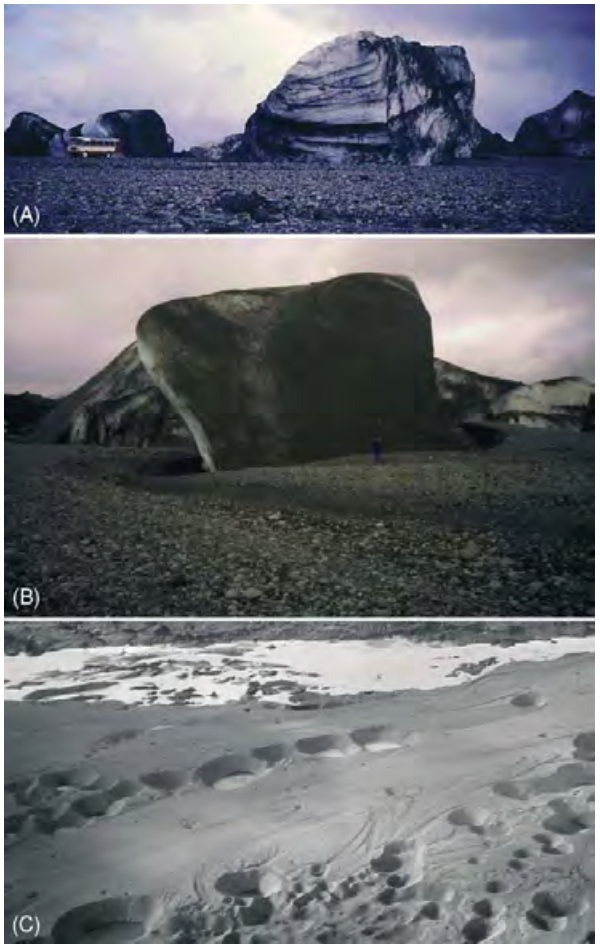


Figure 14 (A) Large ice block transported over 1 km by the November 1996 jökulhlaup on Skeiðarársandur, Iceland. (B) Deeply embedded ice block acts as an obstacle to the flow generating lateral scour channel and a leeside 'tail' of sediment. (C) Aerial view of kettle holes and obstacle marks four years after the jökulhlaup.

characteristic of high-magnitude events, capable of simultaneous channel erosion and infill.

Catastrophic flood units vary in sorting, texture, and matrix content, depending upon the size distribution of available sediment as well as its rate and mode of emplacement. Traction deposition generally results in very poorly sorted, cohesionless, clast-supported deposits, which commonly display poorly defined imbrication (Figure 15A). Stream powers calculated for many catastrophic floods are high enough to carry boulder-sized sediment in suspension. Rapid rainout and deposition from suspension at the mouth of tributary valleys commonly results in more matrix-rich deposits, which are typical of proximal eddy bar deposition (Figure 15B). More distal suspension sedimentation forms stacked upward-fining sequences similar to turbidites within lacustrine and marine environments (Figure 15C). Coarsening-upward successions are commonly deposited on the rising flood stage, reflecting deposition from transport-limited flood flows (Figure 15D).

Floods commonly access large volumes of sediment, allowing their transformation from water flood to hyperconcentrated flows (Figure 16). The presence of laterally extensive deposits consistent with deposition from hyperconcentrated flows indicates catastrophic floods. Although debris-rich floods may be commonplace within alluvial fan and ephemeral systems, within most river systems only exceptional floods have the capacity to generate hyperconcentrated flow conditions. Bedforms within fluvial systems dominated by repeated floods contain reactivation surfaces, which indicate progressive deposition during a series of high-magnitude floods interrupted by interflood pauses. Reactivation surfaces can also be identified by dateable palaeosols and tephra horizons, which can help in the reconstruction of Palaeo flood magnitude and frequency.

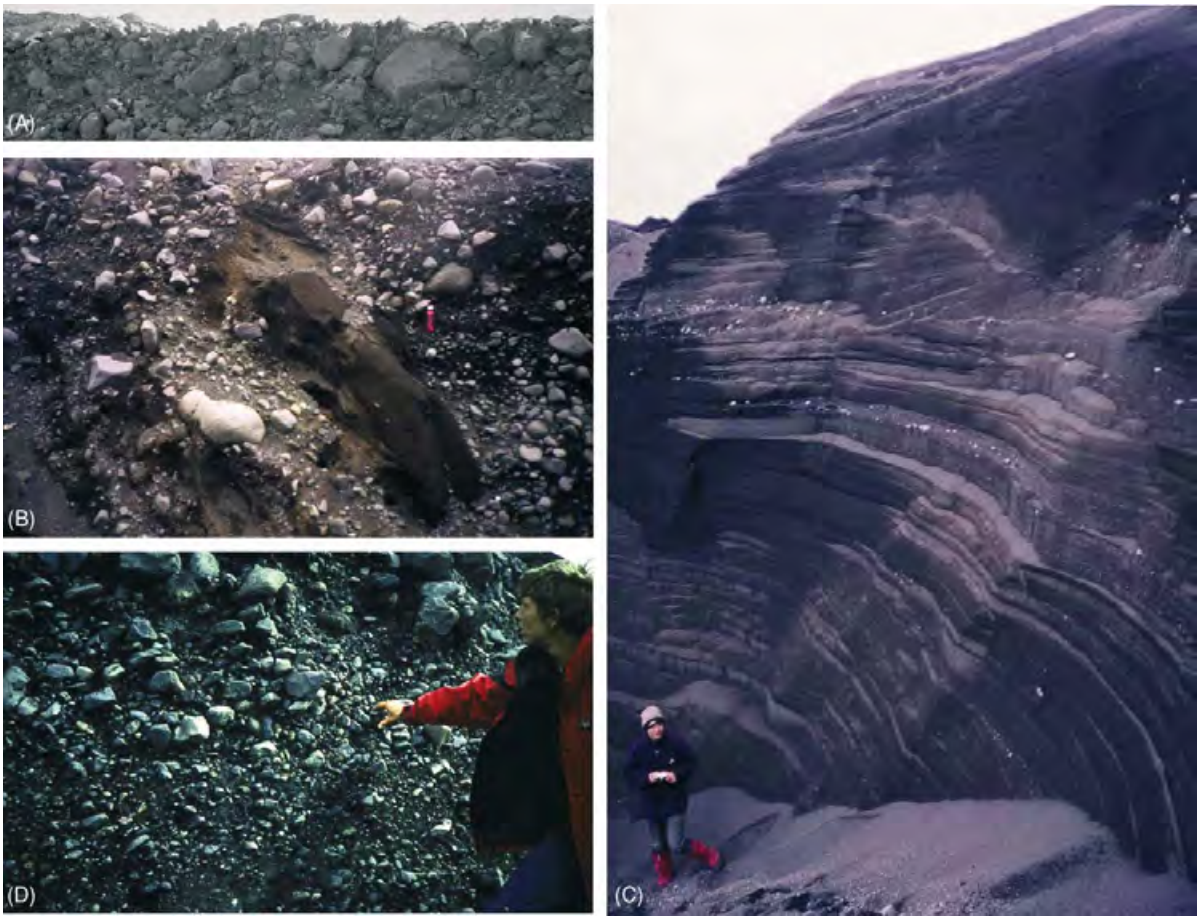


Figure 15 (A) Flood deposits from traction load; (B) proximal eddy bar deposits displaying noncohesive poorly sorted, matrix-supported deposits surrounding a large rip-up clast; (C) rhythmically deposited fine gravel beds; (D) photograph of a coarsening upward sequence reflecting rising-stage deposition.

Figure 16 Massive fine-grained unit on Sólheimasandur, Iceland interpreted as the product of a hyperconcentrated jökulhlaup.

Catastrophic floods deliver large volumes of sediment to the marine environment where they generate turbidity currents capable of transporting sediment offshore for up to 10^3 km. Turbidite units of up to 60 m in thickness identified in the offshore sedimentary record of the north-eastern Pacific Ocean have been attributed to outbursts from glacial Lake Missoula. In Iceland, volcanogenic jökulhlaups play an important role in the supply and transport of sediment to southern Iceland shelf and adjacent marine basins. Giant, sediment-rich jökulhlaups from Katla subglacial volcano dominate the offshore sedimentary record along most of Iceland's southern coast.

Controls on Catastrophic Flood Impact

Flood generation mechanisms, sediment availability, and routeway characteristics are the main controls on flood impact (Figure 1). The long-term impact of any flood may, however, depend upon the ability of more frequent, lower magnitude events to rework flood deposits, thereby allowing channel recovery. If flood recurrence interval is greater than the time required for recovery, then floods may be considered to have not had a long-term impact. Conversely, if flood recurrence interval is less than recovery time, then the channel will not be able to recover from the impact of one flood by the time the next flood occurs. Some flood impacts may, however, be regarded as being irreversible, in instances where high flood sediment calibre makes between-flood reworking impossible or where bedrock erosion occurs. Large-scale flood excavation of bedrock channels, such as those associated with glacial Lake Missoula outbursts, represent a long-lasting impact. Similarly, slackwater deposits associated with extremely high-magnitude flows will also have a high preservation potential if intervening flows with tributary streams are incapable of reworking flood deposits.

Conclusions

Catastrophic floods can be generated by a range of mechanisms within a number of hydroclimatic environments. Although the impacts of floods can be defined in a number of ways it is crucial that we assess carefully whether a flood is truly catastrophic. Continental-scale catastrophic flood landscapes can be identified in association with the growth and decay of ice-sheets during repeated glacial and interglacial cycles. Some floods were large enough to influence ocean circulation, thereby impacting on the climate system. Better understanding of the causes, characteristics, and impacts of catastrophic floods is crucial for the identification and mitigation of hazards and for interpretation of the sedimentary record.

See Also

Biblical Geology. **Engineering Geology:** Natural and Anthropogenic Geohazards. **Sedimentary Environments:** Depositional Systems and Facies; Alluvial Fans, Alluvial Sediments and Settings. **Sedimentary Processes:** Fluvial Geomorphology; Glaciers; Particle-Driven Subaqueous Gravity Processes. **Sedimentary Rocks:** Rudaceous Rocks. **Unidirectional Aqueous Flow**. **Volcanoes**.

Further Reading

Baker VR (1973) Paleohydrology and sedimentology of Lake Missoula flooding in Eastern Washington. Geological Society of America Special Paper 144.

Baker VR (1984) Flood sediments in bedrock fluvial systems. In: Koster EH and Steel RJ (eds.) *Sedimentology of Gravels and Conglomerates*. Canadian Society of Petroleum Geologists Memoir 10, pp. 87–98.

Baker VR and Costa JE (1987) Flood power. In: Mayer L and Nash D (eds.) *Catastrophic Flooding*, pp. 1–25. London: Allen and Unwin.

Baker VR, Benito G, and Rudoy AN (1993) Palaeohydrology of Late Pleistocene superflooding, Altay Mountains, Siberia. *Science* 259: 348–350.

Benito G (1997) Energy expenditure and geomorphic work of the cataclysmic Missoula flooding in the Columbia river gorge, USA. *Earth Surface Processes and Landforms* 22: 457–472.

Björnsson H (1992) Jökulhlaups in Iceland: Prediction, characteristics and simulation. *Annals of Glaciology* 16: 95–106.

Björnsson H (2002) Subglacial lakes and jökulhlaups in Iceland. *Global and Planetary Change* 35: 255–271.

Bull WB (1988) Floods—degradation and aggradation. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 157–165. London: Wiley.

Church M (1988) Floods in cold climates. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 205–229. London: Wiley.

Clague JJ and Evans SG (2000) A review of catastrophic drainage of moraine-dammed lakes in British Columbia. *Quaternary Science Reviews* 19: 1763–1783.

Costa JE (1988) Rheologic, geomorphic and sedimentological differentiation of water floods, hyperconcentrated flows, and debris flows. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 113–122. London: Wiley.

Costa JT and Schuster RL (1988) The formation and failure of natural dams. *Geological Society of America Bulletin* 100: 1054–1068.

Evans SG and Clague JJ (1994) Recent climatic change and catastrophic geomorphic processes in mountain environments. *Geomorphology* 10: 107–128.

Fisher TG, Smith DG, and Andrews JT (2002) Preboreal oscillation caused by a glacial Lake Agassiz flood. *Quaternary Science Reviews* 21: 873–878.

Hirschboeck KK (1988) Flood hydroclimatology. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 27–49. London: Wiley.

Hirschboeck KK (1987) Catastrophic flooding and atmospheric circulation anomolies. In: Mayer LD and Nasch D (eds.) *Catastrophic Flooding*, pp. 23–56. Allen & Unwin.

Kochel RC (1988) Geomorphic impact of large floods: review and new perspectives on magnitude and frequency. In: Baker VR, Kochel RC, and Patton PC (eds.) *Flood Geomorphology*, pp. 169–187. London: Wiley.

O'Connor JE (1993) *Hydrology, hydraulics, and geomorphology of the Bonneville Flood*. Geological Society of America Special Paper 274.

O'Connor JE, Grant GE, and Costa JE (2002) The geology and geography of floods. *Water Science and Application* 5: 359–385.

Shaw J (2002) The meltwater hypothesis for subglacial bedforms. *Quaternary International* 90: 5–22.

Tweed FS and Russell AJ (1999) Controls on the formation and sudden drainage of glacier-impounded lakes: implications for jökulhlaup characteristics. *Progress in Physical Geography* 23: 79–110.

Walder JS and Costa JE (1996) Outburst floods from glacier-dammed lakes: The effect of mode of drainage on flood magnitude. *Earth Surface Processes and Landforms* 21: 701–723.

Ward R (1978) *Floods*. Macmillan: London.

Wolman MG and Gerson R (1978) Relative scales of time and effectiveness of climate in watershed geomorphology. *Earth Surface Processes and Landforms* 3: 189–203.

Wolman MG and Miller JP (1960) Magnitude and frequency of forces in geomorphic processes. *Journal of Geology* 68: 54–74.

Deep Water Processes and Deposits

D J W Piper, Geological Survey of Canada, Dartmouth, NS, Canada

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Deep water beyond the continental shelves covers 63% of the Earth's surface. About one-seventh of this deep-water area, on the continental margins, receives a direct supply of terrigenous sediment from the continents, and this accounts for the majority of deep-ocean sediment deposits, including the largest Quaternary sediment accumulations on the planet, which occur beneath continental slopes and on submarine fans (**Figures 1** and **2**). Beyond the continental margins, the pelagic realm slowly accumulates biogenic material and far-travelled clays (**Figure 3**). The deep oceans are primarily a depositional environment, and that deposition is strongly influenced by ocean-floor morphology. Deep trenches at subduction zones and deep abyssal plains at passive continental

margins trap terrigenous sediment, preventing turbidity currents from reaching the greater part of the ocean floor. The large-scale morphology of the ocean is controlled principally by tectonic processes: seafloor spreading and its associated thermal effects, faulting, and volcanism; the complex rifting of continents in the early stages of ocean formation; and the even more complex effects of subduction and eventual closure of oceans. This article reviews the processes that lead to sediment distribution in the ocean and the large-scale distribution of these sediment deposits.

Technological advances in swath bathymetry in the past decades have greatly improved our understanding of the morphology of the seafloor. Given the vastness of the ocean and the technological challenges involved in monitoring seafloor processes, morphology is an important guide to the processes acting on the seafloor. Seismic-reflection profiling is an essential tool for evaluating the distribution of seafloor facies and hence inferring processes: the growing availability of three-dimensional seismic has revolutionized our understanding of continental-margin sedimentation. Direct sampling of deep-water sediments is limited to piston cores, which rarely penetrate more than 20 m, and some 1000 scientific boreholes, which are hundreds of metres long, drilled by the Ocean Drilling Program and its predecessors.

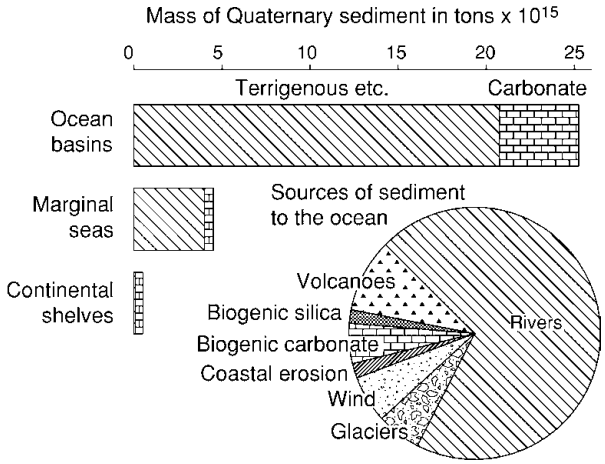


Figure 1 The volume of Quaternary sediment in the ocean, showing that the bulk of the deposition is in the deep ocean. The principal sediment sources are rivers, with volcanoes, glacial and wind transport, and biogenic carbonate also being important.

Sediment Sources

Rivers provide the principal input of sediment into the oceans (Figure 1). At high latitudes, terrigenous

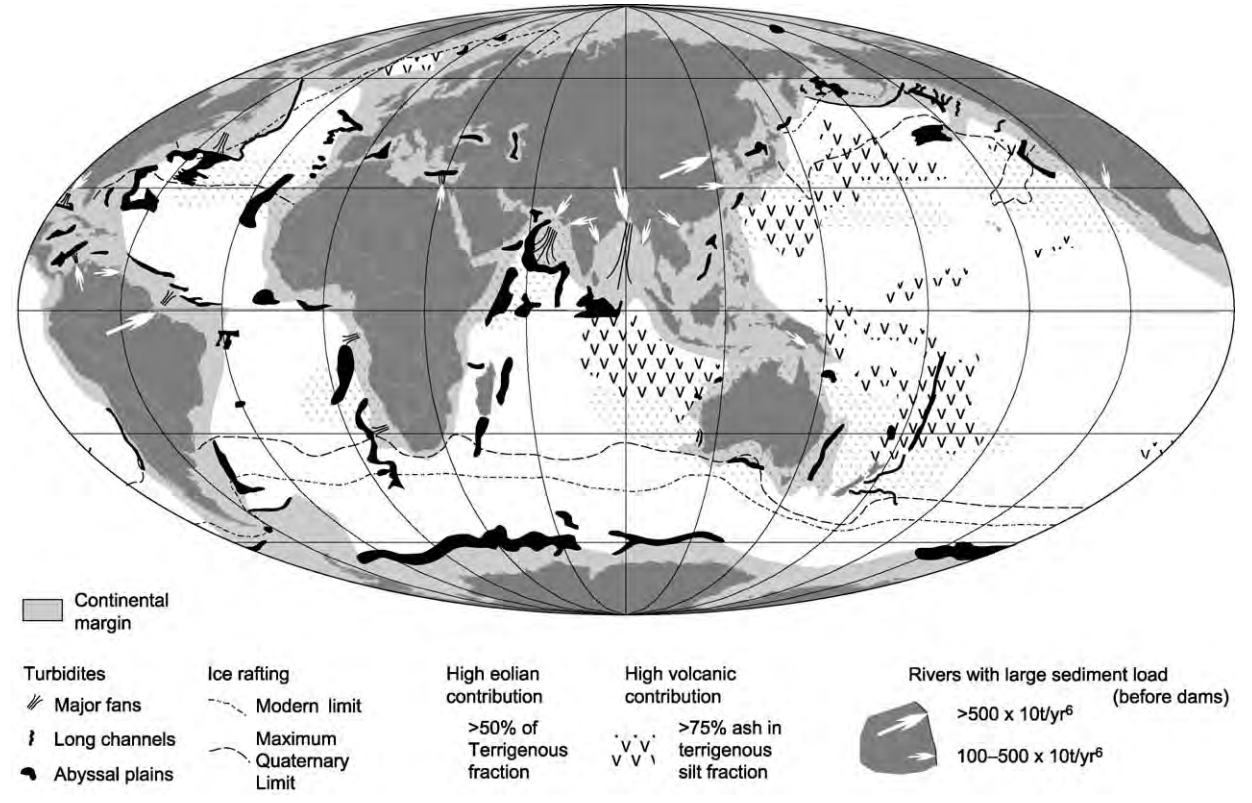


Figure 2 Terrigenous sediment in the deep ocean.

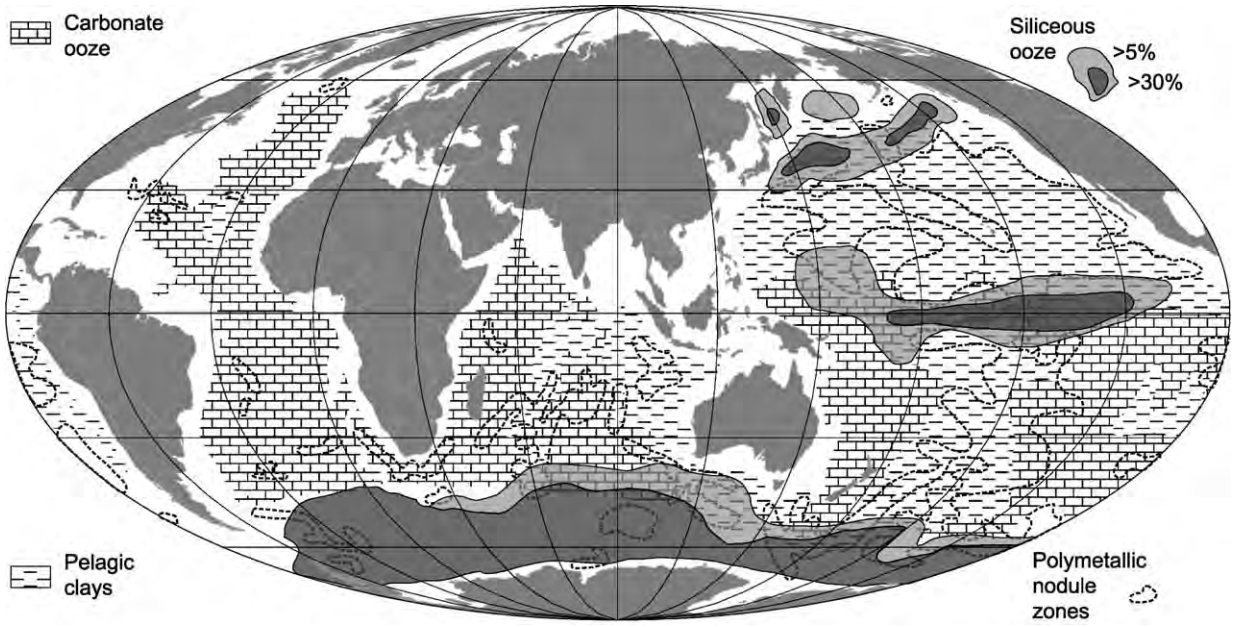


Figure 3 Distribution of the principal types of pelagic sediment.

sediment is supplied by sediment derived from subglacial melt water discharge and directly from ice, including meltout from icebergs thousands of kilometres from their source. In mid-latitudes, wind-blown dust (exceptionally including sand) is a significant component of deep-ocean sediment. Locally, volcanic material is a major component of seafloor sediment. Pelagic sediments (*see* **Sedimentary Rocks:** Deep Ocean Pelagic Oozes) consist principally of the skeletal material of pelagic organisms, notably calcium carbonate from Foraminifera, Coccolithophoridae (coccoliths or nannofossils), and pteropods, and opaline silica from Radiolaria, diatoms, and silicoflagellates. In oceanic areas with low sedimentation rates, various authigenic minerals form *in situ* in the sediments or on the ocean floor and may become significant components of the sediment. These include: phosphorites formed in slow-sedimentation environments in areas of active upwelling; metal-rich sediments and iron oxides associated with hydrothermal discharge in areas of active volcanism, particularly mid-ocean ridges (*see* **Tectonics:** Hydrothermal Vents At Mid-Ocean Ridges); manganese nodules and crusts of iron–manganese oxides, which precipitate in areas of slow sedimentation on the deep ocean floor (*see* **Sedimentary Rocks:** Oceanic Manganese Deposits); zeolites, which form principally from the alteration of volcanic ash and glass; and barite, which may be related to either hydrothermal activity or high organic productivity.

The dispersal of terrigenous sediment from various sources can be tracked by studying sediment petrology. The gross distribution of clay minerals within the ocean has long been used to infer sediment source and dispersion. In recent years, the signature of radiogenic isotopes (particularly lead and neodymium) and mineral geochronology in detrital terrigenous sediment have been used to understand further source and dispersion, including the tracking of ice-rafted detritus and aeolian dust. Tephra beds have been characterized geochemically and mineralogically, both to identify source and as stratigraphic markers.

Sediment Transport Processes in Deep Water

The transport of sediment to deep water involves both the normal processes of oceanic circulation and episodic sediment gravity flows. The resulting terrigenous deposits depend on the supply of sediment from the continent and shallow continental margin and on the complex interaction between flows and ocean morphology.

Surface currents in the ocean are primarily wind-driven, whereas overall circulation in the oceans is a consequence of density differences caused by temperature and salinity variations in the ocean water. Deep circulation is driven by the sinking of cold saline water in the North Atlantic Ocean and around Antarctica, with net upwelling in the North Pacific Ocean. The effect of the Coriolis force is to intensify ocean circulation on the western sides of oceans, where the thermohaline circulation generally

parallels the continental margin. Thus, for example, on the western side of the North Atlantic Ocean a powerful surface current (the Labrador Current) strongly influences the seafloor to water depths of 1500 m, and the deep Western Boundary Undercurrent of cold saline water derived from the Norwegian and Greenland Seas flows southwards at water depths of 3500–5000 m off the coast of Canada and the USA. Constricting topography results in the intensification of currents, for example around and over seamounts, through major fracture zones in the mid-ocean ridge system, and where microcontinental fragments such as Rockall Bank or the Campbell Plateau create constrictions.

The mean flow in both surface currents and the thermohaline circulation is generally a few centimetres per second, but locally the flow can be complex and energetic, for example in the large eddies in the Gulf Stream and Kuroshio, the former creating the 'abyssal storms' measured at the Hebble site in the Western Boundary Undercurrent of the North Atlantic, where velocities exceed $0.2\,m\,s^{-1}$. Similarly high velocities are found in topographical constrictions, for example in the Gulf of Cadiz, where Mediterranean Overflow Water exits the Straits of Gibraltar. Such flows are capable of eroding fine-grained seafloor sediment, may leave lag sands and gravels, and produce a variety of abyssal bedforms. Suspended sediment derived from the continental margin or suspended by ocean currents may be advected long distances as nepheloid layers – zones of higher amounts of particulate matter in the ocean circulation. In lower energy areas, this suspended matter settles to form sediment drifts. Boundaries between different thermohaline water masses are commonly marked by zones of higher turbulence and internal waves, which lead to higher amounts of suspended particulate matter and, in some cases, to erosion where they interact with the continental slope or with seamounts.

Organic matter sinking from the surface waters of the ocean is progressively oxidized, thereby depleting mid-level waters in oxygen and at the same time returning many nutrients to seawater. These nutrients become available in areas of upwelling. Deeper waters therefore have higher concentrations of dissolved carbon dioxide, with the highest concentrations being found in the older deep water of the Pacific Ocean. Deep waters of the Pacific Ocean are consequently undersaturated in calcium carbonate, leading to the dissolution of biogenic carbonate. Most dissolution of opaline silica occurs during oxidation of the protoplasm, but further dissolution takes place in undersaturated deep waters.

The wind-driven surface circulation is sedimentologically important in dispersing surface plumes derived from rivers and glacial margins. In addition, icebergs are transported in the surface circulation and, as they reach warmer waters and melt, they deposit ice-rafted detritus on the seabed. During Pleistocene Heinrich events in the North Atlantic Ocean, surface melt-water plumes were dispersed 2500 km south-eastwards from Hudson Strait and icebergs were dispersed even farther by the surface circulation.

Atmospheric circulation is a lesser agent of sediment transport to the deep ocean. Volcanic eruptions contribute fine volcanic ash through both troposphere plume transport, which may extend for thousands of kilometres, and worldwide distribution of fine ash that reaches the stratosphere. Dust derived from deserts is the principal source of terrigenous material in much of the mid-latitude pelagic realm.

Infrequent gravity-driven processes, including turbidity currents, debris flows, and submarine landslides, transport the largest volumes of sediment from the shallow continental margins to the deep sea (*see* **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes). A turbidity current is a density current in which the denser fluid is a sediment suspension. Turbidity currents may be initiated by many processes: all that are needed are a mechanism to put sediment into suspension and a steep slope to maintain the flow. Thus, hyperpycnal flow of rivers, glacial outburst floods (jökulhlaups), storm resuspension of shelf or beach sediment (*see* **Sedimentary Environments:** Storms and Storm Deposits), and in-mixing of ambient water into a debris flow will all initiate turbidity currents on steep slopes. Potential energy is converted to kinetic energy as the flow accelerates down the continental slope, commonly eroding the seafloor and eventually depositing sediment on the deep ocean floor. Processes within an individual turbidity current may range from hyperconcentrated flow at the head and base of the flow to normal turbulent sediment suspension in the upper part and tail of the flow, ultimately producing a low-density suspension, which contributes to nepheloid-layer dispersion by the thermohaline circulation. Turbidity currents occur every 1–10 years in steep river-fed basins with narrow shelves and off large shelf-crossing rivers, but only every 100–1000 years on most continental margins and every 1000–10 000 years on distal abyssal plains.

The largest submarine landslides involve debris avalanches of blocks tens to hundreds of metres in size, which flow in a laminar fashion and scour deep grooves into the seafloor. Slide blocks overlying sands liquefied by seismic shaking and muddy rotational slumps may break up and in-mix water to evolve into debris flows. Large coherent slides may plough

into seafloor sediments, and outrunner blocks may extend far beyond the main slide; the movements of both are facilitated by hydroplaning.

Subaerial volcanic islands, most importantly in island arcs, contribute sediment to the marine realm by proximal fall-out of ash and by pyroclastic flows entering the sea and forming turbidity currents. Flank collapse of volcanoes, both subaerial and submarine, produces widespread and voluminous debris avalanches around volcanic islands, the largest exceeding 5000 km^3. Submarine volcanic eruptions may be effusive or explosive, producing lava flows, large quantities of hyaloclastite, and pumice. Loose material will tend to move downslope by mass-transport processes and may also be modified by bottom currents.

The biota plays an important role in deep-ocean sedimentology. The skeletons of planktonic organisms are the major component of pelagic sediments. Pelagic sediments are generally thoroughly bioturbated, as are hemipelagic muds. Microbial processes are important in controlling the chemistry of seawater and in the development of authigenic pelagic sediments. Coral reefs and both cold and hot vent communities create large biogenic rock masses in deep water.

Sediment Nomenclature

Different nomenclature schemes are applied to terrigenous sediments of the continental margin and to pelagic sediments of the open ocean. The standard nomenclature for terrigenous sediment is based on grain size, with the Shepard and Folk systems (Figure 4) being the most widely used. The term 'mud' is widely used for sediment comprising principally silt and clay. Standard pyroclastic terms are applied to submarine pyroclastic deposits, and terminology applicable to shallow-water limestones is generally also applied to deep-continental-margin limestones. For pelagic sediments, the Ocean Drilling Program scheme is widely used (Table 1). The term 'ooze' is applied to sediment with a biogenic component exceeding 60% and is preceded by the names of all components making up more than 30% of the sediment, for example foraminifer–nannofossil ooze (often abbreviated to foram–nanno ooze). Minor components (10–30%) are prefixed to the sediment name with the term rich, e.g. zeolite-rich radiolarian ooze. Where the biogenic content is between 30% and 60%, the name is based on the predominant fossil and the appropriate textural name for the siliciclastic component, e.g. foraminiferal clay.

Several genetic names are widely and loosely used to describe marine sediments. Hemipelagic muds are highly bioturbated muds with less than 30% biogenic material that accumulate slowly from suspension

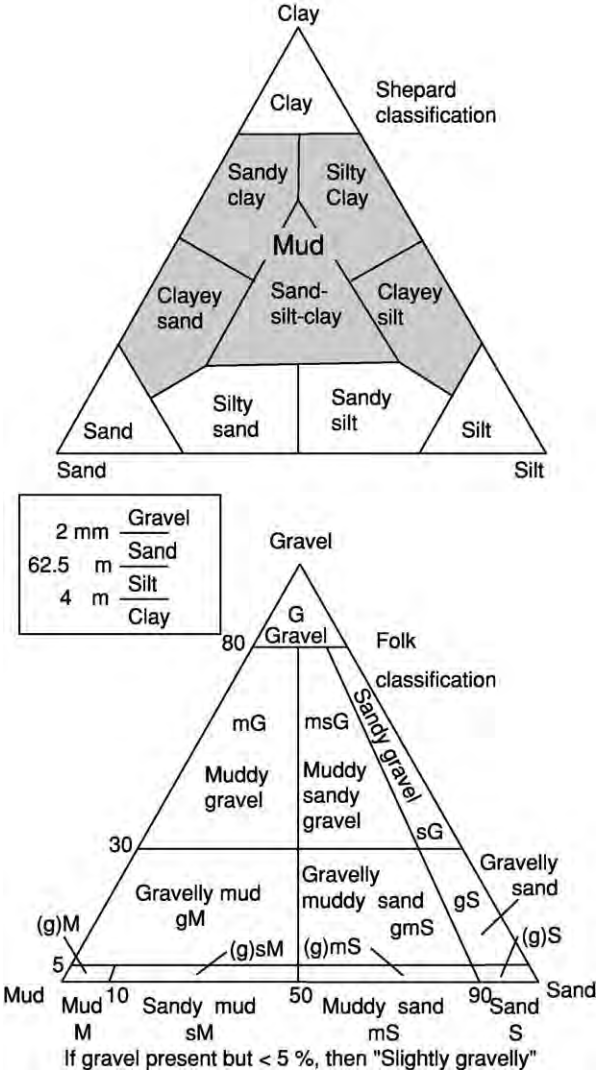


Figure 4 Shepard and Folk nomenclatures for terrigenous sediments.

fall-out of sediment that is either resuspended on the adjacent continental shelf or transported in nepheloid layers. The term turbidite is applied to sands, silts, and muds that are deposited principally from turbidity currents; their deposition is rapid, and bioturbation is sparse or absent. Sorted sands and gravels inferred to be deposited from hyperconcentrated flows are generally included in the term turbidite, although some authors have argued that the transport process resembles that of debris flows. Debrite is used for the blocky and poorly sorted deposits of debris flows, debris avalanches, and similar mass-transport processes. Deposits of contour-following bottom currents are generally called contourites (*see* **Sedimentary Environments:** Contourites); such deposits are most readily recognized from their morphological setting and from their character in seismic-reflection

Table 1 Nomenclature of pelagic sediments

EXAMPLES OF SEDIMENT NAMES					TYPE OF SEDIMENT		PRINCIPAL NAMED COMPONENTS
0 5 10 30 60					100 % biogenic material		
Silty clay, Clayey silt, Clayey sand, Clay (mud = silty clay and clayey silt)					TERRIGENOUS (continental margin)		Sand, silt clay
Clay	Zeolite-bearing clay	Nanno-rich clay			PELAGIC	Mainly abyssal brown clay	Clay, zeolite, ash, nannofossils, Radiolaria
	Foram-bearing clay	Foram-rich clay	Foram clay, nanno clay	Foram ooze, Foram-nanno ooze		Mainly biogenic carbonate	Foraminifera (forams), Nannofossils (nannos), Pteropods
	Diatom-bearing clay	Diatom-rich clay	Diatom clay	Diatom ooze, radiolarian ooze		Mainly biogenic silica	

profiles. The actual deposits are generally hemipelagic muds, although, in areas of strong current flow, coarser bioturbated winnowed deposits occur.

Continental Slopes

Continental slopes are regions of steeply sloping seafloor that lie between continental shelves and the deep ocean basins (Figure 2). Regional gradients are typically 2–5°, but locally slopes may be much steeper. Their large-scale morphology is a consequence of tectonic processes: the different elevations of continental crust and oceanic crust, the details of the original rift tectonics on passive continental margins, and the styles of subduction and accretion on convergent margins. This tectonic frame has been modified, most profoundly on older margins, by the deposition of continental-margin sediment, which locally may be more than 10 km thick over the basement. The position of the shelf break – the generally abrupt increase in gradient where the continental shelf passes into the continental slope – is in many places a direct consequence of sea-level lowering of around 110 m at the Last Glacial Maximum 21 000 years ago and marks the limit of progradation of shallow-marine sediments. In many cases the shelf break represents the seaward limit of delta-top progradation. In high latitudes, the shelf break may mark the seaward limit of continental ice-sheets. Where shelf-edge reefs are developed, the elevation of the shelf break is a consequence of a balance between sea-level rise and vertical reef accretion and is commonly close to present sea-level.

Continental slopes show a complex balance between erosional and depositional processes. The steep gradients promote sediment gravity processes, and the ocean-margin topography leads to strengthening of the oceanic circulation, particularly on the western sides of oceans. Deposition is principally of sediment that crosses the continental shelf and is therefore most important where shelves are narrow, shallow, and energetic and where sediment supply is high, notably seaward of major rivers and in areas of high carbonate productivity, including reefs. Regional deposition on the continental slopes off river mouths is from low-salinity surface plumes of fine-grained suspended sediment and from nepheloid layers developed from the settling of such sediment and resuspension of seafloor sediment. Storms and, on some shelves, tidal currents resuspend shelf sediment and advect it in suspension to the upper slope, where it is deposited as hemipelagic mud.

Slope morphology strongly influences sedimentological processes and products. Submarine canyons are widespread on many continental slopes, ranging in scale from the mighty Monterey Canyon, which is of similar dimensions to the Grand Canyon of the Colorado River, to small-scale slope gullies. Submarine canyons seaward of wide continental shelves are relatively inactive: the morphology appears to be a relict from sea-level lowstands. The canyons concentrate open-ocean water circulation, so that tidal flows may be sufficient to transport fine sands. These flows and the related upwelling also result in higher biogenic productivity than on the open slope. Sediment gravity flows, however initiated, will tend to flow within canyons, where they will accelerate on steep slopes.

Many canyons have their origins in river-mouth processes, both contemporary, as in the case of the Var Canyon in south-eastern France, and more generally at times of sea-level lowstands, when many rivers

discharged near the present shelf break. Direct hyperpycnal flow of sediment-laden floodwaters is the principal erosional process for smaller rivers with high bedload discharge; the role of hyperpycnal flows in very large rivers is uncertain. Glacial outburst floods erode particularly large submarine canyons, such as on the Laurentian Fan off south-eastern Canada. Studies of fjord deltas, a readily accessible modern analogue of shelf-edge rivers, suggest that failure of rapidly deposited river-mouth sediment may be equally important for initiating erosive turbidity-current flows. Canyons initiated by river-mouth processes at sea-level lowstands may undergo headward erosion as sea-level rises, as in the La Jolla Canyon off southern California and the Zaire/Congo Canyon of West Africa. In the La Jolla Canyon, beach sands accumulate in the canyon head under fair-weather conditions and are resuspended by storm waves that set up down-canyon turbidity flows that accelerate and erode canyon-filling sediment. Storm-wave transport of carbonate-platform sediment, especially through tidal passes in reefs, is important in initiating erosive turbidity-current flows on low-latitude carbonate margins. In general, turbidity currents will accelerate and erode on slopes of more than 2°, resulting in erosional undercutting of canyon walls, which promotes sediment failure. On some continental slopes, retrogressive sediment failure is an important process in the development of slope gullies. Slope gullies may also form as a result of density flows, derived from the fall-out of suspended sediment near river mouths and at ice margins, progressively coalescing to produce a badland-like drainage system.

The effects of sediment failure are widespread on many continental slopes, and failures are some of the most prominent morphological features on modern swath bathymetry of the continental slope. Many large-scale features, such as the 200 km wide Storegga Slide on the Norwegian margin, which failed 8000 years ago, are thought to have been triggered by earthquakes, by analogy with the 1929 Grand Banks failure on the Canadian Atlantic margin and the 1998 Papua New Guinea event. In many cases, including Storegga and the Cape Fear Slide on the US Atlantic margin, there is evidence that sediment strength was reduced by excess pore pressure due to the release of gas from gas hydrates. A particular type of large-scale sediment failure occurs on the slopes of oceanic volcanic islands, such as the Canary and Hawaiian islands, where progradation of steep volcanic edifices over weak deep-water sediments results in episodic catastrophic failure of the volcano flanks. Smaller-scale sediment failure is also widespread on continental slopes. On sediment-starved slopes, such as off New England (USA), failure may take place in Tertiary strata and principally involves slides. Retrogressive rotational slumps commonly evacuate many tens of metres of soft sediment on continental slopes producing blocky mass-transport deposits and debris flows. Sediment creep, such as that described from the South Korea Plateau, may eventually lead to failure along décollement zones tens to hundreds of metres below the seafloor.

Larger-scale gravity sliding of the upper few kilometres of the sediment column is the dominant feature of slope evolution seaward of some of the largest deltas in the world, notably those of the Niger and the Amazon, producing slope-parallel ridges and a prominent slope-toe escarpment. A similar large-scale morphology is produced on slopes with active salt tectonics, most spectacularly in the Gulf of Mexico, where the Sigsbee Escarpment is located at the toe of the slope and the slope morphology is dominated by salt diapirs and salt-withdrawal basins. Accretionary prisms at sediment-dominated convergent margins show large-scale morphological similarities to these complex passive-margin slopes, with thrust-controlled ridges and basins. In all of these complex slope settings, submarine canyons and valleys have irregular paths and variable degrees of incision. Slope failures are particularly common, as a result of tectonic oversteepening. Gas hydrates are particularly abundant on many accretionary margins, and fluctuations in the stability field with changes in bottom-water temperature or sea-level may trigger failures.

Overthrusting results in underconsolidated sediment with excess pore pressures, which commonly migrates to the surface to form mud volcanoes. Mud volcanoes are also common in prodeltaic environments with high sedimentation rates. Smaller-scale pockmarks – crater-like seafloor depressions a few tens of metres in diameter – indicate locations where gas or formation fluids are released, either catastrophically or quasi-continuously. In many areas, such pockmarks are developed along faults that leak formation fluids. Leaking hydrocarbons may also provide the energy source for biogenic 'cold-seep' communities on the continental slope.

Low-latitude carbonate platforms supply two main types of sediment to the continental slope. Where the platforms are fringed by rapidly growing reefs, collapse of the reef front results in debris avalanches. Storm transport of shelf sediment off a carbonate platform, particularly if it is focussed through passes in reefs, can produce turbidity currents, depositing well-sorted carbonate-rich sands and muds. At sea-level lowstands, with a narrow continental shelf, carbonate margins may behave like terrigenous margins, with the cutting of slope canyons that lead to small fan channels and lobes. At sea-level highstands,

carbonate production is generally much higher and may lead to progradation of the continental slope, channels on the lower slope, and turbidite deposition of carbonate sediment.

In contrast, the effect of lowered sea-level on mid-latitude continental margins was to increase sediment supply to the deep sea. In many cases, rivers discharged onto and prograded across a much narrower continental shelf, so that sediment was supplied directly to the high-gradient continental slope, down which it was transported by turbidity currents and mass-transport processes. At high latitudes, continental ice streams crossed many continental shelves, delivering glacial diamict directly to trough-mouth fans. Sediment delivery was particularly high during ice retreat in subarctic regions and was aided by abundant melt water. Sediment failures were triggered by glacio-isostatic earthquakes.

The Deep Continental Margin

The continental rise on a passive continental margin is a zone of sediment deposition on slopes that are typically between 1 : 50 and 1 : 500 and occurs beyond the steeper continental slope, which is commonly incised by canyons. The continental rise consists principally of submarine fans. An erosional submarine canyon leads to a submarine fan valley or channel, generally with depositional levees and a downslope decrease in channel depth. Depositional lobes are developed at the downslope termination of the channel and pass gradually into the flat abyssal plain, where the gradient of less than 1 : 1000 would be imperceptible to an observer on the ground. In more complex convergent continental margins there is generally more morphological control of deposition, but the same architectural elements are present: erosional canyons lead to leveed channels, channel-termination lobes, and, finally, basin or trench floors with more sheet-like stratigraphy. Channels may link two morphologically distinct basins, as in the case of the Biscay and Iberia abyssal plains. Indeed, these elements are also present on complex continental slopes with intraslope basins.

The channel systems are pathways for turbidity currents. Many channels are highly meandering; whether a channel is meandering or straight appears to be related to the gradient. On many passive-margin submarine fans, such as the Amazon, Zaire, and Bengal fans, channels change course abruptly through avulsion. Where channels are topographically constrained, for example Bounty Channel east of New Zealand, Surveyor Channel in the north-east Pacific, and the North-west Atlantic Mid-Ocean Channel in the Labrador Sea, an almost constant path has been maintained over thousands of kilometres. The lower parts of the turbidity currents transport sand through the channels and, owing to flow expansion, deposit it rapidly as channel-termination lobes. Only fine-grained sediment is transported to the distal parts of basins. The upper parts of the current entrain water, thicken, and spill over the levees, depositing overbank silts and muds. The cross-sectional areas of channels (kilometres wide, tens to hundreds of metres deep) and the velocities of turbidity currents estimated from cable breaks (2–19 m s^{-1}) testify to the size and power of the turbidity currents.

Sediment drifts are large sediment bodies constructed by deposition, generally of hemipelagic sediment, from the deep-water thermohaline circulation. The most powerful currents, in constrictions or adjacent to steep slopes, will commonly winnow or erode sediment, some of which may be of turbidite origin. This sediment, together with material introduced into nepheloid layers from the tails of turbidity currents or from resuspension on the shelf, is deposited slowly in areas of lesser current velocity. The North Atlantic Ocean, the Mediterranean Sea outlet, the eastern sides of the North and South American continents, the New Zealand margin, and the west side of the Antarctic Peninsula have particularly well-developed sediment drifts. Most have accumulated over many millions of years and have undergone alternating phases of deposition and winnowing or erosion as bottom-water circulation has fluctuated with changing climate. Many sediment drifts support mesoscale bedforms, including erosional furrows and sediment waves, with wavelengths of kilometres and heights of tens of metres.

Sediment Distribution in Pelagic Realms

Half of the Earth's surface is covered by pelagic sediment, yet study of its sedimentology is challenging because of its slow sedimentation rates and intense bioturbation. Some 47% of the pelagic realm is floored by foraminiferal ooze, 15% by siliceous ooze (mostly diatom ooze around Antarctica), and 38% by abyssal brown clay, in areas where there is total dissolution of biogenic material.

The distribution of biogenic oozes is a consequence of surface productivity, which is greatest in areas of upwelling, notably at the equator, on the western margins of continents, and at the Antarctic Convergence. Most biogenic material that reaches the

deep-sea floor does so through accelerated settling in the form of faecal pellets from zooplankton. Calcium carbonate is dissolved in deeper waters, particularly in the Pacific Ocean. The depth at which seafloor dissolution is complete is known as the carbonate compensation depth and fluctuates spatially and through time with changes in the thermohaline circulation.

Thus, foraminiferal and nannofossil oozes are widespread in the Atlantic and Indian oceans and on the East Pacific Rise but are absent over much of the northern and south-western Pacific Ocean (Figure 3). Sedimentation rates are generally 10–50 mm ka^{-1}. Diatom oozes occur beneath the high-productivity areas of the Antarctic Convergence and the extreme North Pacific Ocean, and radiolarian oozes occur beneath the Equatorial Divergence in the Pacific Ocean, with sedimentation rates of 2–10 mm ka^{-1}. Abyssal brown clays are most abundant in the Pacific Ocean, below the carbonate compensation depth, and have sedimentation rates of approximately 1 mm ka^{-1}. The authigenic components of pelagic sediments have been described above. Metalliferous sediments are generally restricted to small areas adjacent to sites of active volcanism but do cover a large area of the East Pacific Rise near the Galapagos Islands. Manganese nodules are particularly common in parts of the Indian and Pacific oceans with slow sedimentation rates. Poorly sorted ice-rafted detritus is an important component of pelagic sediment in the Arctic Ocean, the southern Ocean around Antarctica, and the northern North Atlantic Ocean.

The progressive subsidence of oceanic crust away from spreading centres results in a systematic stratigraphical succession in pelagic sediments. Thin metalliferous sediment and iron–manganese oxides may directly overlie oceanic crust. They are overlain by foraminiferal ooze deposited on the shallower seafloor of the mid-ocean ridge, with the sedimentation rate and the proportion of siliceous organisms depending on latitude. Eventually, subsidence of the oceanic crust may bring the seafloor below the carbonate compensation depth, and abyssal brown clays accumulate. Finally, subsidence and perhaps subduction will bring the ocean floor within reach of terrigenous sediment from the continent.

See Also

Sedimentary Environments: Contourites; Storms and Storm Deposits. **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes; Deposition from Suspension. **Sedimentary Rocks:** Deep Ocean Pelagic Oozes; Oceanic Manganese Deposits. **Tectonics:** Hydrothermal Vents At Mid-Ocean Ridges; Ocean Trenches.

Further Reading

Fisher RV and Smith GA (1991) *Sedimentation in Volcanic Settings*. SEPM Special Publication 45. Tulsa: Society for Sedimentary Geology.

Gardner JV, Field ME, and Twichell DC (1996) *Geology of the United States' Seafloor: The View from GLORIA*. Cambridge: Cambridge University Press.

Ginsburg RN (2001) *Subsurface Geology of a Prograding Carbonate Platform Margin, Great Bahamas Bank*. SEPM Special Publication 70. Tulsa: Society for Sedimentary Geology.

Hay WW (1994) *Material Fluxes on the Surface of the Earth*. Washington, DC: National Academy of Sciences.

Hovland M and Judd AG (1988) *Seabed Pockmarks and Seepages*. London: Graham and Trotman.

Ivanov MK, Kenyon NH, Suzyumov AE, and Woodside JM (eds.) (2003) Special Issue: Sedimentary processes and seafloor hydrocarbon emissions on deep European continental margins. *Marine Geology* 195: 1–318.

Kennett JP (1982) *Marine Geology*. Englewood Cliffs, NJ: Prentice Hall.

Lisitizin AP (2002) *Sea-ice and Iceberg Sedimentation in the Ocean: Recent and Past*. New York: Springer.

Lisitzin AP, Kennett JP, and Woolhiser C (1996) *Oceanic Sedimentation: Lithology and Geochemistry*. Washington, DC: American Geophysical Union.

Mulder T and Syvitski JPM (1995) Turbidity currents generated at mouths of rivers during exceptional discharges to the world oceans. *Journal of Geology* 103: 285–299.

Pickering KT, Hiscott RN, and Hein FJ (1989) *Deep Marine Environments: Clastic Sedimentation and Tectonics*. London: Unwin Hyman.

Rea DK (1994) The paleoclimate record provided by eolian deposition in the deep sea: the geologic history of wind. *Reviews of Geophysics* 32: 159–195.

Stoker MS, Evans D, and Cramp A (eds.) (1998) *Geological Processes on Continental Margins: Sedimentation, Mass Wasting and Stability*. Special Publication 129. London: Geological Society.

Stow DAV and Mayall M (2000) Deep-water sedimentary systems: new models for the 21st century. *Marine and Petroleum Geology* 17: 125–135.

Tappin DR (ed.) (2004) Special Issue: Submarine-slump-generated tsunamis. *Marine Geology* 203: 199–386.

Weaver PPE, Wynn RB, Kenyon NH, and Evans J (2000) Continental margin sedimentation with special reference to the north-east Atlantic margin. *Sedimentology* 47(Suppl. 1): 229–256.

Wynn RB and Stow DAV (2002) Special Issue: Recognition and interpretation of deep-water sediment waves. *Marine Geology* 192: 1–333.

Fluvial Geomorphology

J Lewin and P A Brewer, University of Wales, Aberystwyth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Fluvial geomorphology is concerned with the creation of landforms by river processes through the removal and transfer of materials on Earth's surface. Process studies are rooted in a number of disciplines, the earliest identifying 'processes' as evolutionary time sequences in landforms. In later work, this has continued in such fields as the study of meander dynamics. Hydrology-based approaches have involved the identification of formative (generally high river flow) events and their frequency, together with empirical relationships between river channel dimensions and river flow parameters. Work in engineering hydraulics, including the need to design stable channels, has also been a source of process understanding. Such work has led on to detailed analysis of water flow structures in river channels and prediction of thresholds for, and amounts of, sediment entrainment, transport, and deposition. Physical modelling both of channel reaches and of the treelike networks of rivers in their drainage basins has also now been joined by numerical modelling at a number of scales: flow in meander bends and confluences, the development of channel patterns, the evolution of drainage networks, and simulation of alluvial stratigraphy and sediment transfers within catchments as a whole. All this is particularly helpful because landform development, responding perhaps to only a few formative events in a year or to a rare flood, is only fractionally observable. In effect, modelling allows both a 'speeding up' of processes and experimentation with presumed controlling factors such as gradient and sediment sizes and mixtures. Results may then be checked against field observations. On the other hand, some field circumstances (such as the effect of needle ice or animal trampling on river bank stability, or real-time discharge fluctuations and their effects on groundwater and sediment flux on floodplains) are less easily modelled.

Fluvial systems exist over a range of scales, from centimetre-wide intertidal channels and other drainage networks on a sandy beach, to the more than 6 million km^2 drainage basin of the Amazon. A fine-textured system is shown in Figure 1. All such systems are complex, self-organizing, and hierarchical in structure. At the microscale, processes involve the movement of individual particles. These may be arranged in bodies of sediment (as in increments of bank erosion, moving gravel sheets on a river bed, or as sedimentation layers following a large flood), and these in turn become part of 'form units' such as channels, bed material forms (dunes or bars), levees and the crevasse channels that may cut across them, and backswamps. These units are combined into what have been called 'architectural ensembles', or types of rivers and their floodplains. These have form units that are present to differing extents according to contrasting process environments, each with their differing hydrological regimes, stream

Figure 1 The Zabriskie Point badlands in Death Valley, California, USA. The landscape is dissected by a hierarchy of channel networks ranging in scale from rills a few centimetres wide to a braided channel belt over 10 m wide.

energies, vegetation, and sediment sizes and supply rates.

At the highest level, drainage basins incorporate a range of ensembles from head to mouth. Higher level dynamic behaviour includes the evolution of drainage networks in the context of changing sea-levels, climate, and tectonic activity. This can become very complex, with forms retaining a kind of 'memory' of past conditions, which in part may affect their active behaviour. For example, cold-climate weathering and erosion products of the last Ice Age make up a major part of channel bed materials in some present-day temperate environment rivers. Elsewhere, bedrock channels may be more apparent, and in fact bedrock erosion has recently been more closely studied, notably in headwater channels that may be the major source of transported catchment sediment, and in the rocky channels and gorges found in steeplands, some semiarid environments, and elsewhere.

Overall, fluvial geomorphology is therefore concerned with a wide range of time and space scales, ranging from rock particles to river catchments. At the lowest levels, studies may be concerned with equilibrium responses to quasi-instantaneous forces; with increasing level, however, response time increases to a length that may parallel global climate change. Even without a change in external factors of this kind, it is likely that there is a need to understand evolutionary (and technically chaotic) behaviour, as well as equilibrium behaviour (as in the response of a particle of given shape and mass to applied fluid force). However, modelling complex behaviour in hierarchical systems does commonly incorporate lower level 'process rules', and this underlies the need in fluvial geomorphology for improved understanding of process mechanisms at all levels. Catchment factors (varying the gradient, sediment, and discharge at individual sites) affect local processes, which in turn may operate to change the character and extent of catchments.

Material Transfer

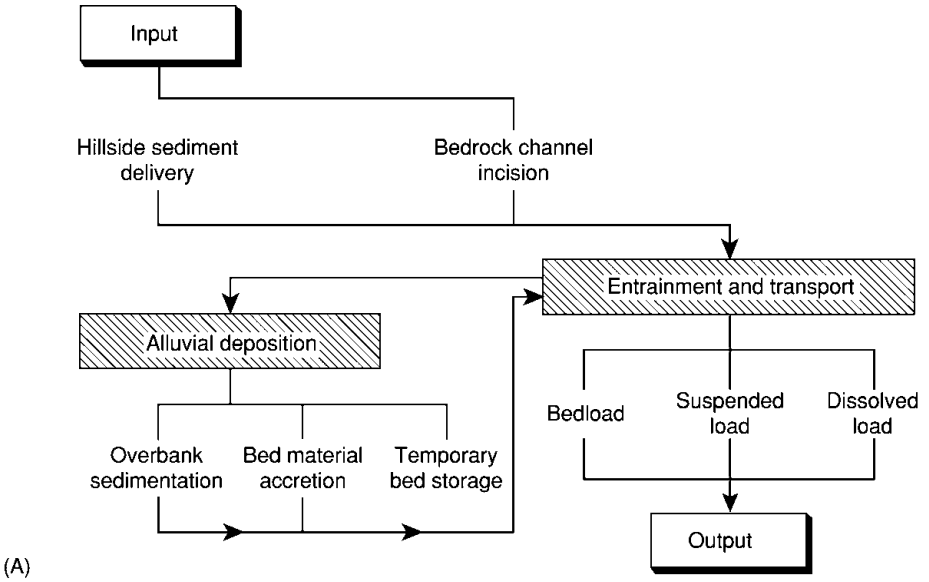
Lower level material transfer in alluvial systems is illustrated diagrammatically in Figure 2A, which emphasizes that periods of storage and reactivation are involved. Coarser material moves incrementally in reach-length steps associated with flood events, whereas fine sediment may move long distances during single events. Solute transfer is continuous, if varying in concentration. Solutes constitute an extremely significant part of the total load carried by rivers, especially, of course, in catchments with soluble limestone and shale lithologies (Table 1). Weathering processes (surface and subsurface) are important both in the general reduction of land surfaces and for making possible the input and transport of rock materials in particulate or solute form. For rivers, cold-climate environments (including glacierized ones) may produce large quantities of coarse, physically weathered materials; low-relief and vegetated tropical environments may contribute clay-rich sediment and solutes to their river systems from weathered soils. Fine sand may dominate river loads in sandstone terrains. Generally speaking, sediment transfer is intermittent; thresholds for movement both in channel and related to sediment input from hillslopes are differentially achieved within catchments, leading to what is called 'complex response' to forcing events.

For submerged or partially submerged particles, it is possible to calculate the physical forces that must be exerted on the particles to move them on a river

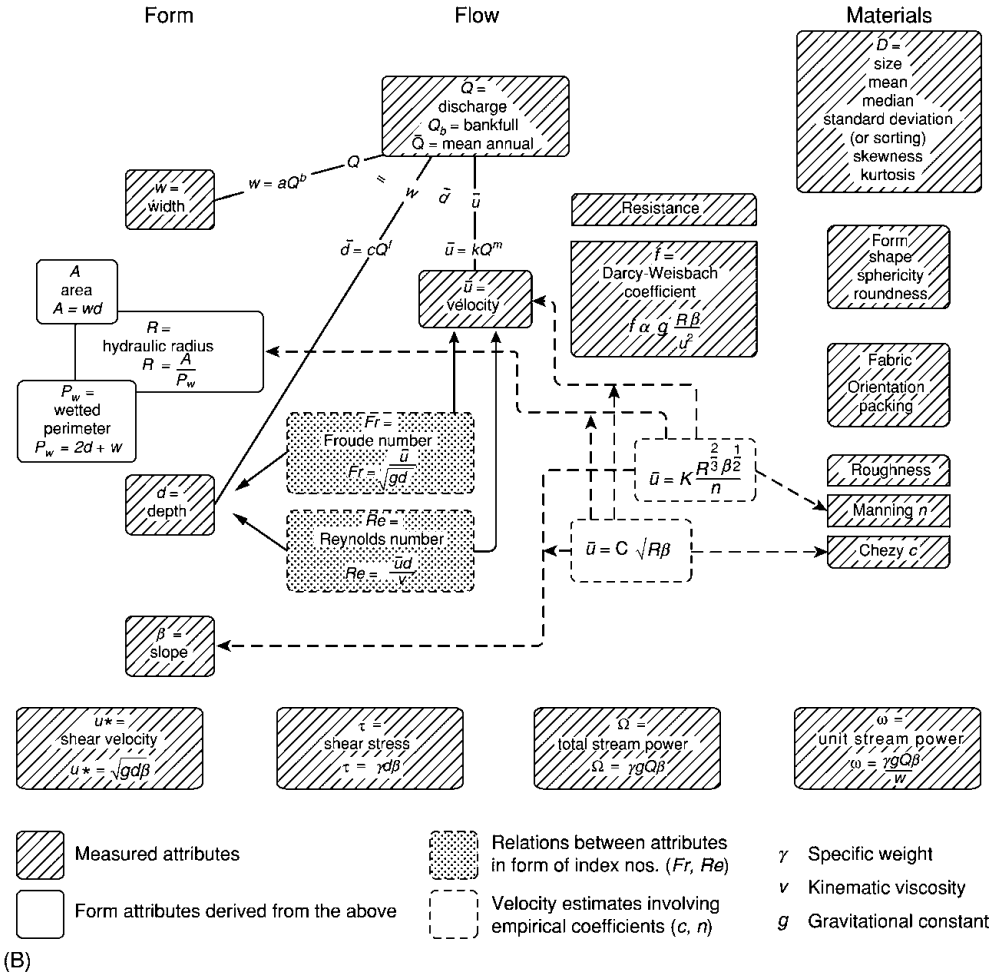
Table 1 Suspended and solute load data for selected rivers[a]

River	*Drainage area (km^2)*	*Mean specific suspended sediment load ($t\,km^{-2}\,yr^{-1}$)*	*Mean specific dissolved load ($t\,km^{-2}\,yr^{-1}$)*	*Percentage of total load carried in solution*
Congo	4 000 000	13	12	48
Nile	3 000 000	0.7	4	86
Brahmaputra	580 000	1370	130	9
Ganges	975 000	537	78	13
Danube	817 000	102	65	39
Mississippi	3 267 000	94	40	30
Colorado	635 000	870	23	3
St Lawrence	1 025 000	5	53	91
Amazon	6 000 000	79	46	37

[a]Data from Milliman JD and Meade RH (1983) World-wide delivery of sediment to the oceans. *Journal of Geology* 91: 1–21; and Meybeck M (1976) Total mineral dissolved transport by world major rivers. *Hydrological Sciences Bulletin, International Association of Scientific Hydrology* 21: 265–284.



(A)



(B)

Figure 2 (A) Material transfer in an alluvial river system. (B) Relationships between river form, flow, and materials. Reprinted with permission of Taylor and Francis.

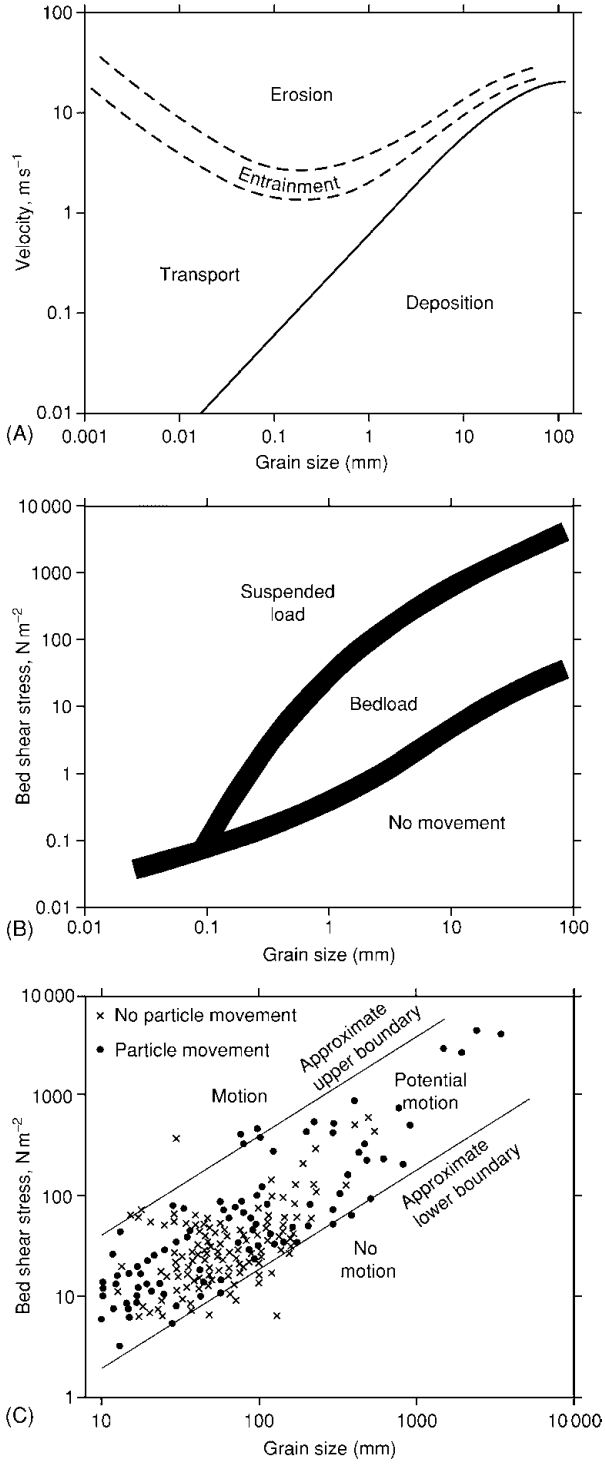


Figure 3 Relationships between sediment transport and bed velocity/shear stress. (A) Hjulström's (1935) critical velocity fields for entrainment, erosion, transport, and deposition. Note that for fine-grained cohesive materials, much higher velocities may be required to erode materials than to transport them in disaggregated form. (B) Summary of experimental observations on the forms of sediment transport. Reproduced with permission from Bridge JS (2003) *Rivers and Floodplains*. Blackwell: Oxford. (C) The more complex relationships actually observed in field situations. Reproduced with permission from Knighton D (1998) Fluvial Forms and Processes. Arnold: London.

bed, but these forces are usually expressed in terms of mean flow parameters (which are more easily estimated) such as mean velocity, bed shear stress, or unit stream power for a given river level (see Figures 2B and 3).These can then be related to movement thresholds, for particles of given size, as bedload or suspension load. However, field data show great complexity because of factors such as the presence of mixed grain size and the packing and armouring (in which finer material at depth is protected by coarser material above) of bed materials (Figure 3). The actual entrainment of particles also involves eddy, sweep, and burst flow phenomena in flowing water. Nevertheless, in general, turbulent flow (turbulence being expressed in terms of the Reynolds number) is capable of transporting fine material in suspension, with amounts set by supply rates, whereas coarser materials (available to the river from its bed and banks) are more characteristically competence limited, such that a threshold of river flow has to be reached for coarser sediment entrainment and transport. For coarser load, it has proved possible to develop a range of bedload formulas (for example, relating transport rate to excess shear stress, discharge, or stream power, the 'excess' being the value above the threshold at which particle motion begins). For gravel rivers, such equations have not in general proved very successful when checked against the limited amount of direct bedload transport observations available. Equipment may be of the kind installed on the East Fork River, Wyoming, where moving sediment was trapped in a set of open slots on the river bed, moved to the bank on a conveyor belt to be weighed, and then returned to the river further downstream (Figure 4).

In general, suspended and solute loads have to be measured following water sampling, and yield measures are much more generally available. Data obtained during flow events over periods of years show results that may be complex in detail: for example, sediment concentrations may be higher as waters rise during a flood, compared to when waters recede, and there is no precise and unique relationship between discharge or water stage and concentration. However, it appears that events of moderate frequency (around annual) and stage (about bankfull or channel-full) achieve the most work, though for bed materials, large floods may be necessary both to generate slope failure and coarse-material input and to move the larger particles on a river bed. Floods that are extreme, compared to the annual average, occur every few years in some environments, whereas in other environments the flood expected, say, every 10 years is not that much larger than the annual one (Figure 5). Globally, sediment yields may vary greatly and are high particularly in high-relief/active tectonic

Figure 4 A field facility for bedload sampling and measurement installed by the United States Geological Survey on the East Fork River, Wyoming.

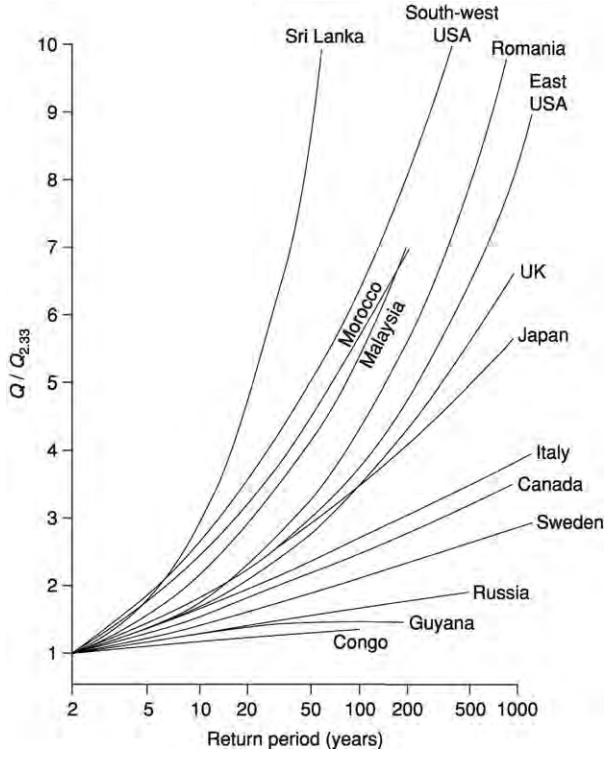


Figure 5 Curves showing flood frequency and magnitude relative to the mean annual flood ($Q_{2.33}$).

environments, on highly erodible terrains (including loess areas), and in regions of human disturbance.

Empirical relationships have also been established between channel measures and flow parameters. Flow velocity can be related to hydraulic radius (A/P; where A is the cross-sectional area of water flow and P is the wetted perimeter) and estimates of bed roughness, whereas river width and depth and velocity are differently related to variation in discharge, either as discharge rises or falls at a given cross-section ('at a station') or for constant flow conditions as discharge increases downstream (Figure 2B). These statistical average 'hydraulic geometry' relations (derived for single-thread rivers and so called because they are power function or 'geometrical' in form) are less apparent for downstream gradient and grain size, though both in general decrease, the former giving concave-up river 'long profiles', and the latter showing the effectiveness of in-channel size sorting and abrasion of particles as they move. In fact, research has rather moved away from detecting empirical relationships of this kind towards physically and theoretically underpinned modelling and field monitoring of bank erosion and channel shape processes. Field and laboratory studies of sorting and abrasion processes have explored topics such as abrupt discontinuities in channel bed grain sizes, and the nature of abrasion processes (see Figure 6).

Form Units

Alluvial rivers generally possess a set of landforms, self-formed in river-transported materials. There is a basic contrast between channels and their floodplains, the latter inundated only during flood flows producing a blanketing of flood deposits, which may

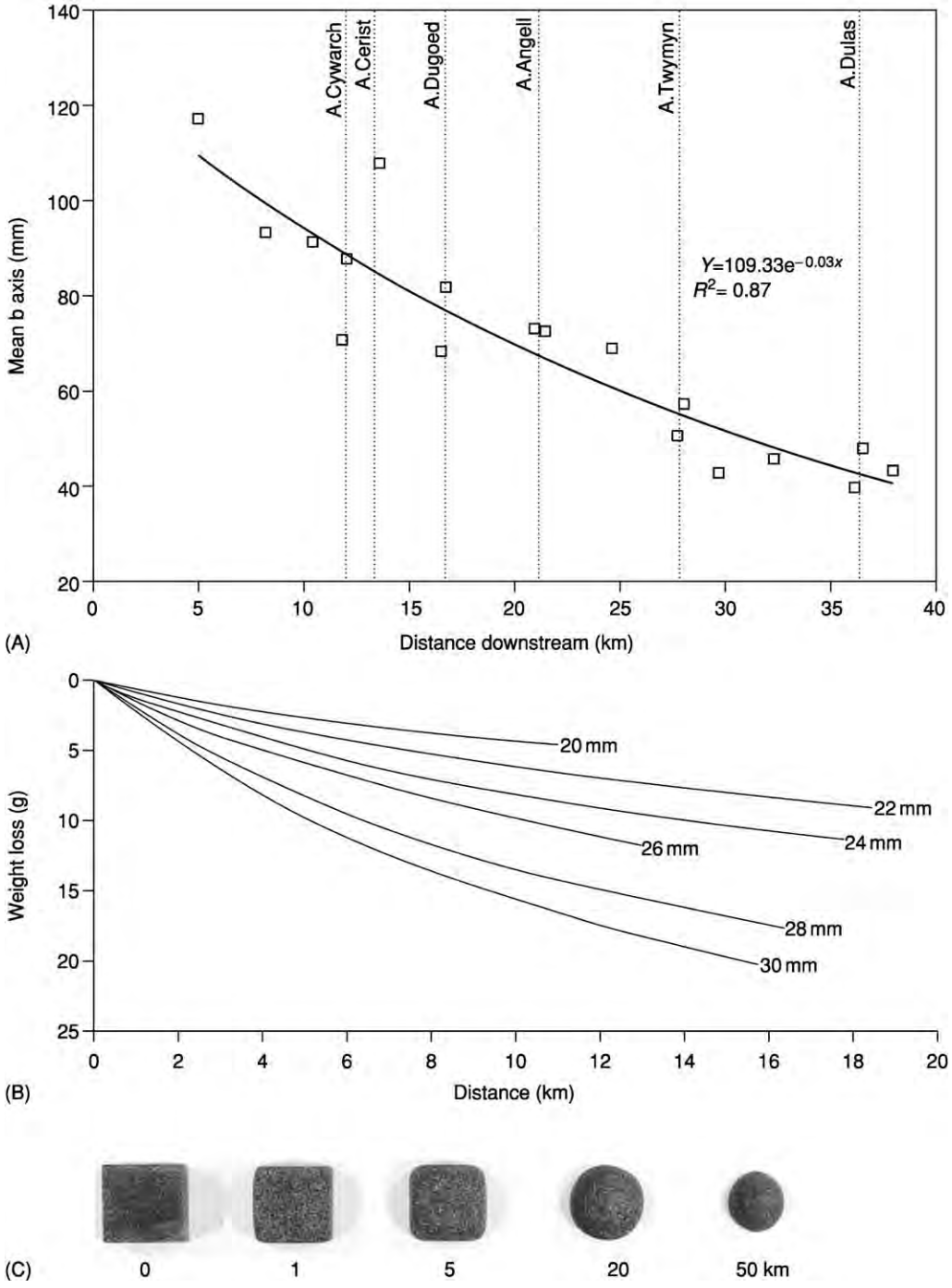


Figure 6 Field and laboratory data illustrating downstream fining and abrasion processes. (A) Downstream decrease in mean grain size on the Afon Dyfi, mid-Wales, UK; abrupt, but generally localized, increases in grain size are attributed to relatively coarse sediment inputs from tributary channels. (B) Weight loss of rock cubes (ranging in size from 20 to 30 mm) in an abrasion tank; weight loss per unit distance is greater for larger cubes because of kinetic energy and surface area effects. (C) Progressive size reduction and roundness increase of a 20-mm chinastone cube during 50 km of travel in an abrasion tank.

be on top of coarser materials laid down in-channel by migrating streams. In more detail, the set of landforms includes channel bars of various types (point bars on the inside of meander bends, together with side- and mid-channel bars exposed at low flows); eroding banks characterized by various types of failure (including particulate removal or mass failure, or beam failure where overhanging banks are undercut); levees (channel-side ridges) and crevasses (dissecting channels) at the channel margin; abandoned channel fills; and backswamp-type floodplain environments in which organic processes may dominate. Each of these has been studied intensively in recent years, with documentation of the style and rate of deposition/erosion and the flow conditions responsible. This has, in the case of in-channel flow and sediment monitoring and modelling, involved painstaking field measurement (sometimes under hazardous flood conditions) and theory development. In other situations, historical records and maps have been used to reconstruct channel changes over a century and longer.

The series of case studies now available exemplifies the general point that hydraulic processes operating in contrasted energy (set by gradient and incident flows) and sedimentary (set by the type and supply volume of sediment of particular size) environments produce contrasted landform developments. For example, rivers with laterally stable courses and high sand load may generate large levees and ponded backswamps behind them, whereas gravel bed streams may have a diversity of bars and bedforms but less well-developed levees and backswamps – especially where rivers are laterally mobile, such as to eliminate the incipient development of such form units.

Architectural Ensembles

Qualitatively recognized channel-pattern types initially included braided, meandering, and straight; anabranching varieties of each (in which the river has more than one branch, with islands or swamps between) are also possible (Figure 7A). The term 'anastomosing' is also used for laterally rather stable anabranching styles. These commonly develop in channels with levees and channel-bed aggradation, with avulsion periodically relocating the river course through a backswamp zone. Braided and meandering channels have also proved to be diverse in style; some of the former have more stable islands between

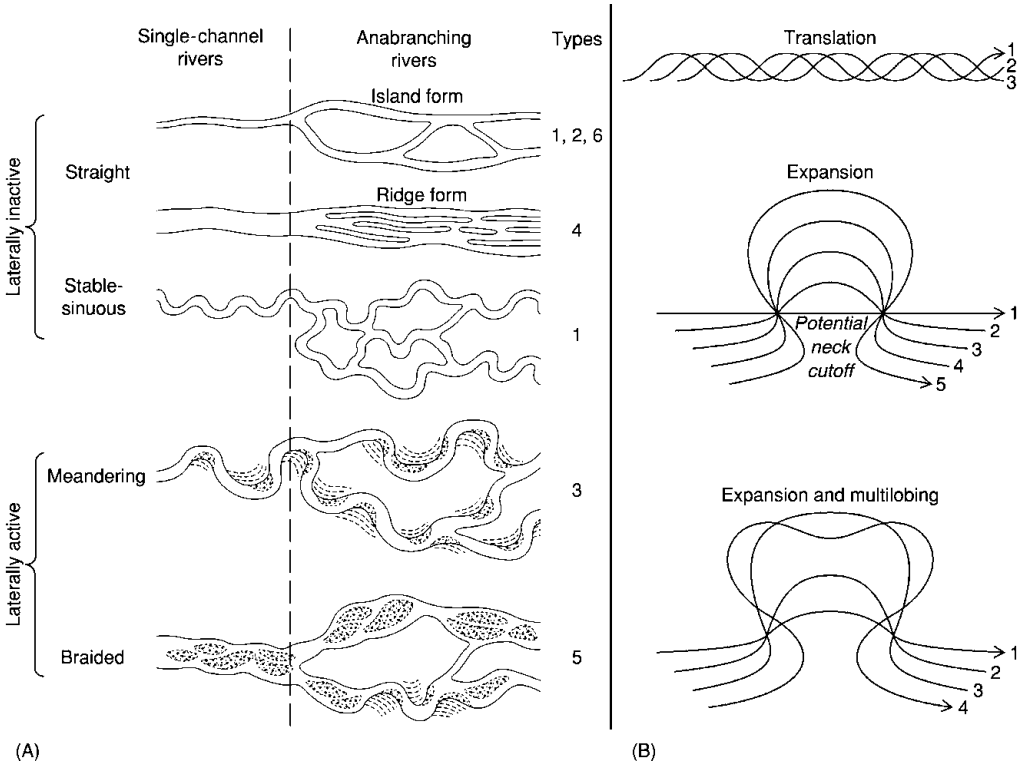


Figure 7 Alluvial river channel patterns. (A) Classification of types. Numbers refer to the styles of anabranching channels plotted in Figure 9. Reproduced with permission from Nanson GC and Knighton AD (1996) Anabranching rivers: their cause, character and classification. *Earth Surface Processes and Landforms* 21: 217–239 © John Wiley and Sons Limited. Reproduced with permission. (B) Forms of development in meandering patterns over time.

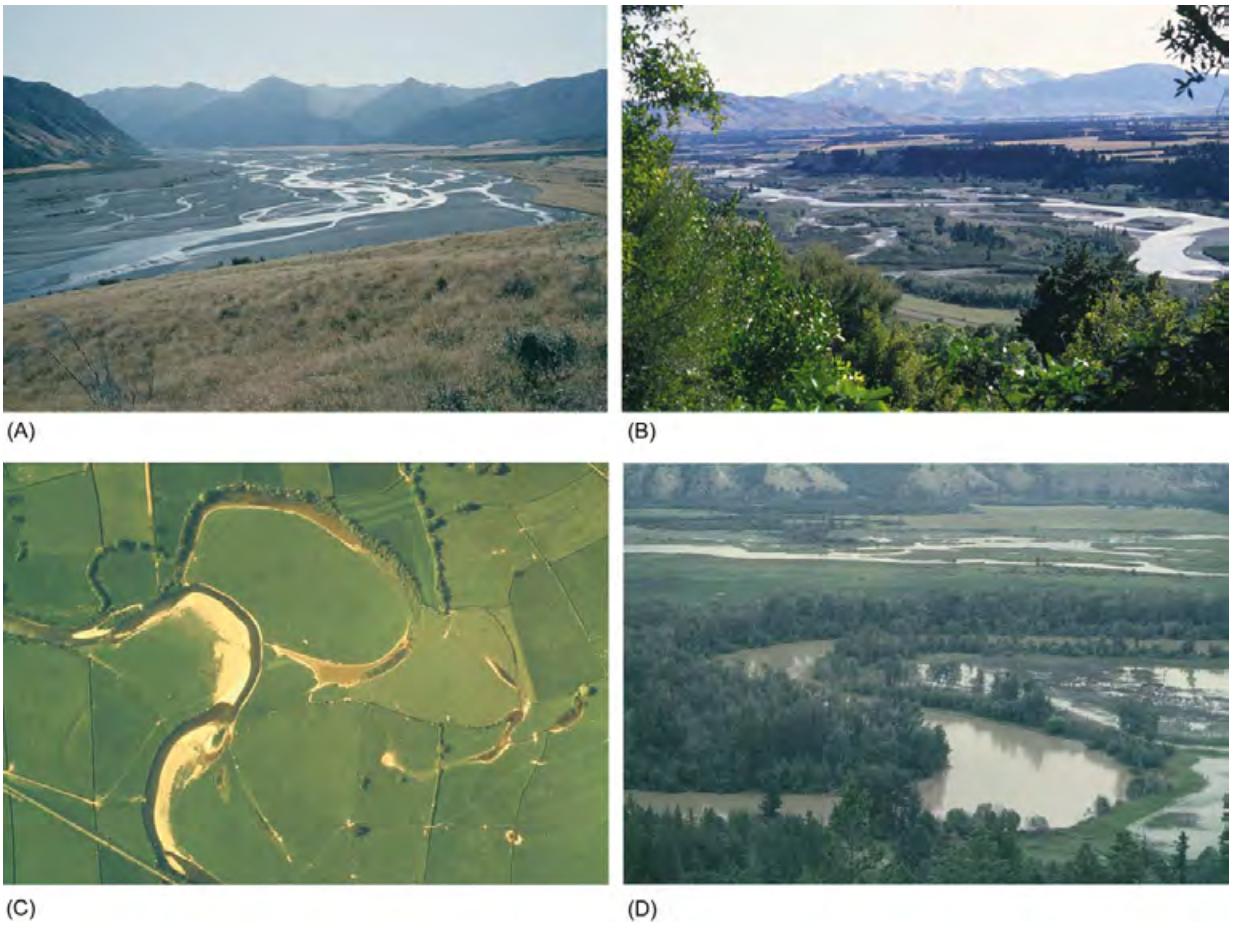


Figure 8 Channel and floodplain styles. (A) The braided course of the Upper Waimakariri River, New Zealand; (B) the Rangitata River at Peel Forest, New Zealand, showing vegetated islands and braid bars; (C) the meandering River Tywi, Wales, UK, showing gravel point bars and water-filled cut-off channels being filled by fine-grained flood sediment; (D) the anastomosing Columbia River, British Columbia, Canada, showing channels, levees with trees, and backswamps (a recent avulsion has created a new disorganized channel in the background).

bar-filled channels (the bars being submerged and migrating at high flows, with scour holes where flow threads join, and deposition at zones of flow divergence) and the latter vary in sinuosity, symmetry, and migration pattern (Figure 7B). Illustrations of four channel-style prototypes are shown in Figure 8.

A similar kind of approach may be taken with architectural ensembles as a whole. Table 2 shows the order and suborder classification of floodplains. The classes are broadly based on energy and sediment size, with subdivision according to factors such as confinement in relatively narrow valleys and the presence of particular sedimentation styles. The observed styles have specified ranges of stream power and grain size, but these may overlap or leave gaps on a stream power/grain size plot (Figure 9). The same also applies to data for rivers designated as braided, meandering, and straight. This is probably because factors other than bankfull power and bed material size are important. Such factors include bank strength (related also to vegetation), the role of overbank flows in bank breaching (as in cut-offs, crevasse formation, and avulsion), and the rate of supply of finer sediment. It is not easy to produce a set of simple process/environment criteria to underpin a qualitative subdivision of ensemble or floodplain types. Such discrimination has been noted, however, on gradient/discharge plots for channel types (Figure 10).

Drainage Basins

Despite local complexities, drainage basins as a whole exhibit general trends in controlling factors from source to mouth (or with increasing drainage area) and therefore in fluvial styles. These are illustrated schematically in Figure 11A, though it should be appreciated that individual catchments may be complex as a result of factors such as lithological

Table 2 A classification of floodplains[a]

Class	*Order*	*Description*
A: High-energy non-cohesive floodplains	A1	Confined coarse textured
	A2	Confined vertical-accretion sandy
	A3	Unconfined vertical-accretion sandy
	A4	Cut and fill
B: Medium-energy non-cohesive floodplains	B1	Braided river
	B2	Wandering gravel bed river
	B3	Meandering, lateral migration
		3a Non-scrolled
		3b Scrolled
		3c Backswamp
		3d Counterpoint
C: Low-energy cohesive floodplains	C1	Laterally stable single channel
	C2	Anastomosing
		2a Organic rich
		2b Inorganic

[a]Data from Nanson GC and Croke JC (1992) A genetic classification of floodplains. *Geomorphology* 4: 459–486.

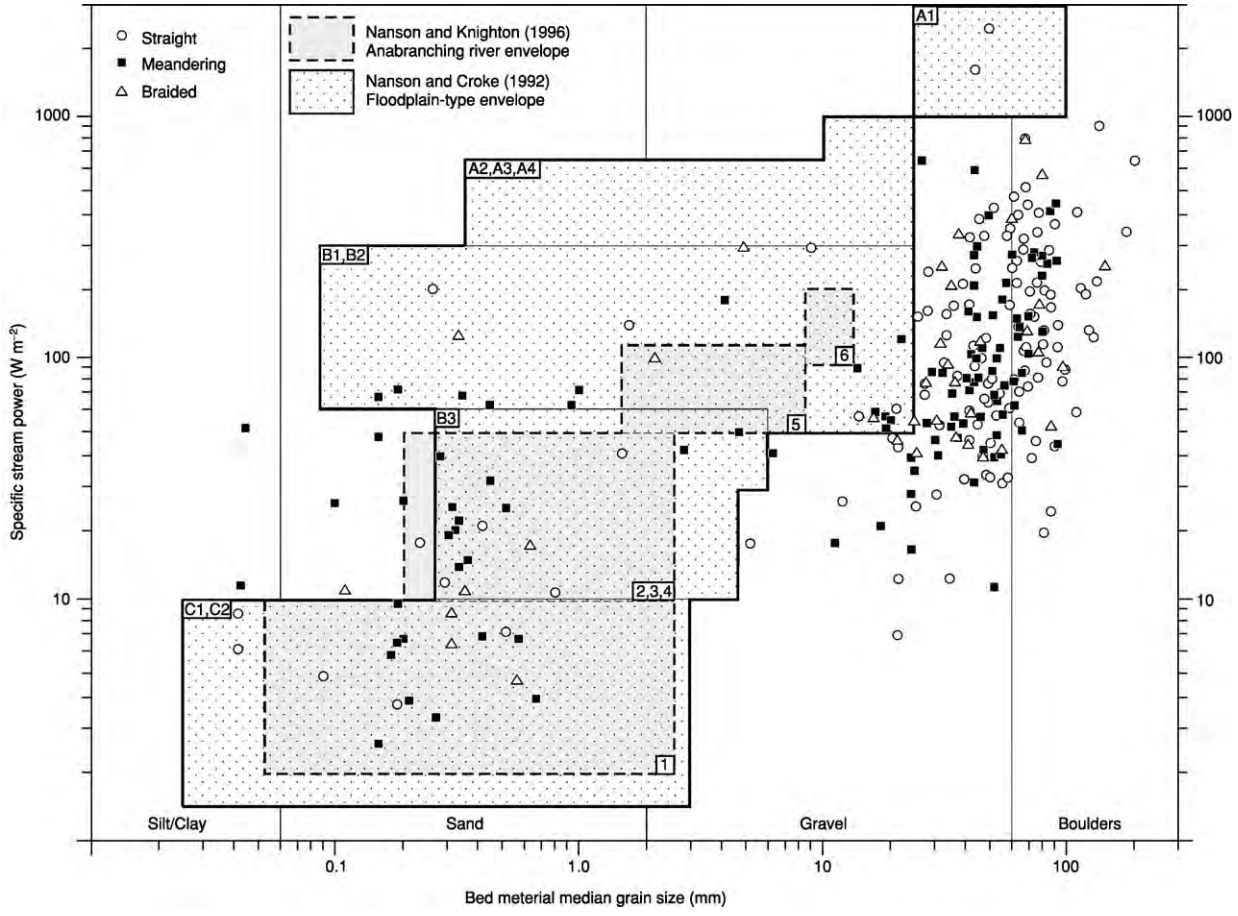


Figure 9 Floodplain types and channel patterns on a bed grain size and unit stream power plot. The envelope boxes overlap and show both the floodplain and the anabranching pattern case-study-based types. Many of the channel-type symbols plot in areas outside the floodplain types, where there may be limited floodplain development, though in part this also reflects a bias in available data towards smaller, steeper streams. The alphanumeric codes and the numbers correlate to the types shown in **Figure 7A** and **Table 2**.

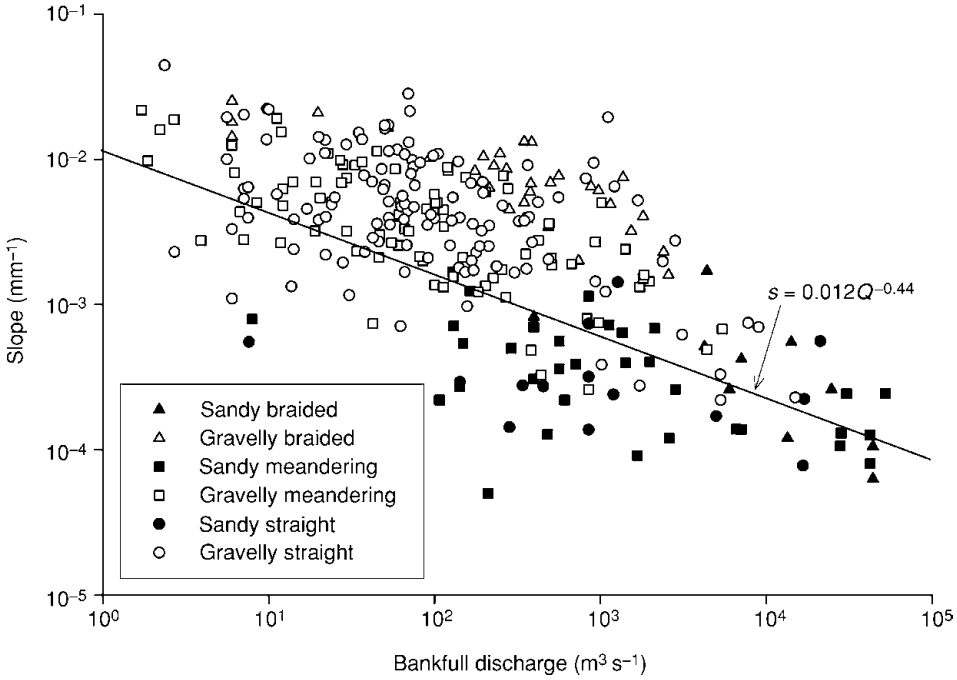


Figure 10 Channel pattern types on a plot of gradient and bankfull discharge in the manner proposed in 1957 by Leopold and Wolman. Their original, but smaller, dataset showed better discrimination, especially between meandering and braided channels. On this plot, sandy channels occur mainly on larger rivers at lower gradients; the threshold between braiding and meandering is higher (and a little more distinct) for gravel bed channels than for sand bed channels. The original Leopold and Wolman threshold is plotted here.

variation or localized tectonic activity leading to gradient modification. But in general, stream gradients decrease and discharge increases with increasing catchment area; stream power (as the product of gradient and discharge) may peak in mid-catchment. Bed grain size decreases downchannel, though this is in practice complicated by factors such as tributary sediment inputs and the existence of step changes in grain size with an abrupt gravel/sand transition. Morphological trends are also apparent (Figure 11B). In headwaters, eroding channels primarily transport hillslope and bedrock erosion materials, which may be coarse (as large or larger than the depth of the channel) and unsorted by river activity. Channels take the form of rapids, pools, and boulder jams. Rare flood events can have dramatic and immediate effects (Figure 12). Lower down-valley, the river sorting and abrasion of transported materials and the increasing dominance of storage and alluvial reworking begin to take effect. Valleys become wider and hillslopes decouple from the river at their foot. In headwaters, channel widths may occupy the whole valley floor, but there is room lower down-valley for rivers to migrate across an increasing body of alluvial storage, with architectural style changing in sympathy with energy, grain size, and other factors. Sediment-transporting events occur more frequently with finer sediment and steadier discharge, though channels may become more fixed in position. In practice, also, the widest valleys may relate to subsiding tectonically defined sedimentation zones in which very thick alluvial (and other) sediments may accumulate, and rivers occupy or migrate across only a fraction of wide valley floors. Here sediment build-up along restricted alluvial ridges allows the periodic relocation of channel belts by avulsion.

Developments in Time

At different scales, it is possible for landforms to achieve quasi-equilibrium with formative factors, exhibit perturbed states in which forms respond and then recover from events/periods of high activity, or evolve unidirectionally towards end-states (in which a highly-organized landscape is reduced to a low-energy and low-relief one following removal of rock mass). These states have been variously approached – initially via qualitative and theoretical proposals (such as WM Davis' influential 'cycle of erosion' proposed more than a century ago, in which uplift initiated drainage development and relief reduction to

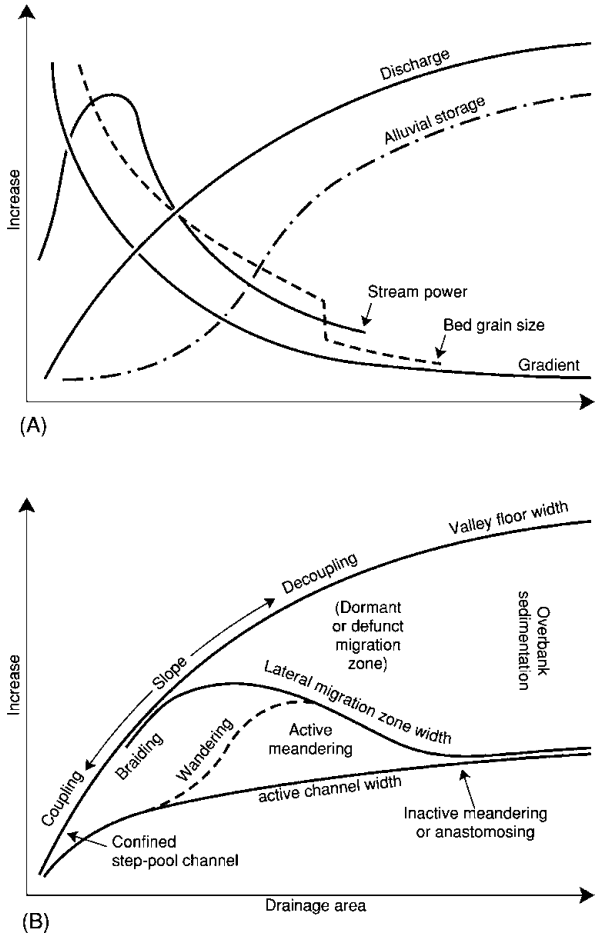


Figure 11 Trends in river catchments (A) in hydraulic factors, based on concepts developed by SA Schumm (in 1977) and M Church (in 2002), and (B) in channel types and activities.

achieve an end-state 'peneplain'). More recently, quantitative modelling has allowed exploration of such developments. Figure 13 illustrates models of drainage evolution; the results can be compared with patterns achieved over very long time-scales in natural landscapes. Such studies may use 'process laws' (based on empirical and/or physically justified relationships) applied iteratively to construct evolving forms. These may be shown to achieve stable pattern states during experimental runs, even though the prediction of individual forms may not be possible (as in the case of individual river bend form and location).

Though unidirectional trajectories in forms have been proposed for higher level landforms, at intermediate scales there has been great concern with perturbations produced by episodic changes in environments (response to a rise or fall in sea-level, tectonic effects, or climate fluctuations that may alter river discharges, sediments, and vegetation cover). Such 'external' changes have also been produced by human agency (including the effects of deforestation and agriculture, mineral extraction, urbanization, and river impounding and channelization). Perturbations over decades to centuries have produced effects that have been closely studied, and these may be ameliorated by good management.

Fluctuations in the rate of catchment erosion and local rates of sediment transfer can result in phases of sediment accumulation and storage, but also of river incision and stored sediment removal. Enhanced delivery may follow climatic change, tectonic uplift, or

Figure 12 Results of an extreme flood in Big Thompson Canyon, Colorado, USA, in July 1976. Photograph by WC Bradley.

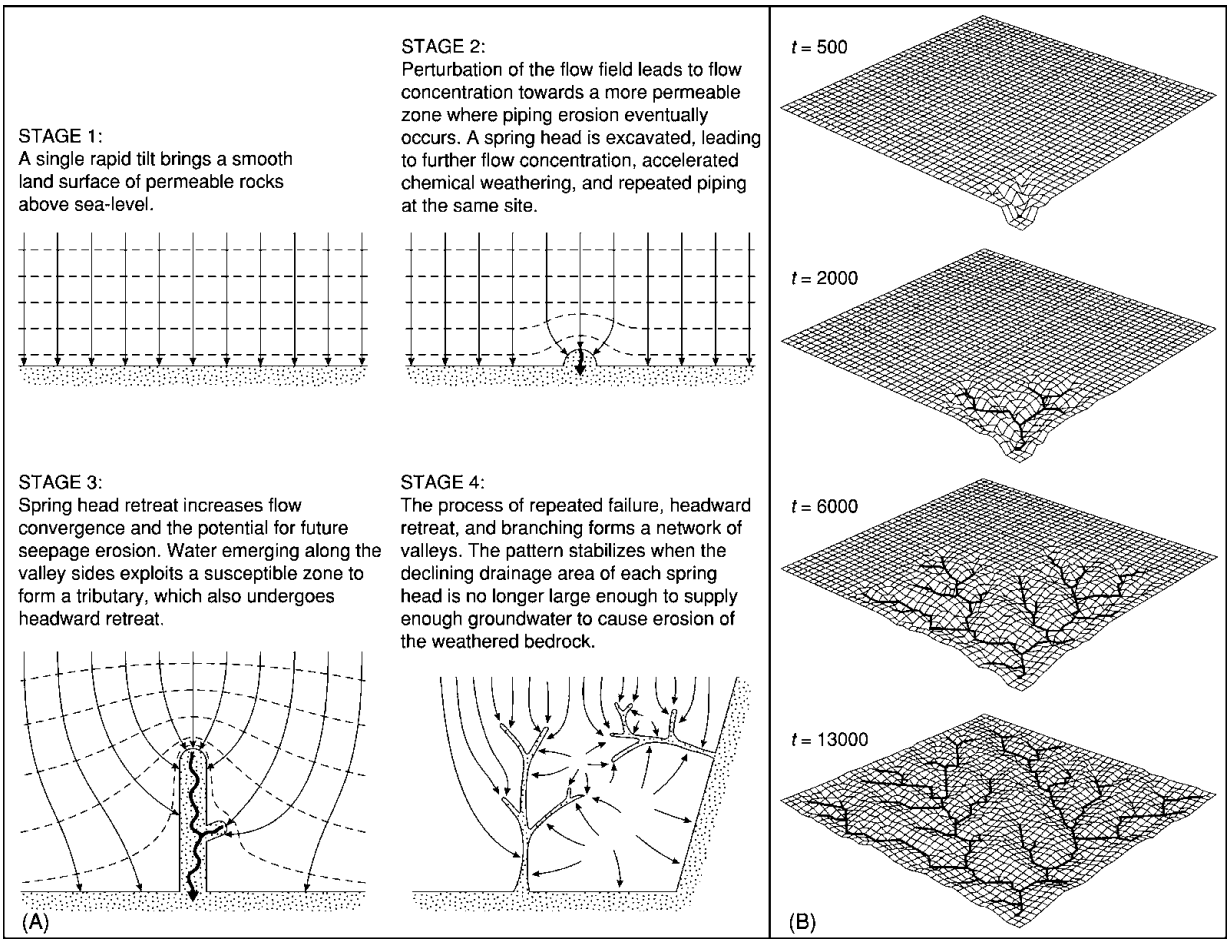


Figure 13 Models of river network development by (A) seepage erosion and (B) a numerical simulation of headward growth. (A) Reproduced with permission from Dunne T (1980) Formation and controls of channel networks. *Progress in Physical Geography* 4: 211–239. (B) Reproduced with permission from Willgoose GR, Bras RL, and Rodriguez-Iturbe I (1991) Results from a new model of river basin evolution. *Earth Surface Processes and Landforms* 16: 237–254.

human activity over a range of time-scales. Channel incision into alluvial deposits or planar bedrock valley floors may produce terraces and inset channels and floodplains. Incision episodes have been related to periods of decreased sediment supply from upstream, to tectonic river steepening, and to sea-level fall. Characteristically, channel steepening tends to propagate up-valley, with a knickpoint or zone moving headward at the expense of the former low-gradient valley floor or floodplain. Downcutting streams can produce terraces at a variety of levels as they migrate ('unpaired terraces'), whereas episodes associated with the occurrence of forcing events may give 'paired' or regionally developed terraces that can be matched over long distances. Major research efforts have been devoted to relating Quaternary climatic fluctuations (as in glacial–interglacial cycles) to the flights of terraces associated with many mid-latitude rivers, and to the relationships with cyclical sea-level change over rather longer time-scales (see Figure 14). Explanatory models have gradually become more complex as the interrelationships between climatic and vegetation change, sea-level fluctuation and tectonics, and river channel activity have been explored.

Studying the effects of environmental change, whether human or natural, has been greatly aided by improving dating techniques for alluvial sediments and by the availability of historical maps and remote-sensing imagery. Episodes of activity can be pinpointed over appropriate time-scales so that, for example, episodes of Holocene and to some extent Pleistocene river activity can be reasonably closely identified. However, alluvial sedimentation is seldom continuous; rivers also rework their own materials to the extent that what is left is only a 'partial archive of remnants' that leaves much activity unrecorded. Nevertheless, evidence is becoming available to allow change/evolutionary models to be compared not simply with static patterns (for example, drainage networks or meander trains) but

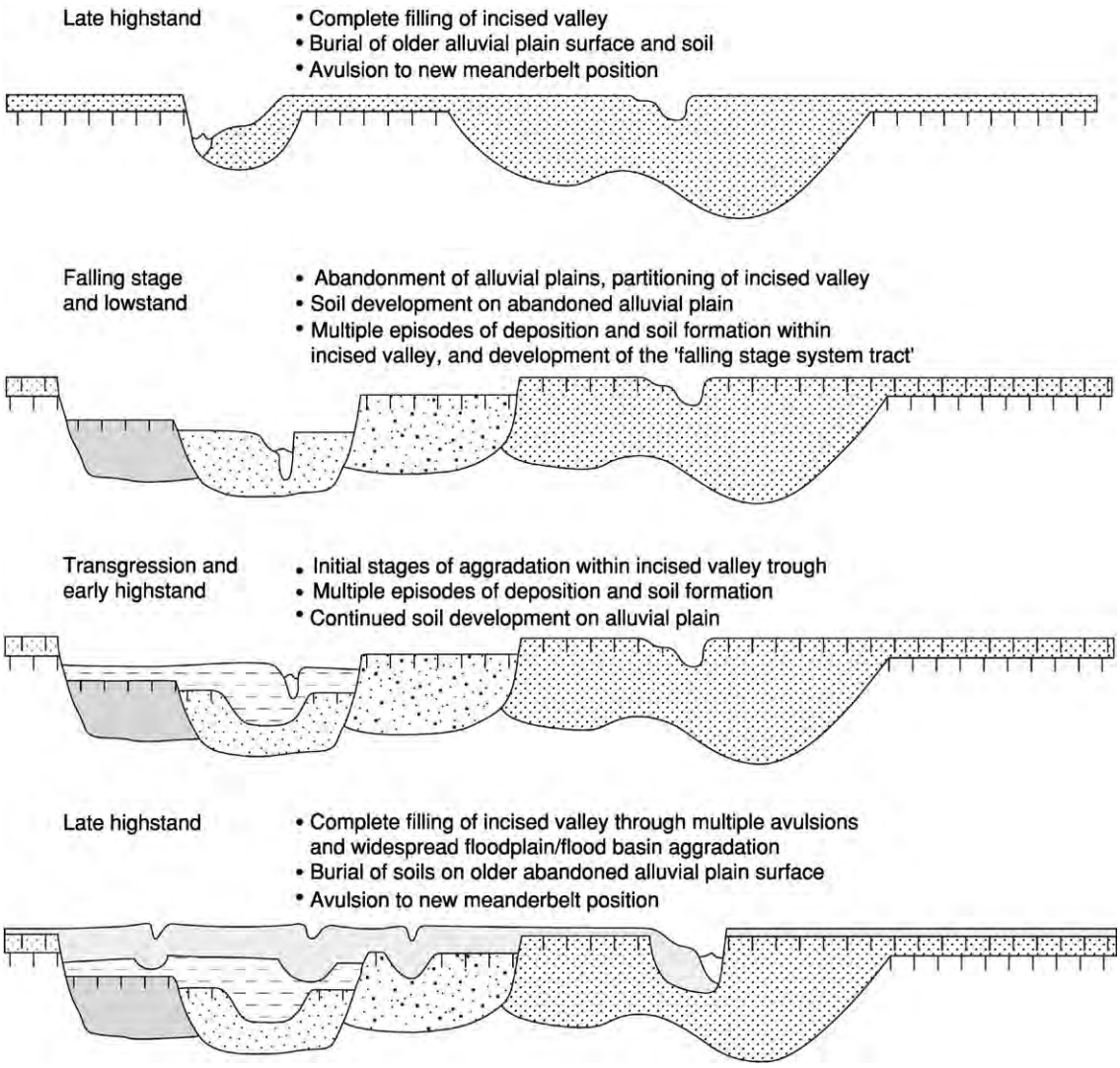


Figure 14 MD Blum and DM Price's 1998 conceptual model of river incision and aggradation in relation to a cycle of sea-level change. Reproduced with permission from Bridge JS (2003) *Rivers and Floodplains*. Blackwell: Oxford.

also with development evidence (as in the case of dated alluvial materials in terraces).

See Also

Europe: Holocene. **Geomorphology**. **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings. **Sedimentary Processes:** Depositional Sedimentary Structures; Catastrophic Floods; Karst and Palaeokarst; Landslides. **Tertiary To Present:** Pleistocene and The Ice Age. **Weathering**.

Further Reading

Bridge JS (2003) *Rivers and Floodplains*. Blackwell: Oxford.

Brown AG and Quine TA (eds.) (1999) *Fluvial Processes and Environmental Change*. Chichester: Wiley.

Church M (2002) Geomorphic thresholds in riverine landscapes. *Freshwater Biology* 47: 541–557.

Dunne T (1980) Formation and controls of channel networks. *Progress in Physical Geography* 4: 211–239.

Gomez B and Church M (1989) An assessment of bedload sediment transport formulae for gravel bed rivers. *Water Resources Research* 25: 1161–1186.

Hjulström F (1935) Studies of the morphological activities of rivers as illustrated by the River Fyris. *Bulletin of the Geological Institute University of Uppsala* 25: 221–527.

Knighton D (1998) *Fluvial Forms and Processes*. London: Arnold.

Leopold LB and Wolman MG (1957) River channel patterns – braided, meandering and straight. *United States Geological Survey Professional Paper* 282B: 39–85.

Lewin J (1989) Floods in fluvial geomorphology. In: Bevan KJ and Carling PA (eds.) *Floods: Hydrological, Sedimentological and Geomorphological Implications*, pp. 265–284. Chichester: Wiley.
Lewin J and Brewer PA (2002) Laboratory simulation of clast abrasion. *Earth Surface Processes and Landforms* 27: 145–164.
Meybeck M (1976) Total mineral dissolved transport by world major rivers. *Hydrological Sciences Bulletin, International Association of Scientific Hydrology* 21: 265–284.
Milliman JD and Meade RH (1983) World-wide delivery of sediment to the oceans. *Journal of Geology* 91: 1–21.
Nanson GC and Croke JC (1992) A genetic classification of floodplains. *Geomorphology* 4: 459–486.
Nanson GC and Knighton AD (1996) Anabranching rivers: their cause, character and classification. *Earth Surface Processes and Landforms* 21: 217–239.
Petts GE, Möller H, and Roux AL (eds.) (1989) *Historical Change of Large Alluvial Rivers: Western Europe*. Chichester: Wiley.
Schumm SA (1977) *The Fluvial System*. New York: Wiley.
Thorne CR, Hey RD, and Newson MD (eds.) (1997) *Applied Fluvial Geomorphology for River Engineering and Management*. Chichester: Wiley.
Willgoose GR, Bras RL, and Rodriguez-Iturbe I (1991) Results from a new model of river basin evolution. *Earth Surface Processes and Landforms* 16: 237–254.
Williams GP (1983) Palaeohydrological methods and some examples from Swedish fluvial environments. *Geografiska Annaler* 65A: 227–244.

Glaciers

M J Hambrey and N F Glasser, University of Wales, Aberystwyth, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Glaciers are amongst the most powerful geological agents on the Earth's surface. They are not only responsible for carving out some of the world's finest landscapes, but also have left widespread deposits that influence agriculture and groundwater flow and provide abundant reserves of sand and gravel for the construction industry. At present, approximately 10% of the Earth's land surface is covered by glacier ice, whilst during Pleistocene glaciations the figure was at least 30%. Continental-scale glaciations occurred in Permo-Carboniferous, Ordovician–Silurian, Neoproterozoic, and Palaeoproterozoic times. The greatest of these glaciations was during the Neoproterozoic, and this has given rise to the concept of the 'snowball Earth', with some geologists arguing for a near-total freeze-up of the planet, followed by severe greenhouse warming, which might have triggered the explosion of life. When the idea of a great Ice Age was first mooted by the Swiss natural historian Louis Agassiz in 1837 (*see* **Famous Geologists:** Agassiz) it met with fierce opposition. In the following decades, increasing numbers of geologists were won over to the concept, as the evidence for glaciation, in the form of characteristic landscapes and extensive surficial deposits ('drift'), throughout Europe and North America became irrefutable. Towards the latter part of the nineteenth century, glacial deposits were being discovered in rocks as old as the Precambrian, although the validity of a glacial interpretation of rocks of Neoproterozoic age was challenged by many until as recently as the 1970s. It was only after glaciologists began examining the processes of erosion and deposition in modern glaciers, from the late 1960s onwards, that the older records of glaciation could be evaluated systematically.

Global Distribution of Glaciers

The global extent of glacier ice has been documented by the World Glacier Monitoring Service (Table 1). The two largest ice-masses, Greenland (10.9%) and Antarctica (85.7%), account for the bulk of glacier ice on Earth, although it is the remaining 3.4% that impinges directly on human activity (Figure 1). The remaining ice-masses occur throughout the Arctic islands and on high mountains, even in the tropics. It is the major ice-masses that are most likely to

Table 1 Distribution of glaciated areas of the world (data from the World Glacier Monitoring Service, 1989)

Region	*Area (km^2)*
Africa	10
Antarctica	13 593 310
Asia and Eastern Europe	185 211
Australasia (i.e. New Zealand)	860
Europe (Western)	53 967
Greenland	1 726 400
North America excluding Greenland	276 100
South America	25 908
World total	15 861 766

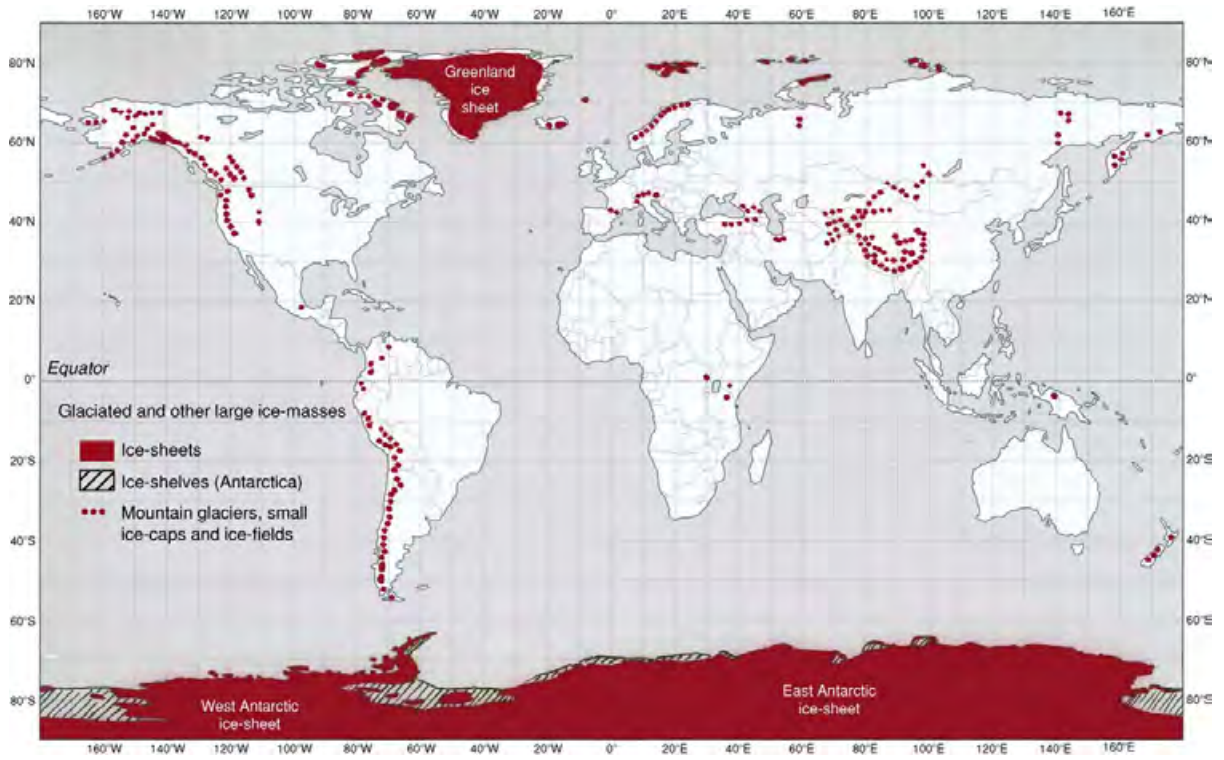


Figure 1 The present-day distribution of glaciers; locations are based in part on *Satellite Image Atlas of Glaciers of the World* (US Geological Survey, (2002) Satellite Image Atlas of Glaciers of the World).

contribute to sea-level rise over the next 100 years as the climate warms. Published figures for the potential contribution of glaciers to sea-level rise have large errors, but Table 2 gives an indication of the relative importance of the different ice-masses. The polar ice-sheets will continue to grow and decay in different places at the same time, until the increase in temperature exceeds the threshold at which they can remain in equilibrium.

Table 2 Potential contribution of the world's ice-masses to sea-level rise (US Geological Survey website, 2001); note that there are considerable uncertainties in the accuracy of these estimates

Ice-masses	*Sea-level rise (m)*
West Antarctic Ice Sheet	8
East Antarctic Ice Sheet	65
Antarctic Peninsula	0.5
Greenland Ice Sheet	6.5
Mountain glaciers	0.5

Characteristics of Glaciers and Ice-Sheets

Glacier Morphology

Glaciers range in size from the great ice-sheets of Greenland and Antarctica to tiny ice-masses just a few hundred metres across. The generally accepted definition of an ice-sheet is an ice-mass exceeding 50 000 km^2 in area. The Greenland Ice Sheet forms an elongated dome, with mountains projecting through it near the periphery, whilst the Antarctic Ice Sheet consists of three dynamically separate regions, of which the dome of East Antarctica has by far the greatest volume (Figure 2). Ice-caps are also dome-like masses, but measure less than 50 000 km^2 (Figure 3). Both ice-sheets and ice-caps discharge the bulk of their ice via ice-streams – zones of accelerated flow with slow-moving ice on either side. As ice-streams enter the sea, coalescence leads to the formation of ice-shelves – floating slabs of ice several hundred metres thick. Like ice-caps, highland ice-fields also bury much of the topography, but mountains project through them as 'nunataks'. Valley glaciers are typical of mountain regions, such as the Alps, and are the best studied of all ice-masses (Figure 4). Above them are upland amphitheatre-like basins where cirque (or corrie) glaciers accumulate. Lastly, hanging glaciers are perched on steep mountain rock faces and are so prone to avalanches that below them are regenerated glaciers.

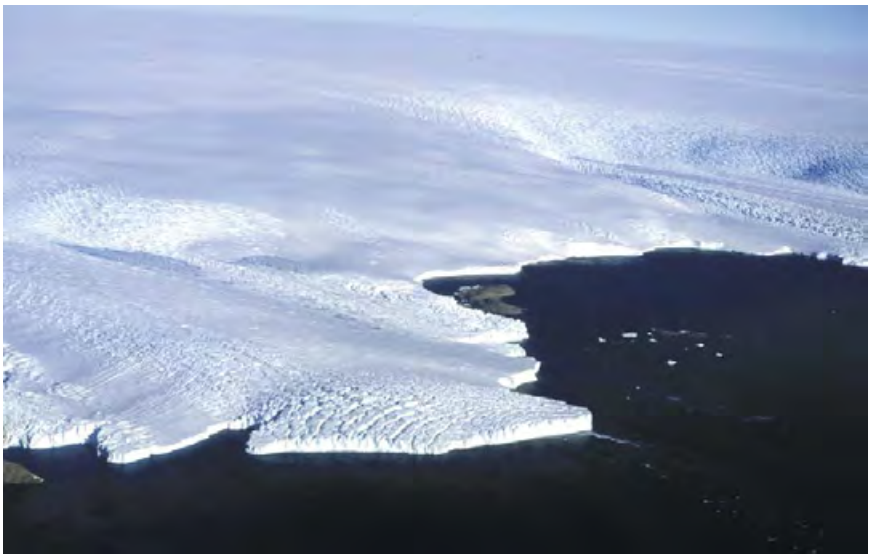


Figure 2 The East Antarctic Ice Sheet near Casey Station, with two small ice-streams (zones of crevassed fast-flowing ice) discharging into the sea. © 2005, Photograph, M J Hambrey.

Figure 3 A small ice-cap on the north-east side of James Ross Island, northern Antarctic Peninsula. © 2005, Photograph, M J Hambrey.

Glacier Mass Balance

Mass balance is the study of the inputs and outputs to the glacial system, i.e. the net gain or loss of snow and ice across a glacier (Figure 5). The inputs to the glacial system are known as accumulation and include all materials that add mass to the glacier, such as snow, refrozen slush, hail, frost, rain, and avalanched snow or ice. The outputs from the system are known as ablation and include anything that removes mass from the glacier, such as direct ice melt, iceberg calving, wind erosion, and sublimation. Mass balance is largely dependent on the prevailing climate since this determines the rates of accumulation and ablation. In general, ablation tends to be dominant in the warm summer months, and accumulation is dominant in the winter months (Figure 5B). The exception to this rule is where iceberg calving occurs throughout the year where an ice-sheet or glacier terminates in water. The rate of growth or recession of a glacier is determined by the difference between the accumulation and ablation rates (called the net mass balance and usually measured over a 1 year period). The point

Figure 4 An alpine valley glacier, the Mer de Glace, in the French Alps. © 2005, Photograph, M J Hambrey.

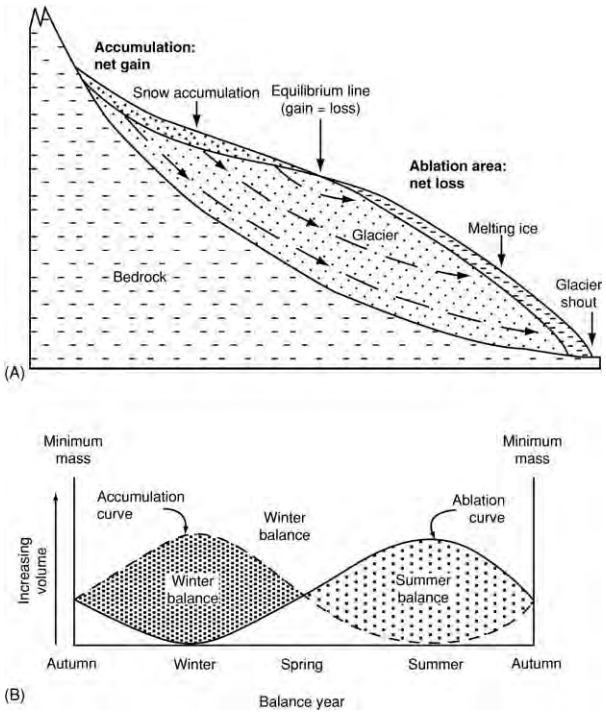


Figure 5 Mass-balance regime of a typical glacier in the temperate regions of the Arctic. (A) Schematic cross-section through the glacier, illustrating zones of accumulation and ablation. (B) How the volumes of accumulation and ablation vary over the year, assuming steady-state conditions.

on a glacier where accumulation equals ablation is known as the equilibrium line and is normally expressed as an altitude in metres above sea-level. Accumulation exceeds ablation in the upper part of a glacier, whereas the converse is true in the lower part (Figure 5A). Exceptions to this rule occur in parts of Antarctica, where costal snowfall compensates for losses in the middle (drier) reaches of a glacier.

Glacier Thermal Regime

Temperatures within glacier ice vary both vertically and horizontally. Variations in temperature at the base of a glacier are known as the basal thermal regime and are determined by the balance between the heat generated at the base of the glacier and the temperature gradient within the ice, which governs the rate at which heat is drawn towards the ice surface (Figure 6). The most important factors in determining the basal thermal regime are ice thickness, ice surface temperature, geothermal heat, and frictional heat. Heat is generated beneath a glacier by geothermal heat, which enters the base of the glacier from the

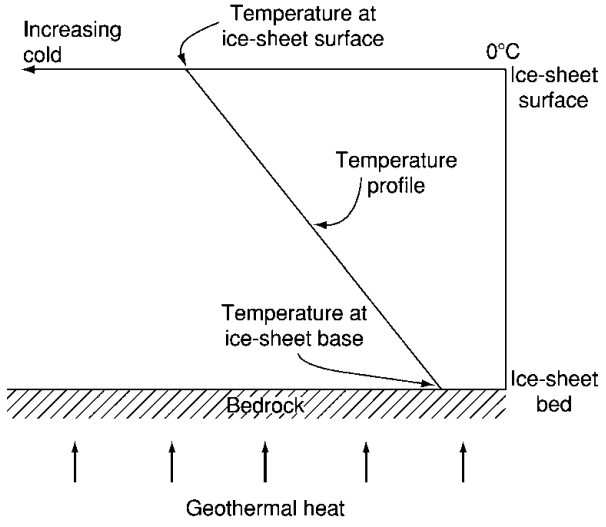


Figure 6 The vertical variation in thermal regime through a polar ice-sheet.

rocks beneath, and by frictional heat, which is given off as a result of basal sliding and internal deformation. The distribution of ice temperature is vital in understanding the landscapes produced by glaciers because the basal thermal regime is a primary factor controlling the patterns of erosion and deposition.

This is because the temperature at the base of the glacier controls which of the flow processes (ice creep, basal sliding, and subglacial deformation) can occur and therefore the dynamics of the glacier.

There are three fundamental types of glacier: cold-based, warm-based, and polythermal. Cold-based glaciers are frozen to their beds and produce little basal melt water. Basal sliding cannot occur, and these glaciers flow by internal deformation alone. In contrast, warm-based glaciers produce abundant meltwater that finds its way to the bed; thus basal sliding is an important component of their flow. Warm-based glaciers therefore have the potential for greater flow rates than cold-based glaciers and as a result have a greater ability to modify the underlying landscape. Polythermal glaciers have basal melting in their accumulation zones, basal freezing at their margins, and a transition zone between the two.

Glacier Flow

Glaciers flow primarily because the ice within them deforms under the influence of gravity. Glacier flow is achieved by three mechanisms: internal deformation, basal sliding, and subglacial bed deformation (Figure 7). Internal deformation is achieved by the processes of ice creep, large-scale folding, and faulting. Ice creep is the mutual displacement of ice crystals relative to one another in response to the applied shear stress and results in slow forward motion in the direction of the ice-surface slope. Folding and faulting are common in situations where ice creep alone cannot adjust sufficiently rapidly to the stresses set up within the ice. Basal sliding is the name given to the process whereby a glacier slides over its bed, often on a thin film of lubricating melt water. Sliding is accompanied by a combination of enhanced basal creep and regelation (refreezing) (Figure 8). Subglacial bed deformation is the process by which unfrozen sediment beneath the glacier deforms under the weight of the overlying ice. This occurs when the water pressure in the pore spaces between the sediment grains rises sufficiently to overcome the resistance between individual grains, allowing them to move relative to one another. This is an important process beneath temperate glaciers that overlie poorly consolidated sediments. For example, it has been estimated that 90% of the velocity of Breiðamerkurjökull in Iceland can be accounted for by subglacial bed deformation.

The fastest moving glaciers are typically ice-streams or outlet glaciers terminating in the sea, which may have velocities of several kilometres a year. Mountain-valley glaciers commonly move a few hundred metres a year, whilst small cirque glaciers may move only a few metres a year. Some glaciers alternate between slow-moving ('quiescent') phases and surge phases (when velocities are several orders of magnitude greater than normal). A surge may last for just a few months (as in Alaska) or for several years (as in the high Arctic).

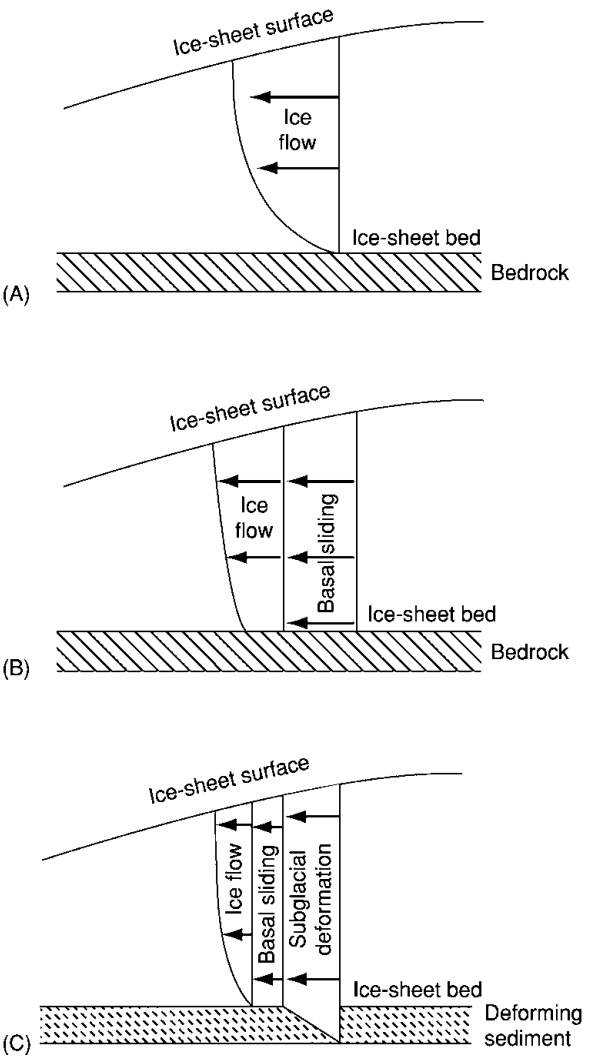


Figure 7 Relative vertical velocity profiles through (A) a cold-based ice-sheet on bedrock, (B) a warm-based ice-sheet on bedrock, and (C) a warm-based ice-sheet on deformable sediment, showing the different components of glacial flow under the different thermal regimes.

Glacier Structure

Glacier ice resembles other geological materials in that it starts as a stratified sequence, in this case mostly of snow, and then undergoes diagenesis (to firn and then to ice) and deformation, producing a wide range of structures. Primary structures include sedimentary stratification, unconformities created in years of

Figure 8 The debris-bearing basal ice layer in a cave beneath Glacier de Tsanfleuron, Switzerland, formed by regelation and subsequently modified by folding. © 2005, Photograph, M J Hambrey.

heavy ablation (Figure 9A), and regelation layering resulting from pressure melting and refreezing at the base of the glacier (Figure 8). Secondary structures are the result of internal deformation and include brittle structures, such as crevasses (Figure 9B), crevasse traces, faults, and thrusts, and ductile structures, typified by foliation (Figure 9C), folding (Figure 9D), and boudinage. Structural development in glaciers can be examined in the context of strain rates (usually measured over a 1 year period) and cumulative strain (estimated from velocities and flow-line maps or by finite-element modelling) over several decades.

Glacial Erosion

Processes of Glacial Erosion

Glacial erosion involves the removal and transport of bedrock or sediment by three main processes: quarrying (also known as plucking), abrasion, and melt water erosion. Quarrying involves two separate processes: the fracturing or crushing of bedrock beneath the glacier, and the entrainment of this fractured or crushed rock. Fracturing of bedrock takes place when a glacier flowing over bedrock creates pressure differences in the underlying rock, causing stress fields that may be sufficient to induce rock fracture. Fluctuations in basal water pressure may help to propagate bedrock fractures beneath a glacier. Evacuation of rock fragments or sediment is possible where localized basal freezing occurs, where the tractive force is sufficient to entrain individual clasts, or where debris is elevated along thrusts.

Abrasion is achieved by bodies of subglacial sediment sliding over bedrock or by individual clasts contained within the ice. The progressive breakdown of material in traction provides a mechanism for the continued replenishment of debris that can carry out erosion beneath ice-masses, even where they flow across hard rocks of uniform lithology. Thus, glacial erosion is most efficient beneath temperate ice. Although cold glaciers may achieve limited amounts of erosion, the release of melt water to lubricate basal sliding beneath temperate ice is a prerequisite for widespread and efficient quarrying and abrasion.

Erosion by glacial melt water beneath glaciers is the result of either mechanical or chemical processes. The effectiveness of melt water as an agent of erosion depends on the susceptibility of the bedrock involved (in particular the presence of structural weaknesses or susceptibility to chemical attack), the discharge regime (in particular the water velocity and the level of turbulent flow), and the quantity of the sediment in transport.

Landforms of Glacial Erosion

The landforms produced by glacial erosion are commonly described on three spatial scales: the microscale, the mesoscale, and the macroscale.

Microscale landforms of glacial erosion are those that are below 1 m in size. They are of low-relief amplitude and are often found superimposed on larger landforms. Striae or striations (Figure 10A) are lines or scratches on a rock surface that are produced by glacial abrasion as debris or clasts entrained

Figure 9 Structures in glacier ice: (A) stratification and unconformity in ice cliffs bordering Granite Harbour, Ross Sea, Antarctica; (B) crevasses in Glacier des Bossons, French Alps; (C) foliation in Gornergletscher, Swiss Alps, with the Matterhorn in the background; and (D) 'similar-style' folding associated with foliation in Griesgletscher, Swiss Alps. © 2005, Photographs, M J Hambrey.

in the basal ice are dragged over the rock. Microscale crag-and-tails are small elongated bedrock protuberances that form where the rock is preferentially protected from glacial abrasion in the lee of resistant grains or mineral crystals on the surface. Crescentic gouges (Figure 10B) and chattermarks are small cracks, gouges, and indentations created on bedrock surfaces. A variety of different morphologies have been recorded, varying in shape from simple fractures or cracks in the bedrock to larger crescentic gouges in which small chips of bedrock have been physically removed from the bed. Finally, p-forms, s-forms, and microchannel networks include a variety of smooth sinuous depressions and grooves sculpted on glaciated bedrock surfaces (Figure 10C).

Mesoscale landforms of glacial erosion are between 1 m and 1 km in size. Streamlined bedrock features (also known as whalebacks or rock drumlins) are bedrock knolls that have been smoothed and rounded on all sides by glacial action. Stoss-and-lee forms are bedrock features with a pronounced asymmetrical profile: they have an abraded slope on the up-glacier or stoss side and a steeper rougher quarried slope on the down-glacier or lee side. Individual landforms

Figure 10 Microscale erosional landforms: (A) striated and ice-moulded limestone at the margin of Glacier de Tsanfleuron, Swiss Alps; (B) crescentic gouges in dolerite beneath Neogene glacigenic sediments, Roberts Massif, Antarctica (ice flow was from left to right); and (C) p-forms (melt water-eroded and glacially striated channels) cut in limestone, Glacier de Tsanfleuron, Swiss Alps. © 2005, Photographs, M J Hambrey.

are commonly known as *roches moutonnées* (Figure 11A). Rock grooves are similar in morphology to striae, but have greater size and depth (Figure 11B). They range from tens of metres to hundreds of metres in length and may be up to several metres wide and a few metres deep. Rock basins are individual depressions carved in the bedrock, often developed along lines of structural weakness. Melt-water channels span a large spatial scale from centimetre- or metre-scale channels cut in soft sediment or bedrock to large valley-scale melt-water channel systems. Landforms associated with glacial melt water form wherever sufficient quantities of water are released in subglacial, ice-marginal, and proglacial settings.

Macroscale landforms of glacial erosion are those features that are at least 1 km in size. They are significant landscape components, which may contain within them many of the smaller landforms described above. Zones of areal scouring are scoured bedrock consisting of an assemblage of streamlined bedrock features, rock basins, and stoss-and-lee forms. Glacial troughs are deep linear features carved into the bedrock, which represent the effects of glacial erosion where ice flow has been confined by topography and channelled along a valley or fjord (Figures 12A and 12B). Cirques are large bedrock hollows that open down-slope and are bounded up-slope by a cliff or steep slope known as a headwall (Figure 12C). Headwall erosion that continues from two opposing sides until they meet produces a sharp ridge – an arête (Figure 12D). If the process takes place from three or more sides, a sharp isolated peak – a horn – may be the product (Figure 12E). Large stoss-and-lee forms (giant *roches moutonnées*) are between hundreds of metres and several kilometres in size. These valley spurs and hills have a pronounced asymmetric profile, with an abraded slope on the up-glacier (stoss) side and a steeper rougher quarried slope on the down-glacier (lee) side.

Bathymetric Forms Resulting from Glacial Erosion

Various erosional phenomena, mainly associated with grounded ice or subglacial melt water, are found in marine settings. The larger-scale forms are filled by sediment and may be recognizable only in

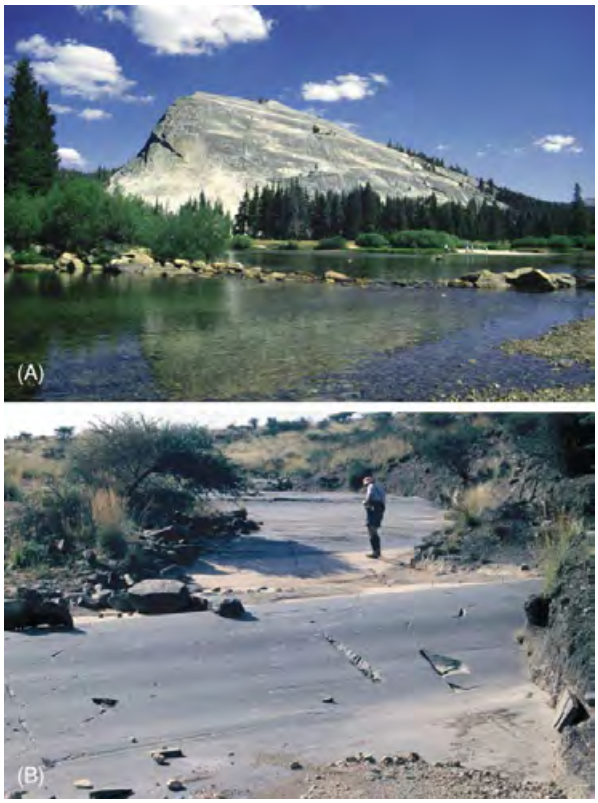


Figure 11 Mesoscale landforms of glacial erosion: (A) *roche moutonnée*, the Lembert Dome, Yosemite, California, USA; (B) shallow striated grooves from the Permo-Carboniferous glaciation, near Douglas, Karoo, South Africa. © 2005, Photographs, M J Hambrey.

seismic profiles. Submarine troughs formed by ice-streams are found on continental shelves and are genetically equivalent to fjords and other glacial troughs but are generally much broader. The largest occur in Antarctica, where they can reach over 400 km in length, 200 km in width, and 1100 m in depth. Steep-sided channels 2–3 km wide, over 100 m deep, and up to 100 km long, carved out by subglacial melt water and subsequently filled by sediment, are known as tunnel valleys. These are well known from the north-west European continental shelf around Britain, the Scotian Shelf off Canada, and Antarctica. Icebergs can also cause considerable erosion if they become grounded on the seafloor. Large tabular icebergs can scour the bed of the sea for several tens of kilometres, leaving impressions up to 100 m wide and several metres deep.

Debris Entrainment and Transport

Glacial erosion and subaerial erosion above a glacier produce a range of sediment types that can be entrained and transported in a variety of ways. The manner in which sediment is transferred through the glacial system is related to the dynamic and thermal characteristics of the glacier. Broadly speaking, polythermal glaciers tend to carry a high basal debris load but little on their surfaces, whereas the reverse is true for temperate glaciers.

As noted above, entrainment of debris at the bed is a product of regelation. Basally derived debris is subject to comminution at the ice–bedrock interface and typically consists of clasts up to boulder size, with subangular and subrounded shapes and faceted and striated surfaces. In addition, there are varying proportions of sand, silt, and clay. Debris on the surface of the glacier (supraglacial debris) is mainly derived from frost-induced rock fall from the crags above and is mainly angular with few fines.

During flow, both supraglacial and basal debris is reorganized. In addition to regelation, the following mechanisms, some of which are illustrated in Figure 13, have been identified

- incorporation of angular rock-fall material within the stratified sequence of snow, firn, and refrozen slush, followed by folding (this debris emerges near the snout at the hinges and upper limbs of the folds, forming medial moraines);
- falling or washing of supraglacial debris into crevasses and moulins;
- incorporation of both supraglacial and basal debris within longitudinal foliation;
- thrusting, allowing debris-rich basal ice and subglacial sediments to be uplifted into an englacial or even supraglacial position; and
- squeezing of subglacial debris into basal crevasses.

Glacial Deposition

Processes of Glacial Deposition

Glacial deposition involves the release of sediment from supraglacial, englacial, and basal-glacial positions. Debris that has been in contact with the bed is heavily modified by intraclast collision, subglacial deformation, and contact with water. Debris that follows a passive transport path (supraglacially or englacially) tends to retain its original characteristics.

Sediment may be deposited directly beneath the glacier or at its margins, or it can be transported long distances away from the glacier by other agents, such as rivers (producing outwash plains) or icebergs. During release from the ice, sediment is reworked by numerous processes, including fluvial processes, debris-flowage, marine or lacustrine currents, and aeolian processes (Figure 14A). The resulting sedimentary facies associations are therefore exceedingly complex, but particular combinations of facies can be

Figure 12 Macroscale glacial erosional landforms: (A) Yosemite glacial valley, California, USA; (B) Kejser Franz Josef Fjord, central east Greenland; (C) a pair of deeply incised cirques on Cuernos del Paine, Chilean Patagonia; (D) arêtes carved in granite peaks, Stauning Alper, East Greenland; and (E) the horn-like peak of Ama Dablam, Khumbu Himal, Nepal. © 2005, Photographs, M J Hambrey.

used to determine the glacial thermal regime and tectonic setting. A temperate glacial system in an area of high relief, especially if it is tectonically active, will have a high proportion of supraglacial debris that on release is modified by streams; little sediment of basal character will be preserved. A polythermal glacier tends to produce a mainly basal glacial deposit, only part of which may be reworked by outlet streams.

Glaciers that enter water bodies behave differently according to their thermal regime. Polythermal glaciers tend to have a floating tongue if the water is deep enough, as in the fjords of Greenland, and the dominant facies produced is poorly sorted sediment ('diamicton') derived from the base of the glacier. The ice in temperate glaciers is usually too weak to float, and the glacial terminus is a cliff resting on the seabed. Sediments produced in these situations, typified by the fjords of Alaska, are dominated by subglacial water output (**Figure 14B**). For the coldest glaciers, terminating as ice-shelves on the continental shelf of Antarctica, direct deposition is restricted to a narrow grounding zone because freeze-on of saline ice at the base of the ice-shelf seals in any remaining sediment. Melt-water production in the grounding zone is generally limited, and the dominant sediment

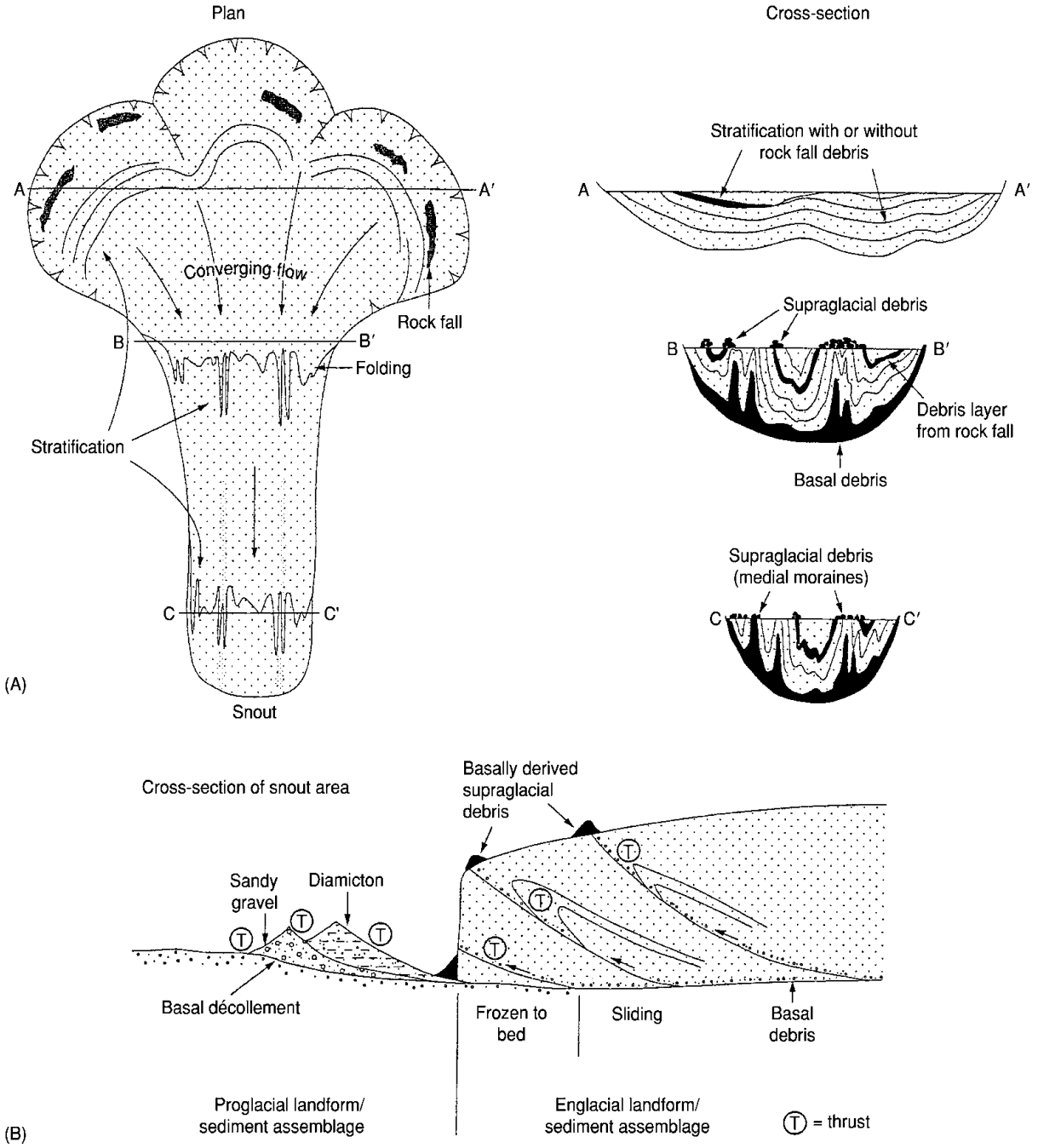


Figure 13 Conceptual models of debris entrainment by (A) folding and (B) thrusting, based on glaciers with multiple accumulation basins in Svalbard. (Reproduced from Hambrey MJ and Glasser NF (2003) Glacial; sediments: processes, environments and facies. In: Middleton GV (ed.) *Encyclopedia of Sediments and Sedimentary Rocks*, pp. 316–331. Dordrecht: Kluwer.)

immediately beyond this area is biogenic and is typically derived from the settling out and current transfer of diatom remains (Figure 14C). This sediment is carried away from the Antarctic coast within icebergs, and some may be released onto the ocean floor hundreds of kilometres from the source.

Glaciotectonic deformation is a widespread phenomenon that affects materials for up to several hundred metres adjacent to and beneath glaciers. Processes such as thrusting and folding within the ice are mirrored by the same processes in subglacial and proglacial settings. All types of material are affected, including wet, dry, and frozen sediments and bedrock. Deformation may detach blocks of rock and frozen sediment that are sometimes hundreds of metres across, leaving depressions in the

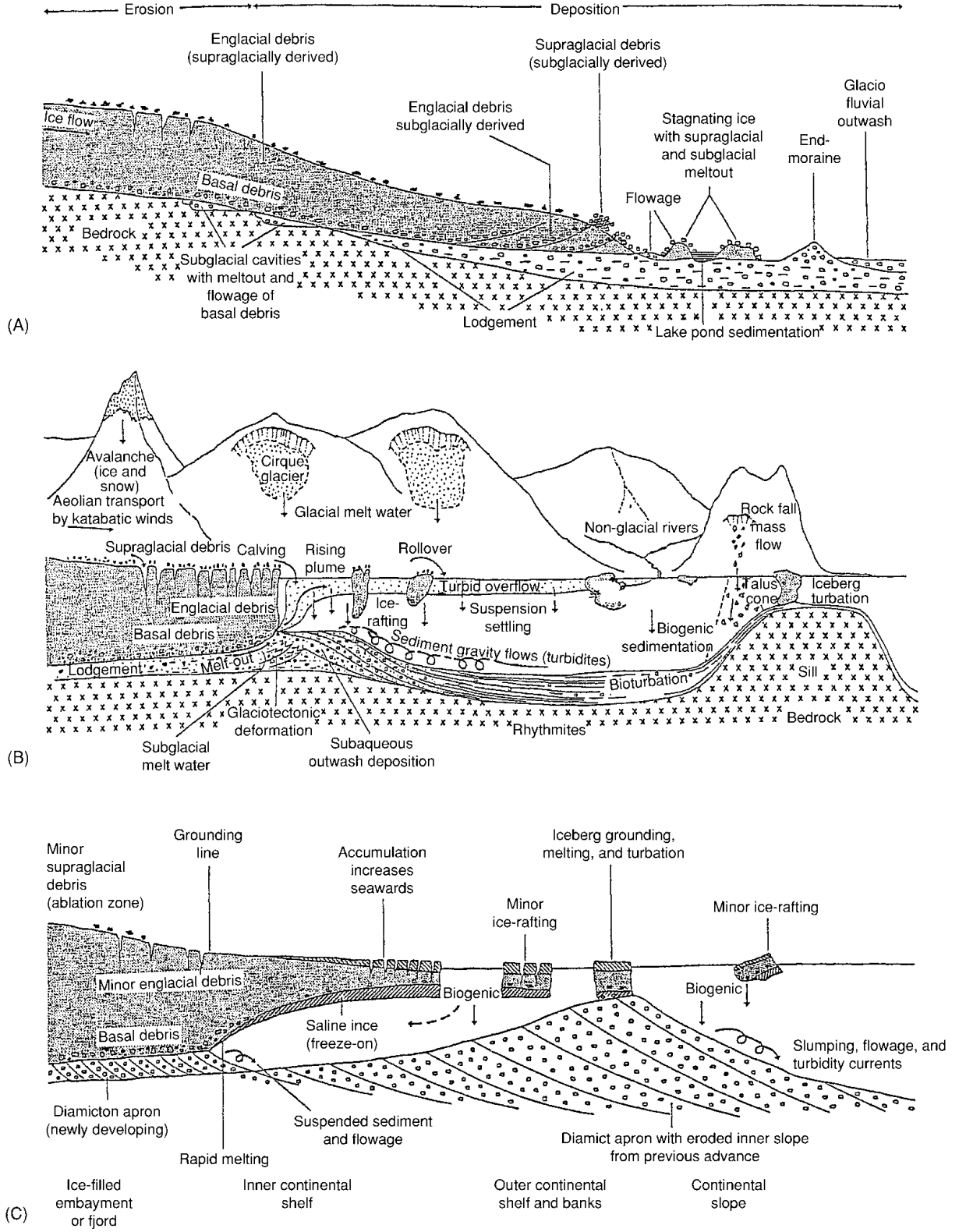


Figure 14 Schematic cross-sections through glaciers showing deposition in different geographical and climatic settings: (A) a terrestrial temperate polythermal glacier; (B) a temperate tidewater glacier in a fjord; and (C) an ice-shelf and continental shelf (Antarctica). (Reproduced with permission from Hambrey MJ (1999) The record of Earth's glacial history during the last 3000 Ma. In: Barrett PJ and Orombelli G (eds.) *Geological Records of Global and Planetary Changes*, pp. 73–107. Terra Antartica Report 3. Siena: Museo Nazionale dell'Antartide.) Kluwer Academic Press.

Figure 15 Typical glacigenic facies: (A) a diamicton interpreted as a basal till on the Isle of Arran, Scotland; and (B) sand and gravel interpreted as a glaciofluvial sediment in Rhosesmor Quarry, North Wales (the reddish colour derives from eroded New Red Sandstone in the Irish Sea). © 2005, Photographs, M J Hambrey.

ground and producing complex ridge forms in front of the glacier. Much of the material is brecciated, but, if it is wet and fine grained, it may deform in a ductile manner.

Glacigenic Sediments

A number of general terms have been applied to deposits left behind by glaciers and the associated melt water. 'Glacigenic sediment' means 'sediment of glacial origin' in the broad sense, embracing sediments that have been influenced by glacier ice even if only slightly. 'Glacial drift' refers to all material being transported by glacier ice, all deposits released by glaciers, and all deposits of predominantly glacial origin deposited from glacial melt water or in the sea by icebergs.

Traditionally, Quaternary researchers have tended to give genetic terms to glacigenic sediments in the field, but this has led to difficulties where the meanings of the terms have not been agreed and where the origin of the sediment is disputed. The increasingly preferred approach is to describe sediments objectively using textural terms, and classifications are available for the typical range of poorly sorted sediments that exist in glacial settings. The terms 'diamicton' (unlithified), 'diamictite' (lithified), and 'diamict' (both) are used for 'a non-sorted or poorly sorted terrigenous sediment that contains a wide range of particle sizes'. In other words diamicts contain particles resulting from weathering and erosion ranging in size from clay to boulders in various proportions (Figure 15A). We use a range of 1% to 50% for the proportion of clasts within diamicts (from a detailed textural classification presented elsewhere).

The International Quaternary Association (INQUA) has proposed a genetic classification that, although somewhat dated, embraces the terminology used by most glacial geologists. The principal terms are as follows.

- Lodgement till is deposited by 'plastering' of glacial debris from the sliding base of a moving glacier as a result of pressure melting or other mechanical processes. Texturally, the deposit is a diamicton or a muddy or sandy gravel.
- Meltout till is deposited by the slow release of glacial debris from ice that is not moving. Typically, meltout tills inherit the characteristics of the sediment within the basal ice layer and are diamictons.
- Basal till embraces both the above.
- Sublimation till is released by the sublimation of debris-rich ice in subzero temperatures. Texturally this is diamicton, but it is found only in arid parts of Antarctica.
- Deformation till comprises weak rock or unconsolidated sediment that has been deformed or disaggregated, with some foreign material admixed. Texturally this is variable and often breccia-like; rafts of rock or sediment may be many metres in length.
- Flow till is till that has been remobilized by gravity flow, although it is better known as glacigenic sediment flow.

The above INQUA classification refers to modern terrestrial glacigenic sediments. The lithified equivalent of till is 'tillite'. All these deposits are prone to reworking by melt water and wind. Thus extensive well-sorted sand and gravel deposits are a major feature of glacial environments (Figure 15B). There is no agreed genetic classification for glaciomarine sediments. The following terms, however, are commonly used.

- Ice-proximal glaciomarine sediment is sediment that is released either from floating basal glacier

ice or by continuous rainout from icebergs, without subsequent winnowing by currents and waves. Texturally this is normally a diamicton.

- Ice-distal glaciomarine sediment is sediment that is released from icebergs, but is subject to winnowing and admixed with other marine sediment, including biogenic components. Texturally these sediments are usually sandy muds or muddy sands with dispersed stones, and in modern environments they are often rich in diatoms.
- Iceberg turbate is sediment that is deposited on the seafloor and subsequently reworked by grounded icebergs. The end product is commonly a massive diamicton.
- Cyclopels (silt–mud couplets) and cyclopsams (graded sand–mud couplets) are rhythmically laminated sediments derived from the turbid overflow plumes of subglacial discharge, especially in fjords. Iceberg rafting may produce dropstone structures.

Glaciolacustrine facies are similar to those in glaciomarine environments. In addition, lakes commonly have rhythmically laminated sands and muds called varves (or varvites if lithified). Sand laminae represent input to the lake during summer, and silt or clay laminae represent settling out from suspension in winter. Dropstones may disrupt the laminae. Other laminated sediments in lakes or the sea may be deposited from turbidity currents and also include dropstone structures.

Depositional Landforms

Glaciers and ice-sheets produce a large range of depositional landforms that are commonly grouped into ice-marginal and subglacial landforms.

Ice-marginal landforms Ice-marginal landforms are produced by advancing, static, or receding ice-margins, as well as during seasonal fluctuations of an ice-front. They may be deposited directly from glacier ice or be composed of sediments previously deposited by other processes. Ice-marginal landforms are commonly used to reconstruct changes in glacier size, morphology, and extent over time.

Glaciotectonic moraines encompass a broad range of different types of moraine formed by the deformation of ice, sediment, and rock. Of these, the most common are push moraines, which are formed when a glacier flows into sediment and bulldozes material into a ridge.

Dump moraines are formed at stationary or near-stationary ice-fronts, where debris accumulates along the margin or front of the glacier to form a ridge of sediment (Figure 16A).

Ablation moraines (sometimes referred to as ice-cored moraines) form wherever ice-melt is retarded beneath a cover of supraglacial debris. This supraglacial debris may be derived from rock falls and avalanches in the accumulation area, in which case it is often arranged in flow-parallel medial moraines, or it may be composed of subglacial or englacial debris elevated to the ice surface by folding, thrusting, or upward-directed flow lines in the ablation area. Uneven melting gives rise to hummocky moraines (Figure 16B).

Outwash fans and outwash plains (or sandar) are formed where glacial melt water emerges from the glacier and sediment is deposited at or beyond the ice margin, producing extensive braided river tracts composed of sand and gravel. Kettle holes form where sediment is deposited over buried ice.

Kame terraces are formed when sediment is deposited by melt water flowing laterally along an ice-margin. Kames are more fragmentary features, which are formed in a similar manner but often in ice-walled tunnels and against steep valley sides.

Subglacial landforms Subglacial landforms (sometimes referred to as subglacial bedforms) are produced beneath actively flowing ice. The most common of these are a family of ice-moulded landforms, all of which are elongated parallel to the direction of ice flow.

Flutes are low flow-parallel ridges that are rarely continuous for less than 100 m (Figure 16C). Megascale lineations are much larger (tens of kilometres in length and hundreds of metres in width) and are often visible only on satellite imagery. Drumlins are typically smooth oval or elliptical hills that are tens of metres high and long. The origin of drumlins is unclear, but they have been attributed to deposition by actively moving ice and to subglacial sheet floods.

Ribbed moraines (also known as Rogen moraines) are large closely spaced moraine ridges, which are often curved or anastomosing but generally orientated transversely to the direction of ice flow. They may represent the fracturing of a pre-existing subglacial till sheet at the transition from cold-based to warm-based ice.

Geometric ridge networks and crevasse-fill ridges are low subglacial landforms that are composed of material derived from the basal debris layer of the glacier, or material beneath the glacier. They form as a result of the squeezing of subglacial material into basal crevasses or former subglacial tunnels, commonly during surges. Alternatively, geometric ridge networks can also form beneath glaciers as a result of the intersection of foliation-parallel ridges and englacial thrusts.

Figure 16 Glacial depositional landforms: (A) terminal dump moraine (left) of Chuchung Glacier (breached by a stream) and lateral moraine (right) of Ama Dablam Glacier, Khumbu Himal, Nepal (an older moraine with person for scale in the foreground); (B) hummocky moraine formed by the thrusting of subglacial sediment into an englacial position before release, Kronebreen, Svalbard; (C) flutes on proglacial area of Austre Levénbreen, north-west Spitsbergen; (D) esker (sinuous ridge of sand and gravel) emerging from beneath Comfortlessbreen, a tidewater glacier in western Spitsbergen. © 2005, Photographs, M J Hambrey.

Eskers are glaciofluvial landforms that are created by the flow of melt water in subglacial, englacial, or supraglacial channels. They are usually sinuous in plan view and composed of sand and gravel (Figure 16D).

Bathymetric Forms Resulting from Glacial Deposition

The morphology and sediment composition of subaquatic features, particularly in fjords and on continental shelves, are less well known than their terrestrial counterparts. As on land, subaqueous depositional assemblages reflect the interaction of a wide range of processes.

Ice-contact features form when a glacier terminus remains quasi-stationary in water, particularly in fjords. Morainal banks form as a result of a combination of lodgement, meltout, dumping, push, and squeeze processes combined with glaciofluvial discharge; poorly sorted deposits are typical of such features. Grounding-line fans extend from subglacial tunnels that discharge melt water and sediment into the sea, and are typically composed of sand and gravel. Developing out of grounding-line fans are ice-contact deltas, which form when the terminus remains stable long enough for sediment to build up to the surface of the fjord. Where a glacier becomes disconnected from the water body, alluvial sediments may prograde to form fluviodeltaic complexes, again dominated by sand and gravel. In addition to these large-scale forms, there are small-scale features that are found particularly on beaches, such as iceberg toolmarks, ridges, depressions from the melting of buried ice, bounce marks, chatter marks, and roll marks. Icebergs can also churn up submarine sediment, particularly on shoals, producing iceberg turbates.

Depositional forms on continental shelves are best known from a combination of deep drilling and seismic profiling. Some, as in Alaska, are simply larger-scale analogues of fjordal features, such as fan-delta complexes, but others show evidence of having formed under floating tongues that are typical of the colder ice of the polar regions. These include subglacial deltas, till tongues, diamicton aprons, and the immense (up to 400 km wide) trough-mouth fans.

Other features are similar to those on land and include shelf moraines, flutes, and transverse ridges.

Conclusions

Glaciers and ice-sheets can be highly effective agents of landscape change, eroding, transporting, and depositing large amounts of material at a variety of spatial scales, from a valley scale to a continental scale. The rates of operation of these processes are strongly controlled by glacier mass balance and thermal regime, which together determine glacier velocity, melt water production, ice deformation styles, and sediment routing through the ice-mass. Glaciers produce a diverse range of landforms and sediments. The classic glacial erosional landforms, such as cirques, glacial troughs, stoss-and-lee forms, and areas of striated bedrock, are generally carved in solid bedrock, although glacial erosional landforms such as melt water channels and tunnel valleys can also form in unconsolidated sediments. Glacial depositional landforms take a variety of forms and include ice-marginal landforms, glaciotectonic moraines, hummocky moraines, outwash fans and plains, kames and kame terraces, and subglacial bedforms. These landforms are typically composed of a range of sediment types, ranging from poorly sorted sediments such as diamicton to well-sorted sand and gravel. Glacial sediments may be deposited in both terrestrial and marine settings, although those deposited on continental shelves and slopes probably have the best preservation potential in the geological record.

See Also

Famous Geologists: Agassiz. **Geomorphology**. **Palaeoclimates**. **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings; Deltas; Lake Processes and Deposits. **Sedimentary Processes:** Erosional Sedimentary Structures. **Sedimentary Rocks:** Mineralogy and Classification; Rudaceous Rocks.

Further Reading

Anderson JB (1999) *Antarctic Marine Geology*. Cambridge: Cambridge University Press.

Benn DI and Evans DJA (1999) *Glaciers and Glaciation*. London: Arnold.

Bennett MR and Glasser NF (1996) *Glacial Geology: Ice Sheets and Landforms*. Chichester: John Wiley and Sons.

Dowdeswell JA and O'Cofaigh C (eds.) (2002) *Glacier-Influenced Sedimentation in High-Latitude Continental Margins*. Special Publication 203. London: Geological Society.

Dreimanis A (1989) Tills, their genetic terminology and classification. In: Goldthwait RP and Matsch CL (eds.) *Genetic Classification of Glacienic Deposits,* pp. 17–84. Rotterdam: Balkema.

Eyles N (1993) Earth's glacial record and its tectonic setting. *Earth Science Reviews* 35: 1–248.

Hambrey MJ (1994) *Glacial Environments*. Vancouver and London: University of British Columbia Press and UCL Press.

Hambrey MJ (1999) The record of Earth's glacial history during the last 3000 Ma. In: Barrett PJ and Orombelli G (eds.) *Geological Records of Global and Planetary Changes,* pp. 73–107. Terra Antartica Report 3. Siena: Museo Nazionale dell'Antartide.

Hambrey MJ and Alean J (2004) *Glaciers,* 2nd edn. Cambridge: Cambridge University Press.

Knight P (1999) *Glaciers*. Cheltenham: Stanley Thornes (Publishers) Ltd.

Maltman AJ, Hubbard B, and Hambrey MJ (eds.) (2000) *Deformation of Glacial Materials*. Special Publication 176. London: Geological Society.

Menzies J (ed.) (1995) *Modern Glacial Environments: Processes, Dynamics and Sediments*. Oxford: Butterworth-Heinemann.

Menzies J (ed.) (1996) *Past Glacial Environments: Sediments, Forms and Techniques*. Oxford: Butterworth-Heinemann.

Paterson WSB (1994) *Physics of Glaciers*, 3rd edn. Oxford: Pergamon.

Post A and LaChapelle ER (2000) *Glacier Ice*. Seattle and Cambridge: University of Washington Press and The International Glaciological Society.

Karst and Palaeokarst

M J Simms, Ulster Museum, Belfast, UK

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Karst is a term applied to terrain with distinctive landforms and underground drainage systems that form through the greater solubility in water of certain rock types, particularly limestone. Karst landscapes are sculpted largely by solution, other rock types largely by mechanical erosion. The word 'karst' has its origins in pre-Indo-European languages, from *kar*, meaning 'rock'. In Slovenia the word 'kras' (or 'krs'), subsequently germanicized as 'Karst', derives from the name of a barren stony limestone area near Trieste, which is still considered the type area for limestone karst. Gypsum and rock salt may form karst,

with extensive gypsum karsts known from Russia and the Ukraine, but their greater solubility renders such landforms more dynamic and, for rock salt, ephemeral in all but the most arid climates. Karst features may also develop, though rarely, on very weakly soluble rocks, such as basalt, granite, or quartzite.

Rock solubility and water are the primary factors in karst development. Arid climates, whether hot or cold, support little karst. Physical rock properties also are important. Highly porous rocks seldom support well-defined karst features, which instead are favoured by low porosity and good secondary permeability, in the form of fractures, focusing the drainage into specific conduits through the karst rock. The removal of rock in solution allows the development of drainage through the rock, rather than just across its surface as happens largely with rocks removed by mechanical erosion. Consequently, karst landscapes generally lack well-developed surface drainage but have underground drainage conduits, or caves. Hence a significant component of karst terrains typically lies beneath the surface, sometimes extending to depths of hundreds, or even thousands, of metres. Intimately associated with the dissolutional aspects of karst are depositional ones. The latter include clastic sediments within the caves and, particularly, minerals deposited by precipitation from karst waters both above and below ground.

Many subdivisions of karst have been proposed. Relict karst is used to denote landforms inherited from earlier climatic or drainage regimes but still subject to modification by the current conditions. Palaeokarst refers to karst features buried by younger rocks and so largely isolated from current karst modification; where uncovered by later denudation, this isolated karst is called exhumed karst. Biokarst encompasses small-scale sculpting of limestone by animals and plants, although the distinction between dissolutional sculpting (true biokarst) and mechanical excavation (bioerosion) is seldom made. Pseudokarst is, as its name implies, 'false karst'. Such features superficially resemble karst but form by quite different processes, such as lava tubes, soil piping, and thermokarst, or cryokarst, formed by localized melting of permafrost.

Often karst geomorphology is regarded as a specialist discipline that is of limited general application to geology or geomorphology. However, ~12% of Earth's terrestrial, ice-free surface is composed of limestone, with 7–10% supporting some form of karst landscape. Furthermore, as much as 25% of the world's population may depend to some extent on karst water supplies. Consequently, the study of karst is crucial to understanding landscape and drainage development over a significant area of Earth's surface.

Karst Processes

The basic process of karst dissolution involves ion dissociation. For rock salt (NaCl), gypsum ($CaSO_4{\cdot}2H_2O$), and quartz (SiO_2), this requires only the presence of water but, on a global scale, the outcrop area of evaporites and the low solubility of quartz render them of only minor and localized significance for karst development. Limestone ($CaCO_3$) and dolomite ($CaMg[CO_3]_2$) are by far the dominant karst rocks but experience very low rates of dissociation in pure water. The addition of free H^+ ions greatly increases the rate of carbonate dissociation and hence even weak acids become effective solvents. In most karst environments, carbonic acid, derived from atmospheric or soil CO_2, is the main source of H^+ ions, although other organic or inorganic acids may be significant locally. The solubility of CO_2 increases with decreasing temperature, in common with other gases. The same is true also for limestone, in marked contrast to most solids for which solubility increases with temperature. Nonetheless, the availability of liquid water and biogenic CO_2 is far more significant for karst development than are low temperatures.

Water with an excess of H^+ ions is commonly referred to as 'aggressive', and continues to take up HCO_3^- ions until an H^+/HCO_3^- equilibrium is reached at saturation, as in the following equation:

$$CO_3^{2-} + 2H^+ \rightleftarrows HCO_3^- + H^+ \rightleftarrows H_2CO_3$$

Different karst waters may reach saturation at different concentrations, determined by the initial CO_2 concentration, but this is not a simple straight-line relationship and mixing of different karst waters may increase aggressivity. This phenomenon, called mixing corrosion, may be significant in certain karst environments. Carbonate solubility also is increased by the foreign ion effect, the addition of ions such as Na^+, K^+, and Cl^-. Seawater is saturated and cannot directly dissolve limestone, but mixing with freshwater can considerably increase carbonate solubility and is of major significance in certain environments. In the same way that an increase in CO_2 concentration increases carbonate uptake, so degassing of CO_2 from a saturated solution causes reprecipitation of calcite.

Dissolution can and does occur in static or laminar flow conditions, though constrained by diffusion rates through the boundary layer. Permeable soil or sediment cover, even when vegetated, may offer only limited resistance to downward percolation of water to the limestone beneath. Although this subsoil water movement may be slow, its dissolutional efficacy is enhanced by higher CO_2 concentrations generated by

respiring plant roots and soil organisms. With faster flow, generally associated with increased water depth and/or conduit width, laminar flow gives way to turbulent flow and there is a rapid (~fivefold) increase in dissolution rate. This laminar–turbulent transition, or turbulent threshold, is very significant for the development of many karst features.

Time is a major factor to consider in the development of karst, reflecting dissolution rates. Features on a millimetre to centimetre scale may form within a few years on limestone, whereas the largest scale landforms may require hundreds of thousands, or even millions, of years to reach their present form. Timescales for karst development on gypsum are much faster, whereas on quartz they are very much slower.

Figure 1 Solution flutes (rillenkarren) on gypsum, Devon coast, England. These may form on gypsum within a few months or years in a humid climate.

Surface Karst (Exokarst)

At the surface of the Earth, karst features are manifest at scales from tens of kilometres down to millimetres. Large-scale karst landscapes and landforms invariably support a diversity of smaller scale features. The type and scale of karst features that develop at a particular location are determined by several factors that can be grouped broadly: (1) intrinsic rock properties such as composition, texture, porosity, and fracturing; (2) extrinsic factors, particularly climate (especially the amount of rainfall), specific environment (terrestrial, lacustrine, or marine), the type and extent of plant cover, sediment cover (if any), and tectonism (uplift or subsidence); and (3) time. In general, karst is best developed when pure limestones with low porosity but good secondary permeability are subject to frequent, rather than heavy, rainfall. Hence karst landforms are better developed in humid temperate or tropical climates than in strongly seasonal ones. Duration of exposure to dissolution is critical; for particular intrinsic and extrinsic conditions there is an optimum time-scale for specific karst features to develop. Beyond this time-scale, dissolution may continue to modify and, ultimately, even destroy the features.

Small-Scale Karst (Karren)

Small-scale solutional sculpting (millimetres to metres), often termed 'karren', develops purely as a function of rock solubility and run off, independently of the underground drainage characteristic of karst terrains. The development time-scale for most karren ranges from a few years to tens of thousands of years under optimum conditions. On exposed limestone solution pans (kamenitzas) form by ponding of water on level surfaces, while solution flutes (rillenkarren; Figure 1), runnels (rinnenkarren), and ripples develop by water flow across inclined surfaces. Fractures, commonly joints, act as initial conduits for flow into the karst aquifer and are widened by dissolution to form grikes (kluftkarren). The limestone between grikes may, depending on factors such as lithology, soil cover, and climate, form flat-topped blocks (clints or flachkarren), rounded blocks, domes or teeth (rundkarren), or sharp pinnacles (spitzkarren), or may break up into fragments (shillow, felsenmeer, or trümmerkarren). Glacially stranded boulders may protect limestone directly beneath them from direct dissolution by rainfall, eventually leaving the boulder perched on a pedestal surrounded by a limestone surface lowered by solution. The height of such pedestals provides a measure of the surface lowering rate (Figure 2). Clints, grikes, solution runnels, and solution pans are common features beneath mineral or organic soils, though typically they are more rounded than those formed subaerially.

Freshwater lake shores support their own unique karren assemblages, strongly influenced by lake water chemistry. Small hemispherical solution pits (Figure 3) are particularly characteristic on flat limestone surfaces around lakes, and notches (swamp notches) may indicate former water levels. Even if the lake water is permanently saturated with carbonate, the limestone can dissolve if seasonal changes in water level trap air pockets in which condensation corrosion occurs; this process forms upward-tapering tubes (röhrenkarren). Similarly shaped, but larger, features called bell holes are found in the entrance zone of some tropical caves and are considered to form by condensation corrosion associated with daily changes in air temperature.

Most karren are essentially abiogenic in origin, but living organisms may contribute directly to limestone

Figure 2 Limestone pedestal beneath a sandstone erratic. The pedestal provides a measure of postglacial (~14 ka) surface lowering of the surrounding limestone. Conspicuous solution runnels (rinnenkarren) have formed on the top of this sandstone erratic despite its low solubility. County Cavan, Ireland.

Figure 3 Abundant solution pits, typical of lacustrine karst, on limestone surfaces exposed in the zone of seasonal submergence. Lough Mask, Ireland.

dissolution on a small scale. Certain algae and invertebrates adopt an endolithic habit, living within cavities that they excavate in the rock. Cyanophytes constitute a significant karstic agent in some climates. Where created with acids, these cavities are true biokarst (Figure 4). However, many organisms excavate cavities mechanically, by bioerosion, and these are not strictly biokarst. Algal biokarst has been termed phytokarst and is ubiquitous in intertidal marine environments, where endolithic organisms are the dominant karst agent. In cave entrances or beneath overhangs, light-orientated phytokarst pinnacles, or

Figure 4 Temperate marine biokarst, encrusted with *Mytilus*, in the intertidal zone. The Burren, Ireland.

photokarren, may develop where biokarstic algae are illuminated from just one direction.

Medium-Scale Karst (Karst Landforms)

Medium-scale features (metres to kilometres) can perhaps be grouped together as karst landforms and develop mainly over time-scales of thousands to hundreds of thousands of years. In any karst landscape, dissolution is greatest near the surface where water first comes into contact with the rock, creating a shallow zone of epikarst above less karstified rock. Over time, and dependent on climate, the depth and dissection of this epikarst zone increases. The most characteristic surface karst landform is the doline (sometimes also known as a sinkhole, especially in North America and by engineers everywhere), an enclosed depression from which drainage is underground; on non-karst lithologies such depressions would form lakes. Dolines can form by one or a combination of mechanisms, e.g., by enhanced dissolution through focusing of drainage, by collapse into a cavity beneath, or by downwashing of a weakly consolidated soil or sediment cover into cavities beneath (Figure 5). Collapse dolines may intercept the regional water table to form cenotes; they are particularly characteristic of some lowland karst plains. Irregular, flat-floored depressions may develop by the lateral solution and planation of limestone inundated by seasonal rises in the regional water table. Known as poljes in their type area of the Dinaric Alps, they are fed by springs and drain via sinks, or may fill and empty via the same conduit, known as an estavelle, depending on flood stage or groundwater level.

Drainage in karst terrains is not exclusively subterranean and elements of fluvial geomorphology are common. Valleys and gorges may develop where rivers cross from adjacent non-karst (allogenic) rocks and travel across the limestone surface before sinking.

Figure 5 Suffosion dolines on glacial till, formed by downwashing of sediment into cavities in the underlying limestone. Black Mountain, South Wales.

Many are now largely dry and were formed under different climatic regimes, reflecting either greater rainfall or the blocking of underground drainage by permafrost; Cheddar Gorge in south-west England is a fine example of the latter. So-called blind valleys end abruptly downstream against an often steep slope where a surface stream sinks underground, though the former surface continuation may continue beyond at a higher level. Conversely, headward excavation by a spring may form a valley ending abruptly upstream at a steep slope, called a steephead or pocket valley. Emerging saturated karst waters may precipitate calcite, often under biogenic mediation, to form sometimes extensive deposits of tufa or travertine, such as those of Plitvice, Croatia.

Residual limestone hills at scales from a few metres to hundreds of metres high are an inevitable consequence of focused drainage from surrounding limestone, particularly where this is sufficient for turbulent flow, and the enhanced dissolution rates associated with it, to occur. Although individual hills can be regarded as landforms, they are more logically considered here as elements of the largest scale of karst development, karst landscapes.

Large-Scale Karst (Karst Landscapes)

Karst landscapes can be broadly defined as areas, typically covering many square kilometres, encompassing characteristic assemblages of karst landforms and karren. Generally, though not invariably, they reflect prolonged karstification over several million years and large-scale interactions between geology, tectonics, and climate. The youngest distinct type of karst landscape, characteristic of many cool temperate regions subjected to repeated glaciation during the Pleistocene, is termed glaciokarst (Figure 6). The Burren of western Ireland is a classic example and shows features typical of glaciation, such as ice-plucked

Figure 6 Temperate glaciokarst landscape with typical joint-controlled clints and grikes, rinnenkarren, rillenkarren, and kamenitzas. The Burren, Ireland.

crags, scoured rock surfaces, limestone pavements, and erratic boulders. Commonly these are superimposed on pre-existing karst landforms/landscapes, but smaller scale karst landforms and karren, particularly grikes (kluftkarren), small dolines, and caves, have commonly been superimposed on the glacial landforms since the ice retreated.

Fluviokarst landscapes are dominated by valleys cut by surface rivers, either before significant underground drainage had developed or under different climatic conditions, notably periglacial environments, which favoured surface rather than underground drainage. Many fluviokarst valleys are now largely dry because drainage has been captured underground.

Cone karst (Fencong) is dominated by conical or hemispherical hills, typically a few tens of metres high and several hundred metres across, separated by dolines and other irregular depressions. Cone karst develops only in wet tropical climates such as in parts of Java, some Caribbean islands, and particularly in China's Ghizou Province. Tower karst (Fenglin),

Figure 7 Relict karst towers, accentuated by marine undercutting, in Halong Bay, Vietnam. These testify to active tectonic uplift in a terrestrial environment, prior to drowning by sea-level rise or regional subsidence. Photo by Sally Hunt.

perhaps the most spectacular of all karst landscapes, is dominated by often precipitously steep limestone towers, from tens to hundreds of metres high, and is restricted to areas in south-east Asia and the adjacent islands (Figure 7). Cone and tower karsts develop where rates of tectonic uplift exceed, for the former, or are equalled, for the latter, rates of surface lowering at points of focused dissolution, leaving residual hills or towers between. In contrast, prolonged denudation with minimal uplift will, ultimately, produce a low-relief corrosion plain, as typified by the Gort lowlands to the south of Galway, Ireland.

Various classifications describe stages of 'maturity' in karst landscapes; the most useful, and comprehensive, scheme (Figure 8) was developed primarily for engineers and encompasses a range of parameters, including karst class (a qualitative measure based on the scale and diversity of karst landforms), mean doline density, rockhead relief, and typical cave size.

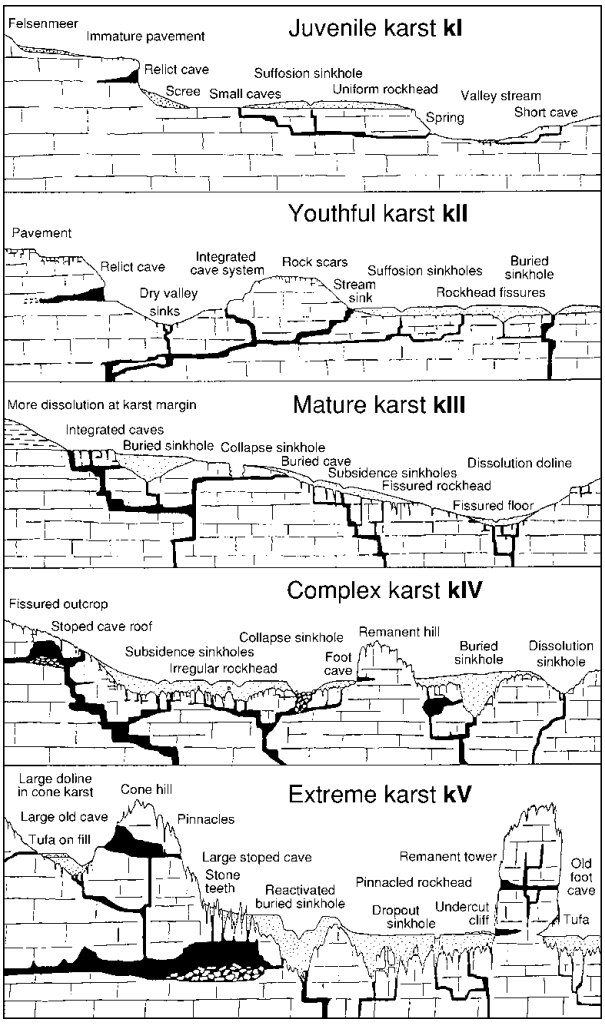


Figure 8 Typical landforms encountered within the five categories of ground conditions recognized by Waltham and Fookes in their 2003 engineering classification of karst.

Karst Denudation and Landscape Development

In landscapes encompassing both soluble and insoluble lithologies, differences between erosional and dissolutional processes have a considerable effect on landscape development in tropical and temperate regions. Karst dissolution is a low-energy process that may occur even in static water and, in the case of limestone, is enhanced by a vegetation cover, which increases soil CO_2 levels. Analogous weathering of non-karst rocks may alter their composition and mechanical strength without significant volume change or transport, forming saprolites sometimes many metres thick. These are removed only by erosion, a higher energy process requiring fast water flow but hindered by vegetation, which binds and protects weathered rock and soil. In temperate and tropical environments, limestone may be removed at a steady rate, proportional to mean annual rainfall, whereas erosion of rock and saprolite is more sporadic and depends on the frequency of intense rainfall events and the extent of vegetation cover. In arid regions, limestones commonly prove more resistant to erosion than do other, less well-cemented, rocks. Conversely, in tectonically stable humid climates, the long-term denudation rate of limestones may be significantly greater than is the overall erosion rate on adjacent rocks, reducing limestone outcrops to lowlands flanked by hills of more slowly denuded lithologies.

Karst Drainage

A significant element of drainage in karst landscapes is underground. Drainage into a karst aquifer is either

autogenic, where derived from a karst catchment, or allogenic, originating on non-karst rocks. Commonly, input to a karst aquifer is mixed. Autogenic recharge is diffuse, as rainfall across the entire karst outcrop, whereas allogenic recharge is concentrated at discrete stream sinks or swallets. Most negotiable caves are associated with such swallets. Water re-emerges from a karst aquifer at a rising (also known as a resurgence or spring). Typically, this occurs at the lowest point that the karst rock is exposed at the surface, although impermeable layers within a karst aquifer may cause a rising to be perched at a higher level. Resurgences may be free-draining streams or may rise from depth under hydrostatic pressure. The more than 300 m-deep Fontaine de Vaucluse in France is an example of the latter; the term 'vauclusian spring' has been coined to refer to the point of emergence of water under pressure.

Caves (Endokarst)

Underground drainage conduits are a fundamental element of karst. Those large enough for a person to enter are called caves, an artificial definition constrained by the size of the caver! In fact, water can move slowly along submillimetre fractures, but at widths of ~5 mm laminar flow gives way to turbulent flow (the turbulent threshold, or hydraulic jump) and dissolution rates increase significantly. Such 'proto-caves', often anastomosing networks of small tubes, commonly are found on fracture planes adjacent to larger caves that have developed from them. The development of a cave, from inception to destruction, is called speleogenesis.

Fractures commonly represent the initial routes for water flow, and the configuration of many passages is strongly influenced by the spatial distribution and orientation of joints, faults, and bedding planes. Rectilinear passage networks arise where inception is dominated by vertical fractures, whereas meandering passages imply inception on subhorizontal ones. Passage shape may be influenced by these fractures but also is strongly controlled by passage location with respect to the 'water table'. Above, in the vadose zone (see Figure 9), water occupies only the lower part of the passage and flows downwards, as in surface streams and waterfalls; dissolution here forms canyonlike passages and vertical shafts. Sequences of deep vadose shafts and canyons, sometimes descending a kilometre or more, are characteristic of high mountain karst regions such as in the Alps. Below, in the phreatic, or 'saturated', zone, fractures and

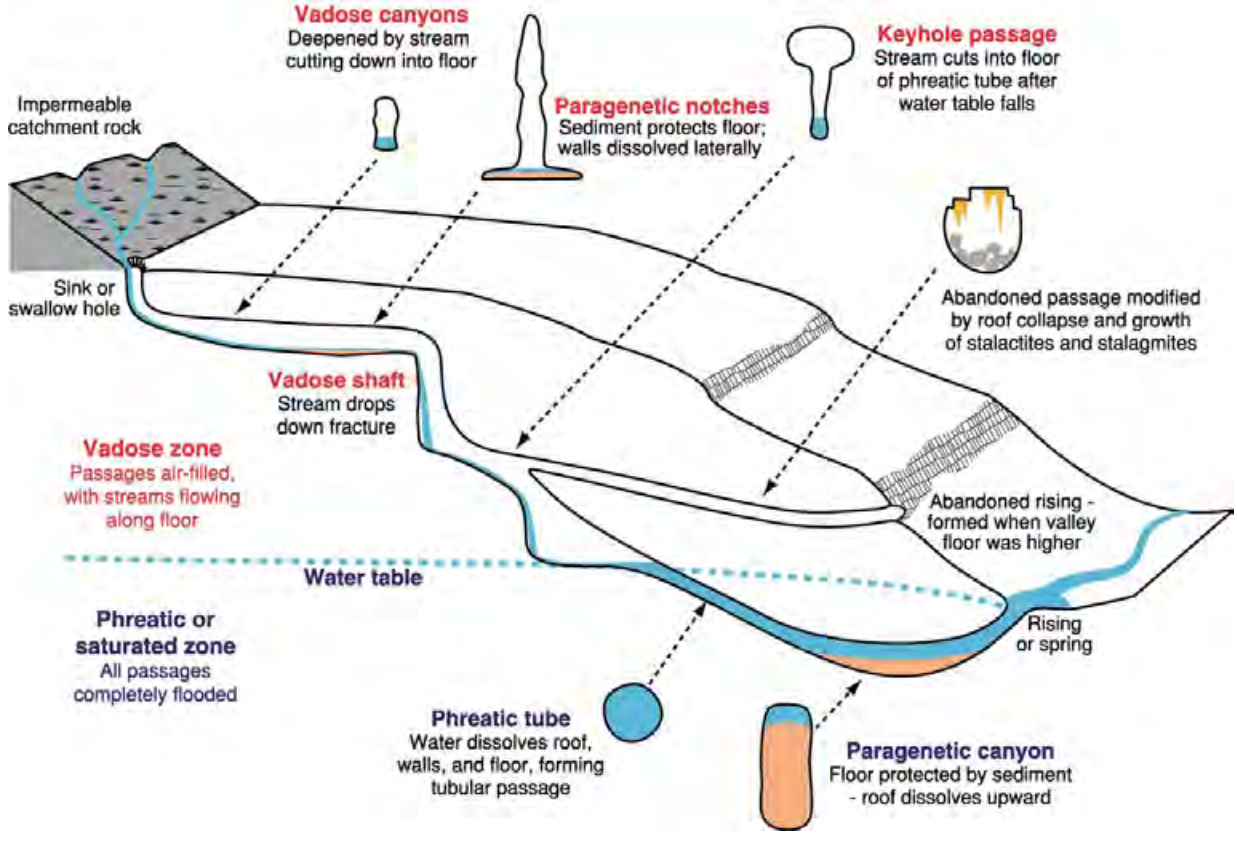


Figure 9 Diagrammatic section through a cave system to show the various features discussed in the text.

Figure 10 Relict paragenetic canyon, with sediment removed, extending upwards from the bedding plane on which the anastomosing 'proto-caves' first developed. The Burren, Ireland.

Figure 11 Active vadose canyon incised below the inception level at which the cascade is entering. Doolin River Cave, The Burren, Ireland.

cavities are water filled, and water flow follows the path of least resistance, often flowing upwards if necessary. Dissolution here occurs in all directions, forming passages with circular or elliptical cross-sections, as seen in the drained tubes of Kentucky's Mammoth Cave and still forming in the active phreas of Mexico's Yucatan, often with blind pockets formed by eddies or by mixing corrosion where percolation water emerges from a minor fracture into a flooded passage. Mixing of karst waters from separate sources may enhance dissolution locally, influencing both conduit configuration and shape. Sediment aggradation is common both in vadose and in phreatic passages and may armour the floor against downward dissolution, causing lateral or upward dissolution by paragenesis (this is a process quite distinct from mineral paragenesis). In vadose passages, lateral dissolution may form paragenetic notches at stream level, spectacularly seen in some of the huge river caves of Mulu, Sarawak. In phreatic environments, dissolution may raise the roof, forming a largely sediment-filled paragenetic canyon. Paragenetic canyons (Figure 10) extend upwards from their initiation level, such as a bedding plane, whereas vadose canyons (Figure 11) cut downwards from it. A drop in aquifer base level may flush out sediment fill from paragenetic canyons and/or form 'keyhole' passages by vadose incision in the floor of former phreatic tubes. The two types of canyon passages can be distinguished by the presence of either undulating wall notches, indicating phreatic paragenesis associated with undulating sediment banks, or horizontal notches, indicating vadose conditions. Keyhole passages, with a vadose trench incised in the floor of a phreatic tube, are particularly significant for interpreting the history of karst aquifers, indicating rejuvenation when base level drops below the cave passage.

Water in a karst aquifer may be confined by an impermeable rock (aquifuge), whereas impermeable or weakly permeable layers (aquicludes or aquitards) may influence water movement through the rock, allowing the development of perched aquifers above regional base level. Hence the altitude of a particular phreatic passage may reflect only local conditions. Eddies caused by turbulent water flowing across limestone in caves and surface streams commonly form asymmetric concavities, called scallops, which are invaluable flow indicators in relict cave passages (Figure 12). Their steeper slopes face downstream, and their size is a function of flow velocity. Smaller scallops indicate faster flow than larger ones; vadose scallops are sometimes only 1–2 cm across, whereas phreatic scallops may be 1 m across! An approximate rule of thumb for calculating flow velocities (in metres per second) is to divide 3.5 by the mean length of the scallops in centimetres.

Figure 12 Heavily scalloped relict phreatic pressure tube entering a larger cave passage. The strongly asymmetric scallops provide clear evidence of upward flow, under pressure, into a larger water-filled passage. Deflection of these scallops above the pressure tube indicates flow from left to right in the main passage. Ogof Llyn Du, North Wales.

Figure 13 Typical cave fills in an active vadose cave. Speleothems (stalactites) are deposited by carbonate-saturated percolation water emerging along joints on roof. Coarse clastic sediments are carried into the cave by fast-flowing water. Marble Arch Cave, Northern Ireland. Photo by Tony Waltham.

Most caves form by an overall downward movement of water through the karst aquifer from sink to rising, but a few form by very different processes. Some are created instead by hydrothermal springs or by the action of sulphuric acid formed from hydrogen sulphide rising from hydrocarbon deposits beneath. Encompassing some of the largest and longest caves in the world, such as the Carlsbad and Lechaguilla caves in New Mexico, they contain some of the most extraordinary arrays of speleothems (cave mineral deposits) known, with gypsum a major component.

Over time, cave passages may collapse or be infilled by clastic sediment, transported in from the surface, or by speleothems (Figure 13). Clastic sediments provide important information on environmental histories and potentially can be dated using techniques such as optically stimulated luminescence. Calcite and aragonite speleothems (stalactites, stalagmites, and flowstone) are the most common mineral deposits, precipitated in response to CO_2 degassing of carbonate-saturated percolation water emerging into air-filled cavities and hence providing evidence of vadose conditions. They are immensely important in karst research because the calcite can be dated radiometrically, particularly by uranium-series disequilibrium, and can provide information on palaeoclimate and hydrology.

Figure 14 Spectacular palaeokarst shaft developed in, and capped by, marine Carboniferous limestones. The size of this pothole, filled with non-marine clastics, provides evidence of the scale and duration of this intra-Carboniferous emergent episode. Cookstown, Northern Ireland.

Palaeokarst and Interstratal Karst

Palaeokarst is defined as karst that has been buried by younger rocks. It is a common, though often unrecognized, component of successions in which limestones are present (Figure 14), and serves as a clear indicator of terrestrial environments and, to some extent, duration of emergence. Interpretation

of palaeokarst may be complicated by two factors; it is usually visible only in two, rather than three, dimensions; and burial by younger rocks does not prevent palaeokarst being modified or even destroyed by subsurface dissolution. Small-scale karst microsculpting with relief of a few centimetres, such as rillenkarren or solution pits, is readily destroyed by pressure solution associated with rock overburden. Lithological contrasts commonly act as a focus for water movement and the development of interstratal karst, which may resemble palaeokarst but is younger than the overlying rock. Interstratal karst is especially prevalent in highly soluble rocks such as gypsum and rock salt. Cave passages may remain open as potential conduits for water flow long after burial of the karst surface, and sometimes are invaded and infilled by mineralising fluids. Such occurrences may be difficult to distinguish from syngenetic mineral deposits, in which conduit formation and mineralisation were penecontemporaneous, but mineralized palaeocaves commonly include some clastic sediment infill beneath the mineral deposits. Remains of organisms within caves are protected from destruction by weathering and scavengers; hence palaeokarst caves encompass many important fossil lagerstätten. Although some animals are adapted for life in caves, most remains found are of animals and plants washed in or fallen in from the surface.

See Also

Engineering Geology: Geomorphology; Subsidence; Ground Water Monitoring at Solid Waste Landfills. **Minerals:** Carbonates. **Mining Geology:** Hydrothermal Ores. **Sedimentary Environments:** Carbonate Shorelines and Shelves; Lake Processes and Deposits. **Sedimentary Processes:** Erosional Sedimentary Structures. **Sedimentary Rocks:** Dolomites; Evaporites; Limestones. **Unidirectional Aqueous Flow**.

Further Reading

Ford DC and Williams PW (1989) *Karst Geomorphology and Hydrology*. London: Unwin Hyman.

Gunn J (ed.) (2003) *Encyclopedia of Caves and Karst Science*. London: Routledge.

Klimchouk AB, Ford DC, Palmer AN, and Dreybrodt W (eds.) (2000) *Speleogenesis: Evolution of Karst Aquifers*. Huntsville, AL: National Speleological Society.

Jennings JN (1985) *Karst Geomorphology*. Oxford: Blackwell.

Lowe DJ and Waltham AC (2002) *Dictionary of Karst and Caves*. Bridgewater, UK: British Cave Research Assoc.

Palmer AN (1991) Origin and morphology of limestone caves. *Bulletin of the Geological Society of America* 103: 1–21.

Simms MJ (1994) Emplacement and preservation of vertebrates in caves and fissures. *Zoological Journal of the Linnean Society* 112: 261–283.

Simms MJ (2004) Tortoises and hares: Dissolution, erosion and isostasy in landscape evolution. *Earth Surface Processes and Landforms* 29: 477–494.

Waltham AC and Fookes PG (2003) Engineering classification of karst ground conditions. *Quarterly Journal of Engineering Geology and Hydrogeology* 36: 101–118.

Zhang Z (1980) Karst types in China. *GeoJournal* 4: 541–570.

Landslides

S F Burns, Portland State University, Portland, OR, USA

© 2005, Elsevier Ltd. All Rights Reserved.

Introduction

Landslides are a major form of geohazard that cause millions of dollars of damage and many deaths each year. They result from the movement of masses of rock, earth, and debris down a slope under the influence of gravity. They can occur at all scales, from a few metres across to several kilometres, and at all rates, from metres per year to metres per second. The term landslide is really a misnomer because the material can move not only by sliding, but also by flowing and/or falling. Landslides are important superficial processes for geomorphology, moving sediment down the slope where water in streams can eventually carry it away. They are generally triggered by water, but sometimes earthquakes and even volcanic eruptions can lead to landsliding. On land the water comes mainly from heavy and persistent rainfall but also can come from rapid snowmelt. Submarine landslides may be triggered by cyclic wave loading as well as by earthquakes, and frequently follow marine trangressions.

Possibly the largest landslide on land occurred some 10 000 years ago in south-western Iran. Called the Saidmarreh Landslide, a mass of limestone 15 km long, 5 km wide, and 300 m thick slid off the underlying marl bedrock and travelled a distance of 18 km while dropping only 1000 m. One of the largest underwater slides has been described to have

occurred on the continental margin of the Atlantic Ocean off the coast of Spanish Sahara in water depths over 2000 m and covering an area of 18 000 km^2. It probably originated after the last glacial advance when sea-level was 100 m lower than it is today. Landslides have also been identified on other planets like Mars, where one with a volume of 100 km^3, 60 km long, and 50 km wide that has travelled 2 km, leaving a scarp 6 km high, has been identified.

Loss of Life and Property due to Landsliding

Human and economic losses from landslides are considerable worldwide and continue to grow as pressures of expanding population cause development to extend into less stable hillslope regions. It is estimated that each year landslides kill nearly 600 people around the world, and cause over $12 billion (US) damage (direct and indirect losses). Direct losses include the cost of rebuilding the structures, whereas indirect losses are concerned with commercial loss, reduced real estate value, and loss of productivity. Japan is one of the most affected countries, suffering over $4 billion (US) loss per year. The United States, Italy, and India each have total annual losses of between $1 and $2 billion (US) each.

The worst landslide disaster in historical times occurred on December 16, 1020 in Kansou, China when over 180 000 people were killed. Many of these people were living in homes excavated in or built on the slopes of loess (silt). This soil has a metastable fabric and collapsed during an earthquake-triggered event. One of the worst disasters in recent years happened in 1998 when an ancient landslide reactivated in Kelso, Washington (USA), and 60 houses on it were destroyed, causing over $25 million (US) of damage to homes and utilities (Figure 1).

Figure 1 Home destroyed in 1998 in Kelso, Washington, by a thrust of soil under it as an ancient landslide reactivated, destroying 60 similar homes over a four-month period. Note the main street in the foreground has also been destroyed.

Factors Determining Slope Stability

The stability of a slope can be considered as two forces working against each other: Driving forces (or shear stress) work to cause slope materials to move down-slope, whereas resisting forces (shear strength) act to keep the materials on the slope. When the ratio of resisting forces over driving forces (called the factor of safety) is greater than 1, the slope is stable. When it is less than 1, the slope usually fails.

'Water' can both increase the driving forces and reduce the resisting forces, thereby reducing stability. Saturation from rainfall can increase the slope mass, thereby increasing the driving forces. Filling soil pores with water also reduces soil strength by increasing pore pressure and reducing the effective stress, and possibly allowing particles to pass by one another so further reducing the internal resistance of the soil to sliding. To reduce the risk of landsliding on a slope, the first consideration should therefore be the removal of excess water.

As the angle of slope increases, the driving forces also increase. Cutting a road into a slope could easily oversteepen it, making that slope more prone to landsliding unless a retaining wall is built. Each geological material has a characteristic 'angle of repose', which is its maximum stable angle, related to its frictional strength. For instance, the angle of repose for an uncemented sand is about 35°, and slopes steeper than this are likely to fail. Clay, shale, serpentinite, and uncompacted fill are all slope materials prone to failure, having significantly lower angles of repose.

'Vegetation' is also important to slope stability for it increases resisting forces through its roots that bind the soil. Trees also act as natural suction pumps that remove water from the soils through evapotranspiration, thereby further increasing slope stability.

If the bedrock is inclined somewhat parallel to the slope, it is called a 'dip slope', and the stability will be determined by the bedding plane strength, which may well be lower than the mass strength. Shales are particularly weak, as are buried 'fossil' soils on these dip slopes. The Gros Vente Landslide that dammed the Gros Vente River in 1925 in Wyoming (USA) and the Vaiont disaster in Italy in 1963 are examples of dip slope failures. Adding a stockpile or a structure to a slope can also increase the driving forces on a slope, reducing its stability.

Types of Landslide

Various classification schemes and names have been given to describe landslides, depending upon the

process that brings the soil and/or rock down the slope. Landslides are generally classified into process groups based on their rates of movement and corresponding amounts of water contained within the landslide mass (Figure 2).

'Falls' refers to the free fall of detached materials, usually rocks, that descend down a steep slope. Frost heave or heavy rains generally dislodge rocks from cliffs, causing a free fall to the bottom of the slope. A collection of this rockfall material at the bottom of the slope is called 'talus'. Rockfalls can be spectacular, such as the 9000-m^3 rockfall that plunged 500 m into the valley at Yosemite National Park in California on 10 July 1996, travelling at 300 km h^{-1}. It generated an air blast that leveled 2000 trees! Another example of a large rockfall happened near the village of Randa in the Pennine Alps of Switzerland in the early 1990s (Figure 3).

A 'topple' is a type of rock displacement where there is forward movement of a block rotating around a fixed hinge. This can occur where jointing in a slope is steeply inclined and parallel to the cliff-face. In some places in the world, this is an important factor in cliff retreat (e.g., Tasmania) and also in recently glaciated upland regions (e.g., Scotland). Toppling is also very common in basalt cliffs where columnar jointing is present.

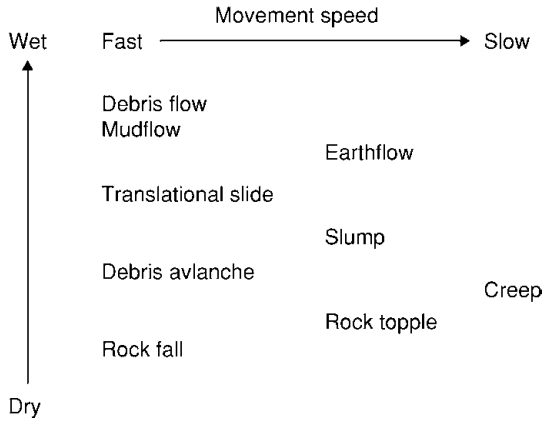


Figure 2 Landslides classification that relates water content to movement speed.

Figure 3 Large rockfall that occurred in the early 1990s in the village of Randa in the Pennine Alps of Switzerland.

'Translational slides' occur along a failure surface in the bedrock and move parallel to the surface. Most of the time the slide fractures at the base and breaks into many pieces. When the landslide moves as a uniform piece, it is called a slab failure or block slide. The failure surfaces upon which such landslides move are generally clay layers, shale layers, or palaeosols (old, buried soils) that parallel the slope. A classic translational slide occurred on 23 June 1925 in Gros Vente, Wyoming, USA. Over many years this slope of sandstone overlying shale (that dips at 20°) had been undercut by the stream at the bottom of the slope. Heavy snowmelt and heavy rains saturated the sandstone that consequently slid on the shale bed into the stream, damming it up and forming Gros Vente Lake.

If the sliding mass occurs along a curved-upward plane and shows rotation, it is called a 'rotational slide', which may develop into a 'slump' if the integrity of the failed mass is destroyed. Such movements are common in mudstones and clay soils or where there is a resistant geologic layer overlying a weaker layer. Such landslides tend to be small, producing gently inclined surfaces on the sides of steep slopes as the mass rotates (backtilts). Many large slumps occurred in Panama between 1914 and 1920 during the construction of the Panama Canal; one was so large that the whole of one side of the canal rotated into the canal itself. The slide was so large that the toe of the slide stuck up as an island in the center of the canal. Note the rotational slide movement affecting a house built on the toe of an ancient landslide that reactivated in 1996 in Tigard, Oregon, USA (Figure 4).

'Flows' occur when material moves down-slope as a slurry, generally due to continuous, permanent deformation of the sediment. 'Debris flows' consist of material ranging in size from clay to boulders mixed with water, displaying plastic flow behaviour. They form steep, lobate fronts with marginal levees. Mudflows are very similar to debris flows but contain finer-grained particles and tend to exhibit more viscous flow characteristics.

Debris flows and mudflows originate in upland source areas from many small soil slips that deliver sediment rapidly to create a stream, often the result of persistent rainfall. Deposition forms a fan-shaped deposit when the gradient becomes flat. Figure 5 shows a house in Dodson, Oregon, along the Columbia River Gorge, USA, that was buried by eight debris flows over a four-day period in February 1996.

Figure 4 Slump destroys a house in Tigard, Oregon, USA; note deformation of supporting pillars. This home was built on the toe of an ancient landslide that reactivated in February 1996.

Figure 5 Debris flow burying a house in Dodson, Oregon, USA, alongside the Columbia River Gorge; eight debris flows occurred here in February 1996.

Velocities can reach 55 km h^{-1}. Debris flows pose a significant hazard to life and property within the debris fans and along the stream valleys that connect the source areas to the debris fans. Debris flows are most common in arid regions, in alpine areas, and on the side of volcanoes. Deposits from debris flows can be differentiated from stream deposits by their high density, due to an abundance of clay and silt in the matrix. 'Debris torrents' are special types of debris flow that commonly consist of large volumes of logs and woody debris in addition to sediment in the slurry, occurring mainly in main drainage channels.

Debris flows that originate during volcanic eruptions are called 'lahars'. One of the world's most devastating debris flow disasters involving a lahar occurred on 13 November 1985 in Armero, Colombia, when a lahar originating 50 km away on the volcano of Nevado del Ruiz roared down the Lagunilla River Valley. It consisted of a 40-m-high wall of mud by the time it hit the town of Armero, burying 23 000 people. Debris flows are so dangerous because they travel at high velocities over long distances, and occur with little warning. America's volcano, Mt. Rainier, is the most dangerous because of its lahar risk: 55 lahars have originated from Mt. Rainier over the past 10 000 years, travelling up to 40 km from the mountain. Thousands of people live within a 40-km radius of Mt. Rainier and face this potential danger. The 1980 eruption of Mt. St. Helens also produced lahars that traveled more than 120 km from the mountain, destroyed 27 bridges over rivers, buried about half of Highway 504 to the west of the volcano, and delivered 34 million m^3 of mud to the Columbia River, 45 km from the volcano, which blocked shipping in the main channel.

'Earthflows' travel more slowly and generally originate on hillslopes as large tongues that break away from scarps; they slow down as they lose water. Heavy and persistent precipitation on steep slopes underlain by weak soils produce such flows. Rates range from 0.17 to 17 km h^{-1}. Once the main movement has occurred, they continue to show creep movement for years afterwards.

'Quick clay landslides' are a special type of flow, characterized by thixotropic clay, especially glaciomarine deposits formed in coastal areas of Norway, Sweden, and eastern Canada. Isostatic rebound of these clays following glacial melting has exposed the sediment to weathering and flushing out of the salts by rainwater-fed groundwater. Sudden vibrations can subsequently cause the deposit to liquefy and flow. A classic example occurred in Rissa, Norway, in April 1978 as 5 million m^3 of quick clays failed, destroying seven farmhouses and moving at a rate of 30 km h^{-1}. One person was killed in this flow, which was triggered by a simple excavation in a local farmyard.

The fastest and most destructive landslides are called 'debris avalanches', or 'sturtzstroms'. They travel at speeds up to 300 km h^{-1}, travel long distances (sometimes in excess of 50 km), and may be very large (some have been in excess of 80 million m^3). Why do these landslides travel such large distances at such high velocities and large volumes? An early idea was that the debris moved on a trapped layer of air. More recent hypotheses stress semielastic collisions between landslide particles: the body moves until all of the original (potential) energy is consumed by friction between the particles. Another hypothesis stresses internal melting and basal melting of the debris, thereby decreasing friction. This last idea comes from a study of the 2.1-km^3 Kofels Landslide in the Otz Valley in Austria where fused rock, resembling obsidian (and called frictionite), is found in layers throughout the base of the debris avalanche,

Figure 6 The May 18, 1980, eruption of Mt. St. Helens in Washington, USA, seen here in 2004, was initiated by a 5.2-magnitude earthquake that also triggered a large debris avalanche seen in the foreground. It has been dissected by the Toutle River. Note Castle Lake, which was formed as the landslide dammed Castle Creek, to the centre right.

Figure 7 Soil creep expressed in trees with bent trunks called 'pistol-butt' trees or 'trees with knees'. The trees are Radiata pines in New Zealand's Southern Alps.

and pumice-like rocks abound at the edges, presumably being shed by this prehistoric landslide as it moved at high velocities.

The world's largest historical landslide, which occurred when Mt. St. Helens erupted in 1980, was a debris avalanche. It traveled over 25 km down the Toutle River Valley at a velocity over 150 km h^{-1} (Figure 6). In Yungay, Peru 22 000 people were killed in 1970 when a debris avalanche of over 80 million m^3 descended from the volcano, Nevado Huascaran, traveling a distance of 14.5 km in 7 minutes at a velocity of about 300 km h^{-1}. It jumped 300 m out of the valley to bury the village that was perched supposedly out of harm's way on a shelf. Another debris avalanche occurred in Elm, Switzerland, in 1881 after heavy rains and caused 10 million m^3 to move 5.5 km in less than a minute at 310 km h^{-1}, killing 115 people. Another 70 people were killed in 1903 in Frank, Alberta, when 30 million m^3 of limestone debris cascaded down the slope above the town at a velocity of 100 km h^{-1}. One the earth's largest debris avalanches dammed the river at Gohna, India, in 1893 during the monsoon rains with 3.6 billion m^3 of debris.

Significant movement on slopes can also occur but at rates almost imperceptible. This is called 'creep', with rates in the range from 0.1 to 100 mm yr^{-1}. Causes can be freeze–thaw or shrink–swell processes operating on the soil sediments, assisting movement down-slope by gravity. Evidence of movement is found in cracked walls, leaning telephone poles and walls, and trees with knees (pistol-butt trees) (Figure 7). Movement occurs mainly in the soil on the slope. Creep may also be a precursor to the failure of a slope during a landslide. It is also a common feature of a landslide mass after it has come to rest, and slopes within the failed mass regain equilibrium. Ancient earthflows have been known to exhibit creep thousands of years after the initial movement of the landslide. Hoofed animals grazing on the sides of slopes when they are wet cause small-scale terraces to form from creep, and these are called 'terracettes'.

Reactivation of Ancient Landslides

If land has moved once, there is a high chance of it moving again given the right conditions. As population pressures increase, people move onto marginal land, which may be more prone to landslides. It is therefore imperative to recognize ancient landslides and assist the planning process, for instance, with maps to raise awareness of their presence prior to building.

Geomorphologists are particularly interested in reading the landscape for signs that tell of past landslides, in order to identify active processes and understand the evolution of the landscape. First indications are scarps (arcuate steep slopes cut into the hill where the landslide has torn away). Second is hummocky topography ('bumpy' ground produced by the landslide mass weathering over time). A hummocky slope with a steep, half-moon slope at the top usually indicates that an old landslide has occurred there. Younger landslides will generally have different vegetation on the slide mass than the surrounding slopes. In coniferous forests, landslides generally are covered with deciduous trees for the first hundred years after the failure. If one can see the actual sediments making up the landslide, the colluvium there is likely to be unsorted and angular in shape as all particles have been mixed together randomly.

Prevention of Landslides and Mitigating Their Impact

Prevention of landsliding or repairing the damage occurring as the result of a landslide necessitates understanding the basic relationship of driving forces and resisting forces. The need is to reduce the driving forces and/or increase the resisting forces on the slope. Decreasing the slope angle to less than the angle of repose for the slope will significantly increase stability. Removing water from the slope through drains or temporary water diversion by the use of tarps in full will decrease the driving forces and increase the resisting forces. Building a buttress at the bottom of the slope by erecting a wall of boulders, building a concrete wall with weep holes (to prevent build-up of water pressure behind it), or building a gabion basket wall will increase the resisting forces of the slope.

Recognizing and mapping old landslides is important. Planning regulations recommend not building on ancient landslides, so preventing disasters like Kelso, Washington. Replacing slope vegetation aids the slope stability and increases the resisting forces through the removal of water. In rockfall-prone areas, the use of screens or sprayed-on shotcrete will keep the rock fragments on the slope, as will rock bolts for the larger boulders. Avalanche shelters will provide additional protection, aided by electronic fences to give warning of rockfall events.

A landslide killed 2600 people in Longarone, Italy, on 9 October 1963. The second highest concrete dam in the world had just been built to a height of 265 m—the Vaiont Dam at the base of Mt. Toc. The reservoir area was prone to landslides, and it was known that Mt. Toc had sedimentary beds that dipped steeply into the valley below and that the slopes had been oversteepened by the river at the base. Filling of the reservoir aggravated the situation, raising the groundwater level (and water pressure) at the base, aggravated by persistent heavy rains that raised the reservoir level 20 m higher than expected within a short period. The slope began to creep at a rate reaching 400 mm day^{-1} the day before failure. On the day of failure, 240 million m^3 of rock broke away on the south side of the valley and slid down into the reservoir, rapidly displacing the water and creating a huge wave 100 m high, which topped over the dam and flooded the Piave River Valley. A total of 2600 people were drowned. Today, the dam still stands, intact. Its foundations were adequate, but the location with respect to the reservoir impounded behind was quite inappropriate. The landslide could have been prevented if the planners had paid attention to the broader context of the location and, in particular, the effect of impounding on slope instability of the reservoir margins, and the movements that had been noted from the past.

See Also

Engineering Geology: Overview; Aspects of Earthquakes; Geological Maps; Natural and Anthropogenic Geohazards; Site and Ground Investigation; Site Classification. **Geological Engineering**. **Geomorphology**. **Sedimentary Environments:** Alluvial Fans, Alluvial Sediments and Settings. **Sedimentary Processes:** Particle-Driven Subaqueous Gravity Processes. **Solar System:** Mars. **Tectonics:** Earthquakes. **Urban Geology**. **Volcanoes**.

Further Reading

Brabb EE and Harrod BL (eds.) (1989) *Landslides: Extent and Economic Significance*. Rotterdam: AA Balkema.

Carson MA and Kirkby MJ (1972) *Hillslope Form and Process*. Cambridge, UK: Cambridge University Press.

Chacon J, Irigaray C, and Fernandez T (eds.) (1996) *Landslides*. Rotterdam: AA Balkema.

Chorley RJ, Schumm SA, and Sugden DE (1984) *Geomorphology*. London: Methuen.

Cruden DM and Fell R (1997) *Landslide Risk Assessment*. Rotterdam: AA Balkema.

Easterbrook DJ (1999) *Surface Processes and Landforms*, 2nd edn. Upper Saddle River NJ: Prentice Hall.

Ritter DF, Kochel RC, and Miller JR (2002) *Process Geomorphology*, 4th edn. Dubuque IA: William C Brown.

Selby MJ (1982) *Hillslope Materials and Processes*. Oxford: Oxford University Press.

Slosson JE, Keene AG, and Johnson JA (eds.) (1992) *Landslides/Landslide Mitigation*, Reviews in Engineering Geology, vol. IX. Boulder: Geological Society of America.

Summerfield MA (1991) *Global Geomorphology*. London: Longman Scientific and Technical.

Turner AK and Schuster RL (eds.) (1996) *Landslides: Investigation and Mitigation*. Special Report 247. Washington DC: Transportation Research Board, National Research Council.

Walker BF and Fell R (eds.) (1987) *Soil Slope Instability and Stabilisation*. Rotterdam: AA Balkema.