

# ArcGIS<sup>®</sup> 9

---

**Geostatistical Analyst**

**Руководство пользователя**



Copyright © 2001 ESRI  
All Rights Reserved.  
Russian Translation by DATA+, Ltd.

The information contained in this document is the exclusive property of ESRI. This work is protected under United States copyright law and the copyright laws of the given countries of origin and applicable international laws, treaties, and/or conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying or recording, or by any information storage or retrieval system, except as expressly permitted in writing by ESRI. All requests should be sent to Attention: Contracts Manager, ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

The information contained in this document is subject to change without notice.

#### **DATA CREDITS**

Carpathian Mountains data supplied by USDA Forest Service, Riverside, California, and is used here with permission.

Radioceasium data supplied by International Sakharov Environmental University, Minsk, Belarus, and is used here with permission. Copyright © 1996.

Air quality data for California supplied by California Environmental Protection Agency, Air Resource Board, and is used here with permission. Copyright © 1997.

Radioceasium contamination in forest berries data supplied by the Institute of Radiation Safety “BELRAD”, Minsk, Belarus, and is used here with permission. Copyright © 1996.

#### **DATA DISCLAIMER**

THE DATA VENDOR(S) INCLUDED IN THIS WORK IS AN INDEPENDENT COMPANY AND, AS SUCH, ESRI MAKES NO GUARANTEES AS TO THE QUALITY, COMPLETENESS, AND/OR ACCURACY OF THE DATA. EVERY EFFORT HAS BEEN MADE TO ENSURE THE ACCURACY OF THE DATA INCLUDED IN THIS WORK, BUT THE INFORMATION IS DYNAMIC IN NATURE AND IS SUBJECT TO CHANGE WITHOUT NOTICE. ESRI AND THE DATA VENDOR(S) ARE NOT INVITING RELIANCE ON THE DATA, AND ONE SHOULD ALWAYS VERIFY ACTUAL DATA AND INFORMATION. ESRI DISCLAIMS ALL OTHER WARRANTIES OR REPRESENTATIONS, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ESRI AND THE DATA VENDOR(S) SHALL ASSUME NO LIABILITY FOR INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES, EVEN IF ADVISED OF THE POSSIBILITY THEREOF.

#### **U. S. GOVERNMENT RESTRICTED/LIMITED RIGHTS**

Any software, documentation, and/or data delivered hereunder is subject to the terms of the License Agreement. In no event shall the U.S. Government acquire greater than RESTRICTED/LIMITED RIGHTS. At a minimum, use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in FAR §52.227-14 Alternates I, II, and III (JUN 1987); FAR §52.227-19 (JUN 1987) and/or FAR §12.211/12.212 (Commercial Technical Data/Computer Software); and DFARS §252.227-7015 (NOV 1995) (Technical Data) and/or DFARS §227.7202 (Computer Software), as applicable. Contractor/Manufacturer is ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

ESRI and the ESRI globe logo are trademarks of ESRI, registered in the United States and certain other countries; registration is pending in the European Community. ArcGIS, ArcInfo, ArcCatalog, ArcMap, 3D Analyst, and GIS by ESRI are trademarks and www.esri.com is a service mark of ESRI. Other companies and products mentioned herein are trademarks or registered trademarks of their respective trademark owners.



# Содержание

## 1 Добро пожаловать в модуль ArcGIS Geostatistical Analyst

- Исследовательский анализ пространственных данных 2
- Моделирование вариограммы 3
- Интерполирование поверхности и моделирование ошибки 4
- Картографирование критических значений 5
- Проверка модели и диагностика 6
- Интерполирование поверхности с использованием кокригинга 7
- Как изучать модуль Geostatistical Analyst 8

## 2 Уроки быстрого обучения

- Введение 12
- Упражнение 1: Построение поверхности с использованием параметров, предложенных по умолчанию 14
- Упражнение 2: Исследование данных 19
- Упражнение 3: Картографирование концентрации озона 26
- Упражнение 4: Сравнение моделей 38
- Упражнение 5: Картографирование вероятности превышения критического значения концентрации озона 39
- Упражнение 6: Создание окончательного варианта карты 42

## 3 Принципы геостатистического анализа

- Что такое детерминистские методы 50
- Что такое геостатистические методы 53
- Проработка проблемы 54
- Основные принципы, лежащие в основе методов геостатистики 59
- Моделирование вариограммы 61
- Кригинг 74
- Модуль Geostatistical Analyst 78



## **4 Исследовательский анализ пространственных данных (ESDA)**

Что такое Исследовательский анализ пространственных данных (ESDA)?	82
Исследовательский анализ пространственных данных (ESDA)	83
Инструменты Исследовательского анализа пространственных данных (ESDA)	84
Изучение распределения данных	95
Изучение распределения ваших данных	98
Поиск глобальных и локальных выпадающих значений в наборе данных	99
Определение глобальных и локальных выпадающих значений	101
Определение глобальных трендов	103
Определение глобальных трендов	105
Изучение пространственной автокорреляции и вариации по направлениям	106
Изучение пространственной структуры и вариации по направлениям	108
Изучение ковариации между несколькими наборами данных	109
Изучение пространственной ковариации между несколькими наборами данных	111

## **5 Детерминистские методы интерполяции пространственных данных**

Как работает интерполяция по методу взвешенных расстояний	114
Как работает интерполяция по методу глобального полинома	120
Создание карты с использованием метода глобального полинома	122
Как работает интерполяция по методу локальных полиномов	123
Создание карты с использованием интерполяции по методу локальных полиномов	125
Как работает интерполяция с использованием радиальных базисных функций	126
Создание карты с использованием интерполяции на основе радиальных базисных функций	129

## **6 Построение поверхности с использованием методов геостатистики**

Что такое геостатистические методы интерполяции?	132
Изучение различных моделей кригинга	133
Изучение типов результирующих поверхностей	135

Создание карты по методу кригинга с использованием параметров, предложенных по умолчанию	136
Изучение преобразований и трендов	137
Что такое ординарный кригинг?	138
Создание карты с использованием ординарного кригинга	139
Что такое простой кригинг?	143
Создание карты с использованием простого кригинга	144
Что такое универсальный кригинг?	150
Создание карты с использованием универсального кригинга	151
Что такое пороговые значения?	153
Что такое индикаторный кригинг?	154
Создание карты с использованием индикаторного кригинга	155
Что такое вероятностный кригинг?	156
Создание карты с использованием вероятностного кригинга	157
Что такое дизъюнктивный кригинг?	159
Создание карты с использованием дизъюнктивного кригинга	160
Что такое кокригинг?	165
Создание карты с использованием кокригинга	166

## **7 Использование аналитических инструментов при построении поверхностей**

Исследование пространственной структуры: вариография	168
Моделирование вариограмм и функций ковариации	175
Определение размера области поиска соседства	181
Определение размера области поиска соседства	185
Выполнение проверки и перекрестной проверки	189
Выполнение перекрестной проверки для оценки выбранных параметров	193
Оценка протокола решений с использованием проверки	195
Сравнение моделей	197
Сравнение моделей	199
Моделирование распределений и определение методов преобразований	200
Использование преобразований (логарифмического, по методу Vox–Сох и арксинуса)	204

Использование преобразований по методу нормальных меток	205
Проверка на двумерное нормальное распределение	206
Проверка двумерного распределения	209
Применение декластеризации для данных, отобранных с различной густотой	211
Декластеризация данных, отобранных с различной густотой	214
Вычитание трендов из данных	216
Удаление глобальных и локальных трендов из данных: вычитание тренда	218

## **8 Отображение геостатистических слоев и управление ими**

Что такое геостатистический слой?	220
Добавление слоев	222
Работа со слоями на карте	223
Управление слоями	224
Просмотр геостатистических слоев в ArcCatalog	225
Отображение геостатистического слоя	227
Изменение символики геостатистического слоя	229
Классификация данных	230
Классификация данных	233
Определение масштабов, при которых будет отображаться геостатистический слой	235
Сохранение и экспорт геостатистических слоев	237

## **9 Дополнительные инструменты геостатистического анализа**

Изменение параметров геостатистического слоя: свойства метода	240
Интерполирование значений для заданных точек	241
Выполнение проверки для геостатистического слоя, созданного на основе поднабора данных	243

## **Приложение А 247**

## **Приложение В 275**

# Добро пожаловать в модуль ArcGIS Geostatistical Analyst

# 1

## В ЭТОЙ ГЛАВЕ

- Исследовательский анализ пространственных данных
- Моделирование вариограммы
- Интерполирование поверхности и моделирование ошибок
- Картографирование критических значений
- Проверка модели и диагностика
- Интерполирование поверхности с использованием кокригинга
- Как изучать модуль Geostatistical Analyst

Вы приступаете к изучению расширения к ArcGIS компании ESRI® - модуля Geostatistical Analyst, предназначенного для усовершенствованного моделирования поверхности с использованием детерминистских и геостатистических методов. Модуль Geostatistical Analyst расширяет возможности ArcMap за счет появления дополнительных инструментов, предназначенных для исследовательского анализа пространственных данных, а также Мастера операций геостатистики, который поможет вам в процессе построения статистически достоверной поверхности. Поверхности, создаваемые с помощью модуля Geostatistical Analyst, могут быть впоследствии использованы в моделях ГИС и для визуализации, в том числе с использованием таких расширений ArcGIS, как ArcGIS Spatial Analyst и 3D Analyst.

Модуль Geostatistical Analyst - революционное средство, поскольку он помогает навести мост между геостатистикой и ГИС. В течение долгого времени вы могли пользоваться инструментами геостатистики, но никогда прежде эти инструменты не были интегрированы в среду ГИС. Такая интеграция важна, поскольку первые профессионалы ГИС смогут количественно описать качество своих моделей путем измерения статистической ошибки интерполированных поверхностей.

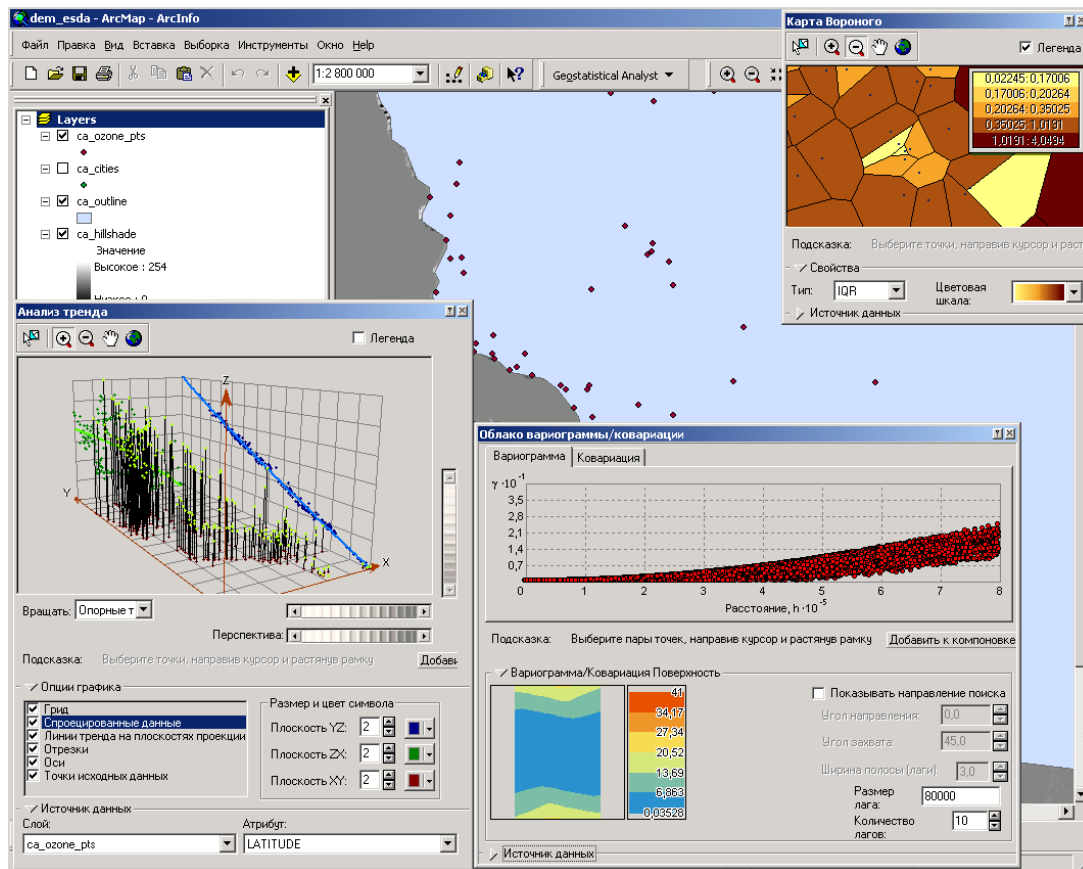
Построение поверхности с использованием модуля Geostatistical Analyst включает три ключевых этапа (проиллюстрированных на следующих страницах):

- Исследовательский анализ пространственных данных
- Структурный анализ (вычисление и моделирование свойств поверхности в соседних точках)
- Интерполирование поверхности и оценка результатов

Программное обеспечение включает серию легких в использовании инструментов и мастеров операций, которые проведут вас по каждому из этих этапов. Оно также включает целый ряд уникальных инструментов для статистического анализа пространственных данных.

# Исследовательский анализ пространственных данных

Используя измеренные значения в опорных точках, с помощью модуля Geostatistical Analyst можно интерполировать значения в других точках в пределах данной территории, для которых измерения не проводились. Инструменты исследовательского анализа пространственных данных, включенные в модуль Geostatistical Analyst, применяются для оценки статистических свойств данных, таких как изменчивость пространственных данных, их зависимость и глобальные тренды.

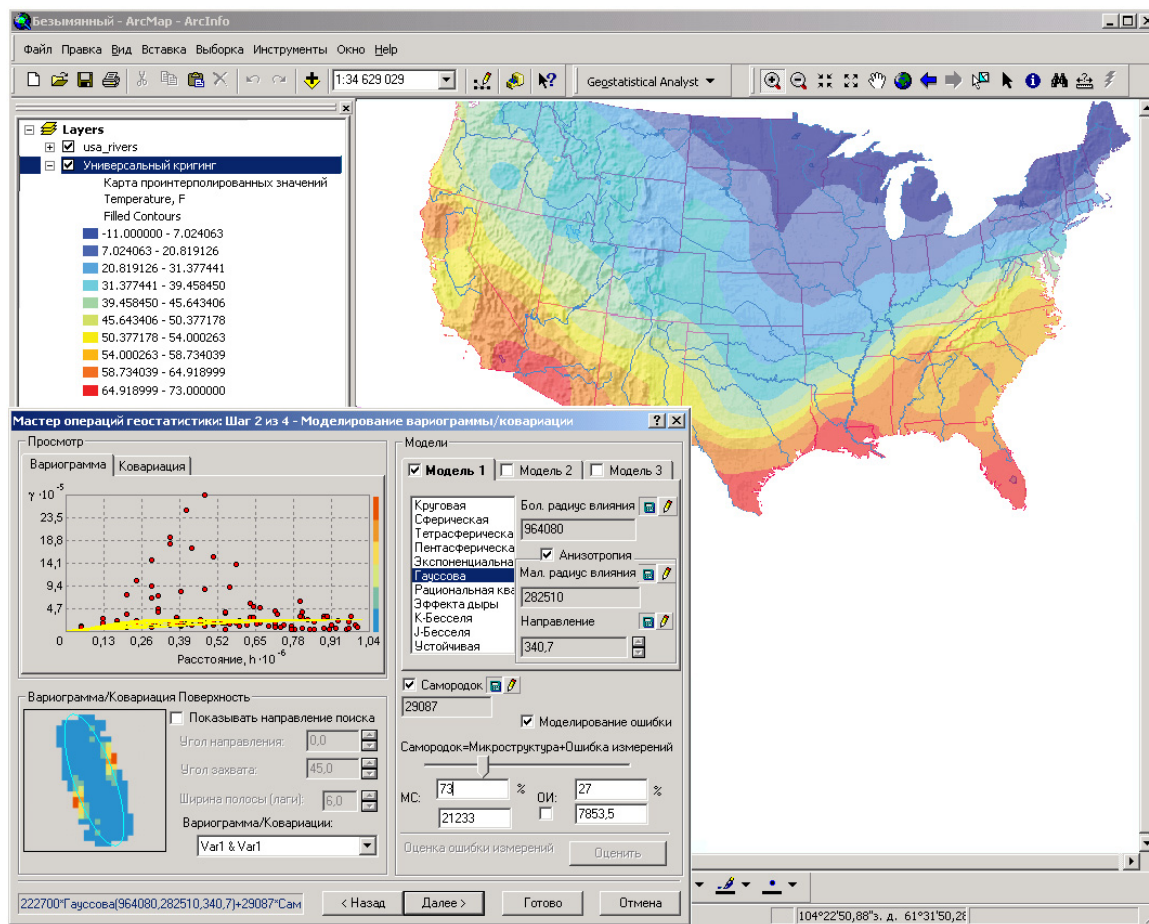


Для изучения свойств значений концентраций озона, измеренных на станциях мониторинга в Карпатах, использован ряд инструментов исследовательского анализа пространственных данных.



# Моделирование вариограммы

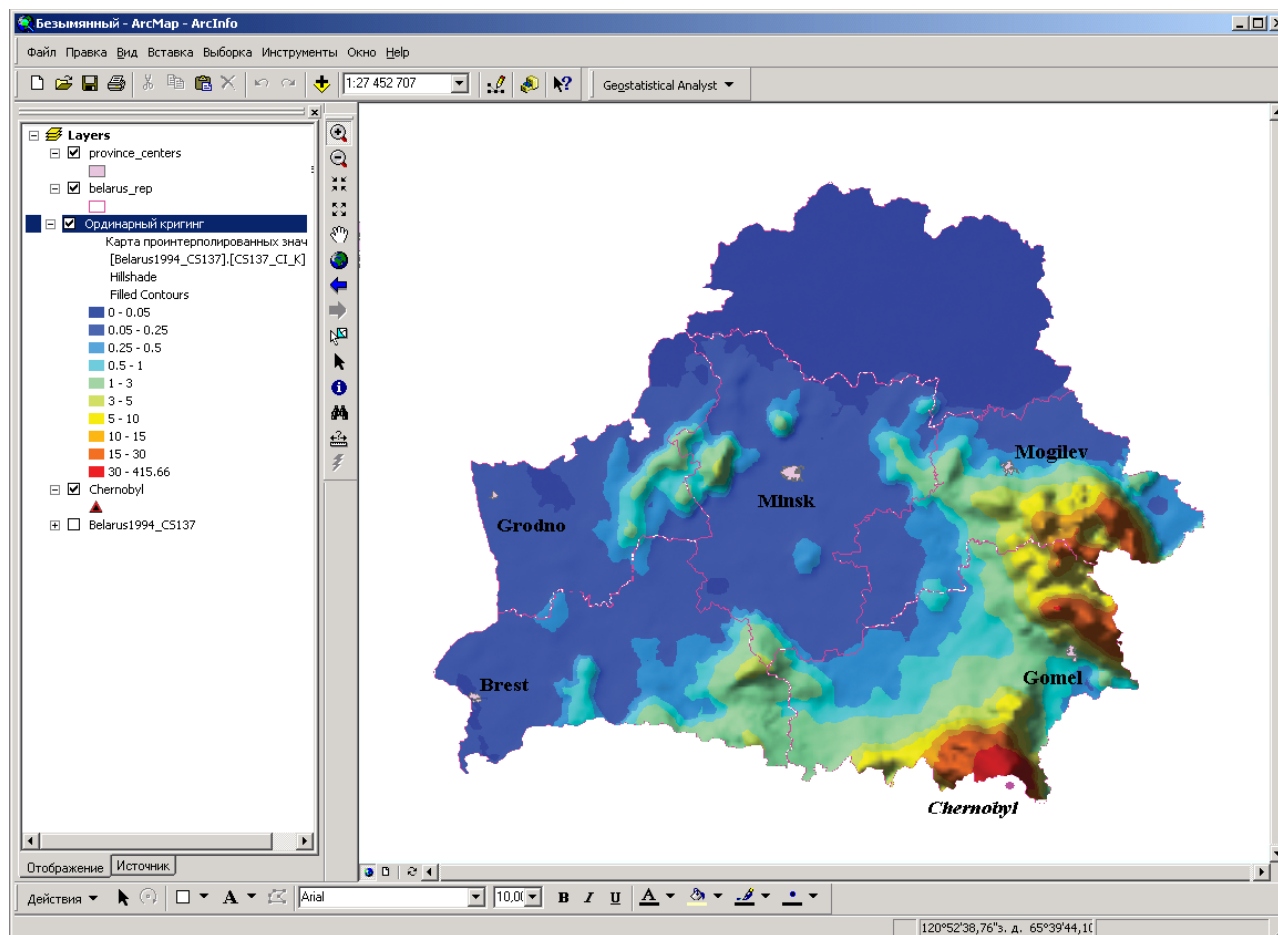
Геостатистический анализ данных происходит в два этапа: 1) моделирование вариограммы или ковариации для анализа свойств поверхности, и 2) кригинг. В модуле Geostatistical Analyst возможно использование целого ряда методов, основанных на кригинге, включая методы ординарного, простого, универсального, индикаторного, вероятностного и дизъюнктивного кригинга.



На рисунке проиллюстрированы две стадии геостатистического анализа данных. На первом этапе для построения модели распределения зимних температур по территории США был использован Мастер построения вариограммы/ковариации. Затем эта модель была использована для составления карты распределения температур.

# Интерполирование поверхности и моделирование ошибки

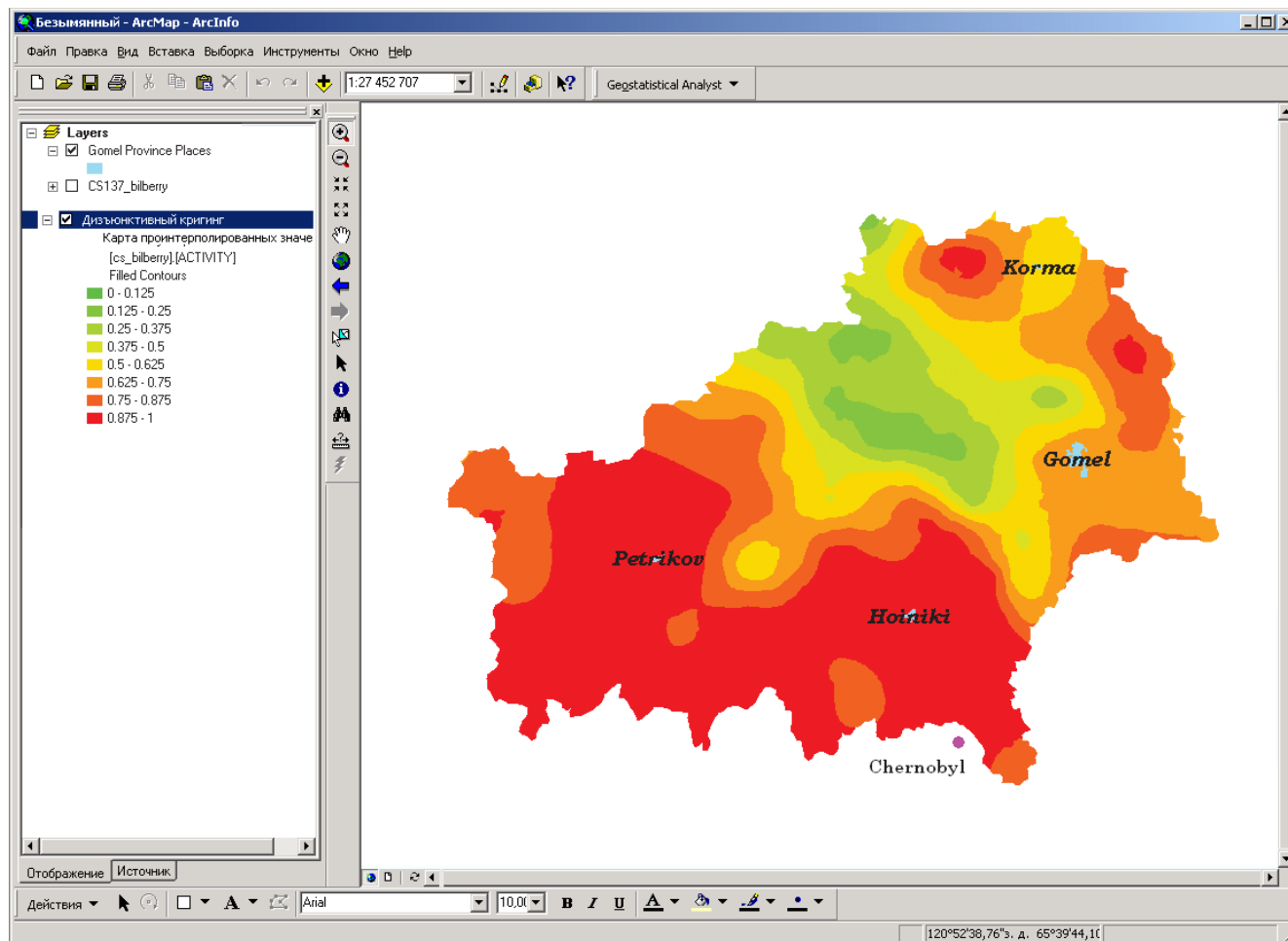
С использованием модуля Geostatistical Analyst могут быть созданы различные картографические слои, включая карты проинтерполированных значений, карты квантилей, карты вероятностей, и карты стандартной ошибки интерполяции.



В данном примере модуль Geostatistical Analyst был использован для создания карты предполагаемых значений уровня загрязнения почв радиоактивным цезием в Белоруссии после аварии на Чернобыльской атомной электростанции.

# Картографирование критических значений

Карты вероятностей могут быть использованы для определения участков, где значения могут превысить критические пороговые величины.

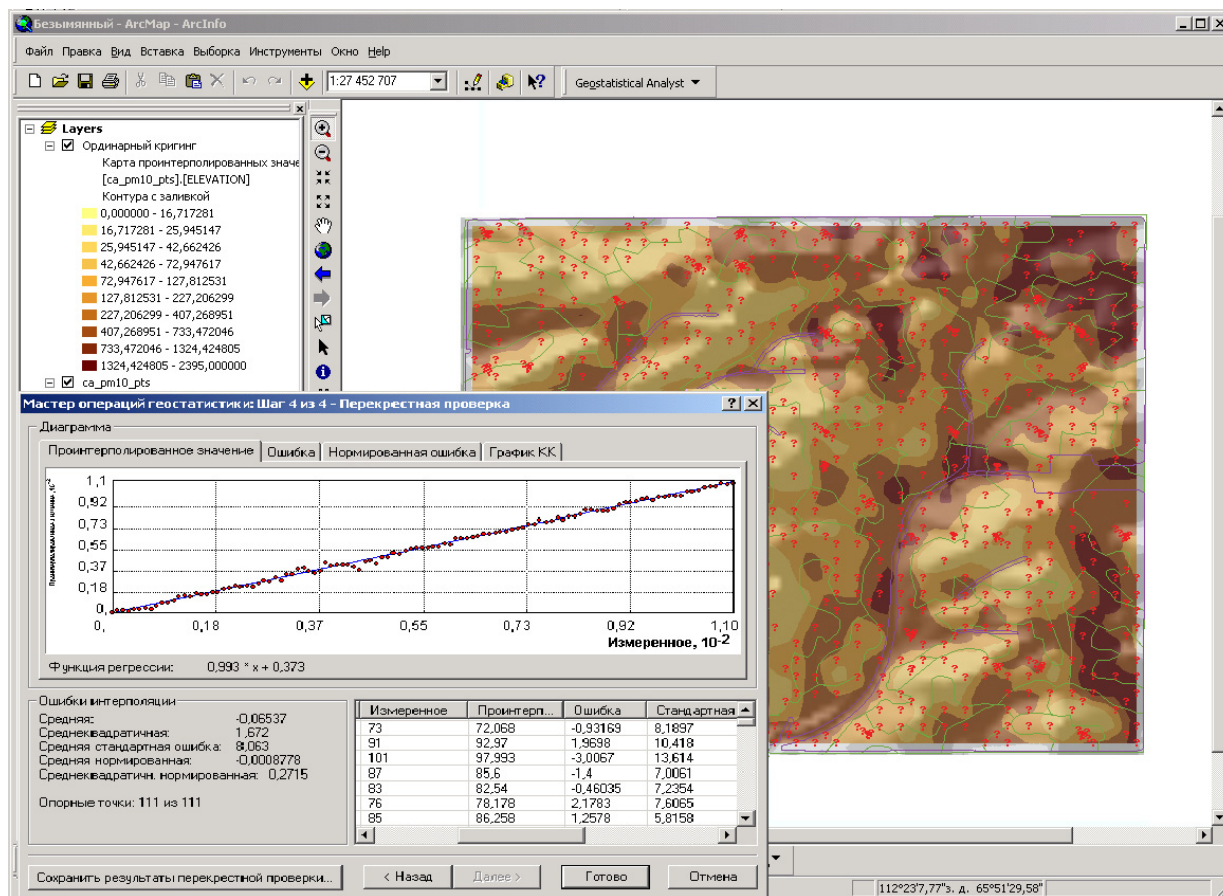


Для участков, показанных темно-оранжевым и красным цветом, вероятность того, что содержание радиоактивного цезия в лесных ягодах превышает верхний допустимый уровень (пороговую величину), составляет 62.5%.



# Проверка модели и диагностика

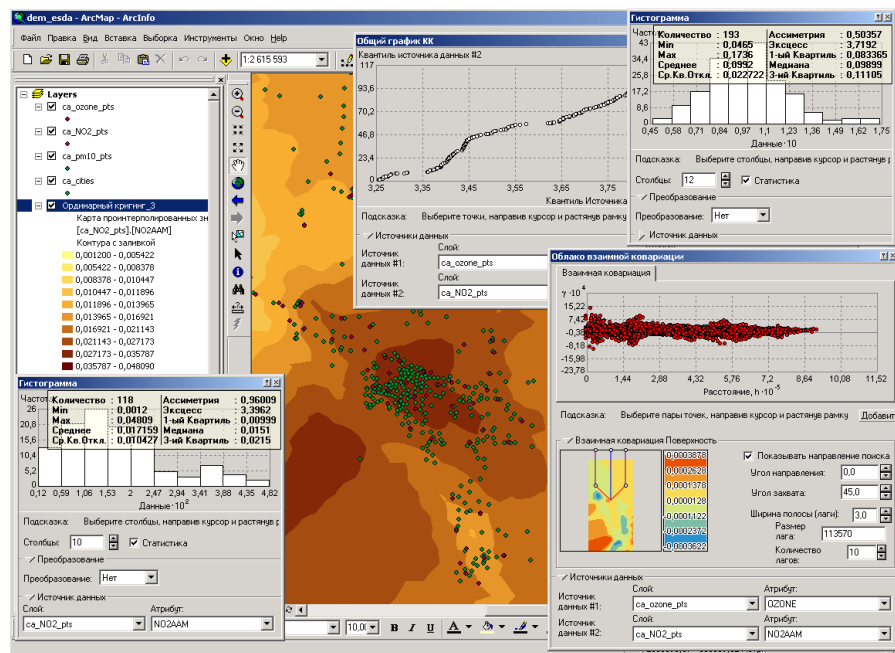
Исходные данные могут быть разбиты на два поднабора. Первый поднабор данных может быть использован для создания модели интерполяции. Вычисленные значения могут быть затем сопоставлены с известными значениями в оставшихся точках с использованием инструмента проверки достоверности модели.



Мастер операций проверки позволяет оценить модель, использованную для интерполяции содержания органического вещества в почве на ферме в штате Иллинойс.

# Интерполирование поверхности с использованием кокригинга

Кокригинг, усовершенствованный метод моделирования поверхности, включенный в модуль Geostatistical Analyst, может быть использован для улучшения качества интерполяции поверхности для одной переменной путем учета значений других переменных, при условии наличия пространственной корреляции между этими переменными.



В данном примере, инструменты исследовательского анализа пространственных данных использованы для изучения пространственной корреляции между концентрациями озона (первая переменная) и двуокиси азота (вторая переменная) в Калифорнии. Поскольку переменные пространственно коррелированы, кокригинг может использовать данные по концентрации двуокиси азота для совершенствования модели интерполяции при картографировании содержания озона в атмосфере.

Кроме того, модуль Geostatistical Analyst содержит целый ряд уникальных инструментов для совершенствования методов интерполяции, включая инструменты: преобразования данных; вычитания трендов из данных с использованием интерполяции по методу локальных полиномов; выявления параметра смещения в модели взаимной ковариации; моделирования ошибки для определения соотношения вариации на микроуровне и ошибок измерений; исследования данных на соответствие двумерному нормальному распределению; оптимального выбора области поиска соседства; составления карт квантилей.

# Как изучать модуль Geostatistical Analyst

Если вы впервые сталкиваетесь с понятием геостатистики, помните, что для того чтобы получить немедленный результат, вам необязательно знать все о модуле Geostatistical Analyst. Начните изучение модуля Geostatistical Analyst с чтения Главы 2, ‘Уроки быстрого обучения’. В этой главе приведены примеры задач, которые вы можете решать с использованием модуля Geostatistical Analyst. Она может служить прекрасной исходной точкой, если вы задумываетесь над тем, как подойти к решению ваших задач, связанных с пространственным анализом. Модуль Geostatistical Analyst поставляется с данными, используемыми в уроках быстрого обучения, поэтому шаг за шагом вы сможете выполнить эти упражнения на вашем компьютере.

Если вы предпочитаете перейти сразу к эксперименту со своими данными, воспользуйтесь Главой 5, ‘Детерминистские методы интерполяции пространственных данных’, и Главой 6, ‘Построение поверхности с использованием методов геостатистики’, для изучения соответствующих понятий и описания шагов, которые следует предпринять при решении конкретной задачи.

## Поиск ответов на вопросы

Как и для большинства людей, ваша цель состоит в том, чтобы выполнить поставленную задачу при минимальных затратах времени и усилий на изучение программного обеспечения. Вы бы хотели иметь интуитивный, легкий в использовании программный продукт, с помощью которого можно получить немедленный результат без чтения документации. В то же время, если у вас возникнут вопросы, вы захотите получить на них быстрый ответ, чтобы иметь возможность выполнить поставленную задачу до конца. Эта книга и предназначена для того, чтобы вы могли при необходимости получить ответы на возникающие вопросы.

В этой книге дано описание задач геостатистического анализа - от базовых до сложных - которые вы можете решать с помощью модуля Geostatistical Analyst. Хотя вы можете и прочесть эту книгу от начала до конца, вы все же будете использовать ее больше как справочную литературу. Если вы хотите знать, как выполнить определенную задачу, например, определить выпадающие (экстремальные) значения в ваших данных, вы можете просто заглянуть в содержание или предметный указатель. В этой книге вы найдете краткое описание того, как шаг за шагом выполнить вашу задачу. Некоторые главы содержат также подробную информацию, которую вы можете изучить, если хотите получить представление о тех концепциях, которые стоят за решением тех или иных задач. Вы также можете обратиться к словарю терминов, помещенному в этой книге, если вы встречаете незнакомые термины по геостатистике или хотите освежить память.

## Об этой книге

Эта книга составлена таким образом, чтобы помочь вам выполнить геостатистический анализ, предоставив вам информацию об основных понятиях и научив вас тому, как выполнять операции, необходимые для решения ваших геостатистических проблем. Темы, затрагиваемые в Главе 2, предполагают, что вы знакомы с основами геоинформационных систем (ГИС) и обладаете базовыми знаниями по ArcGIS. Если вы новичок в ГИС или в ArcMap, желательно, чтобы вы потратили некоторое время на изучение книг “Начало работы с ArcGIS. Руководство пользователя” и “ArcMap. Руководство пользователя”, которые вы получили в пакете документации к ArcGIS. Нет необходимости в прочтении этих книг до того, как вы продолжите изучение данного учебного пособия; просто используйте их как справочную литературу.

В Главе 3 приводятся основные принципы геостатистики, позволяющие вам понять различные методы интерполяции и то, как они работают. В Главе 4 приводится описание различных инструментов ESDA (Исследовательского анализа пространст-

венных данных), которые помогут вам в изучении и понимании ваших данных. В Главе 5 рассказывается о детерминистских методах интерполяции. В Главе 6 обсуждаются различные методы геостатистики, а в Главе 7 перечислены инструменты, которыми вы можете пользоваться при выполнении интерполяции. В Главе 8 дано описание различных инструментов отображения и управления, которые могут быть использованы в работе с геостатистическими слоями. Глава 9 затрагивает целый ряд других понятий и задач, связанных с геостатистическим анализом. В Приложении А приведены подробные математические формулы, относящиеся к функциям и методам, включенным в модуль Geostatistical Analyst. И, наконец, словарь содержит определения различных терминов геостатистики, использованных в этой книге.

### **Как получить помощь на компьютере.**

В дополнение к этой книге, вы можете использовать онлайн-овую систему справки ArcMap для изучения модуля Geostatistical Analyst и ArcMap. Чтобы научиться пользоваться системой справки, обратитесь к книге “*ArcMap. Руководство пользователя*”.

### **Контакты с ESRI**

Если для технической поддержки вам необходим контакт с ESRI, воспользуйтесь картой регистрации и поддержки продукта, которую вы получили вместе с модулем ArcGIS Geostatistical Analyst, или обратитесь к пункту ‘Получение технической поддержки’ в разделе ‘Дополнительная помощь’ в десктоп-системе помощи ArcGIS. Чтобы узнать больше о модуле Geostatistical Analyst и продукте ArcGIS, вы можете посетить сайты ESRI [www.esri.com](http://www.esri.com) и [www.arconline.esri.com](http://www.arconline.esri.com).

### **Решения ESRI в области образования**

ESRI предоставляет возможность получения образования в области геоинформатики, применения ГИС и использования ГИС-технологий. Вы можете выбирать между курсами, проводимыми инструкторами, Web-курсами и самостоятельным изучением учебников, чтобы найти образовательные решения, которые наилучшим образом отвечают вашим возможностям. Для получения подробной информации, посетите сайт [www.esri.com/education](http://www.esri.com/education).





# Уроки быстрого обучения

## 2

### В ЭТОЙ ГЛАВЕ

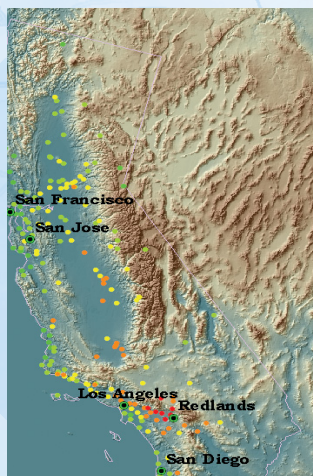
- Упражнение 1: Построение поверхности с использованием параметров, предложенных по умолчанию
- Упражнение 2: Исследование данных
- Упражнение 3: Картографирование концентрации озона
- Упражнение 4: Сравнение моделей
- Упражнение 5: Картографирование вероятности превышения критического значения концентрации озона
- Упражнение 6: Создание окончательного варианта карты

С помощью модуля Geostatistical Analyst вы можете легко построить непрерывную поверхность или карту по опорным точкам, хранящимся в точечном слое, по данным растрового слоя, или воспользовавшись центроидами полигонов. В опорных точках могут быть измерены значения высоты, глубины водоема и уровни загрязнения, как, например, в учебных примерах данного пособия. Модуль Geostatistical Analyst используется совместно с ArcMap и предоставляет обширный набор инструментов для построения поверхностей, которыми вы можете воспользоваться для визуализации, анализа и понимания пространственных явлений.

### Сценарий обучения

Агентство США по защите окружающей среды осуществляет мониторинг концентрации озона в атмосфере над территорией Калифорнии. По всей территории штата на станциях мониторинга выполняются замеры концентраций озона. Расположе-

ние станций показано на рисунке. Уровни концентраций озона известны для всех станций, но мы заинтересованы также в том, чтобы знать значения концентрации в любой точке Калифорнии. Однако, из практических соображений и значительных затрат, станции мониторинга не могут быть размещены повсеместно. Модуль Geostatistical Analyst предоставляет инструменты, которые позволяют выполнить наилучшее из возможных интерполирование значений путем изучения взаимосвязей между всеми опорными точками, а также строить непрерывную поверхность концентрации озона, вычислять стандартные ошибки (неопределенность) интерполяции и определять вероятность того, что в некоторых точках превышены критические значения.



# Введение

Данные, которые понадобятся вам для упражнений, находятся на инсталляционном диске модуля Geostatistical Analyst. Наборы данных были любезно предоставлены Департаментом воздушных ресурсов Калифорнии.

Использованы следующие наборы данных:

Набор данных	Описание
ca_outline	Картографическая основа штата Калифорния
ca_ozone_pts	Контрольные точки измерения концентраций озона (в промилле - ppm)
ca_cities	Местоположение основных городов Калифорнии
ca_hillshade	Карта рельефа Калифорнии с отмывкой

Набор данных по озону (ca\_ozone\_pts) содержит данные за 1996 г., отражающие максимальное значение концентрации озона в промилле (ppm), выбранное из средних значений наблюдений за восьмичасовой период. (Измерения проводились в суточном режиме и группировались в блоки по восемь часов.) Исходные данные были изменены для целей приведенных в пособии упражнений и не могут рассматриваться как точные данные.

Воспользовавшись значениями содержания озона в опорных точках (данными измерений), вы построите две непрерывные поверхности (карты), интерполирующие значения концентрации озона для каждой точки штата Калифорния. Для составления первой карты вы используете опции, предложенные по умолчанию, и увидите, как просто можно построить поверхность по опорным точкам. При составлении второй карты вы сможете учесть пространственные отношения, существующие между точками. При составлении этой второй карты вы будете использовать инструменты исследовательского анализа ESDA, что позволит вам изучить ваши данные. Вы также познакомитесь с некоторыми опциями геостатистики, которыми вы можете воспользо-

ваться при построении поверхности, например, вычитанием трендов и моделированием пространственной автокорреляции. Применение инструментов ESDA и работа с геостатистическими параметрами позволит вам построить более точную поверхность.

Во многих случаях учитываются не фактические значения показателей, а только те, которые представляют риск для здоровья, то есть те значения, которые превышают некий уровень токсичности. В таких случаях должны приниматься немедленные меры. Третья поверхность, которую вы построите, будет отражать вероятность превышения критического значения концентрации озона.

Для данных упражнений принято, что критическое значение составляет 0.12 ppm. Если максимальная средняя концентрация озона за любой восьмичасовой отрезок времени в году превышает данное значение, то данная точка должна находиться под тщательным наблюдением. Вы будете использовать модуль Geostatistical Analyst, чтобы спрогнозировать вероятность появления значений, превышающих этот показатель.

Это учебное пособие разделено на отдельные темы, которые разработаны с учетом того, что возможности модуля Geostatistical Analyst будут изучаться вами постепенно, с удобной для вас скоростью. Для получения дополнительной помощи, обратитесь к онлайн-овой системе помощи ArcMap или к книге *Using ArcMap (Использование ArcMap)*.

- Упражнение 1 поможет вам начать работу с модулем Geostatistical Analyst и выполнить построение поверхности концентрации озона с использованием параметров, предложенных по умолчанию, что позволит продемонстрировать возможности данного модуля.
- Упражнение 2 познакомит вас с процессом изучения данных перед построением поверхности с целью определения экстремальных значений в данных и трендов.

- В Упражнении 3 вы построите вторую поверхность, которая учитывает пространственные взаимосвязи, выявленные в Упражнении 2, и вносит поправки в поверхность, построенную в первом упражнении. В этом упражнении также приведены основные понятия геостатистики.
- В Упражнении 4 показано, как сравнить две поверхности, построенные в первом и третьем упражнениях с тем, чтобы решить, какая из них лучше интерполирует неизвестные значения.
- Упражнение 5 знакомит вас с процессом картографирования вероятности того, что концентрация озона превысит критическое значение, и вы построите третью поверхность.
- В Упражнении 6 показано, как окончательно оформить поверхности, созданные в упражнениях 3 и 5, в виде карты с использованием функций ArcMap.

Для того чтобы выполнить все упражнения из уроков быстрого обучения, вам понадобится несколько часов. Однако, при желании вы можете выполнять упражнения по одному, сохраняя результаты после каждого упражнения.



# Упражнение 1: Построение поверхности с использованием параметров, предложенных по умолчанию

Перед началом работы вы должны запустить ArcMap и подключить модуль Geostatistical Analyst.

## Как запустить ArcMap и подключить модуль Geostatistical Analyst

Нажмите кнопку Пуск на панели задач Windows, выберите Программы, затем ArcGIS, и щелкните по строке ArcMap. В ArcMap, выберите опцию Инструменты, затем строку Дополнительные модули, и в открывшем окне отметьте галочкой опцию Geostatistical Analyst. Нажмите Заккрыть.

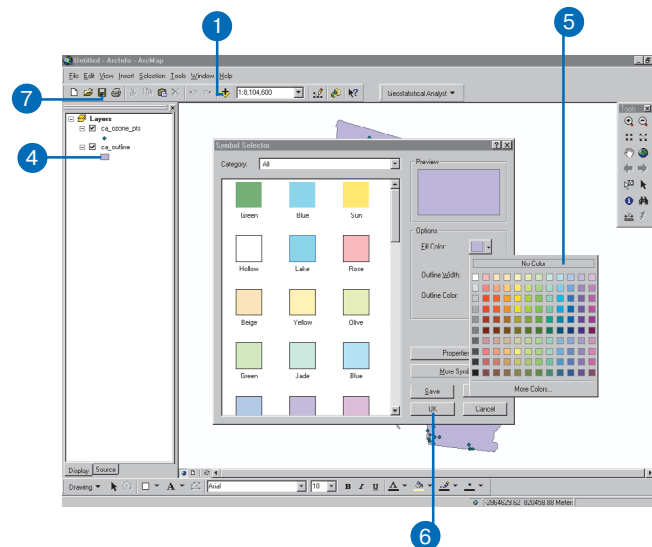
## Как добавить панель инструментов модуля Geostatistical Analyst в ArcMap

Откройте меню Вид, перейдите на строку Панели инструментов, и нажмите Geostatistical Analyst.

## Как добавить слои данных в ArcMap

После того, как вы добавили данные, вы можете использовать ArcMap для отображения данных и, при необходимости, изменения свойств каждого слоя (символов и т.д.).

1. Нажмите кнопку Добавить данные на стандартной панели инструментов.
2. Перейдите к папке, в которую вы записали учебные данные (путь, предлагаемый по умолчанию, - C:\ArcGIS\ArcTutor\Geostatistics), удерживая клавишу Ctrl, выберите наборы данных sa\_ozone\_pts и sa\_outline.
3. Нажмите Добавить.
4. Щелкните по легенде слоя sa\_outline в таблице содержания, чтобы открыть диалог Символ.
5. Откройте палетку Цвет заполнения и выберите опцию Нет цвета.



6. Нажмите OK в диалоге Выбор символа.

Теперь при отображении слоя sa\_outline виден только контур штата. Это позволит вам наложить контур на слои, которые вы будете создавать в ходе выполнения упражнений.

## Сохранение карты

Рекомендуется сохранять вашу карту после каждого упражнения.

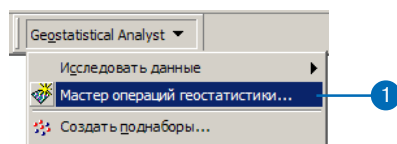
7. Нажмите кнопку Сохранить на стандартной панели инструментов.

Вам понадобится задать название карты, поскольку вы сохраняете ее первый раз (мы предлагаем название Ozone Prediction Map.mxd). В будущем для сохранения просто нажмите кнопку Сохранить.

## Построение поверхности с использованием параметров, предложенных по умолчанию

Далее вы построите (проинтерполируете) поверхность концентрации озона, используя параметры, предложенные в модуле Geostatistical Analyst по умолчанию. В качестве исходного набора данных вы возьмете набор точек с концентрациями озона (ca\_ozone\_pts) и интерполируете значения этих концентраций в тех точках, где их значения неизвестны, используя метод ординарного кригинга. Во многих диалогах вы будете нажимать кнопку Далее, таким образом принимая параметры, предложенные по умолчанию. Не обращайте внимание на элементы диалогов в этом упражнении. Каждый из этих диалогов вы будете открывать снова и снова при выполнении следующих упражнений. Цель этого упражнения - построить поверхность, воспользовавшись параметрами, предложенными по умолчанию.

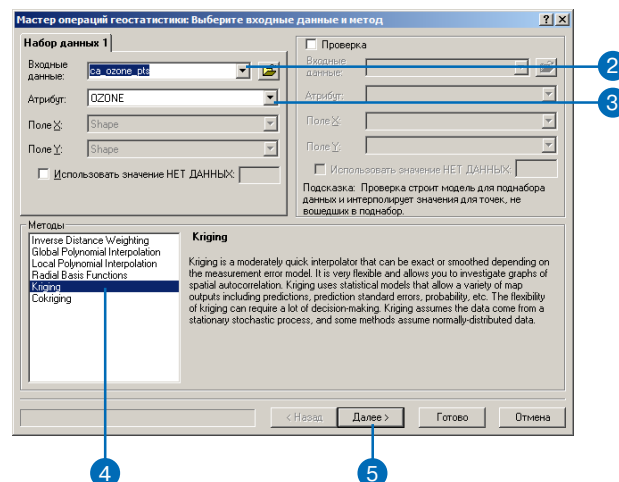
1. На панели инструментов Geostatistical Analyst выберите опцию Мастер операций геостатистики.



2. В строке Входные данные выберите ca\_ozone\_pts.
3. В строке Атрибуты выберите атрибут OZONE.
4. В диалоге Методы выберите Кригинг.
5. Нажмите Далее.

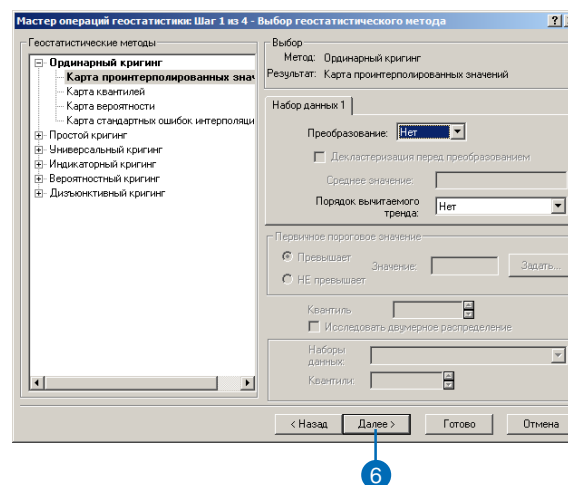
По умолчанию в диалоге Выбор геостатистического метода будут выбраны опции Ординарный Кригинг и Карта проинтерполированных значений.

Обратите внимание, что выбрав метод моделирования поверхности концентрации озона, вы можете нажать кнопку Готово на этом шаге, что позволит построить поверхность с использованием параметров, предложенных по умолчанию. Однако, шаги с 6

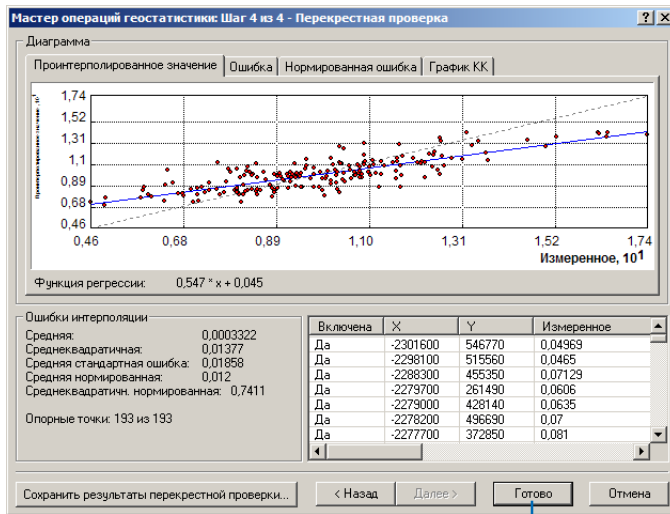


по 10 позволят вам познакомиться с некоторыми другими диалогами.

6. Нажмите Далее в диалоге Выбор геостатистического метода.







Диалог Перекрестная проверка дает вам некое представление о том, “насколько хорошо” модель интерполирует значения в искоемых точках. В Упражнении 4 вы узнаете, как использовать график перекрестной проверки и научитесь понимать статистику.

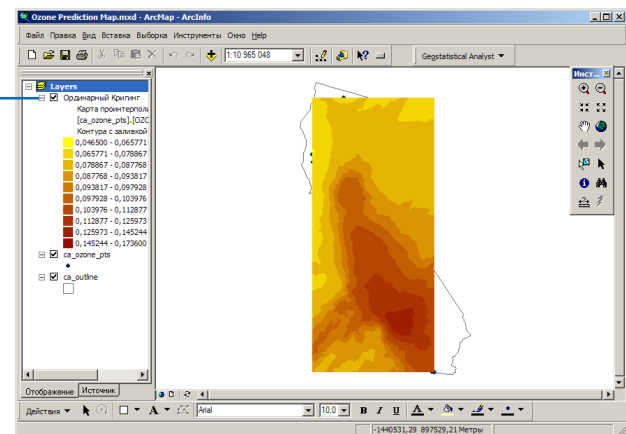
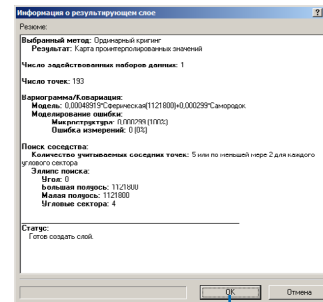
#### 9. Нажмите Готово.

Диалог Информация о результирующем слое резюмирует информацию об использованном при построении результирующей поверхности методе (и связанных с ним параметрах).

#### 10. Нажмите OK.

Карта проинтерполированных значений концентраций озона отобразится как верхний слой в таблице содержания.

#### 11. Выберите этот слой в таблице содержания, затем щелкните на его названии и измените название слоя на “По умолчанию”.



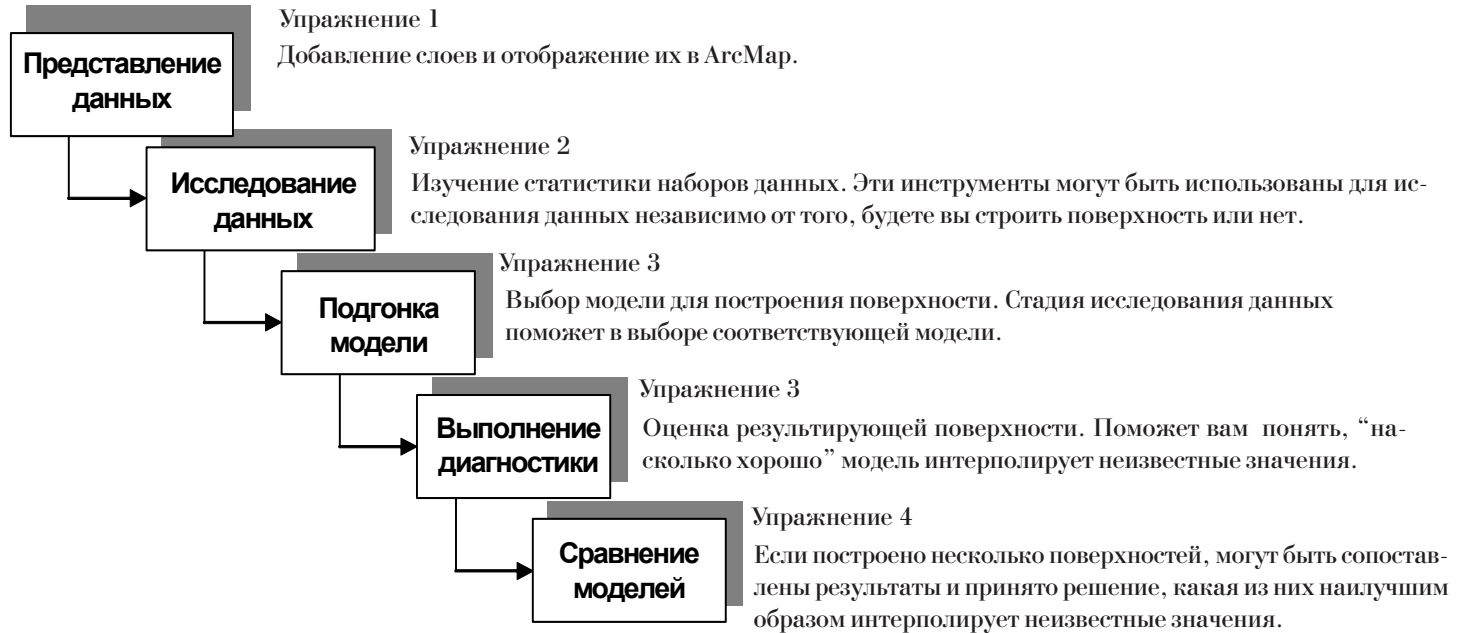
Изменение названия поможет вам отличить этот слой от того, который вы создадите в Упражнении 4.

#### 12. Сохраните карту, воспользовавшись кнопкой стандартной панели инструментов ArcMap.

Обратите внимание, что проинтерполированная поверхность простирается и над океаном. В Упражнении 6 вы научитесь ограничивать поверхность проинтерполированных значений пределами штата Калифорния.

## Методология подбора поверхности

Вы построили карту концентрации озона и выполнили Упражнение 1 данного учебника. Несмотря на то, что задача создания карты (поверхности) с использованием модуля Geostatistical Analyst достаточно проста, важно структурировать процесс в соответствии с диаграммой, приведенной ниже.



В следующих упражнениях учебника вы будете следовать данной структуре. Кроме того, в Упражнении 5 вы построите поверхность для тех точек, значения которых превышают заданное пороговое значение, а в Упражнении 6 создадите окончательно оформленный вариант карты с результатами выполненного в ходе упражнений анализа.

Обратите внимание, что вы уже выполнили первый шаг этого процесса - отображение данных - в Упражнении 1. В Упражнении 2, вы будете исследовать данные.

## Упражнение 2: Исследование данных

В этом упражнении вы будете исследовать ваши данные. В соответствии со структурой процесса, предложенной на предыдущей странице, для принятия оптимальных решений при построении поверхности вы сначала должны изучить свой набор данных, чтобы лучше понимать его. При изучении данных вы должны обратить особое внимание на очевидные ошибки в исходной выборке, которые могут сильно повлиять на результирующую поверхность с проинтерполированными значениями, а также исследовать распределение данных, выявить глобальные тренды и т.д.

Модуль Geostatistical Analyst предлагает множество инструментов для изучения данных.

В этом учебнике вы будете изучать данные тремя способами:

- Исследовать распределение данных.
- Выявлять тренды в ваших данных, если они есть.
- Изучать пространственную автокорреляцию и влияние по направлениям.

Если вы закрыли карту после выполнения Упражнения 1, в меню Файл выберите Открыть. В диалоге в окне Искать в: перейдите к папке, в которой вы сохранили документ карты (Ozone Prediction Map.mxd). Нажмите Открыть.

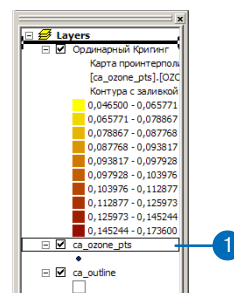
### Изучение распределения данных

#### Гистограмма

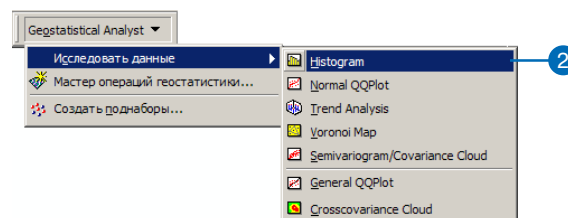
Методы интерполяции, используемые для построения поверхности, дают лучшие результаты при нормальном распределении данных (кривая в форме колокола). Если ваши данные распределены асимметрично (неравномерно), вы можете воспользоваться возможностью преобразования данных, чтобы привести их к нормальному распределению. Таким образом, важно понимать распределение ваших данных перед тем, как вы приступи-

те к построению поверхности. Инструмент Гистограмма создает гистограммы частот для атрибутов в наборе данных, позволяя вам изучать одномерное (для одной переменной) распределение данных по каждому атрибуту. Далее, вы изучите распределение озона для слоя `ca_ozone_pts`.

1. Выделите слой `ca_ozone_pts`, переместите его в верхнюю часть таблицы содержания, затем за ним поместите слой `ca_outline`.

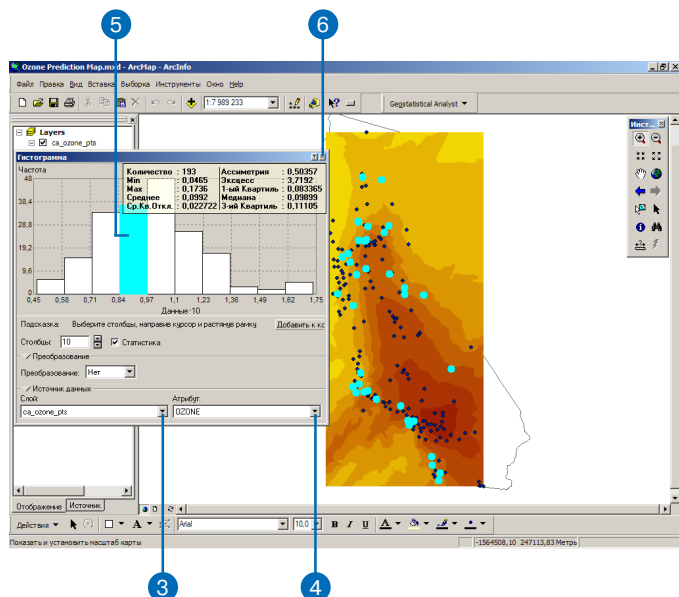


2. Выберите панель инструментов Geostatistical Analyst, перейдите к опции Исследовать данные и затем выберите строку Гистограмма.



Вы можете изменить размер окна диалога Гистограмма таким образом, чтобы видеть одновременно и карту, как это показано на рисунке на стр. 20.

3. В окне Слой выберите sa\_ozone\_pts.
4. В окне Атрибут выберите OZONE.



Распределение концентрации озона показано на гистограмме, на которой все значения объединены в 10 классов. Высота столбцов пропорциональна количеству данных с определенными значениями, попадающих в каждый класс.

В целом, важными характеристиками распределения являются центральное значение, его размах и симметрия. Для быстрой проверки характера распределения: если среднее и медиана имеют приблизительно одно и то же значение, это является одним из подтверждений того, что данные подчиняются закону нормального распределения.

Гистограмма, приведенная выше, показывает, что данные являются унимодальными (одновершинными) и достаточно симметричными. Похоже, что данные имеют распределение, близ-

кое к нормальному. Правый хвост распределения указывает на присутствие относительно небольшого количества элементов выборки с большими значениями концентрации озона.

5. Выберите столбик гистограммы со значениями концентрации озона от 0.162 до 0.175 ppm.

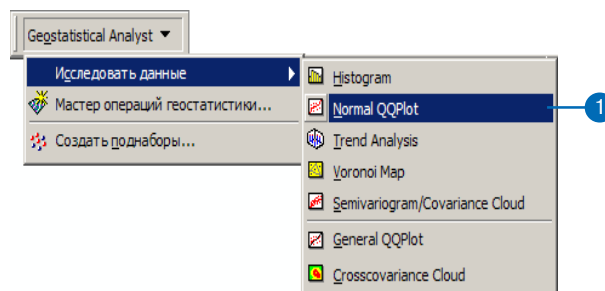
Элементы выборки, имеющие такие значения, будут выделены на карте. Обратите внимание, что эти точки расположены в районе Лос-Анджелеса.

6. Закройте диалоговое окно.

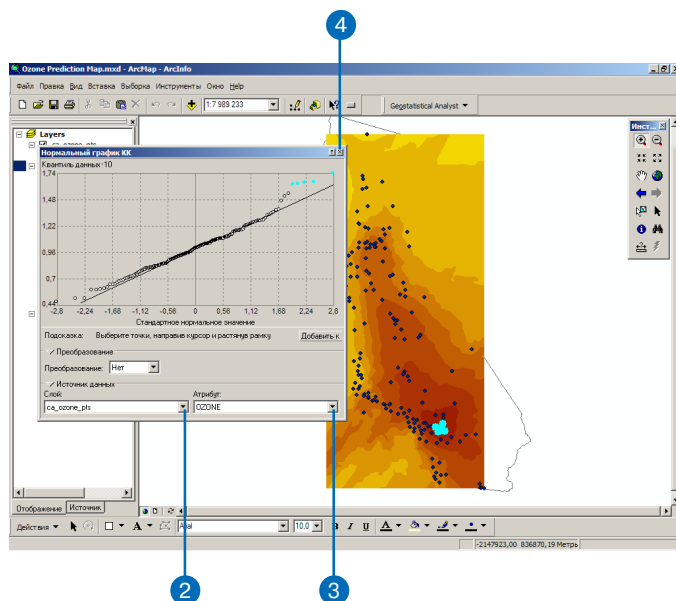
## Нормальный график КК (Квантиль-квантиль)

График КК позволяет вам сравнить распределение ваших данных со стандартным нормальным, предлагая другие параметры измерения нормальности данных. Чем более точно по точкам можно построить прямую линию, тем ближе распределение к нормальному.

1. На панели инструментов Geostatistical Analyst, выберите опцию Исследовать данные, а затем строку Нормальный график КК.



2. В окне Слой выберите ca\_ozone\_pts.
3. В окне Атрибут выберите OZONE.



В целом график КК представляет собой график, на котором квантили из двух распределений расположены по двум осям координат (даны относительно друг друга). Для двух идентичных распределений график КК будет представлять собой прямую линию. Таким образом, нормальность распределения данных по концентрации озона можно проверить, используя на графике квантили этих данных и квантили стандартного нормального распределения. На графике КК, приведенном выше, вы можете видеть, что график очень близок к прямой линии. Главное отклонение от этой линии приходится на высокие значения концентрации озона (которые были выделены цветом на гистограмме, поэтому выделены и на этом графике).

Если данные не отражают нормального распределения ни на гистограмме, ни на графике КК, перед тем как использовать кригинг в качестве метода интерполяции, может возникнуть необходимость преобразовать данные, чтобы привести их к нормальному распределению.

4. Закройте диалоговое окно.

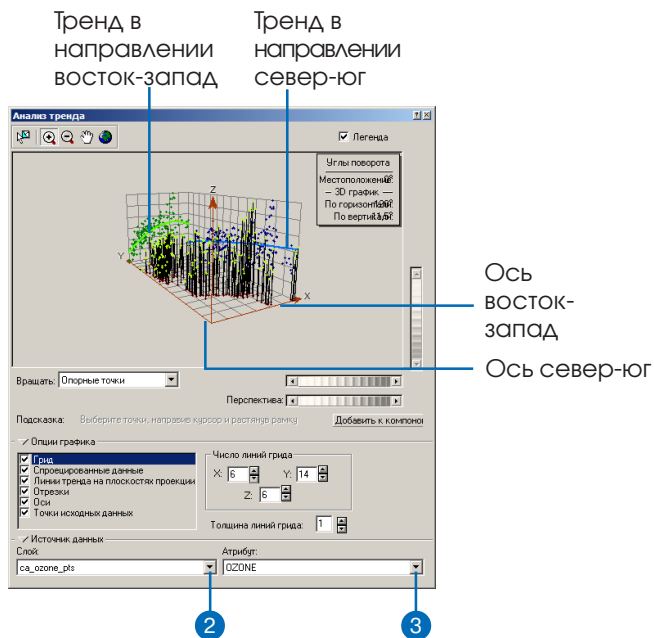
## Определение глобальных трендов в ваших данных

Если в ваших данных существует тренд, он представляет собой неслучайную (детерминистскую) составляющую поверхности, которая может быть описана какой-либо математической формулой. Например, пологий склон может быть представлен как плоскость. Долина может быть описана более сложной формулой (полиномом второго порядка), которая строит U-образную форму. Эта формула может дать описание требуемой поверхности. Однако во многих случаях формула слишком сглажена, чтобы она могла точно описать поверхность, поскольку ни склон не является правильной плоскостью, ни долина не имеет правильной “U”-образной формы. Если вы считаете, что поверхность тренда неадекватно отражает поверхность, с которой вы работаете, вы можете вычесть его и продолжить анализ, моделируя остатки или значения в опорных точках после вычитания тренда. При моделировании остатков, вы будете анализировать изменения на поверхности в узком интервале. Это та часть, которая не может быть точно описана плоской или U-образной поверхностью.

Инструмент Анализ тренда позволяет вам определить отсутствие/наличие трендов в исходном наборе данных.

1. На панели инструментов Geostatistical Analyst выберите Исследовать данные, а затем опцию Анализ тренда.
2. Из списка Слой выберите ca\_ozone\_pts.





### 3. В окне Атрибут выберите OZONE.

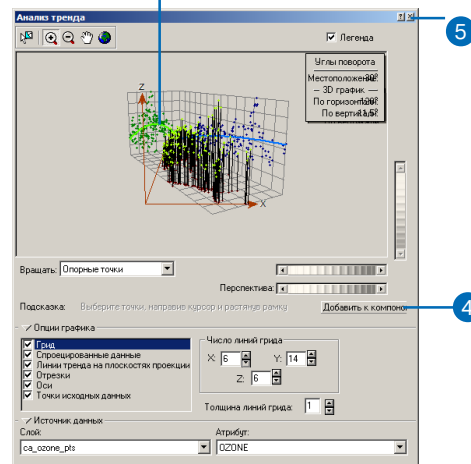
Каждый вертикальный столбец (отрезок) на диаграмме анализа тренда представляет местоположение, а его высота пропорциональна значению каждой точки из набора данных. Точки проецируются на перпендикулярные плоскости, соответствующие направлениям восток-запад и север-юг. Через спроецированные точки проведена линия (полином), наилучшим образом описывающая их расположение, которая моделирует тренды в определенных направлениях. Если линия близка к прямой, это означает, что в данных нет тренда. Однако, если вы посмотрите на светло-зеленую линию на верхнем рисунке, вы можете увидеть, что она начинается с низких значений и увеличивается по мере продвижения на восток, пока не начнет снижаться. Это указывает на то, что данные, возможно, имеют сильный тренд в направлении восток-запад и более слабый в направлении север-юг.

### 4. На шкале прокрутки Перспектива доведите угол поворота до 30°, перемещая указатель влево.

Такое вращение позволит вам лучше рассмотреть форму тренда в направлении восток-запад. Вы можете видеть, что проекция показывает перевернутую U-образную кривую. Поскольку тренд имеет U-образную форму, для описания глобального тренда хорошо подойдет полином второго порядка. Хотя, тренд и показан на плоскости проекции, соответствующей направлению восток-запад, поскольку мы повернули точки на 30°, действительный тренд имеет направление с северо-востока на юго-запад. Выявленный тренд возможно вызван тем фактом, что загрязнение является низким на побережье, а при движении внутрь материка появляются крупные населенные пункты, количество которых снова уменьшается в горах. Вы будете вычитать эти тренды в Упражнении 4.

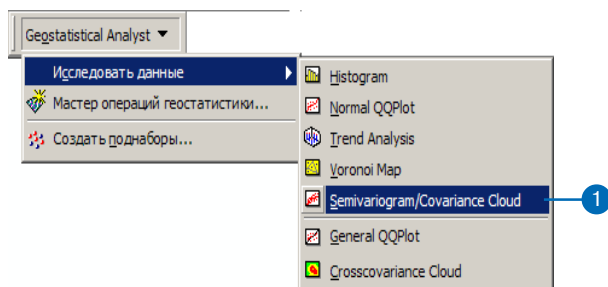
### 5. Закройте диалоговое окно.

### U-образный тренд

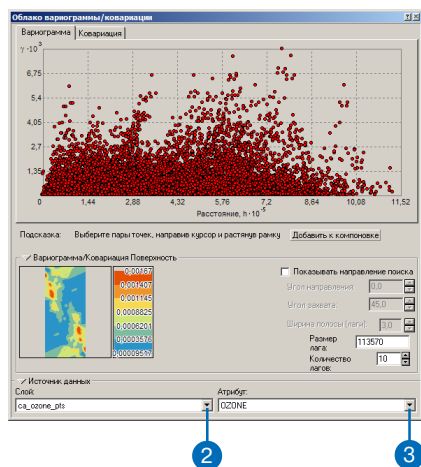


## Понимание пространственной автокорреляции и влияний по направлениям

1. На панели инструментов Geostatistical Analyst выберите Исследовать данные, а затем опцию Облако вариограммы/ковариации.



2. В окне Слой выберите sa\_ozone\_pts.
3. В окне Атрибут выберите OZONE.

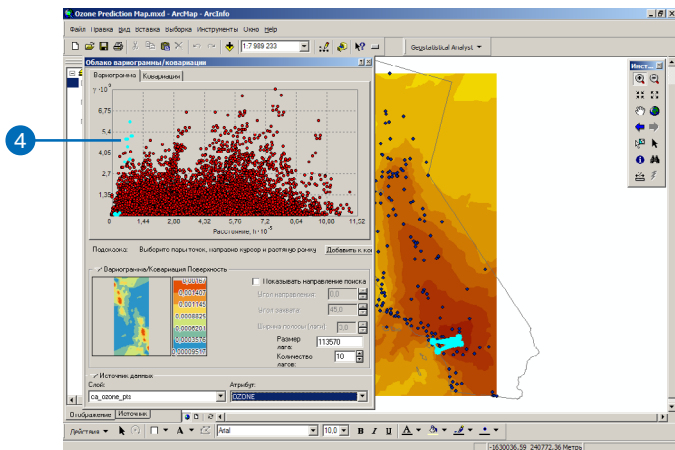


Опция Облако вариограммы/ковариации позволяет вам изучить пространственную автокорреляцию между опорными точками. Пространственная автокорреляция опирается на предположение, что объекты, которые расположены ближе всего друг к другу, в наибольшей степени похожи. Опция Облако вариограммы/ковариации позволяет вам изучить эту взаимосвязь. Для этого по оси y откладывается значение вариограммы, равное квадрату разности значений для каждой пары точек, а по оси x - расстояние, на которое пары точек отстоят друг от друга.

Каждая красная точка на графике Облако вариограммы/ковариации представляет пару значений. Поскольку близкие по расположению точки должны быть больше всего похожи, на вариограмме близкие местоположения (крайние левые значения по оси x) должны иметь низкие значения на вариограмме (находиться в нижней части оси y). По мере того, как расстояние между парами точек увеличивается (величины смещаются вправо по оси x), значения вариограммы должны также увеличиваться (расти по оси y). Однако, при определенном расстоянии облако выравнивается, что указывает на то, что между парами точек за пределами этого расстояния нет корреляции.

При рассмотрении вариограммы, если вы обнаружите, что для некоторых точек, расположенных близко друг к другу (около нуля по оси x) значение вариограммы выше, чем вы ожидали (большое значение по оси y), вы должны исследовать эти пары точек, чтобы определить, не являются ли данные в этих точках неточными.

4. Инструментом выбора укажите область, в которую попадают эти точки, чтобы выделить их цветом. (Используйте приведенную диаграмму как пример. Вам необязательно выделять те же точки, что показаны на диаграмме.)



Пары точек, выбранные на вариограмме, выделены цветом на карте; отрезками показаны связи, указывающие на то, что точки образуют пару.

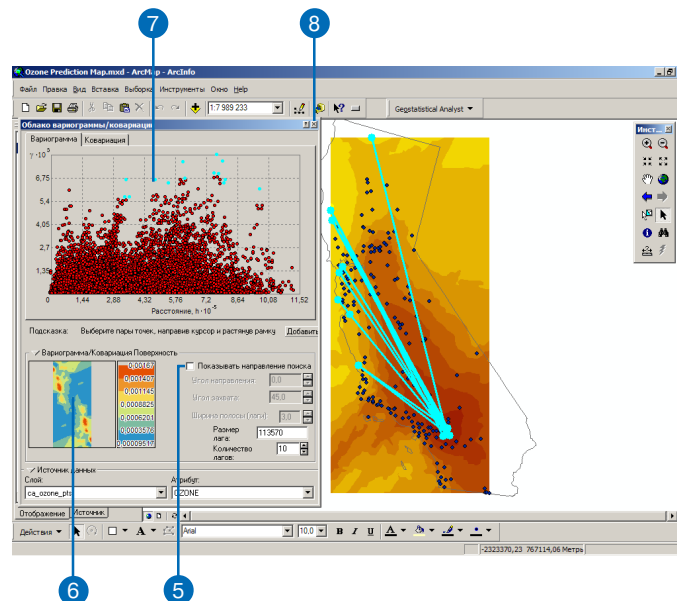
Существует множество причин, по которым значения отличаются в большей степени между точками в районе Лос-Анджелеса и другими территориями. Одна из них - то, что в районе Лос-Анджелеса больше, чем в других районах, машин, что несомненно приводит к большому загрязнению и способствует появлению более высоких концентраций озона в районе Лос-Анджелеса.

Помимо глобальных трендов, обсуждавшихся в предыдущем разделе, на данные могут также оказывать влияние направленные воздействия. Причины этого могут быть неизвестны, но они могут быть описаны статистически. Эти направленные воздействия могут влиять на точность поверхности, которую вы будете строить в следующем упражнении. Однако, если вы знаете, что такие воздействия существуют, модуль Geostatistical Analyst предоставляет инструменты, позволяющие учитывать их в процессе построения поверхности. Для изучения влияний по направлениям по облаку вариограммы, воспользуйтесь инструментом Направление поиска.

- Отметьте галочкой опцию Показывать направление поиска.
- Выберите и перемещайте указатель направления на любой угол.

Направление указателя определяет, какие пары опорных точек будут отображаться на вариограмме. Например, если стрелка указывает в направлении восток-запад, только те пары опорных точек, которые расположены к востоку или к западу друг от друга, будут отображаться на вариограмме. Это позволяет исключить те пары точек, которые вас не интересуют, и изучить влияние, которые данные испытывают в различных направлениях.

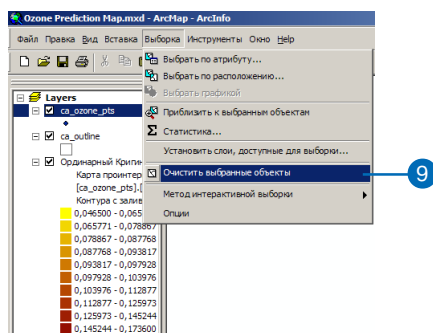
- Инструментом выбора выделите область с точками, имеющими наибольшие значения на вариограмме, чтобы выделить их цветом на графике и на карте. (Используйте приведенную диаграмму как пример. Вам необязательно выделять те же точки, что показаны на диаграмме, или использовать то же направление поиска.)



Обратите внимание, что большинство связанных местоположений (представляющих пары точек на карте), независимо от расстояния, соответствуют одной из опорных точек, расположенных в районе Лос-Анджелеса. Если учитывать другие пары точек, удаленных на любое расстояние, то можно увидеть, что высокие значения на вариограмме имеют не только пары точек, одна из которых расположена в районе Лос-Анджелеса, а другая - ближе к побережью. Многие из пар, в которых одна из точек расположена в районе Лос-Анджелеса, а другая - во внутренних районах штата, также имеют высокие значения на вариограмме. Это происходит из-за того, что значения концентрации озона в районе Лос-Анджелеса намного превышают любые значения в других регионах Калифорнии.

8. Закройте диалоговое окно.

9. В меню Выборка выберите опцию Очистить выбранные объекты, чтобы снять выделение с точек на карте.



В этом упражнении мы узнали, что:

1. Данные по концентрациям озона имеют распределение, близкое к нормальному. Они являются унимодальными (одновершинными) и практически симметричными относительно линии среднего/медианы, что видно на гистограмме.
2. Нормальный график КК (Квантиль-квантиль) еще раз подтвердил, что данные имеют нормальное распределение, поскольку точки на графике образуют почти прямую линию; следовательно, нет необходимости в преобразовании данных.
3. Воспользовавшись инструментом Анализ тренда, вы увидели, что данные характеризуются наличием тренда, и что тренд наилучшим образом может быть описан полиномом второго порядка в направлении с юго-востока на северо-запад (угол 330 градусов).
4. Из диаграммы Облако вариограммы/ковариации мы узнали, что высокие значения концентрации озона в Лос-Анджелесе дают высокие значения вариации как для близлежащих точек, так и для удаленных от данного района точек.
5. Поверхность вариограммы показывает, что в данных существует пространственная автокорреляция.

Зная, что в наборе данных нет экстремальных (или ошибочных) значений опорных точек и что распределение является близким к нормальному, вы с уверенностью можете переходить к интерполяции поверхности. Кроме того, вы сможете построить более точную поверхность, поскольку знаете, что в ваших данных существует тренд, который вы сможете учесть в вычислениях.

## Упражнение 3: Картографирование концентрации озона

Для картографирования концентрации озона в Упражнении 1 вы использовали параметры, предложенные по умолчанию. Однако, вы не учитывали статистические свойства выборки. Например, при изучении данных в Упражнении 2, выяснилось, что в данных существует тренд. Этот факт может быть использован в процессе интерполяции.

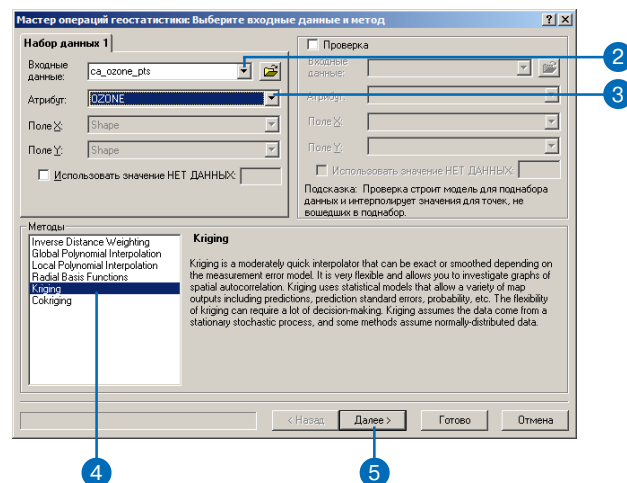
В этом упражнении вы:

- Усовершенствуете карту концентрации озона, созданную в Упражнении 1.
- Познакомитесь с некоторыми основными концепциями геостатистики.

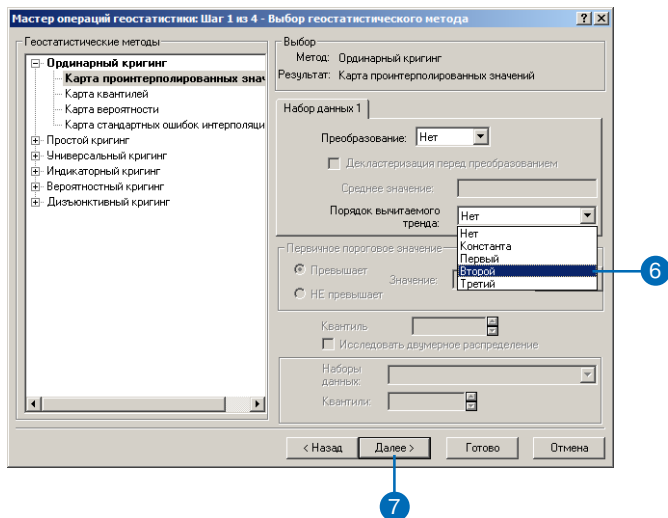
Вы снова воспользуетесь для интерполяции методом ординарного кригинга. Кроме того, для получения более точных проинтерполированных значений, в вашу модель вы включите тренд.

1. На панели инструментов Geostatistical Analyst выберите опцию Мастер операций геостатистики...
2. В окне Слой выберите `ca_ozone_pts`.
3. В окне Атрибут выберите `OZONE`.
4. В окне Методы выберите Кригинг.
5. Нажмите Далее.

По умолчанию будут выбраны Ординарный кригинг и Карта проинтерполированных значений).



При исследовании данных в Упражнении 2 вы узнали, что в ваших данных есть глобальный тренд. После обработки инструментом анализа тренда, вы определили, что тренд наилучшим образом может быть описан полиномом второго порядка и что тренд имеет направление с юго-востока на северо-запад. Этот тренд может быть представлен математической формулой и вычтен из данных. После вычитания тренда статистический анализ будет выполнен для остатков или той составляющей поверхности, которая соответствует вариации на микроуровне. Тренд будет автоматически добавлен обратно перед построением окончательной поверхности, и, таким образом, результат интерполяции будет более значимым. После вычитания тренда, он перестанет оказывать влияние на выполнение анализа данных, а после того, как он будет снова добавлен в данные, поверхность будет более точно представлять интерполируемые значения.

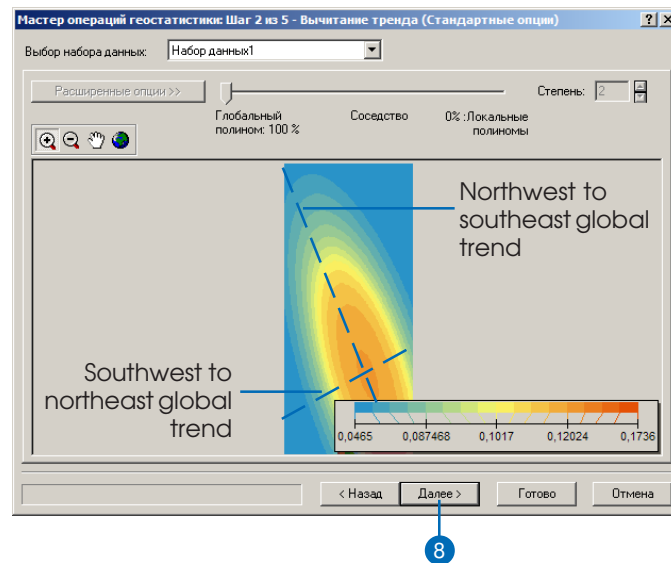


6. В диалоговом окне Выбор геостатистического метода, в окне Порядок вычитаемого тренда выберите Второй.

Для описания тренда будет использован полином второго порядка, поскольку в диалоге Анализ тренда в Упражнении 2 была получена U-образная кривая, расположенная в направлении с юго-запада на северо-восток.

7. В диалоге Выбор геостатистического метода нажмите Далее.

По умолчанию, модуль Geostatistical Analyst картографирует глобальный тренд для набора данных. Поверхность показывает наиболее быстрое изменение в направлении с юго-запада на северо-восток и более плавное изменение в направлении с северо-запада на юго-восток (что приводит к образованию эллипса).



Тренд следует вычитать только в том случае, если для этого есть основание. Тренд в направлении с юго-запада на северо-восток, характеризующий качество воздуха, может быть отнесен за счет сосредоточения озона между горами и побережьем. Высоты и преобладающее направление ветра также являются факторами относительно низких значений концентрации в горах и на побережье. Высокая концентрация населения приводит к высоким уровням загрязнения между горами и побережьем. Тренд в направлении с северо-запада на юго-восток меняется значительно медленнее благодаря высокой концентрации населения вокруг Лос-Анджелеса и уменьшению его плотности в районе Сан-Франциско. Следовательно, мы можем обоснованно вычесть (удалить) эти тренды.

8. В диалоге Вычитание тренда нажмите Далее.

## Моделирование вариограммы/ковариации

По облаку вариограммы/ковариации в Упражнении 2 вы исследовали общую пространственную автокорреляцию опорных точек. Для этого вы изучили вариограмму, которая показывает квадрат разностей значений в каждой паре точек в зависимости от расстояния между точками. Цель моделирования с помощью вариограммы/ковариации - подобрать лучшую модель, которая пройдет через точки на вариограмме (желтая линия на графике).

Вариограмма - это функция, которая связывает дисперсию (или различие) опорных точек и расстояние, на которое они отстоят друг от друга. Ее графическое представление может быть использовано для получения картины пространственной корреляции опорных точек и их соседей.

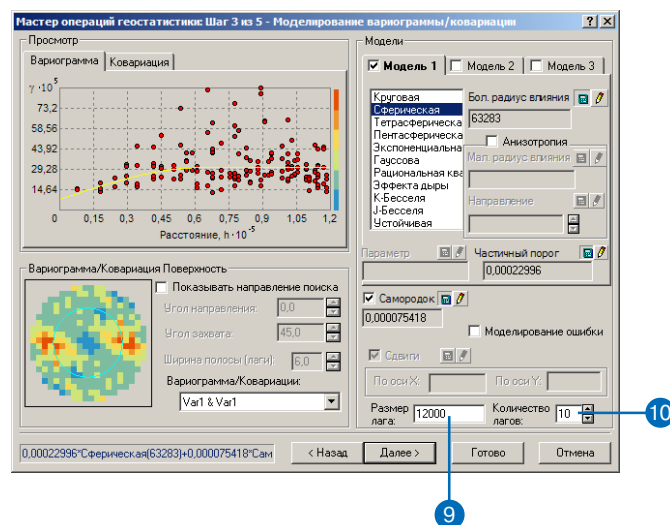
Диалоговое окно Моделирование вариограммы/ковариации позволяет вам моделировать пространственные взаимосвязи в наборе данных. По умолчанию рассчитаны оптимальные параметры для сферической модели вариограммы. Модуль Geostatistical Analyst сначала определяет оптимальный размер лага для группировки значений вариограммы. Размер лага - это размер класса расстояний между точками, в который сгруппированы пары точек, с тем чтобы сократить большое количество возможных комбинаций. Эта операция носит название биннинга (binning) - группирования по классам. Обратите внимание, что в результате биннинга на этой вариограмме отображается меньшее количество точек, чем на той, что приведена в Упражнении 2. Правильно определенный размер лага может также помочь в выявлении пространственной корреляции. Диалоговое окно отображает значения вариограммы как поверхность и как точечный график, по одной из осей которого откладывается расстояние. Затем по умолчанию строится сферическая модель вариограммы (наиболее подходящая для всех направлений) и определяются связанные с ней параметры, которые обычно носят название самородка, радиуса влияния и частичного порога.

Попробуйте подобрать вариограмму для небольших значений лагов. Можно использовать различные размеры бинов (классов) и изменить сферическую модель, предложенную по умолчанию, меняя размер лага и количество лагов.

9. Наберите новое значение размера лага, равное 12000.

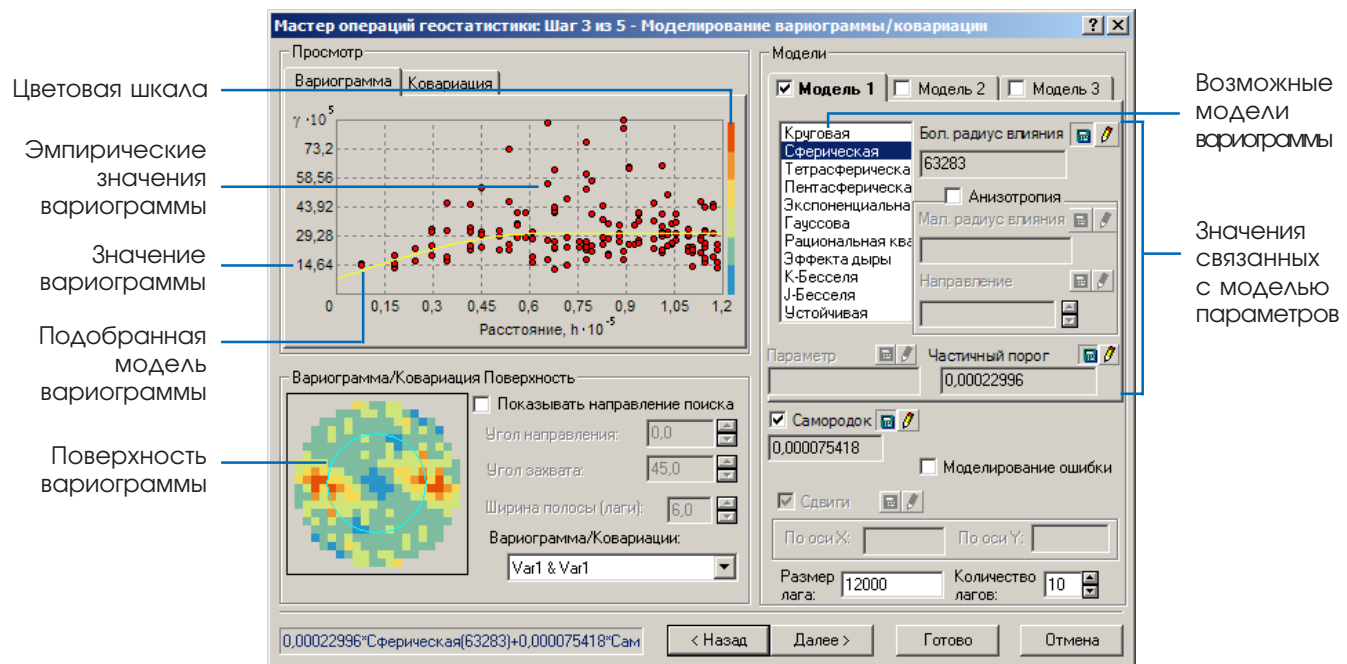
10. В окне Количество лагов введите 10.

Сокращение размера лага означает, что вы увеличиваете детальность модели, что позволяет изучать локальные различия между соседними опорными точками. Обратите внимание, что с уменьшением размера лага, подобранная вариограмма (желтая линия) быстро поднимается, а затем выравнивается. Радиус влияния (range) - это расстояние, при котором линия выравнивается. Такое выравнивание вариограммы указывает на то, что за пределами значения радиуса влияния автокорреляция небольшая.



После того, как вы вычтете тренд из данных, вариограмма будет моделировать пространственную автокорреляцию для опорных точек без учета тренда. Тренд будет автоматически добавлен в вычисления перед окончательным построением поверхности.





Цветовая шкала, которая представляет вычисленное значение вариограммы, обеспечивает прямую связь между эмпирическими значениями вариограммы на графике и теми же значениями на поверхности вариограммы. Значение каждой “ячейки” на поверхности вариограммы обозначено цветом, при этом меньшие значения показаны голубым и зеленым цветом, а более высокие значения - оранжевым и красным цветом. Среднее значение каждой ячейки поверхности вариограммы нанесено на графике вариограммы. По оси x графика вариограммы откладывается расстояние от центра ячейки до центра поверхности вариограммы. Значения вариограммы отражают различия точек. В нашем примере, вариограмма начинается с маленьких значений (объекты, расположенные поблизости, в наибольшей степени похожи) и увеличивается по мере удаления точек друг от

друга (чем дальше отстоят объекты друг от друга, тем больше они непохожи друг на друга). Обратите внимание, что в соответствии с поверхностью вариограммы различия между объектами быстрее нарастают в направлении с юго-запада на северо-восток, чем в направлении с юго-востока на северо-запад. Ранее мы вычли тренд, присутствующий в данных для всей территории. Теперь видно, что существуют составляющие автокорреляции и на более детальном уровне, которые мы и будем моделировать далее.



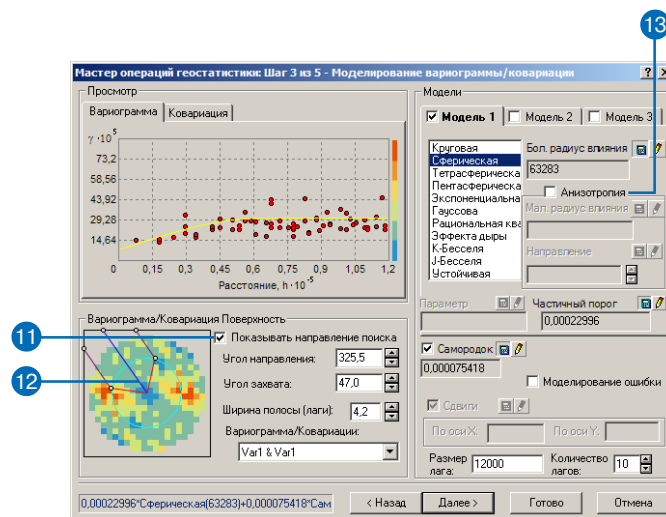
## Вариограммы по направлениям

Влияние по направлениям будет затрагивать как значение точек на вариограмме, так и подбираемую модель. В определенных направлениях близко расположенные друг к другу объекты могут быть более похожи, чем в других направлениях. Направленные влияния носят название анизотропии и могут учитываться модулем Geostatistical Analyst. Анизотропия может быть вызвана ветром, переносом, геологической структурой и многими другими процессами. Направленное влияние может быть описано статистически и учтено при составлении карты.

Вы можете исследовать различия в данных для определенного направления с помощью инструмента Направление поиска. Это позволит вам изучить направленное влияние по графику вариограммы. Эта операция не влияет на результирующую поверхность. Далее приведены шаги, которые помогут вам получить требуемый результат.

11. Отметьте галочкой опцию Показать направление поиска. Обратите внимание на сокращение числа значений вариограммы. Отображаются только те точки, которые соответствуют направлению поиска.
12. Удерживайте курсор на центральной линии инструмента Направление поиска. Меняйте направление инструмента поиска. Обратите внимание, как по мере изменения направления поиска, меняется вариограмма. Только значения поверхности вариограммы, расположенные в направлении поиска, отображаются на верхнем графике вариограммы.

Чтобы действительно учесть направленные влияния для модели вариограммы при интерполяции поверхности, вы должны вычислить анизотропную модель вариограммы или ковариации.

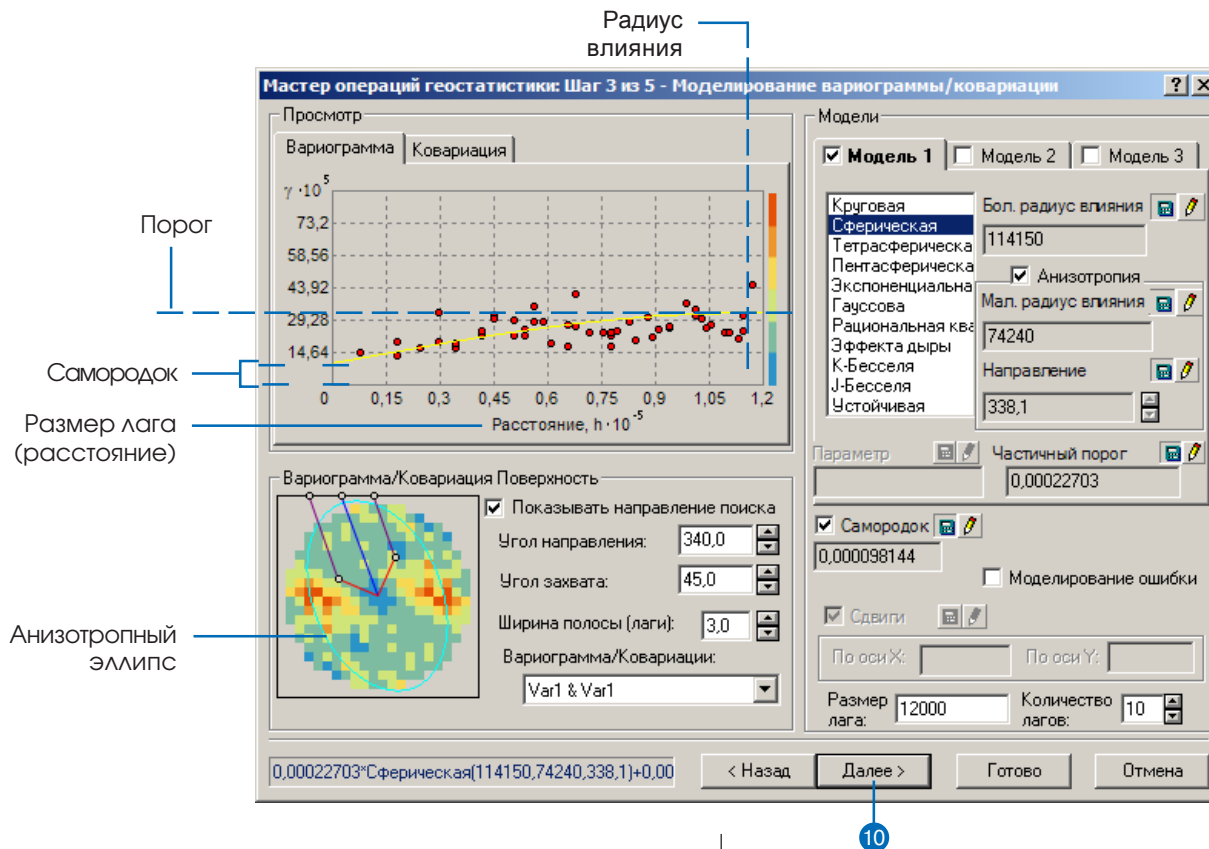


13. Включите опцию Анизотропия.

Голубой эллипс на поверхности вариограммы указывает на радиус влияния вариограммы в различных направлениях. В нашем случае большая ось эллипса направлена приблизительно с северо-северо-запада на юго-юго-восток.

Теперь анизотропия будет включена в модель с тем, чтобы учесть в результирующей поверхности различия в автокорреляции по направлениям.





ленные на расстояние лага, большее, чем радиус влияния, пространственно не коррелируют. Самородок представляет ошибку измерений и/или вариацию на микроуровне (отклонения на очень маленьком пространстве, слишком трудные для определения). Можно оценить ошибку измерений, если для каждой точки у вас есть множественные измерения, или вы можете разложить самородок на ошибку измерений и вариацию на микроуровне, выбрав опцию Моделирование ошибки (самородка).

16. Нажмите Далее.

Теперь у вас есть модель, которая описывает пространственную автокорреляцию и учитывает возможность вычитания тренда и направленное влияние в данных. Эта информация, наряду с информацией о расположении опорных точек вокруг искомой и их измеренных значениях, будет использоваться при выполнении интерполяции. Но как следует использовать в расчетах измеренные человеком значения в опорных точках?

## Поиск соседства

Обычно, чтобы ограничить данные, применяемые в интерполяции, задают область в форме окружности (или эллипса). Попадающие в заданную окрестность точки и используются при интерполяции значений в точках, в которых измерения не проводились.

Кроме того, чтобы избежать смещения в определенном направлении, окружность (или эллипс) делят на сектора, в каждом из которых выбирается одинаковое количество точек. Воспользовавшись диалогом Поиск соседства, вы можете определить количество точек (максимум 200), радиус (или малую и большую оси эллипса), и количество секторов круга (или эллипса), которые будут использованы при выполнении интерполяции.

Цвет выделенных точек в окне просмотра данных указывает на то, какие веса будут присвоены каждой точке при вычислении неизвестных значений. В нашем примере, четыре точки (красные) имеют веса более 10 процентов. Чем больше вес, тем боль-

шее влияние данная точка будет оказывать в процессе выполнения интерполяции.

17. Щелкните мышью на графическом изображении, чтобы выбрать точку, для которой будет выполняться интерполяция (эта точка будет находиться на перекрестье). Обратите внимание на то, что изменилась выборка опорных точек (наряду с присвоенными им весами). Именно эти точки будут использованы при вычислении значения в искомой точке.

18. Для целей данного учебного пособия, наберите следующие координаты в диалоге Искомая точка:

$X = -2044968$  и  $Y = 208630.37$ .

19. Отметьте галочкой опцию Форма и в окне Угол наберите 90. Обратите внимание, как меняется форма. Однако, чтобы учесть направленные влияния, измените значение угла обратно на 338.1.

Использованные точки и присвоенные веса

Сектор поиска соседства

Перекрестье определяет положение искомой точки

Периметр области поиска соседства

Предварительный просмотр поверхности или соседних точек

Количество точек, используемых в расчетах, для каждого сектора поиска

Минимальное количество точек, которое будет учитываться в каждом секторе поиска

Геометрия и количество секторов, используемых в поиске

18

17

21

20. Отключите опцию Форма — модуль Geostatistical Analyst будет использовать параметры, предложенные по умолчанию (ранее вычисленные в диалоге Вариограмма/Ковариация).

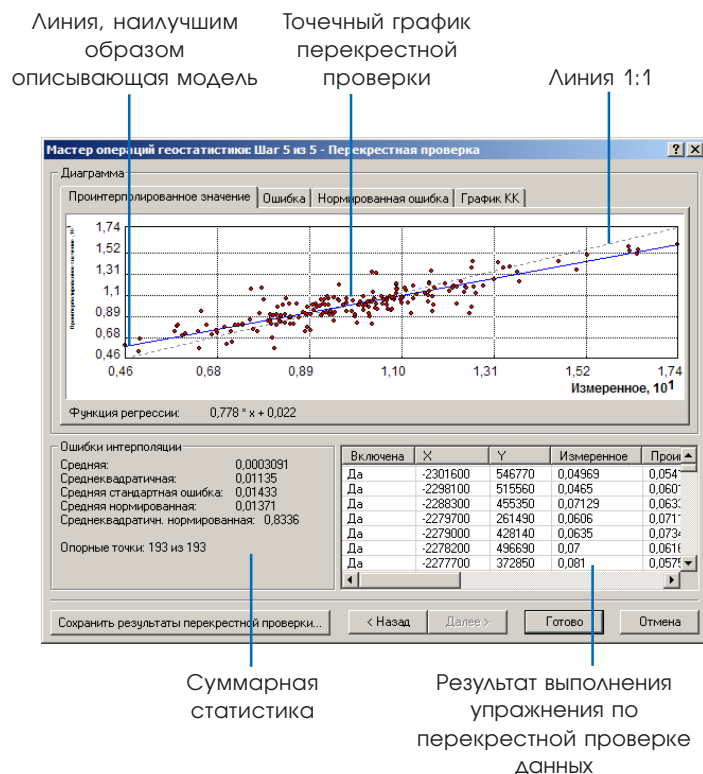
21. Нажмите Далее в диалоге Поиск соседства.

Перед тем, как перейти к построению поверхности, вы воспользуетесь диалогом Перекрестная проверка, чтобы выполнить диагностику параметров и определить, “насколько хорошей” будет ваша модель.

## Перекрестная проверка

Перекрестная проверка дает вам представление о том, “насколько хорошо” модель интерполирует неизвестные значения.

Для всех точек в ходе выполнения перекрестной проверки выполняется следующая операция: точки последовательно исключаются из выборки, затем вычисляется значение в этой точке с использованием оставшихся данных, а затем измеренное и вычисленное значение сравниваются. По суммарной статистике перекрестной проверки можно определить, подходит ли модель для составления карты.



Помимо визуализации отстояния точек от линии 1:1, для оценки поведения модели могут быть использованы статистические показатели. Цель перекрестной проверки - помочь вам принять обоснованное решение о том, какая из моделей наиболее точно интерполирует значения. Для модели, которая точно интерполирует значения, средняя ошибка должна быть близка к 0, среднеквадратичная ошибка и среднее из стандартных ошибок интерполяции должно иметь наименьшее из возможных значение (это полезно при сравнении моделей), а среднеквадратичная нормированная ошибка должна быть близка к 1.

В данном случае термин “ошибка интерполяции” используется для обозначения разницы между интерполированным и фактическим значением. Для модели, выполняющей точную интерполяцию, средняя ошибка интерполяции должна быть близка к 0, если интерполированные значения являются несмещенными; среднеквадратичная нормированная ошибка интерполяции должна быть близка к 1, если стандартные ошибки небольшие; и среднеквадратичная ошибка вычислений должна иметь маленькие значения, если проинтерполированные значения близки к измеренным.

Диалог Перекрестная проверка также позволяет вам просмотреть графики рассеяния, на которых будут показаны: Ошибка, Нормированная ошибка, а также график КК (Квантиль-квантиль) для каждой точки выборки.

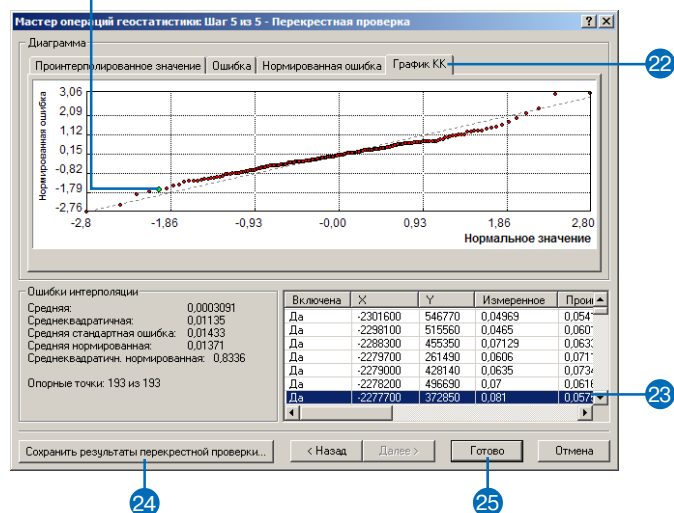
22. Выберите закладку График КК, чтобы просмотреть график Квантиль-квантиль.

Из данного графика (КК) видно, что некоторые значения расположены немного выше линии, а некоторые - немного ниже линии, но большинство точек расположены очень близко к прямой пунктирной линии, что указывает на то, что ошибки интерполяции близки к нормальному распределению.

23. Чтобы выделить цветом местоположение определенной точки, выберите строку в таблице, соответствующую интересующей вас точке. Выбранная точка на диаграмме будет выделена зеленым цветом.

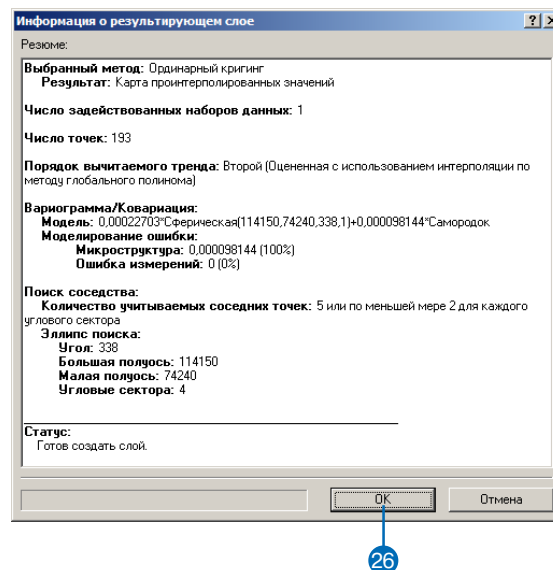
24. Кроме того, нажмите кнопку Сохранить результаты перекрестной проверки... с тем, чтобы использовать таблицу для дальнейшего анализа.

Выбранная точка



25. Нажмите Готово.

Диалог Информация о результирующем слое содержит краткую информацию о модели, которая будет использоваться для построения поверхности.



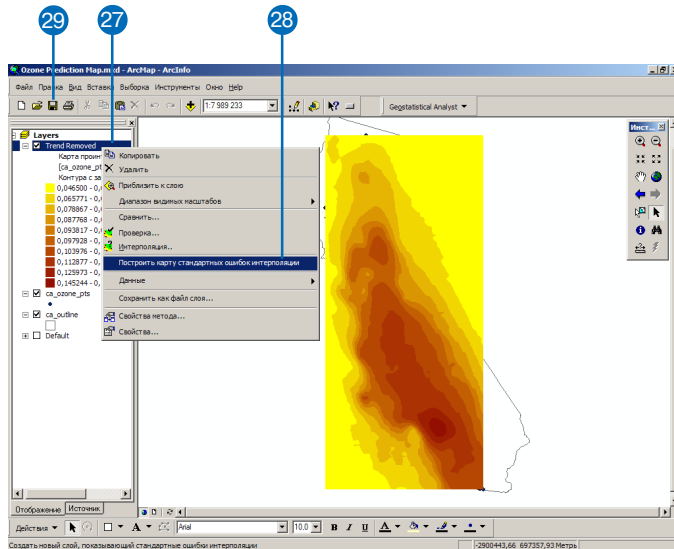
26. Нажмите OK.

Карта проинтерполированных значений концентраций озона появится как верхний слой в таблице содержания ArcMap.



По умолчанию, слою присваивается название метода кригинга, использованного для построения поверхности (например, Ординарный кригинг).

27. Щелкните по названию слоя, чтобы выделить его цветом, затем еще раз щелкните по нему и измените название на “Вычтенный тренд”.



Вы можете также построить поверхность стандартной ошибки интерполяции, чтобы оценить качество полученных после выполнения вычислений значений.

28. Нажмите правую клавишу мыши на созданном вами слое “Вычтенный тренд” и выберите в открывающемся меню опцию Построить карту значений стандартной ошибки интерполяции .

29. На стандартной панели инструментов нажмите Сохранить.

Стандартные ошибки количественно характеризуют неопределенность интерполяции для каждой точки построенной вами поверхности. Простое правило большого пальца гласит, что в 95 процентах случаев, истинное значение поверхности будет равно проинтерполированному значению  $\pm 2$  двукратная стандартная ошибка интерполяции, если данные подчиняются закону нормального распределения. Обратите внимание, что на поверхности стандартной ошибки интерполяции точки, расположенные ближе к точкам выборки, как правило, имеют более низкую ошибку.

Поверхность, построенная вами в Упражнении 1, просто использовала значения, предлагаемые в модуле Geostatistical Analyst по умолчанию, без учета трендов поверхности, возможности применения лагов меньших размеров или анизотропной модели вариограммы. Поверхность проинтерполированных значений, построенная вами в этом упражнении, учитывает глобальные тренды в данных, возможность изменения размера лага и направленное влияние (анизотропию) на вариограмме.

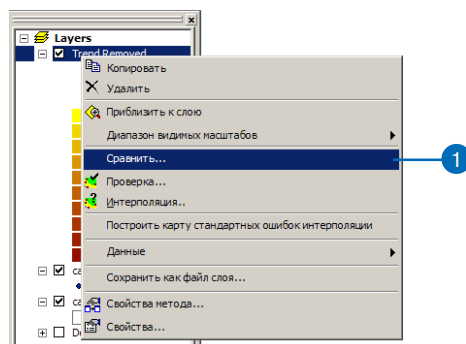
В Упражнении 4, вы сравните две модели, чтобы определить, какая из них наилучшим образом интерполирует неизвестные значения.

Примечание: Как вы снова можете видеть, интерполяция распространяется и на океан. В Упражнении 6 вы узнаете, как ограничить поверхность проинтерполированных значений территорией штата Калифорния.

## Упражнение 4: Сравнение моделей

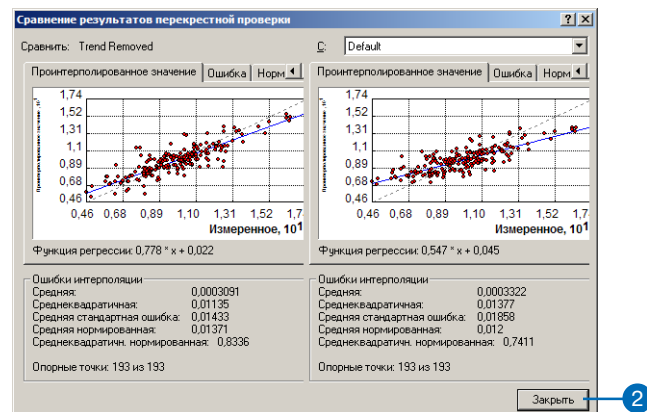
Используя возможности модуля Geostatistical Analyst, вы можете сравнить результаты двух построенных и отображенных на картах поверхностей. Это позволит вам, на основе изучения статистики перекрестной проверки, принять взвешенное решение о том, какая из них более точно интерполирует значения концентрации озона.

1. Правой клавишей мыши щелкните на слое “Вычтенный тренд”, выберите опцию Сравнить... Вы будете сравнивать этот слой со слоем, созданным вами в упражнении 2 на основе параметров, предложенных по умолчанию.

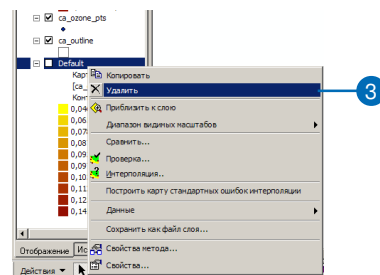


Поскольку для слоя с вычтенным трендом среднеквадратичная ошибка интерполяции меньше, среднеквадратичная нормированная ошибка интерполяции ближе к единице, и средняя ошибка интерполяции ближе к нулю, чем для слоя, построенного с параметрами, предложенными по умолчанию, вы можете с определенной уверенностью утверждать, что модель с вычтенным трендом лучше и является более достоверной. Таким образом, вы можете удалить слой, созданный с использованием параметров, предложенных по умолчанию, поскольку он вам больше не нужен.

2. Закройте диалог Сравнение результатов перекрестной проверки.



3. Нажмите правую клавишу мыши на слое “По умолчанию” и выберите Удалить.



4. Выберите слой “Вычтенный тренд” и переместите его в нижнюю часть таблицы содержания с тем, чтобы вы могли видеть точки выборки и контурную карту штата Калифорния.
5. Нажмите Сохранить на стандартной панели инструментов.

Итак, вы определили лучшую поверхность интерполяции. Но существуют и другие типы поверхностей, которые вы, возможно, захотите построить.

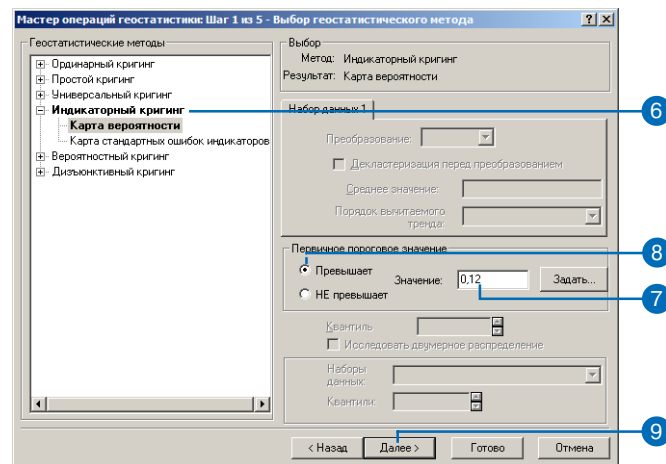
## Упражнение 5: Картографирование вероятности превышения критического значения концентрации озона

В Упражнениях 1 и 3 для картографирования концентраций озона в Калифорнии вы воспользовались методом ординарного кригинга с применением различных параметров. В процессе принятия решений вы должны также подумать о том, как следует использовать карту проинтерполированных значений концентраций озона для выявления территорий, подверженных опасности, поскольку существует неопределенность в вычислении значений для всей поверхности. Например, предположим, что критическое значение концентрации озона составляет 0.12 ppm для восьмичасового периода, и вы хотите узнать, есть ли точки, в которых это значение превышено. Чтобы помочь в процессе принятия решений, вы можете воспользоваться модулем Geostatistical Analyst для картографирования вероятности того, что значения концентрации озона превышают критические.

Хотя модуль Geostatistical Analyst предоставляет целый ряд методов, с помощью которых можно решить эту задачу, в данном упражнении мы воспользуемся методом индикаторного кригинга. Для этого метода не существует ограничений на нормальность распределения данных. Значения данных преобразуются в серии нулей и единиц, в соответствии с тем, находятся ли значения данных ниже или выше критического (порогового) значения. Если в качестве порогового используется значение равное 0.12, любому значению ниже этого будет присвоено значение 0, а значениям, превышающим данную величину, будет присвоено значение 1. Затем индикаторный кригинг использует модель вариограммы, рассчитанную на основе преобразованного в нули-единицы набора данных.

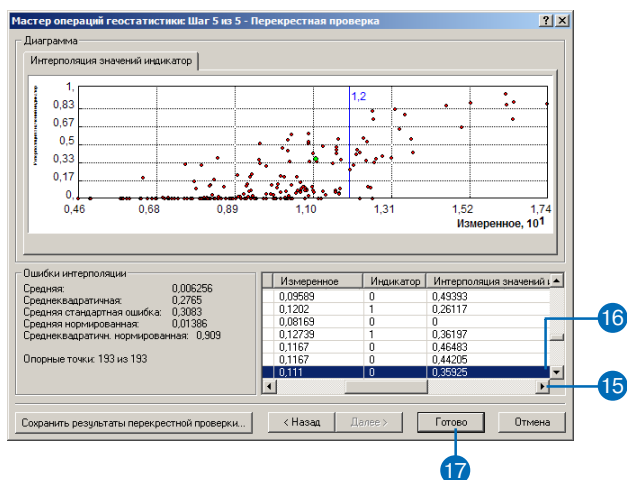
1. На панели Geostatistical Analyst выберите Мастер операций геостатистики....
2. В окне Входные данные (слой) выберите sa\_ozone\_pts.
3. В окне Атрибут выберите OZONE.
4. В окне Метод выберите Кригинг.

5. В диалоге Выберите входные данные и метод нажмите Далее.
6. Выберите Индикаторный кригинг; обратите внимание, что выбрана опция Карта вероятности.
7. Установите первичное пороговое значение равным 0.12.
8. Отметьте опцию Превышает.



9. В диалоге Выбор геостатистического метода нажмите Далее.
10. В диалоге Выбор дополнительных отсекающих нажмите Далее.
11. Выберите опцию Анизотропия, чтобы учесть направленность данных.
12. Задайте размер лага равным 25000 и количество лагов, равным 10.

13. Нажмите Далее в диалоге Моделирование вариограммы/ковариации.
14. Нажмите Далее в диалоге Поиск соседства.  
Голубая линия на диаграмме в следующем диалоге показывает пороговое значение 0.12 ppm. Точкам слева от этой линии присвоен индикатор 0, а точкам справа от линии - индикатор 1.
15. Прокрутите таблицу вправо, пока не будут отображаться столбцы Измеренное, Индикатор и Интерполяция значений индикатора.
16. Выделите цветом строку в таблице, со значением индикатора, равным 0. Эта точка будет выделена зеленым цветом на графике, слева от голубой линии, соответствующей пороговому значению.



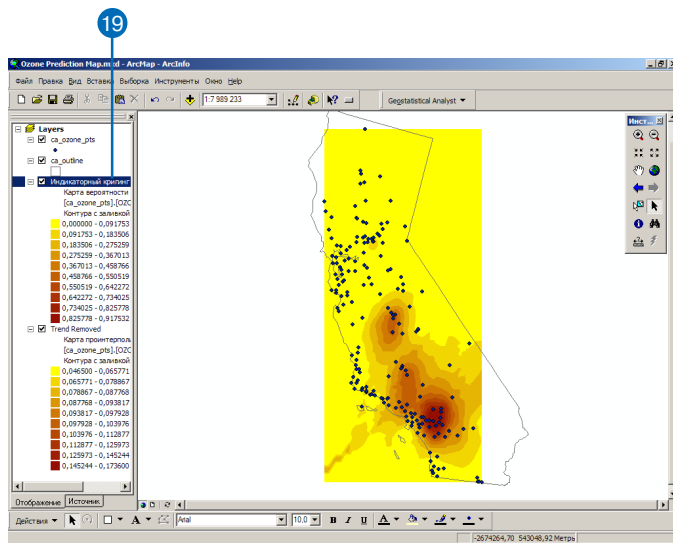
Столбцы с измеренными значениями и значениями индикатора показывают действительное и преобразованное значение каждой точки выборки. Проинтерполированные значения индикатора могут быть интерпретированы как вероятность превышения пороговых значений. Предполагаемые значения индикатора вычисляются с использованием модели вариограммы на основе бинарных (0, 1) данных, полученных индикаторным преобразованием ваших исходных данных. Перекрестная проверка последовательно исключает точку, а затем для каждой рассчитывает предполагаемое значение индикатора.

Например, наибольшее измеренное значение равно 0.1736. Если бы измерения в этой точке в действительности не проводились, на основе метода индикаторного кригинга с вероятностью 85 процентов было бы определено, что значение концентрации в данной точке превышает пороговое значение.

17. Нажмите кнопку Готово в диалоге Перекрестная проверка.
18. Нажмите ОК в диалоге Информация о результирующем слое.

В виде данных ArcMap верхним слоем отобразится карта вероятности.

На карте показаны предполагаемые значения индикатора, которые могут быть интерпретированы как вероятность того, что в 1996 году в течение одного или нескольких дней было превышено пороговое значение концентрации в 0.12 ppm.



По нашей карте видно, что вероятность того, что значения концентрации в районе Лос-Анджелеса (в среднем, составляющие менее 0.12 ppm для каждого восьмичасового интервала в году) превысят установленное нами пороговое значение, довольно высока.

19. Выделите слой Индикаторный кригинг и, удерживая, перетяните его вниз, расположив между слоями “ca\_outline” (картографическая основа штата Калифорния) и “вычтенный тренд”.

Нажмите Сохранить на стандартной панели инструментов, чтобы сохранить карту. В Упражнении 6 вы узнаете, как с использованием возможностей ArcMap картографически правильно оформить карту поверхности проинтерполированных значений, построенной вами в Упражнении 3, и поверхности вероятности, построенной в данном упражнении.

## Упражнение 6: Создание окончательного варианта карты

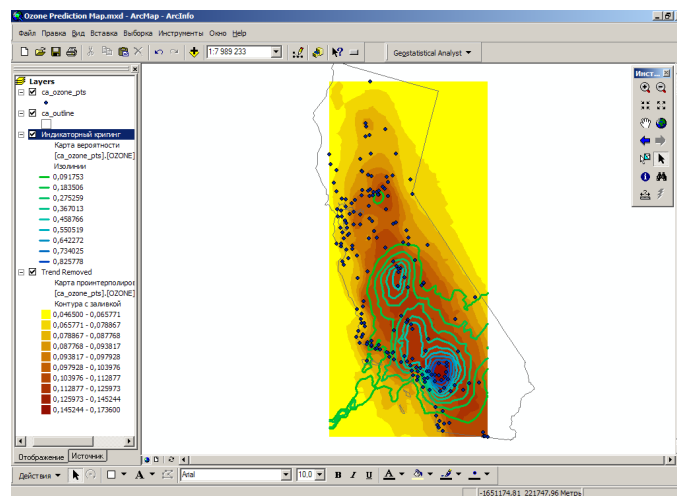
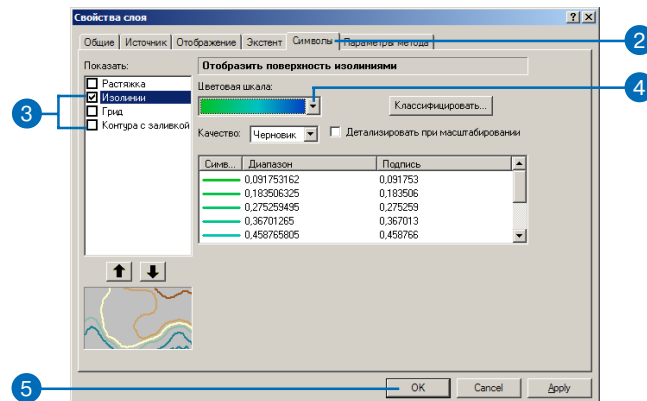
Вы приступаете к созданию окончательного варианта карты, который вы будете использовать в презентации. Для создания такой результирующей карты, на которой будут показаны поверхность проинтерполированных значений и поверхность со значениями вероятности, вы воспользуетесь возможностями ArcMap.

### Отображение обеих поверхностей

Вы можете изменить отображение карты вероятности таким образом, чтобы вы могли видеть обе карты - проинтерполированных значений и вероятности - одновременно. Уровни значений вероятности будут показаны изолиниями.

1. Нажмите правую клавишу мыши на слое Индикаторный кригинг. Выберите опцию Свойства...
2. Выберите закладку Символы.
3. Снимите выделение с опции Контур с заливкой, затем отметьте галочкой строку Изолинии.
4. В окне Цветовая шкала выберите другую цветовую палитру.
5. Нажмите ОК.

Теперь, как показано на рисунке справа, вы видите одновременно и карту значений вероятности (изолинии), и карту проинтерполированных значений.



## Экстраполяция значений концентраций озона

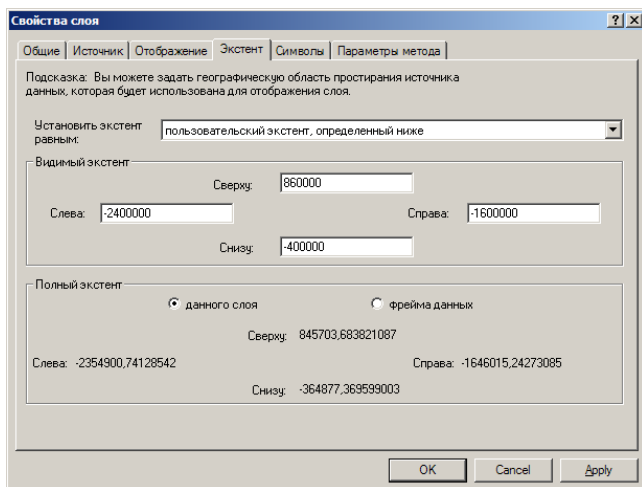
По умолчанию, модуль Geostatistical Analyst интерполирует значение выбранной переменной в любой точке, которая лежит в пределах территории, ограниченной расположением опорных точек в направлениях север-юг и запад-восток. Однако, карта проинтерполированных значений концентраций озона не покрывает географически всю территорию штата Калифорния (слой sa\_outline). Чтобы преодолеть эту проблему, вы экстраполируете значения (т. е., вычислите значения за пределами ограничивающей рамки) для обеих поверхностей.

1. Нажмите правую клавишу мыши на слое Индикаторный кригинг в таблице содержания и выберите Свойства... Выберите закладку Экстент (Область простираения). В окне Установить экстент равным: выберите опцию “пользовательский экстент, определенный ниже”, и наберите следующие значения для опции Видимый экстент, затем нажмите ОК:

Левая: -2400000      Правая: -1600000

Верхняя: 860000      Нижняя: -400000

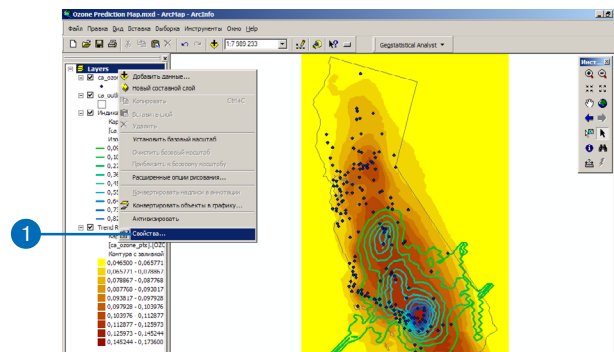
Повторите этот шаг для слоя Вычтенный тренд.



## Вырезание слоев по контуру штата Калифорния

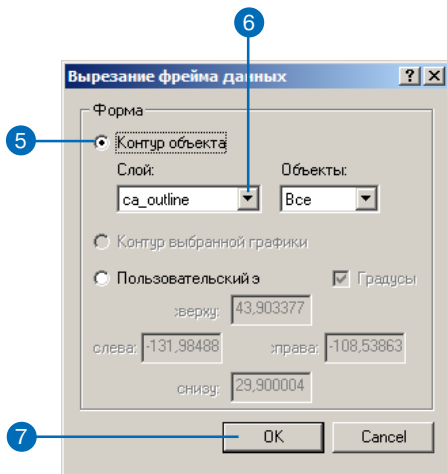
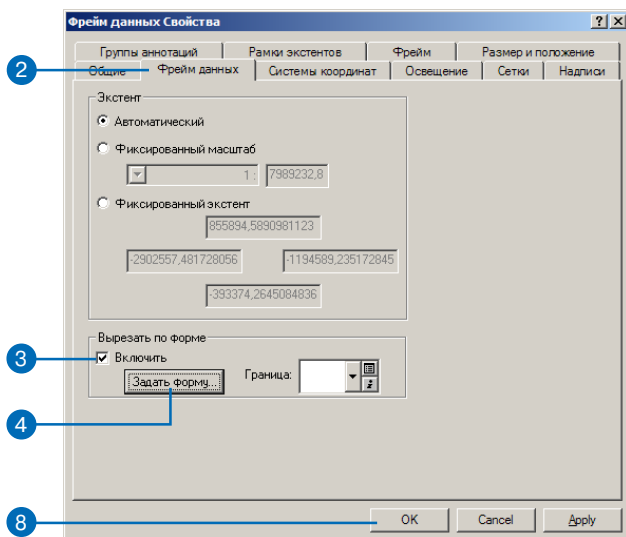
Теперь вы вырежете слои по лекалу контура слоя sa\_outline, поскольку вас интересует только картографирование уровней концентрации озона в пределах штата Калифорния. Это сделает вашу карту визуально более привлекательной.

1. Нажмите правую клавишу мыши на строке Слои и выберите Свойства.

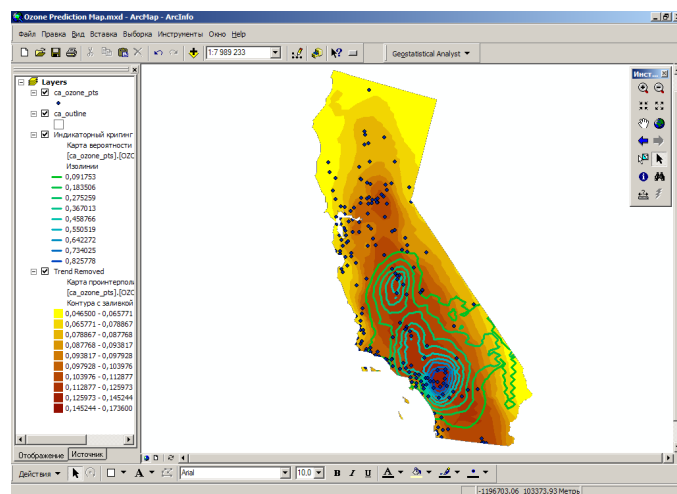


2. Выберите закладку Фрейм данных.
3. Отметьте галочкой Включить в разделе Вырезать по форме.
4. Нажмите кнопку Задать форму.
5. Выберите опцию Контур объектов.
6. В списке Слои выберите sa\_outline.
7. Нажмите ОК.
8. Нажмите ОК, чтобы закрыть диалог Вырезание фрейма данных.





Вырезанная карта должна выглядеть, как показано на рисунке.

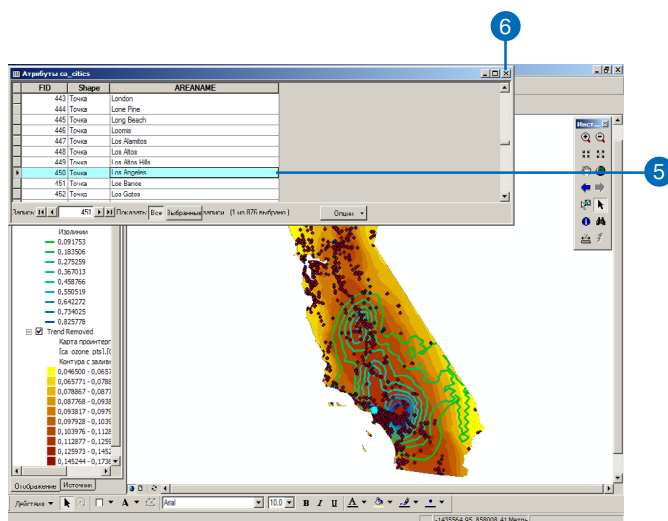


## Размещение на карте Лос-Анджелеса

1. На стандартной панели инструментов нажмите кнопку **Добавить данные**.
2. Перейдите к папке, в которую вы сохранили учебные данные (путь, предлагаемый по умолчанию - C:\ArcGIS\ArcTutor\Geostatistics), затем выберите **ca\_cities**.
3. Нажмите **Добавить**.

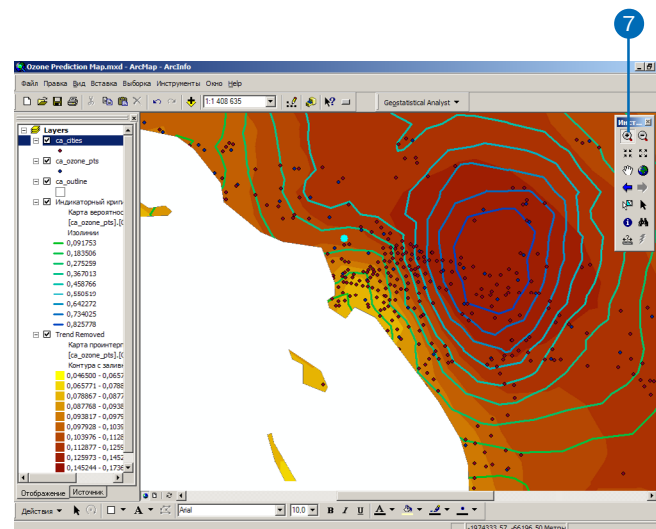
На экране отобразится карта городов Калифорнии.

- Выбрав слой `ca_cities`, нажмите правую клавишу мыши и выберите опцию Открыть таблицу атрибутов.
- Прокрутите таблицу и найдите в столбце `AreaName` название Лос-Анджелес (Los Angeles). Выберите эту строку.  
Город Лос-Анджелес будет выделен на карте цветом.
- Закройте таблицу атрибутов.



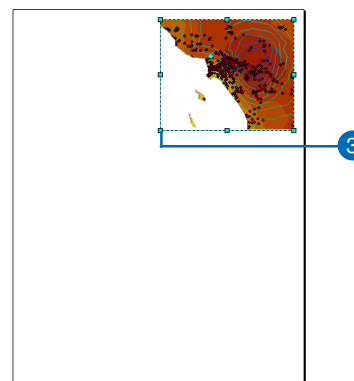
- На панели инструментов выберите инструмент Фиксированное увеличение и увеличьте изображение в районе Лос-Анджелеса.

Обратите внимание, что область с наиболее высокими значениями концентрации озона расположена в действительности немного к востоку от Лос-Анджелеса.



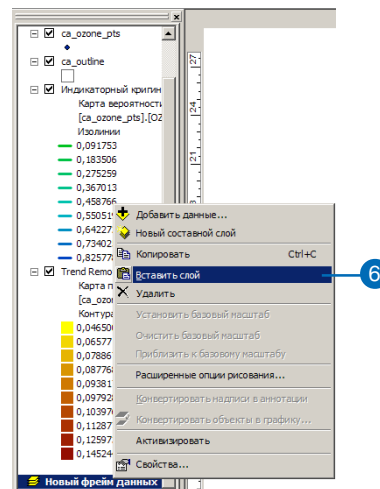
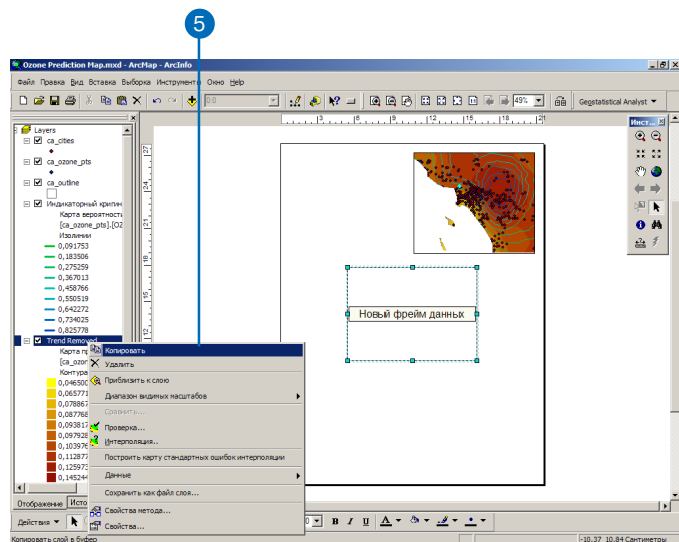
## Создание компоновки

- В Главном меню выберите Вид, а затем Вид компоновки.
- Щелкните по карте, чтобы выбрать ее.
- За нижний левый угол растяните рамку данных, чтобы изменить размер карты.



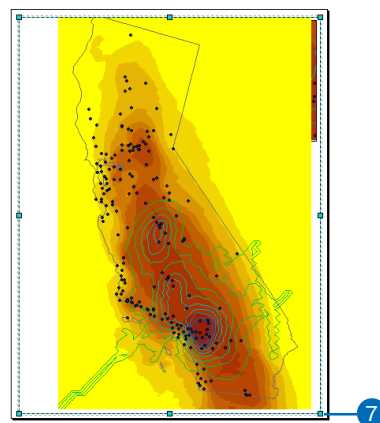
4. В Главном меню выберите Вставка и выберите опцию Фрейм данных.

На карту помещена новая рамка данных. Теперь вы можете скопировать все слои из первой рамки в эту рамку с тем, чтобы отобразить карту значений концентраций озона для всей Калифорнии наряду с картой, показывающей в более крупном масштабе район Лос-Анджелеса.



Повторите шаги 5 и 6 для всех остальных слоев.

7. Растяните Новый фрейм данных на всю страницу.

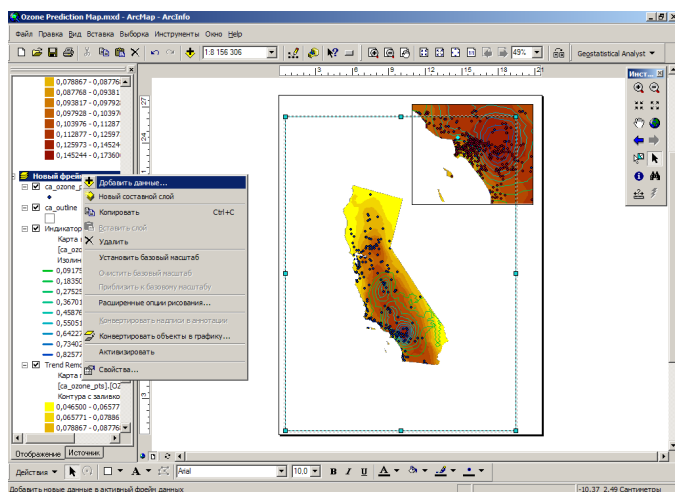


5. Нажав правую клавишу мыши на слое “Вычтенный тренд”, выберите опцию Копировать.
6. Нажав правую клавишу мыши на строке Новый фрейм данных в таблице содержания, выберите опцию Вставить слой.

- Чтобы отобразить карту целиком в Новом фрейме данных, на панели инструментов выберите кнопку Полный экстенд.
- Нажав правую клавишу мыши на Новом фрейме данных, выберите Свойства.
- Выберите закладку Фрейм данных и аналогично тому, как вы проделали это для первой рамки, включите опцию Включить в разделе Вырезать по форме и нажмите кнопку Задать форму... Выберите в качестве лекала, по которому вы будете вырезать, слой `sa_outline`, затем нажмите ОК.

## Добавление отмывки и прозрачности

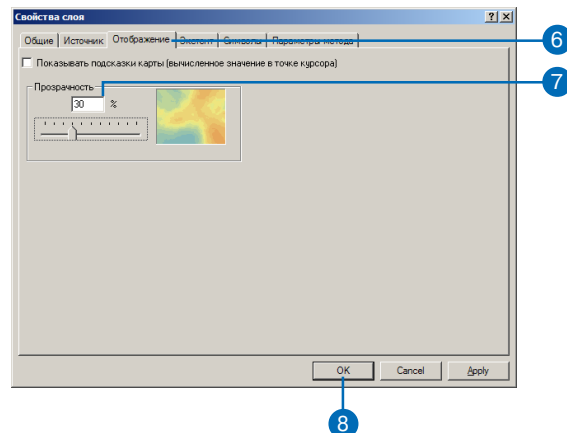
- Нажав правую клавишу мыши на строке Новый фрейм данных, выберите опцию Добавить данные.



- Перейдите к папке, в которой вы установили учебные данные (путь, предлагаемый по умолчанию - `C:\ArcGIS\ArcTutor\Geostatistics`), затем выберите файл `sa_hillshade`.
- Нажмите Добавить.

На экране отобразится карта рельефа Калифорнии.

- Перетащите слой `sa_hillshade` в нижнюю часть таблицы содержания.
- Нажав правую клавишу мыши на слое “Вычтенный тренд” в таблице содержания Нового фрейма данных, выберите Свойства.



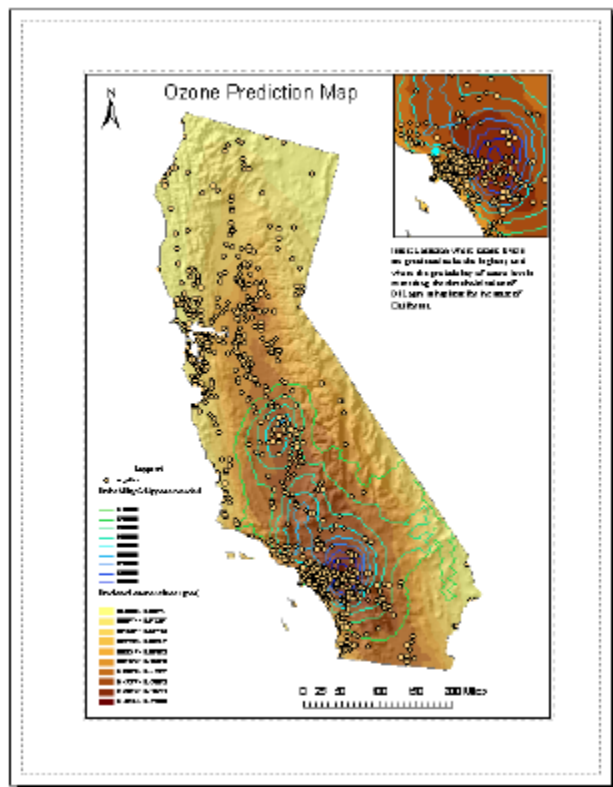
- Выберите закладку Отображение.
- В разделе Прозрачность наберите значение 30 (процентов).
- Нажмите ОК.

Рельеф с отмывкой будет теперь частично отображаться как подстилающий слой для слоя “Вычтенный тренд”.

## Добавление элементов карты

- В Главном меню выберите Вставка, а затем Легенда.
- Поместите легенду в левый нижний угол макета.
- Дополнительно, выбрав опцию Вставка, вы можете добавить стрелку севера, масштабную линейку или текст масштаба.

На следующем рисунке приведен пример окончательной карты, которую вы можете создать, воспользовавшись возможностями ArcMap. Если вам необходимо узнать больше о том, как добавлять элементы на карту, обратитесь к книге *Использование ArcMap*.



На карте видно, что для территории к востоку от Лос-Анджелеса характерны наиболее высокие проинтерполированные значения концентрации озона и наиболее высокая вероятность того, что эти значения превысят критическое пороговое значение (0.12 ppm) в хотя бы один восьмичасовой период в течение 1996 года. Поскольку таковы результаты анализа (но помните, что исходные данные были изменены), возможно, вы захотите сфокусировать внимание на этих районах и проанализировать временные серии измерений концентраций озона с тем, чтобы точно определить районы потенциального риска.

# Принципы геостатистического анализа

# 3

## В ЭТОЙ ГЛАВЕ

- Что такое детерминистские методы
- Что такое геостатистические методы
- Проработка проблемы
- Основные принципы, лежащие в основе методов геостатистики
- Моделирование вариограммы
- Интерполирование неизвестных значений с использованием кригинга
- Модуль Geostatistical Analyst

Модуль Geostatistical Analyst использует значения в опорных точках, расположенных в различных частях ландшафта, и строит (интерполирует) непрерывную поверхность. Опорные точки - это точки, в которых измерены значения какого-либо явления, например, размеры утечки радиации с атомной электростанции, или объемы пролива нефти, или значения высот. Модуль Geostatistical Analyst восстанавливает поверхность, используя значения в измеренных точках для интерполирования значений в каждой точке ландшафта.

В модуле Geostatistical Analyst возможно применение двух групп методов интерполяции: детерминистских и геостатистических. Все методы построения поверхности основаны на сходстве точек, которые расположены близко к опорным. Детерминистские методы используют для интерполяции математические функции. Геостатистика опирается как на статистические, так и на математические методы, которые могут быть использованы для построения поверхности и для оценки ошибки интерполяции.

Модуль Geostatistical Analyst, помимо различных способов интерполяции, предлагает также множество сопутствующих инструментов анализа. Эти инструменты позволяют изучить данные и глубже понять их с тем, чтобы на основе имеющейся информации, вы могли построить более достоверные поверхности.

В этой главе дан обзор теоретических положений, которые лежат в основе детерминистских и геостатистических методов интерполяции. Первая часть главы познакомит вас с детерминистскими методами интерполяции. Затем на примере вы узнаете, что такое геостатистические методы и познакомитесь с принципами, концепциями и предположениями, которые образуют основу геостатистики.



## Что такое детерминистские методы

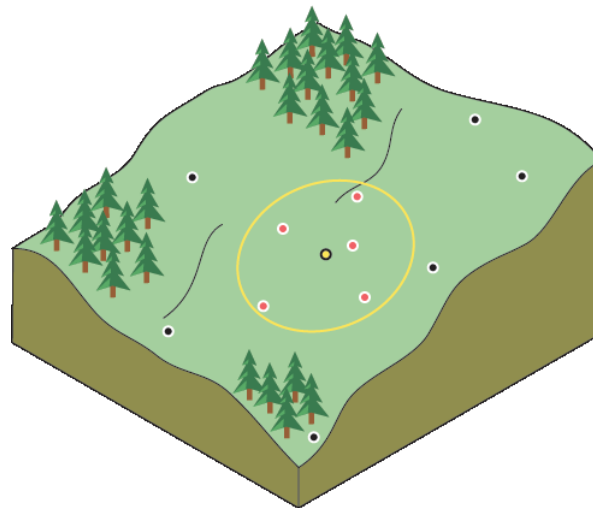
Построение непрерывной поверхности для представления определенных измерений - одно из ключевых требований, предъявляемых к большинству ГИС-приложений. Возможно, наиболее часто используемый тип поверхности - это цифровая модель рельефа. Такие наборы данных в мелком масштабе есть в готовом виде для различных территорий мира. Однако, как уже говорилось, любое измеренное значение в точке ландшафта, земной коры или атмосферы может быть использовано для построения непрерывной поверхности. Главная проблема, с которой сталкиваются специалисты, занимающиеся моделированием в ГИС, - построение наиболее точной из возможных поверхностей на основе существующих опорных точек, наряду с оценкой ошибки интерполяции и отклонений в значениях проинтерполированной поверхности. Вновь построенные поверхности впоследствии используются в ГИС-моделировании и анализе, наряду с их трехмерной визуализацией. Понимание качества этих данных может значительно улучшить эффективность и направленность ГИС-моделирования. Эта роль и отводится модулю Geostatistical Analyst.

### Анализ свойств поверхности в окрестностях опорной точки

В целом, объекты, расположенные ближе друг к другу, как правило, более похожи между собой, чем удаленные друг от друга объекты. Это один из основных принципов географии (Tobler, 1970). Представьте себе, что вы занимаетесь планированием в городе, и перед вами стоит задача разбить в своем городе живописный парк. У вас есть несколько предполагаемых мест, и вам необходимо смоделировать обзор для каждой проектируемой точки. Для этого вам необходим подробный набор данных с высотами поверхности на изучаемую территорию. Предположим, что у вас уже есть данные о высотах для 1000 точек, расположенных по всему городу. Эти данные вы можете использовать для построения новой поверхности высот.

При построении поверхности высот, вы можете предположить, что значения высот в точках, для которых выполняется интерполяция, и значения в опорных точках, расположенных к ним ближе всего, будут похожи. Но возникает вопрос: сколько опорных точек следует рассматривать? И следует ли значения во всех опорных точках учитывать одинаково?

По мере того, как вы удаляетесь от искомой точки, влияние опорных точек будет уменьшаться. Учет в вычислениях точки, удаленной на значительное расстояние от опорной, может быть ошибочным, поскольку точка может находиться на участке местности, кардинальным образом отличающимся от того, на котором расположена искомая точка.



Одно из решений - учитывать достаточное количество точек, образующих небольшую, но репрезентативную выборку. Число точек будет варьировать в зависимости от общего количества опорных точек и их расположения в пространстве, а также от характера поверхности. Если выборки с высотами относительно равно-



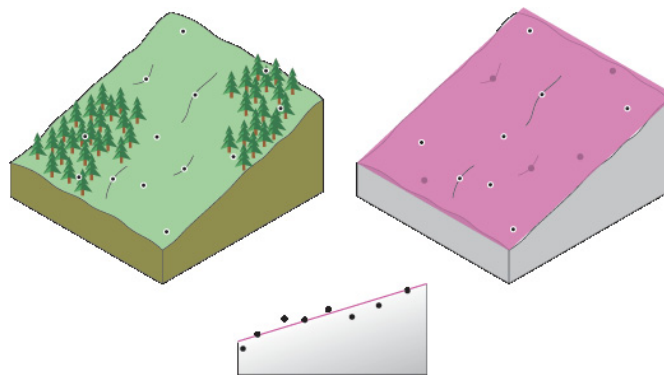
мерно распределены, и характеристики поверхности не меняются в различных частях ландшафта, вы можете с достаточной точностью интерполировать значения поверхности на основе значений в близлежащих точках. Чтобы учесть различную удаленность точек от искомой точки, значениям опорных точек, расположенных ближе к ней, присваивается больший вес.

Это основа метода интерполяции, известного как Метод (обратных) взвешенных расстояний - Inverse Distance Weighting (IDW). Как следует из названия, вес значения уменьшается по мере увеличения расстояния от искомой точки.

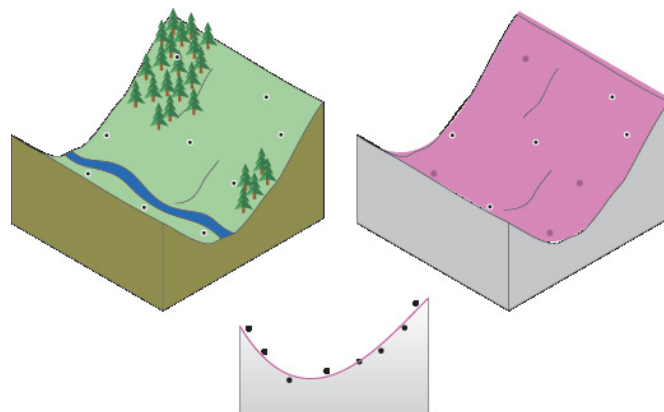
## Визуализация интерполяции по методу глобального полинома

Есть и другие решения для интерполяции значений в точках, в которых не проводились измерения. Одно из предполагаемых мест для парка - пологий склон горы. Поверхность горы - наклонная плоскость. Однако, опорные точки расположены в небольших понижениях или на небольших возвышенностях (локальная вариация). Использование ближайших соседних опорных точек для интерполирования значений может исказить искомое значение в сторону занижения или завышения значений из-за влияния таких понижений и возвышений. В дальнейшем, вы научитесь учитывать локальную вариацию, вычитая из поверхности тренд (для данной территории трендом является наклонная плоскость). Способность определить и смоделировать локальные структуры и тренды может увеличить точность создаваемой поверхности.

Чтобы взять за основу вашей интерполяции поверхность тренда для всей территории, вы можете подобрать плоскость, которая будет проходить через опорные точки. Плоскость может быть определена математической формулой, которая является частным случаем семейства математических формул, известных как полиномы (многочлены). Затем вы сможете определить неизвестное значение высоты в интерполируемой точке по значению



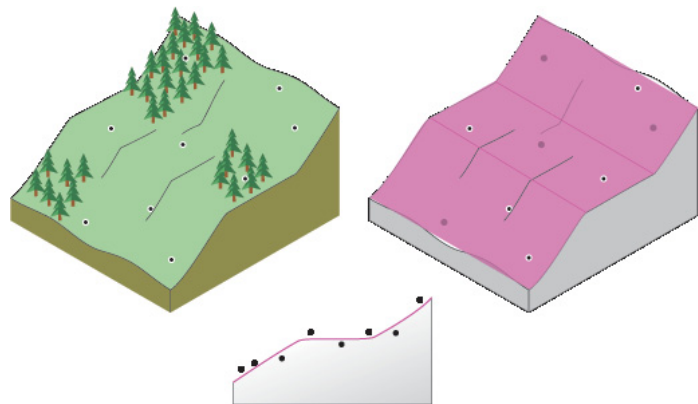
соответствующей точки на плоскости. Плоскость может проходить выше или ниже определенных точек. Цель интерполяции - минимизировать ошибку. Вы можете измерить ошибку путем вычитания значения каждой опорной точки из ее проинтерполированного значения на плоскости, нахождения квадрата этой величины, и последующего суммирования результатов. Такой метод носит название подбора по методу наименьших квадратов. Этот процесс составляет теоретическую основу для интерполяции по методу глобального полинома первого порядка.



Но что если вы захотите описать плоскость, которая соответствовала бы долине? Получить достоверную поверхность, используя для ее описания плоскость, было бы довольно сложно. Однако, если у вас только один перегиб плоскости (см. нижний рисунок), вы сможете подобрать более точную поверхность, чем плоскость (то есть, получить более близкие значения для большего количества значений). Возможность описания одного перегиба поверхности - основа интерполяции по методу глобального полинома второго порядка. Два перегиба плоскости могут быть описаны полиномом третьего порядка, и так далее. Перегибы могут быть расположены в двух направлениях, и тогда можно будет описать поверхность “в форме чаши”.

### Визуализация интерполяции по методу локальных полиномов

Теперь рассмотрим, что происходит, когда поверхность сначала является наклонной, затем выравнивается, а затем снова образует склон. Попытка описать этот участок плоскостью даст плохие результаты. Однако, если вы сможете использовать несколько меньших перекрывающихся плоскостей, а затем использовать центр каждой плоскости как интерполируемое значение для каждой

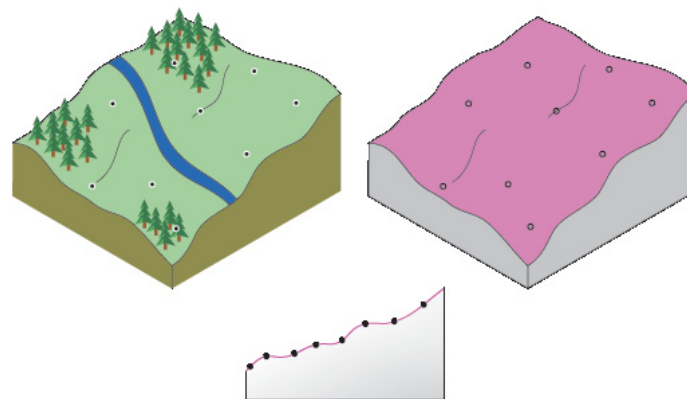


точки на этой плоскости, результирующая поверхность будет более гибкой и возможно более точной. Это концептуальная основа для интерполяции по методу локальных полиномов.

### Визуализация радиальных базисных функций

Радиальные базисные функции позволяют строить поверхности, учитывающие глобальный тренд наряду с локальными вариациями. Это помогает в тех случаях, когда подбор плоскости для значений в опорных точках не дает точного описания поверхности.

Чтобы построить поверхность, допустим, что у вас есть возможность изогнуть и растянуть интерполируемую поверхность таким образом, чтобы она прошла через все опорные точки. Существует много способов для определения формы поверхности в промежутках между опорными точками. Например, вы можете заставить поверхность образовывать изящные изгибы (плоский сплайн), или вы можете контролировать, насколько крепко вы тяните за края поверхности (сплайн с натяжением). Это концептуальная основа интерполяции, основанная на радиальных базисных функциях.



# Что такое геостатистические методы

## Геостатистические решения

Итак, способы, рассмотренные нами, относятся к детерминистским методам интерполяции, поскольку они напрямую основаны на измеренных значениях опорных точек в окрестностях искомой или на заданных математических формулах, которые определяют сглаживание результирующей поверхности. Второе семейство методов интерполяции состоит из геостатистических методов, которые основаны на статистических моделях, учитывающих автокорреляцию (статистические взаимоотношения между опорными точками). С помощью этих методов можно не только построить искомую поверхность, но и получить некую количественную оценку точности интерполяции.

Далее вы рассмотрите основные шаги применения методов геостатистики на примере ординарного кригинга.

Кригинг аналогичен методу взвешенных расстояний (IDW) в том, что опорным точкам из окрестности искомой точки, для получения ее значений, присваиваются веса. Однако веса основаны не только на расстоянии между измеренными точками и искомой точкой, но и на распределении опорных точек в пространстве в целом. Чтобы учесть расположение в пространстве, придав точкам весовые коэффициенты, необходимо количественно определить пространственную автокорреляцию.

Чтобы решить задачу геостатистического анализа, необходимо пройти несколько этапов.

**Расчет эмпирической вариограммы**—кригинг, как и большинство способов интерполяции, построен на предположении, что объекты, расположенные поблизости, более похожи друг на друга, чем удаленные друг от друга объекты (это предположение в данном примере количественно определено как пространственная автокорреляция). Эмпирическая вариограмма - это средство для исследования связей между точками. Пары, расположенные на близком расстоянии, должны иметь меньшую разницу в измеренных значениях, чем те опорные точки, которые

удалены друг от друга. То, насколько это предположение верно, может быть изучено по эмпирической вариограмме.

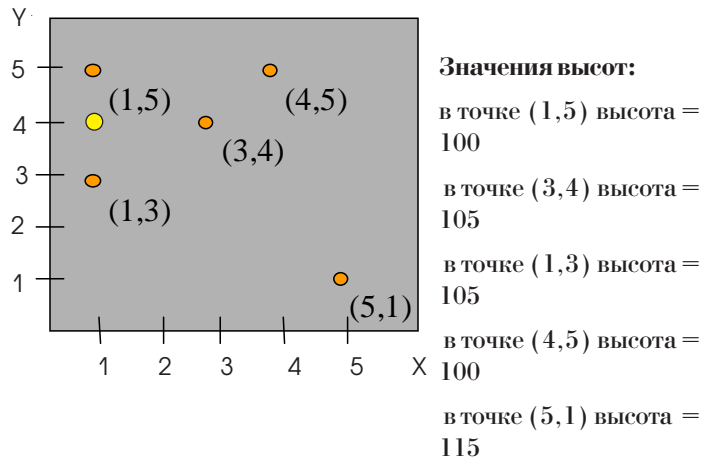
**Подбор модели**—осуществляется путем подбора линии, которая наилучшим образом проходит через точки на графике эмпирической вариограммы. То есть, вы должны определить линию таким образом, чтобы (взвешенный) квадрат разницы между каждой точкой и линией был как можно меньше. Такой способ носит название подбора по методу (взвешенных) наименьших квадратов. Эта линия и будет моделью, количественно определяющей пространственную автокорреляцию в ваших данных.

**Создание матриц**—уравнения для ординарного кригинга содержатся в матрицах и векторах, которые зависят от пространственной автокорреляции между опорными и искомыми точками. Значения автокорреляции могут быть получены из модели вариограммы, описанной выше. Матрицы и векторы определяют веса кригинга, присваиваемые каждому измеренному значению.

**Выполнение интерполяции**—на основе весов кригинга, вы можете вычислить предполагаемое значение в искомой точке с неизвестным значением.

## Проработка проблемы

Представьте, что вы на местности измерили значения высоты в пяти точках исследуемого вами ландшафта. Эти опорные точки показаны на нижнем рисунке оранжевым цветом. Рядом с каждой точкой даны ее координаты (X,Y).



### Уравнения кригинга

Для интерполяции значения в точке с координатами (1,4) ( $X=1$  и  $Y=4$ ), называемой искомая точка (желтая точка на рисунке), вы воспользуетесь методом ординарного кригинга. Модель ординарного кригинга определяется формулой

$$Z(s) = m + e(s)$$

где  $s = (X,Y)$  - положение точки; например, координаты одной из опорных точек  $s = (1,5)$ , и  $Z(s)$  - значение измеренной величины для данной точки; например, для данной точки -  $Z(1,5) = 100$ . Модель основана на постоянном среднем  $m$  для данных (нет тренда) и случайных ошибках  $e(s)$  с пространственной зависимостью. Предположим, что случайный процесс  $e(s)$  является внутренне стационарным. Эти допущения обсуждаются в следующих разделах. Интерполятор может быть получен

как взвешенная сумма данных,

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

где

$Z(s_i)$  - измеренное значение в  $i$ -ой точке, например,  $Z(1,5) = 100$ ;

$\lambda_i$  - неизвестный вес для измеренного значения в  $i$ -ой точке;

$s_0$  - координаты искомой точки, например, (1,4); и

$N = 5$  - число опорных точек.

Формула аналогична формуле интерполяции по методу взвешенных расстояний (IDW). Однако, в методе IDW, вес  $\lambda_i$  зависит исключительно от расстояния до искомой точки. В ординарном кригинге, вес  $\lambda_i$  зависит от вариограммы, то есть от расстояния до искомой точки, и от пространственных взаимосвязей между опорными точками, расположенными вокруг искомой точки.

При выполнении интерполяции для нескольких точек, некоторые из искоемых значений окажутся выше или ниже фактических значений величин. В среднем, разность между проинтерполированными значениями и фактическими значениями должна быть равна нулю. Такое условие называется “условием несмещенности интерполятора”. Для гарантии того, что интерполятор является несмещенным для неизвестного измерения, сумма весов  $\lambda_i$  должна быть равна единице. Воспользовавшись этим ограничением, убедитесь, что разница между истинным значением,  $Z(s_0)$ , и интерполятором,  $\hat{Z}(s_0)$ , имеет наименьшее из возможных значений. Это минимизирует статистическое ожидание следующей формулы, на основе которой получаются урав-

$$\left( \hat{Z}(s_0) - \sum_{i=1}^N \lambda_i Z(s_i) \right)^2$$

нения кригинга. Путем минимизации ожидания, в среднем, интерполятор кригинга близок настолько это возможно, к неизвестному значению. Решение для минимизации, ограниченное условием несмещенности, дает уравнения кригинга,

$$\begin{matrix} \mathbf{Г} & * & \mathbf{л} & = & \mathbf{g} \\ & or & & & \\ \begin{pmatrix} \gamma_{11} & \Lambda & \gamma_{1N} & 1 \\ \mathbf{M} & \mathbf{O} & \mathbf{M} & \mathbf{M} \\ \gamma_{N1} & \Lambda & \gamma_{NN} & 1 \\ 1 & \Lambda & 1 & 0 \end{pmatrix} & * & \begin{pmatrix} \lambda_1 \\ \mathbf{M} \\ \lambda_N \\ m \end{pmatrix} & = & \begin{pmatrix} \gamma_{10} \\ \mathbf{M} \\ \gamma_{N0} \\ 1 \end{pmatrix} \end{matrix}$$

Эти уравнения станут более понятными, когда в следующем разделе вы подставите значения в матрицы. Помните, что цель состоит в решении уравнений для всех  $\lambda_i$  (весов), поэтому интерполятор может быть образован с использованием  $S_i Z(s_i)$ .

Большинство элементов могут быть подставлены, если вам известна вариограмма. В нескольких последующих разделах вы узнаете, как рассчитать значения вариограммы. Гамма-матрица  $\mathbf{G}$  содержит смоделированные значения вариограммы для всех пар опорных точек, где  $g_{ij}$  означает смоделированные значения вариограммы, основанные на расстоянии между двумя опорными точками, определенными как точки  $i$ -ая и  $j$ -ая. Вектор  $\mathbf{g}$  содержит смоделированные значения вариограммы для каждой пары из опорной точки и искомой точкой, где  $g_{i0}$  обозначает смоделированные значения вариограммы, основанные на расстоянии между  $i$ -ой точкой и искомой точкой. Также оценивается неизвестное значение  $m$  в векторе  $\mathbf{l}$ ; оно растет из-за условия несмещенности.

### Расчет эмпирической вариограммы

Для расчета значений матрицы  $\mathbf{G}$  мы должны изучить структуру данных путем создания эмпирической вариограммы. На вариограмме показаны значения половины квадрата разности для пар точек (откладывается по оси  $y$ ) в зависимости от расстояния между ними (откладывается по оси  $x$ ).

Первый шаг в создании эмпирической вариограммы - вычисление расстояния и квадрата разности между значениями для каждой пары точек. Расстояние между двумя точками рассчитывается с использованием Евклидова расстояния:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Эмпирическая дисперсия - это половина квадрата разности  $0.5 * \text{среднее}[(\text{значение в точке } i - \text{значение в точке } j)^2]$ .

Locations	Distance Cal.	Distances	Difference <sup>2</sup>	Semivariance
(1,5), (3, 4)	$\text{sqrt}[(1-3)^2 + (5-4)^2]$	2.236	25	12.5
(1,5),(1,3)	$\text{sqrt}[2^2 + 2^2]$	2	25	12.5
(1,5), (4, 5)	$\text{sqrt}[3^2 + 0^2]$	3	0	0
(1,5),(5,1)	$\text{sqrt}[4^2 + 4^2]$	5.657	225	112.5
(3, 4), (1,3)	$\text{sqrt}[2^2 + 1^2]$	2.236	0	0
(3, 4), (4, )	$\text{sqrt}[1^2 + 1^2]$	1.414	25	12.5
(3, 4), (5,1)	$\text{sqrt}[2^2 + 3^2]$	3.606	100	50
(1,3), (4, 5)	$\text{sqrt}[3^2 + 2^2]$	3.606	25	12.5
(1,3),(5,1)	$\text{sqrt}[4^2 + 2^2]$	4.472	100	50
(4, 5), (5,1)	$\text{sqrt}[1^2 + 4^2]$	4.123	225	112.5

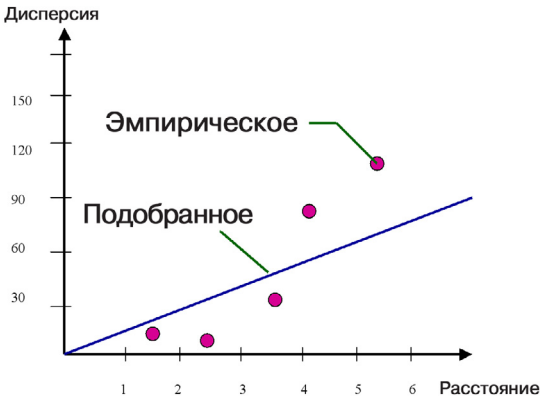
Как видите, для больших наборов данных (с большим количеством измеренных опорных точек) число пар точек будет быстро расти и станет неуправляемым. Следовательно, вы можете сгруппировать пары точек; такой процесс известен как биннинг. В данном примере, бин - это заданный диапазон расстояний. Это означает, что все точки, находящиеся на расстоянии от 0 до 1 сгруппи-

рованы в первый бин, те точки, которые находятся на расстоянии от 1 до -2 метров друг от друга сгруппированы во второй бин, и т.д. Берется средняя эмпирическая дисперсия для всех пар точек. В нашем примере, данные распределены в пять бинов.

Бининг эмпирической вариограммы					
Размер лага	Расстояние ме	Ср.расстояние	1/2 дисперсии	Среднее	
1+-2	1.414, 2	1,707	12.5, 12.5	12,5	
2+-3	2.236, 2.236, 3	2,491	12.5, 0, 0	4,167	
3+-4	3.606, 3.606	3,606	50, 12.5	31,25	
4+-5	4.472, 4.123	4,298	50, 112.5	81,25	
5+	5,657	5,657	112,5	112,5	

### Подбор модели

Теперь вы можете отобразить на графике - эмпирической вариограмме - среднюю дисперсию в зависимости от среднего расстояния для бина. Но значения эмпирической вариограммы не могут быть использованы непосредственно в матрице G, поскольку вы можете получить отрицательные стандартные ошибки вычислений; вместо этого вы должны подобрать модель для эмпирической вариограммы. После подбора модели, вы будете использовать полученную модель для определения значений вариограммы для различных расстояний.



Для простоты, подобранная вами модель - линия регрессии, построения с использованием метода наименьших квадратов, имеющая положительный угловой коэффициент и проходящая через ноль. В модуль Geostatistical Analyst включено большее количество моделей, которые можно использовать для подбора.

Формула для определения дисперсии для любого заданного расстояния в нашем примере выглядит следующим образом:

Дисперсия = Угловой коэффициент \* Расстояние

Угловой коэффициент определяет наклон подобранной прямой. Расстояние - это расстояние между парами точек, оно обозначается как *h*. В нашем примере, дисперсия для любого расстояния может быть определена по формуле:

Дисперсия = 13.5 \* *h*

Теперь создадим матрицу значений G. Например, значение *g*<sub>12</sub> для точек (1, 5) и (3, 4) в уравнении равно:

Дисперсия = 13.5 \* 2.236 = 30.19

	(1, 5)	(3, 4)	(1, 3)	(4, 5)	(5, 1)	
G Матрица (Гамма)						
(1, 5)	0	30.19	27.0	40.5	76.37	1
(3, 4)	30.19	0	30.19	19.09	48.67	1
(1, 3)	27.0	30.19	0	48.67	60.37	1
(4, 5)	40.5	19.09	48.67	0	55.66	1
(5, 1)	76.37	48.67	60.37	55.66	0	1
	1	1	1	1	1	0

В приведенном выше примере, для пары точек (1, 5) и (3, 4), величина лага была вычислена с использованием расстояния между двумя точками (см. предыдущую таблицу). Значение вариограммы определяется путем умножения углового коэффициента, равного 13.5, на расстояние. Единицы и нули в нижней строке и крайнем правом столбце определены в соответствии с условием несмещенности.

Формула значений матрицы для ординарного кригинга следующая:

$$G \cdot l = g$$

Теперь, когда построена матрица **G**, необходимо найти значения вектора **l**, содержащего значения весов, которые будут присвоены измеренным значениям, находящимся вокруг искомой точки. Поэтому, выполним простую операцию из матричной алгебры и получим следующую формулу:

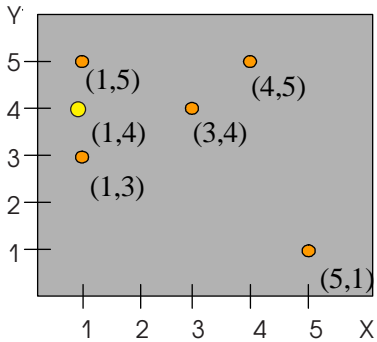
$$l = G^{-1} \cdot g$$

где  $G^{-1}$  - обратная матрица **G**. Обратную матрицу **G** получаем, выполнив операцию линейной алгебры.

Инверсия Г Матрицы (Гамма)					
-0.02575	0.00704	0.0151	0.00664	-0.00303	0.3424
0.00704	-0.04584	0.01085	0.02275	0.0052	-0.22768
0.0151	0.01085	-0.02646	-0.00471	0.00522	0.17869
0.00664	0.02275	-0.00471	-0.02902	0.00433	0.28471
-0.00303	0.0052	0.00522	0.00433	-0.01173	0.42189
0.3424	-0.22768	0.17869	0.28471	0.42189	-41.701

Далее, для искомой точки строится вектор **g**. Например, рассмотрим точку (1,4). Вычисляем расстояние от точки (1,4) до каждой из точек с измеренными значениями (1,5), (3,4), (1,3), (4,5), и (5,1). Исходя из этих расстояний, определяем подобранную дисперсию по формуле Дисперсия =  $13.5 \cdot h$ , выведенной ранее. Вектор **g** для точки (1,4) приведен в следующей таблице.

Точка	Расстояние	g Вектор для (1,4)
(1,5)	1	13.5
(3,4)	2	27.0
(1,3)	1	13.5
(4,5)	3.162	42.69
(5,1)	5	67.5
		1



Теперь, после того, как были созданы матрица **G** и вектор **g**, вычислим вектор весов кригинга:  $l = G^{-1} \cdot g$ . Для этого, воспользуемся линейной алгеброй. Веса приведены ниже в таблице.

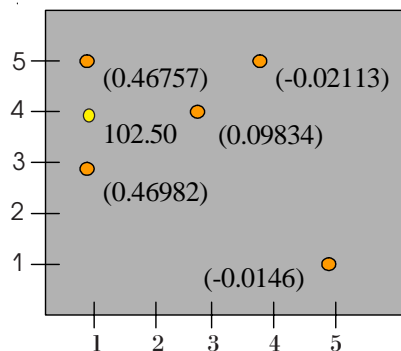
### Выполнение интерполяции

Теперь, когда у вас есть значения весов, умножьте вес каждого измеренного значения на это значение. Сложите результаты и получите искомое значение для точки (1,4).

Веса	Значения	Результат	
0.46757	100	46.757	
0.09834	105	10.3257	
0.46982	105	49.3311	
-0.02113	100	-2.113	
-0.0146	115	-1.679	
-0.18281		102.6218	Кригинг

Далее, изучите полученный результат. На следующем рисунке показаны веса (в скобках) точек с измеренными значениями, использованными для нахождения значений в искомой точке (1,4).





Значения:

(1,5) =

100

(3,4) =

105

(1,3) =

105

(4,5) =

100

(5,1) =

115

Как и ожидалось, веса уменьшаются с расстоянием, но такое уменьшение определено более точно, поскольку при присвоении весов учитывалось не только расстояние между точками, но и пространственное распределение данных. Выполненная интерполяция представляется достаточно достоверной.

## Дисперсия кригинга

Одна из сильных сторон использования статистического подхода - это возможность вычисления статистической меры ошибки интерполяции. Для этого необходимо умножить каждое значение вектора **I** на каждое значение вектора **g** и сложить их все вместе, чтобы получить значение, известное как дисперсия значений, полученных с использованием кригинга. Корень квадратный из дисперсии кригинга носит название стандартной ошибки кригинга.

В данном случае, стандартная ошибка кригинга равна 3.6386. Если принять, что ошибки подчиняются закону нормального распределения, 95 процентов интерполированных значений будут равны:

Интерполятор Кригинга  $\pm 1.96 \cdot \text{корень квадратный}$  (из дисперсии кригинга)

Значение 1.96 получено на основе законов стандартного нормального распределения, где 95 процентов вероятности попадет в интервал от -1.96 до 1.96. Интервал проинтерполированного значения может быть интерпретирован следующим образом. Если интерполяция выполняется снова и снова по одной и той же модели, для длительного периода, в 95 процентах случаев проинтерполированное значение будет равно значению в исходной точке в пределах заданного интервала. В нашем примере, интервал интерполированного значения колеблется от 95.49 до 109.75 ( $102.62 \pm 1.96 \cdot 3.64$ ).

G Вектор	Веса ( $\lambda$ )	g Вектор, Весовые коэффициенты
13.5	0.46757	6.312195
27.0	0.09834	2.65518
13.5	0.46982	6.34257
42.69	-0.02113	-0.90204
67.5	-0.0146	-0.9855
1	-0.18281	-0.18281
	Дисперсия Кригинга	13.2396
	Стандартная ошибка кригинга	3.6386

# Основные принципы, лежащие в основе методов геостатистики

## Случайные (вероятностные) процессы с зависимостью

В отличие от детерминистских подходов к интерполяции, геостатистика предполагает, что все значения, полученные на изучаемой территории, являются результатом случайного (стохастического) процесса. Случайный процесс не означает, что события независимы, как при каждом подбрасывании монеты. Геостатистика базируется на случайных процессах с зависимостью. Для примера, подбросьте три монеты и посмотрите, что выпало - орел или решка. Не подбрасывайте четвертую монету. Правило, определяющее, как положить четвертую монету, заключается в том, что если вторая и третья монеты выпали решкой, положите четвертую монету так же, как выпала первая; в противном случае положите четвертую монету противоположно тому, как выпала первая монета.

В пространственном или временном контексте, такая зависимость носит название автокорреляции.



## Интерполяция для случайных процессов с зависимостью

Какое все это имеет отношение к геостатистике и интерполированию неизвестных значений? В примере с монеткой, мы определили правила зависимости. В действительности они неизвестны. В геостатистике существует две ключевых задачи: (1) отыскать правила зависимости и (2) выполнить интерполяцию. Как видно из примера, интерполяция основана на том, что вначале должны быть выявлены правила зависимости.

Кригинг базируется на тех же двух задачах: (1) определении функции вариограммы и ковариации (пространственной автокорреляция) и (2) интерполяции неизвестных значений. Поскольку существуют две отдельные задачи, в геостатистике данные используются дважды: сначала для оценки пространственной автокорреляции, а затем для выполнения интерполяции.

## Понятие стационарности

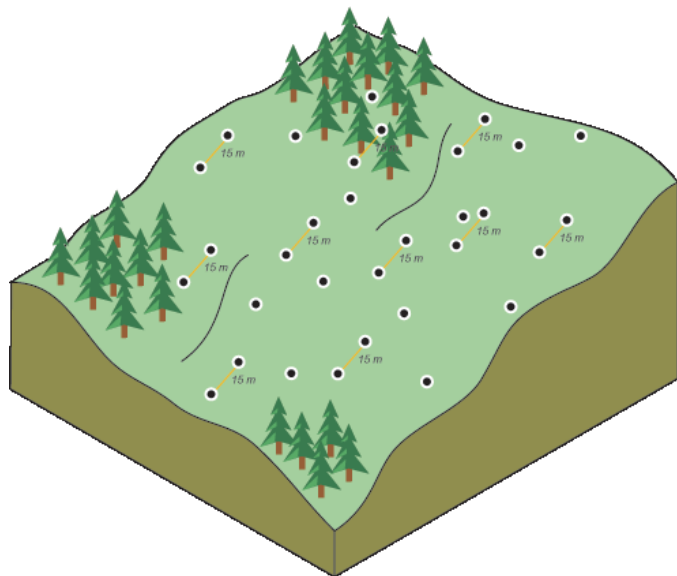
Снова рассмотрим пример с монетой. Для монет существует единственное правило зависимости. Если у вас есть только один набор измеренных значений, вы не сможете узнать правила зависимости, если не найдете их описание. Однако, при продолжающихся наблюдениях нескольких опорных точек, зависимости становятся более понятными. В целом, статистика полагается на некое понятие о репликации, которое предполагает, что на основе повторяющихся наблюдений может быть получена оценка, а также понята изменчивость и неопределенность оценки.

В пространственных задачах, идея стационарности используется для получения необходимой репликации. Стационарность - это предполагаемое свойство пространственных данных, использование которого часто оправдано. Существует два типа стационарности. Один носит название средней стационарности. Предполагается, что при этом типе стационарности среднее является постоянным для опорных точек и не зависит от их расположения.

Второй тип стационарности носит название стационарности второго порядка для ковариации и внутренней стационарности для вариограмм. Стационарность второго порядка - это предположение, что ковариация имеет одно и то же значение между двумя точками, которые расположены на одном и том же расстоянии и в одном и том же направлении, независимо от того, какие две точки вы выбираете. Ковариация зависит от расстояния между любыми двумя точками, а не от их взаимного расположе-

ния. Для вариограмм, внутренняя стационарность - это предположение, что дисперсия разности имеет одно и то же значение для двух точек, расположенных на одном и том же расстоянии и в одном и том же направлении, независимо от того, какие две точки вы выбираете.

Стационарность второго порядка и внутренняя стационарность - предположения, используемые для получения необходимой репликации, позволяющей определить правила зависимости, на основе которых выполняется интерполяция и оценивается неопределенность полученных результатов. Обратите внимание, что репликация обеспечивается за счет использования пространственной информации (аналогичного расстояния между любыми двумя точками). Пример с монетой является зависимым (первая и вторая монеты независимы, но первая и четвертая зависимы), таким образом этот случайный процесс не обладает стационарностью второго порядка.



# Моделирование вариограммы

В следующих разделах мы обсудим подробнее, как построить вариограмму. Предположив, что данные обладают свойством стационарности, можно изучить и количественно оценить автокорреляцию. В геостатистике этот процесс носит название пространственного моделирования, или структурного анализа (вариографии). При пространственном моделировании вариограммы, начните с графика эмпирической вариограммы, вычисляемого по формуле,

Вариограмма(расстояние  $h$ ) =  $0.5 * \text{среднее} [( \text{значение в точке } i - \text{значение в точке } j )^2]$

для всех пар точек, удаленных на расстояние  $h$ . В формулу входит вычисление половины квадрата разности между значениями пар точек. Нанести быстро на график все пары точек становится нереальным. Вместо того, чтобы наносить на график каждую пару точек, пары группируются в бины (классы) для определенных лагов (расстояний). Например, вычислите среднюю дисперсию для всех пар точек, расположенных далее, чем в 40 метрах, и ближе, чем 50 метрах, друг от друга. Эмпирическая вариограмма - это график, на котором по оси  $y$  отложены усредненные значения вариограммы, а по оси  $x$  - расстояние (или лаг) (см. диаграмму).



Снова обратите внимание на предположение о внутренней стационарности, которое позволит вам выполнить репликацию. Таким образом, возможно использовать “среднее” в формуле вариограммы, приведенной выше.

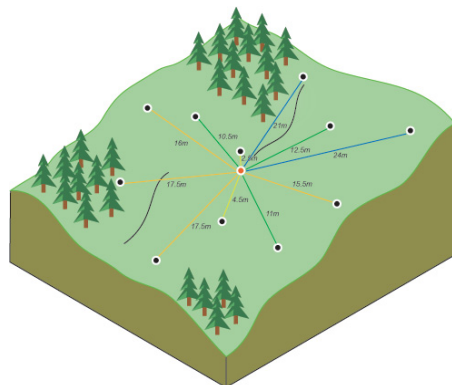
После того, как вы создали эмпирическую вариограмму, вы можете провести через точки линию, описывающую модель эмпирической вариограммы. Моделирование вариограммы аналогично подбору линии по методу наименьших квадратов в регрессионном анализе. Выбирается некая функция, которая служит моделью, например, куполообразная кривая, поднимающаяся вначале и опускающаяся при увеличении расстояний между точками, если они превышают определенное значение.

Основная цель - вычисление параметров кривой, минимизирующей отклонение от точек в соответствии с определенным критерием. Вы можете выбирать из большого количества моделей вариограммы. Для получения более подробной информации и рекомендаций по выбору моделей, обратитесь к Главе 7. А сейчас мы детально рассмотрим каждый из этапов моделирования вариограммы.

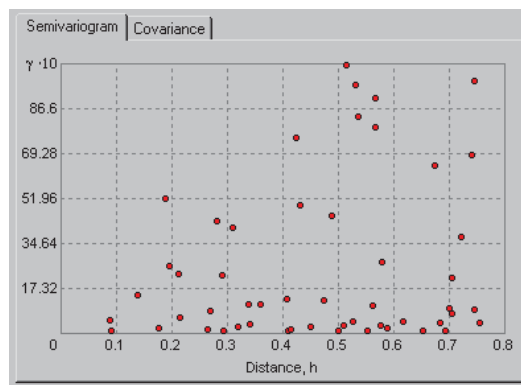
## Построение эмпирической вариограммы

Для создания эмпирической вариограммы необходимо вычислить квадрат разности между значениями для всех пар опорных точек. Если вы нанесете эти значения на график, отложив по оси  $y$  половину квадрата разности, а по оси  $x$  - расстояние, на которое отстоят опорные точки, вы получите диаграмму, которая носит название облака вариограммы. На нижнем рисунке показаны пары одной из опорных точек (которая показана красным цветом) с 11 другими опорными точками.

Одна из основных целей вариографии - изучение и количественная оценка пространственной зависимости, которая также носит название пространственной автокорреляции. Пространственная автокорреляция количественно оценивает предположе-



ние, что объекты, расположенные ближе друг к другу, более похожи чем удаленные друг от друга. Таким образом, пары опорных точек, которые расположены ближе (на облаке вариограммы это точки, расположенные в начале оси x) будут иметь сходные значения (низкие значения на оси y облака вариограммы). Поскольку пары опорных точек становятся удаленными на все большее расстояние (движение вправо по оси x облака вариограммы), точки должны становиться более непохожими и иметь более высокое значение квадрата разности (увеличение значений по оси y облака вариограммы).

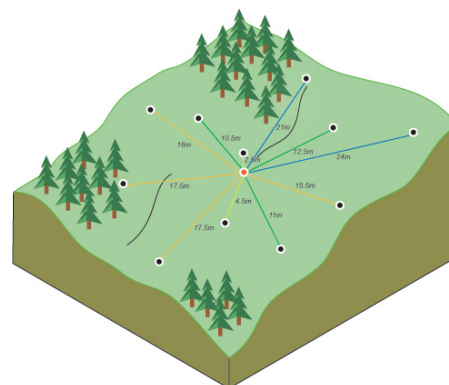


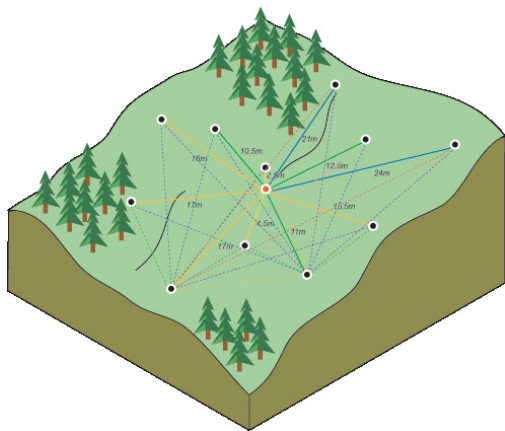
## Бининг (группировка значений) эмпирической вариограммы

Как вы видите из рисунка на предыдущей странице, на котором показано расположение точек в ландшафте, и из облака вариограммы на верхнем рисунке, задача нанесения на график каждой пары опорных точек трудновыполнима из-за большого объема данных. Большое количество точек на графике делает его перегруженным, и вы не можете адекватно интерпретировать его. Чтобы сократить количество точек на эмпирической вариограмме, пары опорных точек должны быть сгруппированы на основе их удаленности друг от друга. Этот процесс группировки значений известен, как бининг.

Бининг - это двухстадийный процесс. Во-первых, формируются пары точек, а во-вторых, эти пары группируются таким образом, чтобы точки были удалены на одинаковое расстояние и располагались в одинаковом направлении. На изображении ландшафта с 12 опорными точками, вы видите, как подобраны пары для одной опорной точки, показанной красным цветом. Одинаковым цветом соединяющих отрезков показаны пары со сходными бинами.

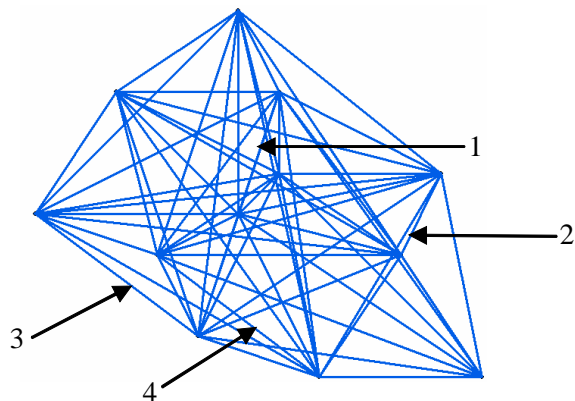
Этот процесс продолжается для всех возможных пар точек. Как вы можете заметить, в процессе образования пар, число пар



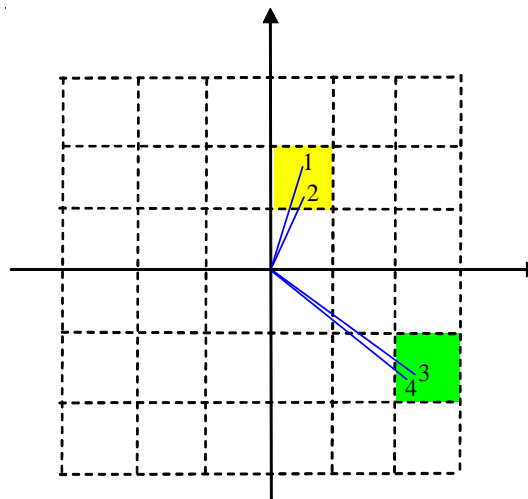


быстро увеличивается с добавлением каждой новой опорной точки. Поэтому для каждого бина на графике облака вариограммы среднее значение расстояния и дисперсии для всех пар этого бина отображается одной точкой.

На рисунке внизу показаны все возможные парные связи между всеми 12 опорными точками. Точки развернуты таким образом, чтобы север был вверху.



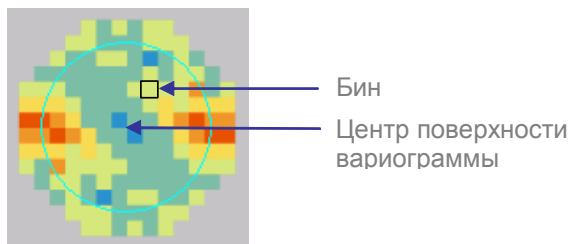
На втором этапе процесса бининга, пары группируются на основе аналогичных расстояний и направлений. Представьте такой график, на котором каждая точка имела бы одно и то же начало координат. Это свойство делает эмпирическую вариограмму симметричной. Всегда откладывайте связующие отрезки справа от вертикальной оси.



Вы видите, что связующие отрезки 1 и 2 имеют приблизительно одно и то же расстояние и направление. Каждая ячейка грида образует бин. Связующие отрезки 1 и 2 попадают в один и тот же бин, показанный желтым цветом. Для связующего отрезка 1 найдите квадрат разности значений двух связанных опорных точек, и то же самое проделайте для связующего отрезка 2. Затем найдите их среднее и умножьте его на 0.5, чтобы получить одно значение эмпирической вариограммы для бина.

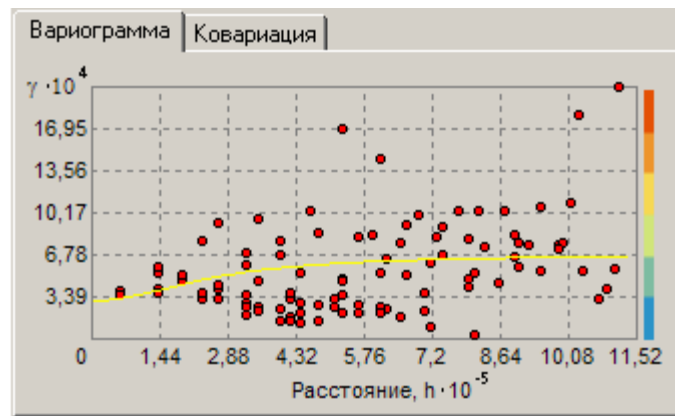
Выполните эту операцию для другого бина, показанного зеленым цветом, со связующими отрезками 3 и 4. Чтобы упростить понимание, показаны только четыре связи, но конечно же, их гораздо больше.

Для каждого бина определяется квадрат разности значений для всех пар связываемых опорных точек, затем вычисляется среднее значение, оно умножается на 0.5, и мы получаем одно значение вариограммы для бина. В модуле Geostatistical Analyst вы можете контролировать размер лага и число бинов. Значение эмпирической вариограммы для каждого бина обозначается цветом, полученное изображение носит название поверхности вариограммы.



На верхнем рисунке показаны семь бинов по горизонтали и по вертикали от центра поверхности вариограммы. Для бинов “холодные” цвета (синие и зеленые) обозначают низкие значения, а “теплые” цвета (красный и оранжевый) обозначают высокие значения. Как видите, в целом, значения эмпирической вариограммы увеличиваются по мере удаления бинов от центра. Это означает, что значения сильнее различаются с увеличением расстояния. Также обратите внимание на симметрию, описанную ранее.

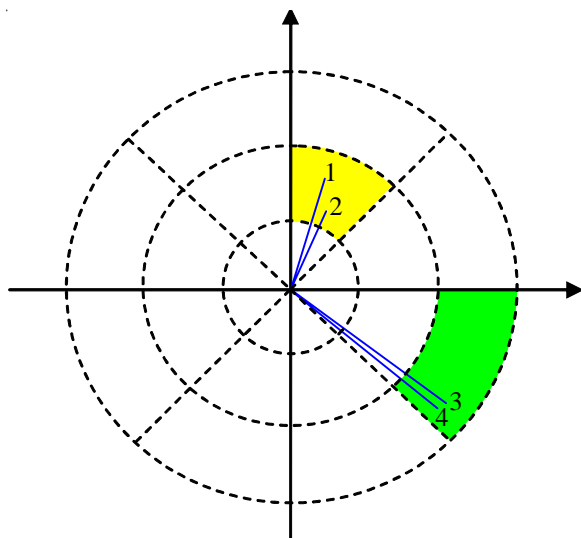
В модуле Geostatistical Analyst можно также построить график эмпирической вариограммы.



На графике, приведенном выше, значение эмпирической вариограммы для каждого бина для каждого направления показано красной точкой (по оси y отложены значения эмпирической вариограммы, а по оси x - расстояние от центра бина до начала координат (центра поверхности вариограммы)). Цветовая шкала справа от графика соответствует цветам на поверхности вариограммы. После выполнения бининга и усреднения значений облака вариограммы, становится более очевидным, что различия увеличиваются с расстоянием. Желтая линия на верхнем рисунке - это подобранная модель вариограммы, которую мы будем обсуждать далее.



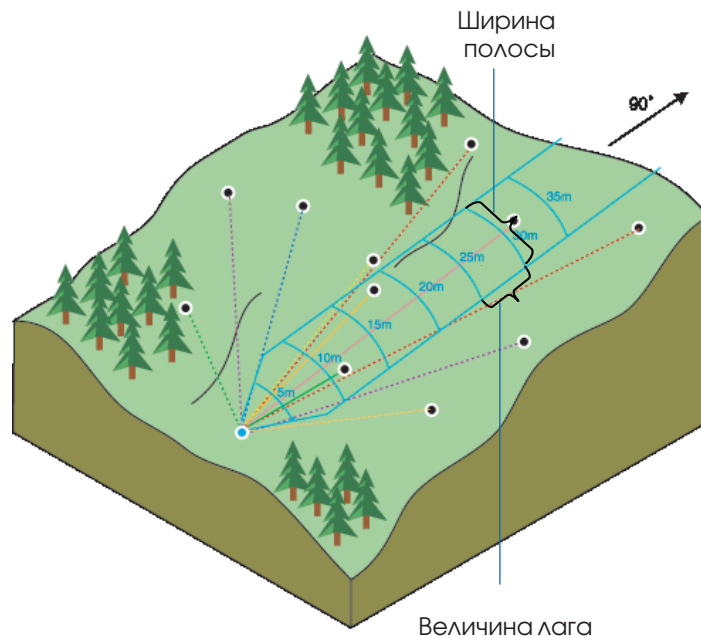
Альтернативный метод, часто используемый для группировки пар в бины, основан на использовании радиальных секторов (см. рисунок). В модуле Geostatistical Analyst этот метод не используется.



### Эмпирические вариограммы для различных направлений

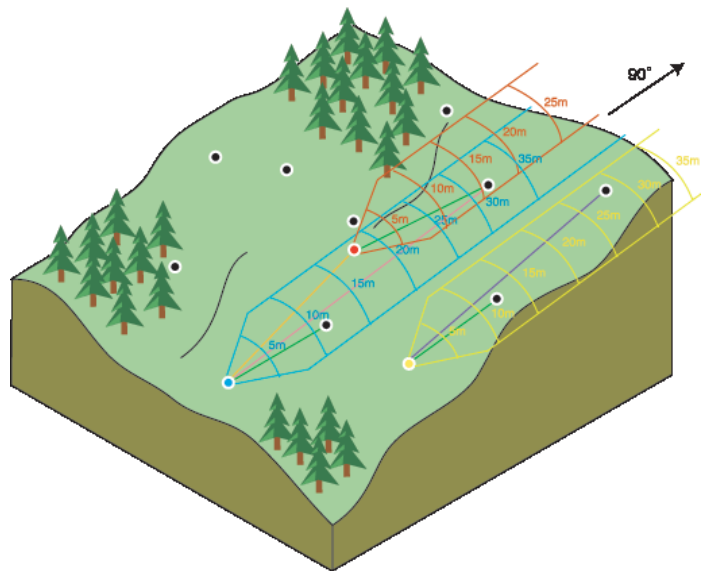
Иногда значения измеренных опорных точек будут содержать составляющую, отражающую влияние по направлениям, которое может быть статистически определено, но не всегда это влияние можно объяснить каким-либо определенным процессом. Такое направленное влияние известно как анизотропия. Угловой допуск задает угол, определяющий ширину полосы захвата, в которую попадут или не попадут близлежащие точки. Ширина полосы устанавливает диапазон поиска соседних точек и то, какие пары точек будут нанесены на график вариограммы.

Точки в бинах - это пары опорных точек, отстоящих друг от друга на определенное расстояние и в определенном направлении. Вы можете изучать бининг по направлениям, либо ограничив пары точек, которые будут нанесены на график в процессе группировки значений, либо исследуя только ту часть диаграммы для всех пар точек, которая относится к одному направлению. На рисунке внизу показан бининг по направлению 90 градусов, с шириной полосы пять метров, угловым допуском в 45 градусов, и величиной лага в пять метров от опорной точки, показанной голубым цветом.



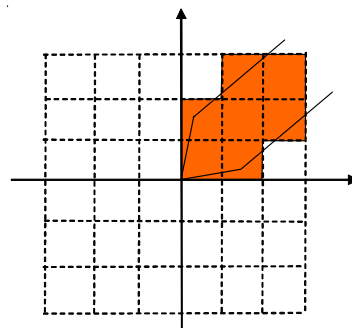
Поиск по направлению продолжается для каждой опорной точки и для каждого направления на поверхности.

На рисунке на этой странице показан бининг по направлению для трех точек. Обратите внимание, что в процесс группировки будет включено меньшее количество пар опорных точек, чем для вариограммы для всех направлений, приведенной в предыдущем примере.



Пары затем формируются в бины в соответствии с одинаковой удаленностью и отстоянием в одинаковом направлении, бины усредняются, и среднее значение для пар каждого бина наносится на вариограмму.

Или же, выполнив бининг по методу грида, описанному ранее, вы можете сформировать в бины все пары, а затем создавать поднаборы по направлениям, как показано на следующем рисунке. Бин будет нанесен на график облака вариограммы, если центр ячейки на поверхности вариограммы включен в направление поиска.



## Выбор размера лага

Выбор размера лага оказывает важное влияние на вид эмпирической вариограммы. К примеру, если размер лага слишком велик, автокорреляция на микроуровне может не проявиться на графике. Если размер лага слишком мал, может быть сформировано много пустых бинов, и количество включенных в бин опорных точек может быть слишком мало для получения репрезентативных “средних значений” для бина.

Когда опорные точки расположены по регулярной сетке, размер ячейки грида обычно служит хорошим индикатором размера лага. Однако, если данные получены с использованием нерегулярной или случайной схемы опорных точек, выбор подходящего размера лага является не столь очевидным. Правило большого пальца состоит в том, что произведение размера лага на количество лагов должно равняться примерно половине максимального расстояния между парами точек. Кроме того, если радиус влияния подобранной модели вариограммы очень мал относительно области отображения эмпирической вариограммы, вы можете уменьшить размер лага. Напротив, если радиус влияния подобранной модели вариограммы большой (по сравнению с областью отображения эмпирической вариограммы), вы можете увеличить размер лага. Далее будут рассмотрены модели вариограммы.

## Подбор модели для эмпирической вариограммы

Моделирование вариограммы/ковариации - ключевой шаг между пространственным описанием и пространственным интерполированием. В предыдущих разделах было рассказано, как подобрать модель вариограммы и как она используется в уравнениях кригинга (гамма- матрица,  $G$  и вектор  $g$ ). Главное применение геостатистики - интерполирование значений атрибутов в искомым точках (кригинг).

Итак, вы узнали, как эмпирическая вариограмма и ковариация предоставляют информацию о пространственной автокорреляции наборов данных. Однако, такая информация может быть получена не для всех возможных направлений и расстояний. По этой причине и для того, чтобы убедиться, что интерполяция по методу кригинга имеет положительные дисперсии, необходимо подобрать модель (т.е. непрерывную функцию или кривую) для эмпирической вариограммы/ковариации.

Теоретически, эта операция аналогична регрессионному анализу, при котором подбирается непрерывная линия или кривая.

### Различные типы моделей вариограммы

В модуле Geostatistical Analyst предусмотрены следующие функции для моделирования эмпирической вариограммы: Круговая, Сферическая, Тетрасферическая, Пентасферическая, Экспоненциальная, Гауссова, Рациональная квадратическая, Эффекта дыры, К-Бесселя, J-Бесселя и Устойчивая. Выбираемая вами модель влияет на получаемые результаты, особенно в том случае, если сильно различается форма кривой около начала координат. Чем круче кривая у начала координат, тем большее влияние на искомый результат оказывают ближайшие соседи (опорные точки).

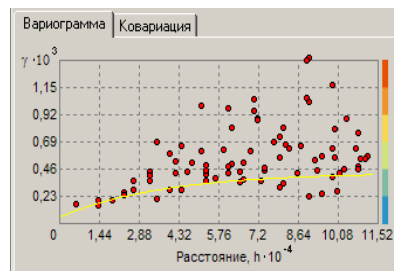
В итоге результирующая поверхность будет менее гладкой. Каждая модель разработана с учетом возможности наиболее точного подбора в зависимости от типа изучаемого явления.

На диаграмме на этой странице показаны две часто используемые модели и то, как различаются эти две функции:



- Сферическая модель

Эта модель показывает постепенное уменьшение пространственной автокорреляции (соответственно, увеличение дисперсии) до некоторого расстояния, за пределами которого автокорреляция равна нулю. Сферическая модель - одна из наиболее часто используемых моделей.



- Экспоненциальная модель

Эта модель применима, когда пространственная автокорреляция уменьшается по экспоненте с уменьшением расстояния между точками, исчезая полностью только при удалении точек друг от друга на бесконечное расстояние. Экспоненциальная модель также используется довольно часто.

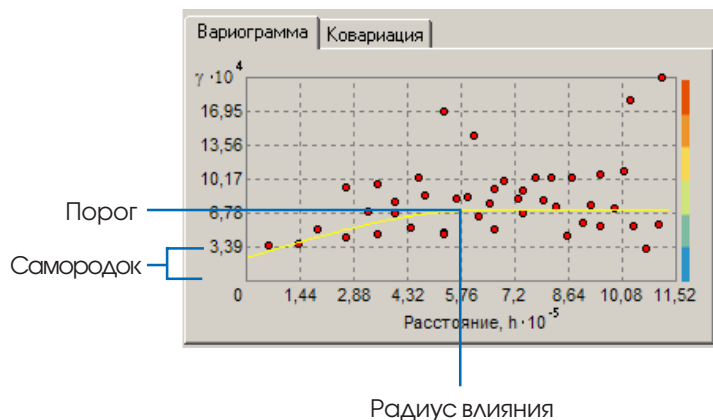
## Понимание вариограммы — радиус влияния, порог и самородок

Как уже упоминалось, вариограмма отражает пространственную автокорреляцию опорных точек. После того, как каждая пара точек (после бининга) нанесена на график, необходимо подобрать модель вариограммы. Для описания таких моделей часто используются определенные параметры.

### Радиус влияния и порог

Если вы посмотрите на модель вариограммы, вы заметите, что при определенном расстоянии кривая, описывающая модель, выравнивается. Расстояние, при котором модель начинает выравниваться, называется радиусом влияния. Опорные точки, отстоящие друг от друга на расстояние, меньшее, чем радиус влияния, пространственно коррелируют, в то время как точки, отстоящие друг от друга на расстояние, большее, чем радиус влияния, - нет.

Значение, которое модель вариограммы, принимает в точке радиуса влияния (значение по оси  $y$ ) носит название “порога”. Частичный порог - это значение порога, из которого вычтено значение самородка.



## Самородок

Теоретически, если расстояние между точками равно нулю (т.е.,  $\text{лаг} = 0$ ), значение вариограммы также должно быть равно нулю. Однако, при бесконечно малых расстояниях, разница между измерениями зачастую не стремится к нулю. Этот факт носит название эффекта самородка. Например, если модель вариограммы пересекает ось  $y$  в точке 2, самородок равен 2.

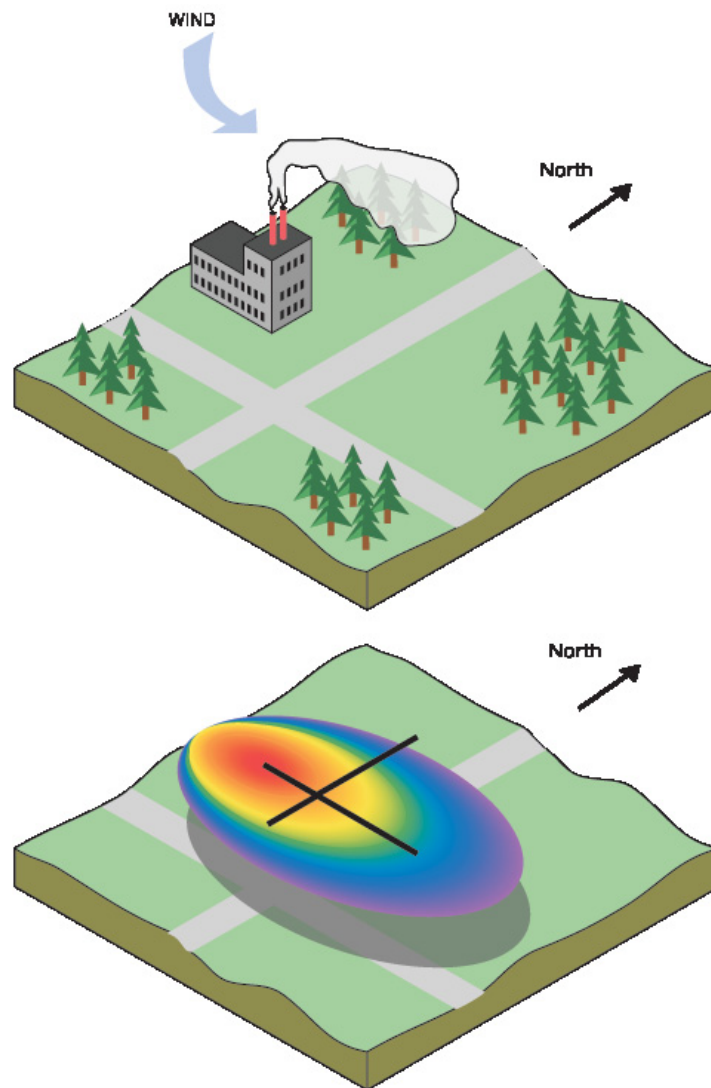
Эффект самородка может быть отнесен за счет ошибок измерений или пространственных составляющих дисперсии на расстояниях меньших, чем интервал выборки (или за счет обоих явлений). Ошибка измерений возникает вследствие ошибок, присущих измерительным приборам. Природные явления могут пространственно варьировать в зависимости от масштаба наблюдений. Вариация на микроуровне меньше, чем те значения расстояний между опорными точками, которые появятся при вычислении значения самородка. Перед сбором данных важно получить некое представление о том, как пространственная вариация будет проявляться на разных масштабных уровнях.

## Учет влияний по направлениям—тренд и анизотропия

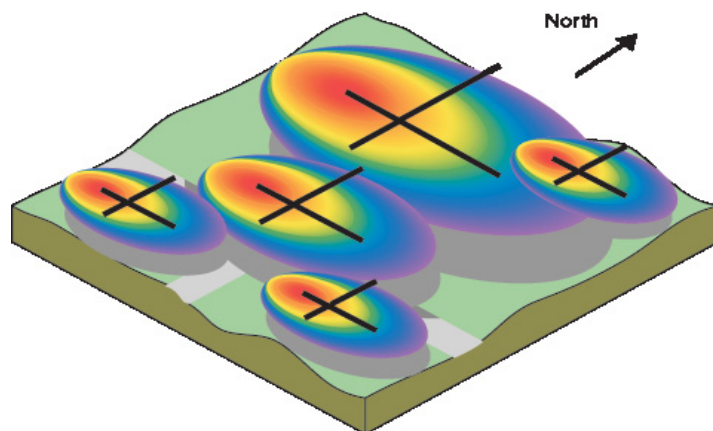
Существует два типа направленных составляющих, которые могут оказывать влияние на интерполирование результирующей поверхности: глобальные тренды и влияния по направлениям, проявляющиеся на вариограмме/ковариации (и известные как анизотропия). Глобальный тренд - это доминирующий процесс, который оказывает детерминистское влияние на все измерения. Глобальный тренд может быть представлен математической формулой (например, полиномом) и вычтен из анализа значений в опорных точках, а затем вновь добавлен перед выполнением интерполяции. Этот процесс носит название “вычитание (или удаление) тренда” (см. Главу 7, “Использование аналитических инструментов при построении поверхностей”).

Глобальный тренд может быть рассмотрен на примере влияния преобладающих ветров на шлейф от дымовой трубы фабрики (рисунок справа). На рисунке самые высокие концентрации загрязняющих веществ показаны теплыми цветами (красными и желтыми), а более низкие концентрации - холодными цветами (зелеными и синими). Обратите внимание, что значения загрязнения в направлении восток-запад меняются медленнее, чем в направлении север-юг. Это происходит из-за того, что преобладающий ветер дует в направлении с запада на восток, в то время, как направление юг-север перпендикулярно направлению ветра.

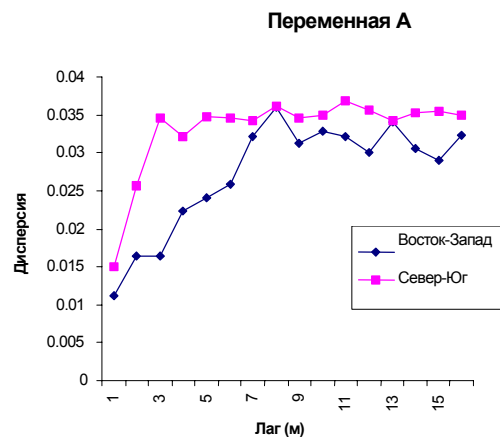
Форма кривой модели вариограммы/ковариации может также меняться в зависимости от направления (анизотропия), если вы вычли глобальный тренд из данных, или если тренда в данных нет. Анизотропия отличается от глобального тренда, рассмотренного выше, поскольку глобальный тренд может быть соотнесен с каким-либо физическим процессом (в предыдущем примере, преобладающим ветром) и описан математической формулой. Причина анизотропии (влияния по направлениям) на вариограмме, как правило, не известна, поэтому моделируется как случайная погрешность. Но даже без знания причины анизотропии, ее можно оценить количественно и учесть при выполнении интерполяции.



Анизотропия, - это, как правило, не детерминистский процесс, который может быть описан единой математической формулой. У нее нет единого источника, и она не имеет предсказуемого влияния на все измеренные значения. Анизотропия - это характеристика случайного процесса, который показывает, что пространственная автокорреляция сильнее проявляется в одном направлении, чем во всех остальных. На следующем рисунке проиллюстрировано, как процесс должен выглядеть теоретически. Снова более высокие концентрации загрязнения показаны теплыми цветами (красными и желтыми оттенками), а более низкие значения концентрации - холодными цветами (зеленым и синим). Случайный процесс показывает колебания, которые в одном направлении короче, чем в другом. Эти колебания могут быть результатом некоего неизвестного или неизмеренного физического процесса, но моделируются как случайный (стохастический) процесс с автокорреляцией в одном направлении.



Поскольку в данном примере присутствует анизотропия, после построения эмпирической вариограммы можно увидеть, что пространственные отношения различны для двух направлений. В направлении север-юг форма кривой вариограммы растет перед выравниванием быстрее.



При анизотропии форма кривой вариограммы может меняться в зависимости от направления. Изотропия - это случай, когда вид вариограммы не зависит от направления.

## Комбинирование моделей вариограммы

Часто существует два (или более двух) процесса, которые будут диктовать пространственное распределение какого-либо явления. Например, количество растительности (биомасса) может зависеть от высоты рельефа и от влажности почвы. Если такая зависимость известна, для интерполирования значений биомассы можно воспользоваться кокригингом. Вы можете использовать измеренные значения величины биомассы, в качестве первого набора данных, значения высот рельефа, в качестве второго набора данных, и значения величины влажности почвы, в качестве третьего набора данных (см. Главу 6, 'Построение поверхности с использованием методов геостатистики'). Для каждого из наборов данных вы можете подобрать различные модели вариограммы, поскольку они отражают различные пространственные структуры. Так для набора данных с высотами рельефа может лучше всего подойти сферическая модель, для набора данных со значениями влажности почвы - экспоненциальная модель, а для набора данных по биомассе - сочетание моделей. Модели могут быть затем скомбинированы, чтобы наилучшим образом охарактеризовать структуру данных.

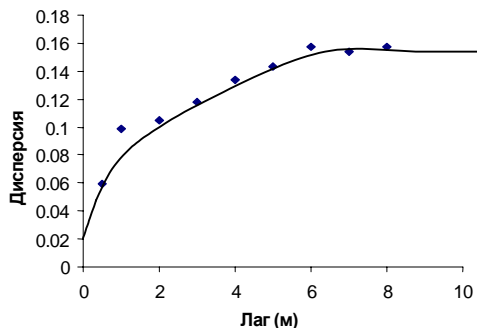
Однако, иногда вам неизвестны причинно-следственные отношения факторов, определяющих пространственную структуру

какого-либо явления. Если воспользоваться тем же примером с биомассой, то может случиться так, что у вас есть только опорные точки, в которых измерены значения величины биомассы. Когда вы изучите вариограмму, вы обратите внимание на точки перегиба.

Точки быстро растут, затем выравниваются, а затем снова образуют перегиб, после которого выравнивание идет до значения порога. Вы можете предположить, что в данных существуют две различные структуры, которые нельзя описать единой моделью. Вы можете подобрать модель вариограммы, используя две отличающиеся модели (например, сферическую и экспоненциальную), а затем скомбинировать их в единую модель.

Представление двух очевидных случайных процессов на одной вариограмме не рекомендуется; если это возможно, следует разделить пространственные процессы. Однако, не всегда можно понять причинно-следственные связи. Выбор нескольких моделей позволяет добавить большее количество параметров для оценки. При этом подбор моделей является субъективным и выполняется "на глаз". Количественно оценить правильность подбора комбинации из моделей вы можете, воспользовавшись методами перекрестной проверки и изучив статистику такой проверки (см. Главу 7, 'Использование аналитических инструментов при построении поверхностей').

Модуль Geostatistical Analyst позволяет выбрать до трех моделей помимо модели с учетом эффекта самородка. В приведенном выше примере, модель скомбинирована из трех компонентов: модели с учетом эффекта самородка и двух сферических моделей с различными радиусами влияния.





## Использование модуля Geostatistical Analyst для подбора модели вариограммы

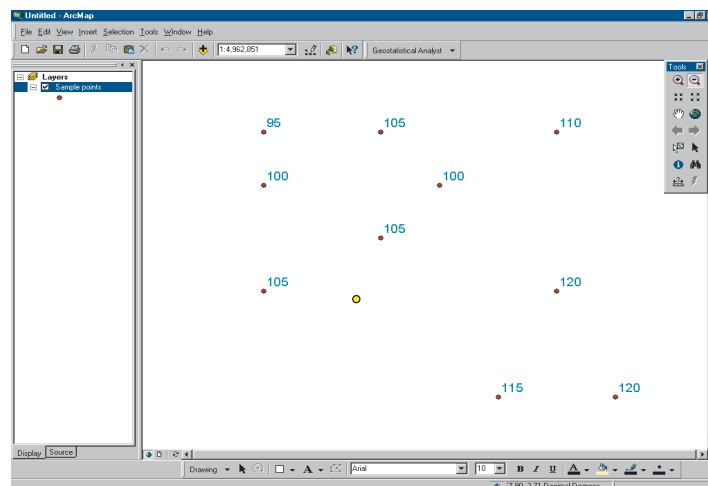
Пример, приведенный ранее в этой главе, был упрощен для облегчения понимания. Чтобы продемонстрировать понятие моделирования вариограммы, мы воспользуемся большим количеством опорных точек.

Из измеренных значений высот рельефа было отобрано десять опорных точек.

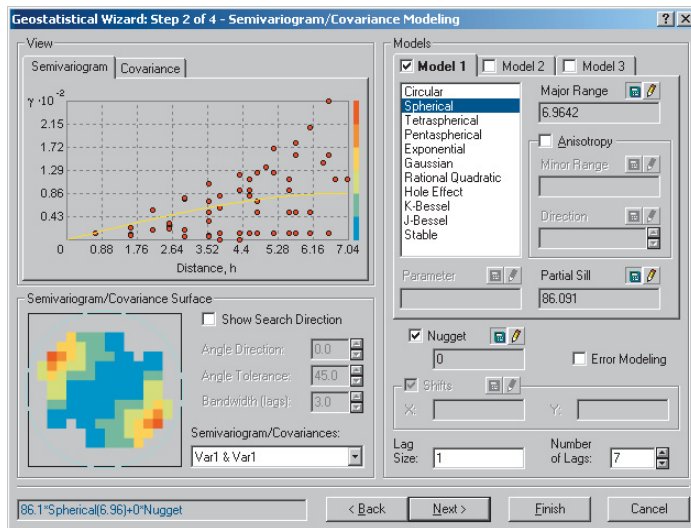
Номер точки	Координата X	Координата Y	Значение
1	1	3	105
2	1	5	100
3	1	6	95
4	3	4	105
5	3	6	105
6	4	5	100
7	5	1	115
8	6	3	120
9	6	6	110
10	7	1	120

В данном примере мы будем интерполировать пока неизвестное значение высоты для точки с координатами  $x = 2.75$ ,  $y = 2.75$  (на рисунке справа точка обозначена желтым цветом).

Пространственное распределение опорных точек, их значения и искомая точка показаны в документе ArcMap, изображение которого приведено ниже.



На первых двух панелях Мастера операций геостатистики, вы определяете набор данных, поле, значение которого вы интерполируете, и метод кригинга (в данном примере будет использован метод ординарного кригинга). Третья панель содержит диалог моделирования вариограммы. На этом этапе наша цель состоит в том, чтобы подобрать модель для эмпирической вариограммы. На рисунке на стр. 73 приведен список возможных моделей. В предыдущем примере для подбора была использована простая прямая линия, но как вы видите, существует и много других возможностей. Каждая из моделей несколько отличается одна от другой, поэтому у вас есть возможность выбрать лучшую. Более подробную информацию по используемым моделям вы можете найти в Главе 7, 'Использование аналитических инструментов при построении поверхностей'.



В данном примере для описания эмпирической вариограммы будет применена сферическая модель. Далее приведена формула сферической модели. Как видите, эта формула более сложная, чем для прямой линии, использованной в предыдущем примере в этой главе. Однако, обе формулы служат одной и той же цели, но с разными результатами.

$$\gamma(\mathbf{h}) = \begin{cases} \theta_s \left[ \frac{3}{2} \frac{h}{\theta_r} - \frac{1}{2} \left( \frac{h}{\theta_r} \right)^3 \right] & \text{for } 0 \leq h \leq \theta_r \\ \theta_s & \text{for } \theta_r < h \end{cases}$$

где

$\theta_s$  - значение порога,

$\mathbf{h}$  - вектор лага, а  $h$  - длина  $\mathbf{h}$  (расстояние между двумя опорными точками),

$\theta_r$  - радиус влияния модели.

Обратите внимание, что параметры сферической модели даны в нижнем левом углу диалога синими буквами. Это означает, что сферическая модель используется со значением порога, равным 86.1, радиусом влияния - 6.96, и нулевым самородком. Следовательно, вычисленные с помощью сферической модели значения вариограммы будут равны:

$g(\mathbf{h}) = 86.1 * (1.5 * (h/6.96) - 0.5 * (h/6.96)^3)$ , для всех значений лага  $h \leq 6.96$

и

$g(\mathbf{h}) = 86.1$ , для всех значений лага  $h > 6.96$

Это аналогично нахождению значения вариограммы для заданного расстояния,  $h$ , по выбранной линии в нашем предыдущем примере; поскольку линия была подобрана, по этой линии определялись значения для матрицы и векторов из уравнения ординарного кригинга. В данном случае то же самое может быть проделано с использованием выбранной сферической модели.

# Кригинг

Как и при выполнении интерполяции по методу взвешенных расстояний (IDW), при кригинге формируются веса для измеренных значений, находящихся в окрестностях искомой точки, которые используются при интерполировании ее значения. Как и в случае с интерполяцией по методу взвешенных расстояний, ближайшие опорные точки будут иметь большее влияние. Однако, веса кригинга для окружающих искомую точку опорных точек будут присваиваться более сложным и обоснованным способом, чем при методе взвешенных расстояний IDW. Метод взвешенных расстояний использует простой алгоритм, основанный на расстоянии, а веса кригинга вычисляются на основе значений вариограммы, учитывающей пространственную структуру данных. Для создания непрерывной поверхности или карты явления, для точек на изучаемой территории интерполяция выполняется с использованием вариограммы и с учетом пространственной организации опорных точек, расположенных в окрестностях искомой точки.

## Поиск соседства

Можно предположить, что при удалении опорных точек от искомой точки, измеренные значения будут меньше пространственно коррелировать с искомой точкой. Таким образом, при поиске соседства можно исключить точки, которые удалены от искомой точки на большое расстояние и оказывают на нее небольшое влияние. Но при удалении точек друг от друга, они не только меньше связаны между собой, - существует вероятность того, что точки, расположенные на значительном расстоянии от искомой, могут исказить результат в том случае, если они расположены на территории, сильно отличающейся от той, где находится искомая точка. Другая причина для использования области поиска соседства - скорость вычислений. Вспомните, как в первом примере мы находили обратную матрицу для матрицы размером 5x5. Если бы у вас было 2 000 опорных точек, такая матрица была бы слишком велика для нахождения обратной матри-

цы. Чем меньше область поиска соседства, тем быстрее может быть выполнена интерполяция. Поэтому обычно ограничивают число точек, используемых при интерполировании, путем определения области поиска соседей.

Заданная форма для поиска соседей ограничивает дальность поиска опорных точек, которые будут использованы в вычислениях, и его направление. Другие параметры поиска соседства ограничивают точки, которые будут использованы внутри заданной области.

Форма области поиска соседства зависит от исходных данных и от поверхности, которую вы хотите построить. Если пространственная автокорреляция ваших данных не испытывает влияний по направлениям, вы захотите учитывать опорные точки равномерно во всех направлениях. Поэтому вы, возможно, выберете круг в качестве формы области поиска. Однако, если в ваших данных присутствует пространственная автокорреляция по направлениям, вы можете выбрать в качестве формы области поиска соседства эллипс, большая ось которого параллельна направлению автокорреляции.

Область для поиска соседства может быть задана в шаге 3 Мастера операций геостатистики. После того, как вы определили область поиска соседства, вы можете ограничить выбор используемых опорных точек. Вы можете задать максимальное и минимальное количество соседей, которые будут включены в анализ. Вы можете также разделить область соседства на сектора с тем, чтобы убедиться, что вы используете значения во всех направлениях. Если вы разделите область соседства на сектора, заданное максимальное и минимальное количество соседей будет применено к каждому сектору.

Может быть использовано несколько различных типов секторов (см. рисунок внизу).



Используя конфигурацию данных в заданной области соседства, в сочетании с подобранной моделью вариограммы, можно определить веса опорных точек. На основе этих весов и значений опорных точек, можно выполнить интерполяцию неизвестного значения в искомой точке. Чтобы построить модель непрерывной поверхности, процесс необходимо повторить для каждой точки пространства.

**Построение поверхности проинтерполированных значений с использованием области поиска соседства**

По мере того, как ваши наборы данных становятся больше и покрывают все большую территорию, вы захотите ограничить число опорных точек, учитываемых при выполнении интерполяции. Если вы учитываете в вычислениях точки, удаленные от искомой на значительное расстояние, нужно принимать во внимание то, что они могут находиться на территории, сильно отличающейся от той, где расположена искомая точка. Вам необходимо учесть при расчетах количество точек, достаточное для формирования репрезентативной выборки, и исключить точки, удаленные от искомой на большое расстояние, поскольку они могут оказывать на нее очень маленькое влияние или находиться в области, сильно отличающейся от той территории, где расположена искомая точка (см. рисунок). В диалоге, приведенном на рисунке, в качестве области поиска соседства задан круг с радиусом 3, а максимальное число учитываемых соседей равно

Число включаемых соседей = 5

Стратегия поиска: круг с четырьмя квадрантами  
Радиус = 3

Координаты тестовой точки (x=2.75, y=2.75)

Точки, которые будут использованы

Веса

Проинтерполированное значение= 107.59

5. Точки, используемые для интерполирования неизвестного значения в искомой точке (2.75, 2.75), выделены на рисунке, и им присвоены цвета (в зависимости от процентного соотношения коэффициентов  $I_i$ ). В область соседства включены следующие

Номер точки области соседства	Исходный номер точки	Координата x	Координата y	Значение
1	1	1	3	105
2	2	1	5	100
3	4	3	4	105
4	6	4	5	100
5	7	5	1	115

точки:

Вычисленное значение для искомой точки ( $x = 2.75, y = 2.75$ ) равно 107.59. Модуль Geostatistical Analyst выполнил интерполяцию, решив уравнения ординарного кригинга.

$$\mathbf{G}^* \mathbf{l} = \mathbf{g} \text{ и решив уравнение для весов } \mathbf{l} = \mathbf{G}^{-1} * \mathbf{g}$$

С помощью сферической модели вариограммы и упорядоченного набора измеренных значений, определенных через поиск соседства, можно найти значения  $\mathbf{l}$  из приведенного выше уравнения. Сначала создадим матрицу  $\mathbf{G}$ . Это можно сделать, вычислив расстояния между парами точек и подставив эти значения в уравнение подобранной сферической модели,

$$g(h) = 86.1 * (1.5 * (h/6.96) - 0.5 * (h/6.96)^3), \text{ для } 0 < h < 6.96$$

Расстояния между опорными точками равны:

Точки	Расстояние	Точки	Расстояние
1,2	2.000	2,4	3.000
1,3	2.236	2,5	5.657
1,4	3.605	3,4	1.414
1,5	4.472	3,5	3.606
2,3	2.236	4,5	4.124

Если подставить расстояние ( $h$ ) между точками 1 и 3  $h = 2.236$ , значение вариограммы равно:

$$g(h) = 86.1 * (1.5 * (2.236/6.96) - 0.5 * (2.236/6.96)^3) = 40.065$$

Чтобы получить матрицу  $\mathbf{G}$ , повторите эту операцию для каждой пары точек. Чтобы сделать запись более понятной, точкам были присвоены другие номера (как показано в таблице на предыдущей странице).

i	1	2	3	4	5	6
1	0.000	36.091	40.065	60.920	71.564	1.000
2	36.091	0.000	40.065	52.221	81.855	1.000
3	40.065	40.065	0.0000	25.881	60.920	1.000
4	60.920	52.221	25.881	0.000	67.559	1.000
5	71.564	81.855	60.920	67.559	0.000	1.000
6	1.000	1.000	1.000	1.000	1.000	0.000

Затем найдем обратную матрицу  $\mathbf{G}^{-1}$ .

i	1	2	3	4	5	6
1	-0.0191	0.01005	0.00776	-0.0021	0.00336	0.2114
2	0.01005	-0.0187	0.00472	0.00402	-0.0001	0.24891
3	0.00776	0.00472	-0.0317	0.01619	0.00304	-0.1038
4	-0.0021	0.00402	0.01619	-0.0214	0.00324	0.27739
5	0.00336	-0.0001	0.00304	0.00324	-0.0095	0.36607
6	0.2114	0.24891	-0.1038	0.27739	0.36607	-47.922

Теперь для решения уравнения ординарного кригинга,  $\mathbf{l} = \mathbf{G}^{-1} * \mathbf{g}$ , необходимо создать вектор  $\mathbf{g}$ . Для этого, вычислите расстояние от каждой из пяти опорных точек до искомой точки (2.75, 2.75). Расстояния равны:

Из точки  $x = 2.75, y = 2.75$

Точки	Расстояние
1	1.768
2	2.850
3	1.275
4	2.574
5	2.850

Вектор **g** можно создать, подставив значение каждого расстояния в уравнение подобранной сферической модели.

Из точки  $x = 2.75, y = 2.75$

Точки Дисперсия, полученная по модели

1	32.097
2	49.936
3	23.390
4	45.584
5	49.936
6	1.000

Дополнительный столбец в векторе **g** (и дополнительные строка и столбец в матрице) были добавлены для того, чтобы сумма весов была равна 1 (т.е., был использован множитель Лагранжа, определение и содержание которого даны в Приложении А).

Теперь найдем веса для вектора **I**. Приведем пример нахождения значения веса для точки 1:

$$I_1 = (-0.019 \cdot 32.097 + 0.01005 \cdot 49.936 + 0.00776 \cdot 23.390 - 0.0021 \cdot 45.584 + 0.00336 \cdot 49.936 + 0.2114 \cdot 1.000) = 0.355$$

Веса всех точек и множителя Лагранжа (точка 6) равны:

Точки	$I_i$
1	0.355
2	-0.073
3	0.529
4	-0.022
5	0.211
6	-0.210

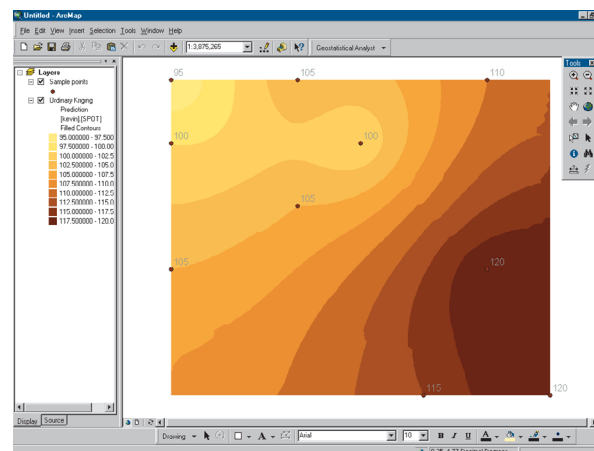
Наконец, найдем значение в искомой точке (2.75, 2.75), умножив веса измеренных точек (за исключением точки 6) на их значение и сложив их между собой.

$$\text{Искомое значение} = 0.355 \cdot 105 - 0.073 \cdot 100 + 0.529 \cdot 105 - 0.022 \cdot 100 + 0.211 \cdot 115$$

$$\text{Искомое значение} = 107.59$$

i	$\lambda_i$	Значение <sub>i</sub>
1	0.355	105
2	-0.073	100
3	0.529	105
4	-0.022	100
5	0.211	115

Повторение этого процесса для нескольких искомых точек и нанесение результатов на карту позволяет построить поверхность, показанную на рисунке внизу.



Результирующие поверхности могут быть построены с использованием модуля Geostatistical Analyst в целом ряде форматов, включая шейп-файл с изолиниями, шейп-файл с полигонами с заливкой, грид, представляющий непрерывную поверхность, и отмыску.

# Модуль Geostatistical Analyst

В этом последнем разделе, вы получите дополнительную информацию о модуле Geostatistical Analyst, как о расширении к ArcMap.

Доступ к модулю можно получить из меню Geostatistical Analyst на панели инструментов ArcMap. В модуль Geostatistical Analyst включены три основных компонента: (1) Исследование данных, (2) Мастер операций геостатистики и (3) Создание поднаборов.

## Исследование данных

Перед тем как воспользоваться методами интерполяции, вы можете исследовать свои данные с помощью инструментов исследовательского анализа пространственных данных (ESDA). Инструменты ESDA помогут вам понять сущность ваших данных и выбрать соответствующие параметры для модели интерполяции. Например, при использовании ординарного кригинга для создания карты квантилей, вам необходимо изучить распределение данных, поскольку для этого метода необходимо, чтобы данные подчинялись закону нормального распределения. Кроме этого, вы можете исследовать данные с помощью инструментов ESDA на наличие тренда, а также не учитывать тренд в процессе интерполяции.

При работе с модулем вы можете использовать следующие инструменты:

- Гистограмма—Исследование одномерного распределения набора данных.
- Карта Вороного—Анализ стационарности и пространственной изменчивости набора данных.
- Нормальный график КК (Квантиль-квантиль)—Проверка нормальности распределения данных.
- Анализ тренда—Определение глобальных трендов в наборе данных.
- Облако вариограммы/ковариации—Анализ пространственных зависимостей в наборе данных.

- Общий график КК (Квантиль-квантиль)—Исследование того, имеют ли два набора данных одно и то же распределение.
- Облако взаимной ковариации—Понимание взаимной ковариации между двумя наборами данных.

## Мастер операций геостатистики

Модуль Geostatistical Analyst предоставляет целый ряд методов интерполяции, которые используют опорные точки для построения поверхности изучаемого явления. Методы интерполяции в модуле Geostatistical Analyst разделены на два основных типа: детерминистские и геостатистические.

### Детерминистские

Детерминистские методы основаны на параметрах, которые контролируют либо (i) область распространения сходных значений (например, метод взвешенных расстояний - Inverse Distance Weighted) или (ii) степень сглаживания поверхности (например, радиальные базисные функции). Эти методы не используют модель стохастических пространственных процессов.

### Геостатистика

Геостатистика предполагает, что хотя бы часть пространственной неоднородности природных явлений может быть смоделирована вероятностными процессами с пространственной автокорреляцией.

Методы геостатистики можно использовать для:

- Описания и моделирования пространственных структур — вариография.
- Интерполирования значений в точках, не имеющих измеренных значений—кригинг.
- Оценки неопределенности, связанной с интерполяцией значений в искомой точке—кригинг.



Кригинг может быть использован для построения следующих поверхностей:

- Карты проинтерполированных значений, полученных на основе кригинга
- Карты стандартных ошибок, связанных с вычисленными значениями
- Карты вероятности, показывающие, был ли превышен заданный критический уровень
- Карты квантилей для предварительно заданных значений вероятности

## Создание поднаборов данных

Наиболее точный способ оценить качество результирующей поверхности - сравнить полученные при выполнении интерполяции значения с полевыми измерениями. Часто нет возможности вернуться в изучаемый район, чтобы получить независимый набор данных, который можно будет использовать для проверки. Одно из решений состоит в разделении исходного набора данных на две части. Одна часть может быть использована для моделирования пространственной структуры и построения поверхности. Другая часть может быть использована для сравнения и проверки качества интерполяции. Диалог Создать поднаборы данных позволяет вам создавать и тестовый, и учебный набор данных.

## Обработка данных

Модуль содержит много инструментов для анализа данных и построения целого ряда результирующих поверхностей.

Хотя цели исследования могут отличаться, при анализе/картографировании пространственных процессов мы рекомендуем придерживаться следующего подхода:

### Отобразить данные

Добавить слои и отобразить их в ArcMap.

### Исследовать данные

Исследовать статистические и пространственные свойства данных.

### Подобрать модель

Выбрать модель для построения поверхности. Для определения и уточнения соответствующей модели используется Мастер операций геостатистики.

### Выполнить диагностику

Оценить качество результирующей поверхности с использованием инструментов перекрестной и простой проверки. Это поможет вам понять, насколько хорошо модель интерполирует значения в искомой точке.

### Сравнить модели

Может быть построено более одной поверхности. Поверхности могут быть сопоставлены с использованием статистики перекрестной проверки.



# Исследовательский анализ пространственных данных (ESDA)

# 4

## В ЭТОЙ ГЛАВЕ

- Что такое Исследовательский анализ пространственных данных?
- Инструменты Исследовательского анализа пространственных данных
- Изучение распределения данных
- Определение глобальных и локальных выпадающих значений
- Определение глобальных трендов
- Изучение пространственной автокорреляции и вариации по направлениям
- Изучение ковариации между несколькими наборами данных

Исследовательский анализ пространственных данных (ESDA) позволяет вам исследовать данные различными методами. Перед построением поверхности, ESDA помогает вам глубже понять изучаемое явление с тем, чтобы вы могли принимать более обоснованные решения при работе с данными. Каждый новый вид данных может быть обработан и изучен, и инструменты ESDA помогут вам по-разному взглянуть на одни и те же данные. Все инструменты ESDA связаны между собой и с отображением данных в ArcMap. То есть, если на гистограмме вы выбираете столбец, точки, относящиеся к этому столбцу, будут также выбраны на графике КК (если он открыт), либо в любом другом открытом диалоге ESDA, либо на карте ArcMap.

Среда ESDA разработана, как следует из ее названия, для целей исследования. Однако, выполнение некоторых задач целесообразно для большинства научных работ. Изучение распределения данных, определение глобальных и локальных выпадающих (экстремальных) значений, поиск глобальных трендов, изучение пространственной автокорреляции и понимание ковариации между несколькими наборами данных, - все это задачи, выполнение которых полезно для понимания структуры данных. Инструменты ESDA могут помочь вам в решении этих и многих других задач.

# Что такое Исследовательский анализ пространственных данных (ESDA)?

Среда ESDA позволяет вам графически исследовать набор данных с тем, чтобы лучше понять его структуру. Каждый инструмент ESDA по-разному представляет данные и отображается в отдельном окне. ESDA включает следующие диалоги: Гистограмма, Карта Вороного, Нормальный график КК, Анализ тренда, Облако вариограммы/ковариации, Общий график КК, и облако взаимной ковариации. Все виды связаны между собой и с картой ArcMap.

## Выбор объектов; выбор группы объектов и динамические связи выборки

Диалоги ESDA связаны между собой, и, если вы выбираете точки в одном из окон, они выделяются цветом на всех картах и графиках. Для выбора объектов в виде данных ArcMap или в диалогах инструментов ESDA вы можете воспользоваться инструментом выбора группы объектов (при использовании этого инструмента будут выбраны объекты, попадающие в прямоугольник заданного вами размера). Любой выбор, осуществленный в виде ESDA или на карте ArcMap, будет продублирован во всех окнах ESDA и на карте ArcMap (см. диаграмму на следующей странице); такое свойство носит название динамической связи выборки. Если вы выбираете точки в диалогах некоторых инструментов ESDA (на гистограмме, карте Вороного, графике КК и в окне анализа тренда), выбранные точки вида связываются с картой ArcMap, и соответствующие точки выделяются на карте цветом. Поскольку точки на графиках вариограммы/ковариации представляют пары опорных точек, при выборе точек в окне инструмента Облако вариограммы/ковариации, на карте цветом выделяются пары точек и отображаются отрезки, соединяющие каждую пару. Когда в виде данных ArcMap выбираются пары точек, точки на графике вариограммы/ковариации также выделяются цветом.

## Взаимодействие слоев ArcMap и инструментов ESDA

Инструменты ESDA взаимодействуют со слоями ArcMap следующим образом:

1. Инструменты ESDA работают со слоями, содержащими точечные и полигональные объекты (например, данные переписи, эпидемиологические или демографические данные).

2. Слой, который был выделен в таблице содержания ArcMap, перед тем, как вы открыли инструмент ESDA, будет использоваться этим инструментом, как слой, предложенный по умолчанию.
3. Если для выделенного в ArcMap слоя не определена возможность выбора объектов, слоем, предложенным по умолчанию при запуске инструмента ESDA, будет верхний слой в таблице содержания.
4. В списке слоев, которые могут быть исследованы с помощью инструмента, в окне соответствующего инструмента отобразятся только те слои с точечными объектами, которые доступны для выбора объектов.
5. Для любого слоя будут сохранены сформулированный запрос.
6. Слой, который исследуется с помощью инструментов ESDA, не обязательно должен отображаться в виде данных ArcMap; но если он не отображается в ArcMap, выборка не будет видна на карте.
7. Если в таблице содержания ArcMap выделен только один точечный слой, и выбран инструмент ESDA, для работы с которым необходимо несколько слоев, выделенный в таблице содержания слой будет по умолчанию первым набором данных для данного инструмента.
8. Если в таблице содержания ArcMap выделено два или большее количество слоев, и выбран инструмент ESDA, работающий с несколькими слоями, первый выбранный в таблице содержания точечный слой служит первым исходным набором данных для этого инструмента ESDA, а второй выделенный слой - вторым набором данных.

## Преобразования

Несколько методов, используемых в модуле Geostatistical Analyst, требуют условия нормального распределения данных. Когда данные асимметричны (распределение неравномерное, одностороннее), возможно, вы захотите преобразовать данные, чтобы привести их к нормальному распределению. Такие инструменты ESDA, как Гистограмма и график КК, позволяют вам изучить влияние различных преобразований на распределение набора данных. Если вы выберете опцию преобразования данных перед построением поверхности с использованием методов геостатистики, для вашей проинтерполированной поверхности вычисленные значения будут преобразованы обратно к исходному виду.

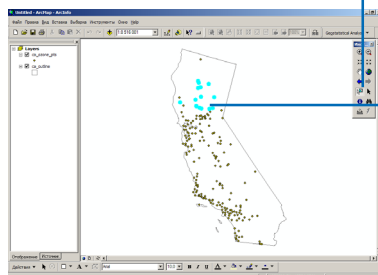
# Исследовательский анализ пространственных данных (ESDA)

## Выбор опорных точек

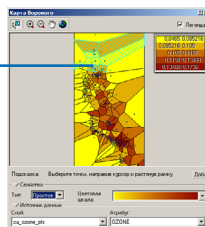
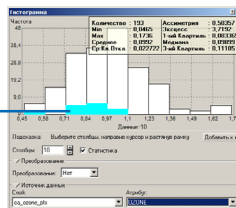
Выбор по местоположению

Инструмент  
Гистограмма

Инструмент выбора



Вид данных ArcMap

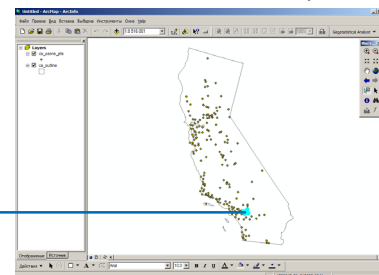


Инструмент построения карты  
по методу Вороного

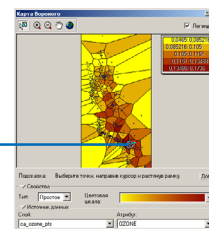
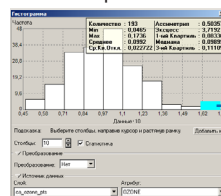
Выберите интересующие вас точки в виде данных ArcMap. Инструменты ESDA отобразят выбранные точки на фоне всех точек (например, как частичную гистограмму).

Выбор с использованием инструмента ESDA

Вид данных ArcMap



Инструмент  
Гистограмма



Инструмент построения карты  
по методу Вороного

Выберите объект, который вас интересует (например, хвост гистограммы) в окне инструмента ESDA. Местоположение выбранных точек отобразится в окне вида ArcMap. Когда вы откроете новый инструмент ESDA, он также покажет соответствующие свойства выбранных точек (например, местоположение ячеек Вороного).

# Инструменты Исследовательского анализа пространственных данных (ESDA)

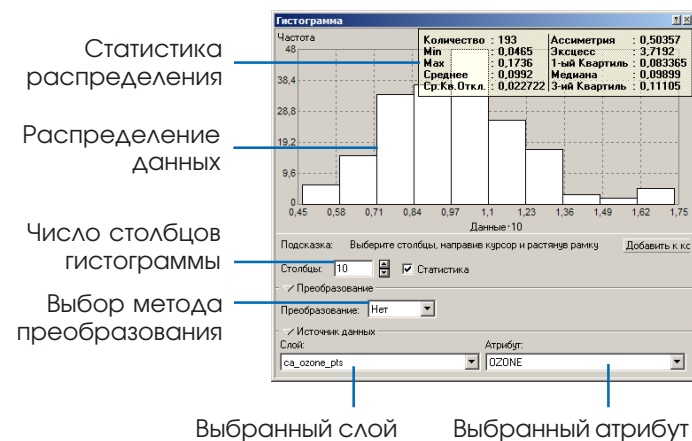
Каждый инструмент ESDA предоставляет вам возможность исследовать данные с различных точек зрения. Каждый инструмент использует отдельное окно, которое взаимодействует с отображением данных в ArcMap и в окнах других инструментов ESDA. Вы можете работать со следующими инструментами: Гистограмма, карта Вороного, Нормальный график КК, Анализ тренда, Облако вариограммы/ковариации, Общий график КК и Облако взаимной ковариации.

## Гистограмма

Инструмент гистограммы в ESDA дает одномерное (по одной переменной) описание ваших данных. Инструмент показывает плотность распределения для интересующего вас набора данных и подсчитывает суммарную статистику.

### Плотность распределения

Плотность распределения - это столбчатая диаграмма, которая показывает, насколько часто наблюдаемые значения попадают в тот или иной интервал или класс. Вы задаете число классов равной длины, которые будут использованы на гистограмме. Количество значений, попадающих в каждый класс, пропорционально высоте столбца. В качестве примера на гистограмме, приве-



денной на этой странице, показана плотность распределения для набора данных со значениями концентрации озона (10 классов).

### Суммарная статистика

Распределение может быть охарактеризовано с помощью некоторых суммарных показателей, описывающих его форму, положение и размах.

### Меры положения

Меры положения дают вам представление о том, где находятся центр и другие параметры распределения.

Среднее (mean) - это арифметическое среднее из значений данных. Среднее характеризует положение центра распределения.

Значение медианы (median) соответствует совокупной доле половины (0.5). Если данные были организованы в порядке возрастания, это означает, что 50 процентов значений будут ниже медианы, а 50 процентов значений - выше медианы. Медиана дает еще одну меру центра распределения.

Первый и третий квартили (quartile) характеризуют совокупные доли 0.25 и 0.75, соответственно. Если данные были организованы в порядке возрастания, 25 процентов значений будет ниже первого квартиля, и 25 процентов значений - выше третьего квартиля. Первый и третий квартили - это частный случай квантилей. Квантили рассчитываются по следующей формуле:

$$\text{квантиль} = (i) - 0.5 / N$$

где  $(i)$  -  $i^{\text{ый}}$  класс упорядоченных значений данных, а  $N$  - количество значений данных.

### Характеристики размаха

Размах точек относительно среднего значения - еще одна характеристика диаграммы плотности распределения. Дисперсия (variance) данных равна среднему из квадратов отклонения значений всех точек от среднего. Единицы измерения дисперсии - квадрат единиц измерения, в которых выполнялись исходные

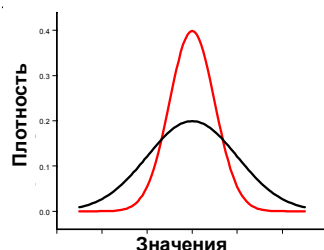


измерения, и поскольку для вычисления дисперсии используются квадраты разностей, она чувствительна к экстремально высоким и экстремально низким значениям.

Стандартное отклонение (standard deviation) равно корню квадратному из дисперсии. Оно характеризует распределение данных относительно среднего в тех же единицах измерения, что и исходные измерения. Чем меньше дисперсия и стандартное отклонение, тем плотнее измеренные значения сгруппированы около среднего.

На нижней диаграмме показаны два распределения с различными стандартными отклонениями. Частотное распределение, показанное черной линией, более пологое (значения данных более разнообразны и их размах шире), чем распределение, показанное красным цветом. Дисперсия и стандартное отклонение для пространственного распределения, показанного черным цветом, больше, чем для частотного распределения, показанного красным цветом.

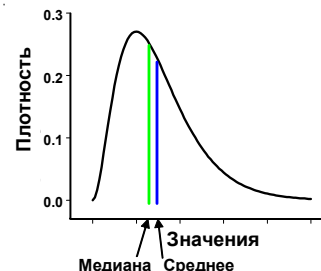
### Меры формы



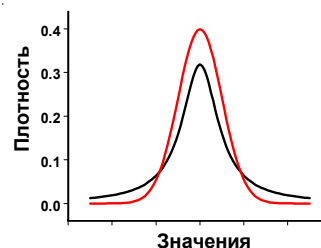
Плотность распределения можно охарактеризовать также формой.

Коэффициент асимметрии - это мера симметрии распределения. Для симметричных распределений этот коэффициент равен нулю. Если у графика распределения есть длинный правый хвост больших значений, распределение имеет положительную симметрию, а если график имеет длинный левый хвост малень-

ких значений, распределение имеет отрицательную симметрию. Для распределений с положительной симметрией, среднее больше медианы, и наоборот, для распределений с отрицательной симметрией, среднее меньше медианы. На нижнем рисунке показано распределение с положительной симметрией.



Экссесс кривой плотности распределения зависит от размера хвостов графика и дает меру того, насколько вероятно, что в распределении будут встречаться резко выделяющиеся, выпадающие значения. Экссесс нормального распределения равен 3. Распределения с относительно толстыми хвостами являются “островершинными” и имеют эксцесс больше 3. Распределения с относительно тонкими хвостами являются “плосковершинными”, и их эксцесс меньше 3. На рисунке, приведенном внизу, нормальное распределение показано кривой красного цвета, а кривая с эксцессом больше нормального (островершинная) дана черным цветом.





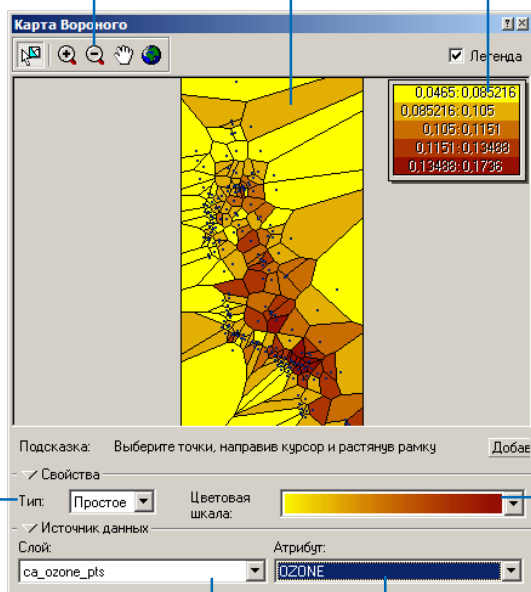
## Карта Вороного

Карты Вороного строятся из серий полигонов, образуемых вокруг опорных точек.

Инструменты  
(масштабирование,  
перемещение по  
карте, и т.п.)

Карта  
Вороного

Значения  
ячеек



Выбор  
метода

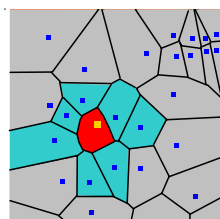
Выбранный  
набор данных

Выбранный  
атрибут

Цветовая  
шкала

Полигоны Вороного создаются таким образом, чтобы каждая точка внутри полигона находилась ближе к рассматриваемой опорной точке, чем к любой другой опорной точке. После того, как полигоны созданы, соседи опорной точки определяются как любая другая опорная точка, чей полигон имеет общую границу с выбранной опорной точкой.

Например, на следующем рисунке яркая желтая опорная точка окружена полигоном, обозначенным красным цветом. Каждая точка, попадающая в красный полигон, ближе к желтой опорной точке, чем к любой другой опорной точке (опорные точки показаны небольшими темно-синими точками). Все полигоны, обозначенные голубым цветом, имеют общую границу с красным полигоном, таким образом, опорные точки, расположенные в голубых полигонах, являются соседями желтой опорной точки.



Используя определение соседей, можно вычислить целый ряд локальных статистических показателей. Например, локальное среднее может быть найдено, как среднее из значений опорных точек, попадающих в красный и голубые полигоны. Затем это среднее значение присваивается красному полигону. После того, как эта операция будет выполнена для всех полигонов и их соседей, с помощью цветовой шкалы будут показаны значения локальных средних, чтобы помочь визуализировать регионы высоких и низких значений.

Инструмент составления карты по методу Вороного предлагает целый ряд методов для присвоения полигонам вычисленных значений.

С использованием простого значения: Значение, присваиваемое ячейке, - это значение опорной точки, попадающей в этот полигон.

С использованием среднего: Значение, присваиваемое ячейке, - среднее, полученное из значений опорной точки ячейки и ее соседей.

С использованием моды: Все значения ячеек группируются в пять классов. Значение, присваиваемое ячейке, - мода (наиболее часто встречающееся значение) ячейки и ее соседей.

По методу кластера: Все ячейки группируются в пять классов. Если интервал класса ячейки отличается от каждого из интервалов класса соседей, ячейка окрашивается в серый цвет, чтобы отличить ее от соседей.

По методу энтропии: Все ячейки распределяются по пяти классам на основе естественной группировки значений данных (т.н., “регулируемые” квантили; обратитесь к Главе 8, ‘Отображение геостатистических слоев и управление ими’). Значение, присваиваемое ячейке - это энтропия, вычисляемая на основе значений опорных точек в ячейке и ее соседях,

$$\text{Энтропия} = -S(p_i * \text{Log } p_i)$$

где  $p_i$  - частота встречаемости ячеек, отнесенных к каждому классу.

Например, рассмотрим ячейку с четырьмя соседями (общее количество ячеек равно 5). Их значения помещены в соответствующие классы:

Class	Frequency	$p_i$
1	3	3/5
2	0	0
3	1	1/5
4	0	0
5	1	1/5

Значение энтропии, присвоенное ячейке, будет равно:

$$E = -[0.6 * \log_2(0.6) + 0.2 * \log_2(0.2) + 0.2 * \log_2(0.2)] = 1.371$$

Значение энтропии будет минимальным, если значения всех ячеек попадают в один и тот же класс. Тогда,

$$E_{\min} = -[1 * \log_2(1)] = 0$$

Значение энтропии будет максимальным, когда все ячейки попадают в разные классы. Тогда,

$$E_{\max} = -[0.2 * \log_2(0.2) + 0.2 * \log_2(0.2) + 0.2 * \log_2(0.2) + 0.2 * \log_2(0.2) + 0.2 * \log_2(0.2)] = 2.322$$

По методу медианы: Значение, присваиваемое ячейке, - медиана, вычисленная для плотности распределения ячейки и ее соседей.

По методу стандартного отклонения: Значение, присваиваемое ячейке, - стандартное отклонение, вычисленное для значений ячейки и ее соседей.

Диапазон между квантилями: Первый и третий квантили рассчитываются для плотности распределения ячейки и ее соседей. Значение, присваиваемое ячейке, вычисляется путем вычитания значения первого квантиля из значения третьего квантиля.

Другие статистические показатели для карты Вороного используются для других целей. Статистические показатели могут быть сгруппированы в следующие функциональные категории:

#### *Локальное сглаживание*

Среднее

Мода

Медиана

#### *Локальные отклонения*

Стандартное отклонение

Диапазон между квантилями

Энтропия

#### *Локальные выпадающие значения*

Кластер

#### *Локальное влияние*

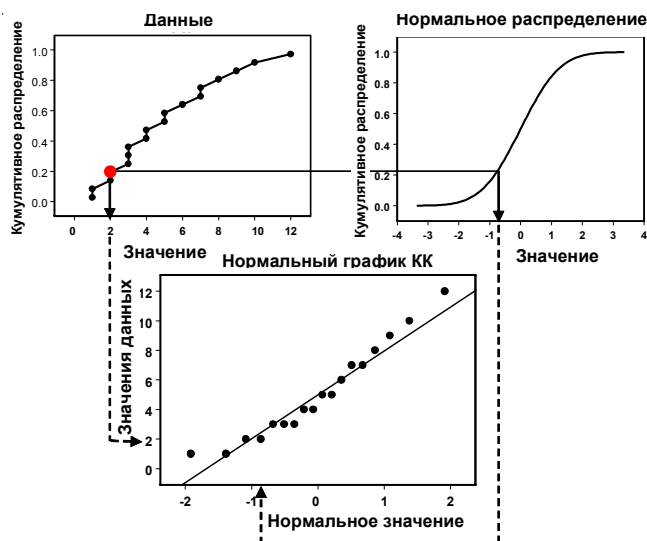
Простое значение

## Нормальный и общий графики КК (Квантиль-квантиль)

Графики КК Plots - графики, на которых квантили из двух распределений рассматриваются относительно друг друга.

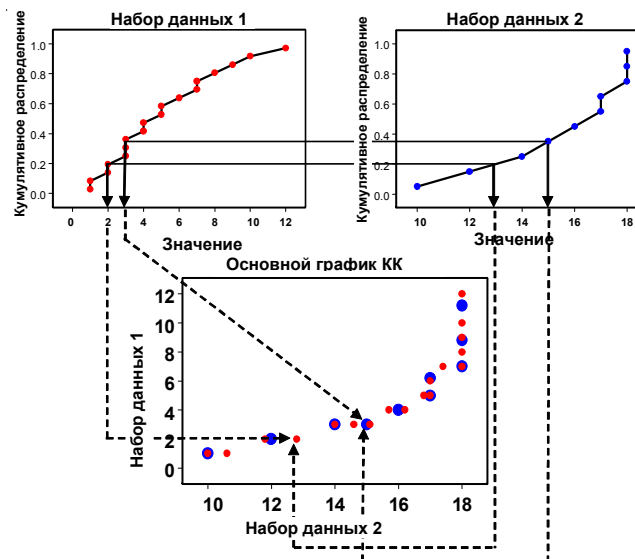
### Построение нормального графика КК

Кумулятивное (совокупное) распределение данных можно получить, упорядочив данные и построив график, по одной оси которого откладываются эти упорядоченные значения, а по другой - значения совокупного распределения, вычисляемого по формуле  $(i - 0.5) / n$  для  $i$ -ого упорядоченного значения из общего числа  $n$  значений (доли данных ниже значения). Между значениями используется линейная интерполяция. Нормальный график КК строится путем нанесения на соответствующие оси координат значений из набора данных и значений, полученных по кривой нормального распределения, соответствующих одинаковому значению кумулятивного распределения (см. нижний рисунок).



### Построение общего графика КК

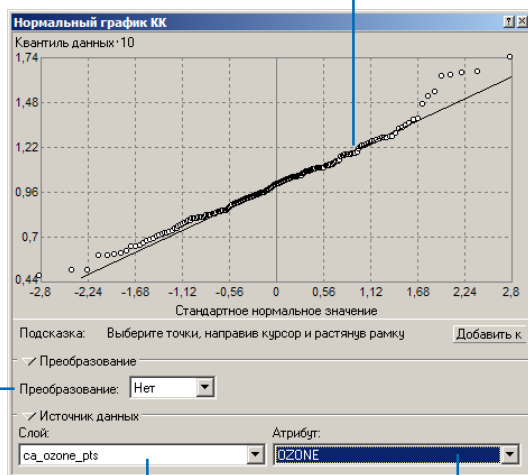
Общий график КК используется для оценки схожести распределений двух наборов данных. Общий график КК создается путем нанесения значений данных, для которых совокупное распределение имеет равные значения (см. нижний рисунок).



## Нормальный график КК (Квантиль-квантиль)

Используемый метод преобразования

График квантилей исходного набора данных относительно квантилей стандартного нормального распределения

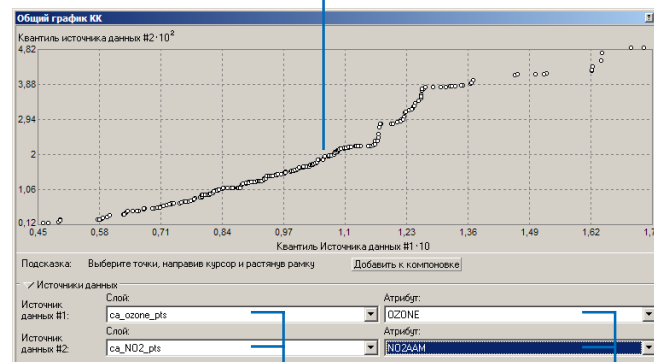


Исходный набор данных

Используемый атрибут

## Общий график КК (Квантиль-квантиль)

График квантилей двух наборов данных



Исходные наборы данных

Используемые атрибуты

## Анализ тренда

Вам может быть также интересно построить карту, отражающую тренд данных, или вычесть тренд из набора данных перед использованием кригинга. Инструмент Анализ тренда может помочь вам в определении глобального тренда в ваших данных.

Инструменты  
(перемещение по  
карте,  
масштабирование,  
т.п.)

Линия тренда в  
направлении  
восток-запад

Линия тренда в  
направлении  
север-юг

Каждая линия  
показывает  
положение и  
значение (высоту)  
каждой опорной  
точки

Углы отображения

Контроль движения  
по вертикали

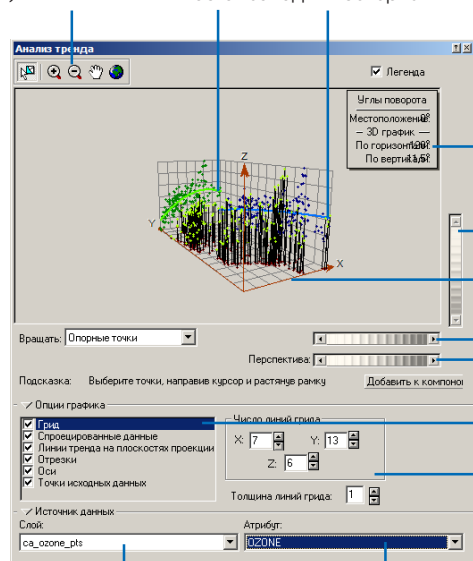
Оси карты

Контроль движения  
по горизонтали

Контроль  
перспективы

Опции  
отображения

Параметры  
отображения

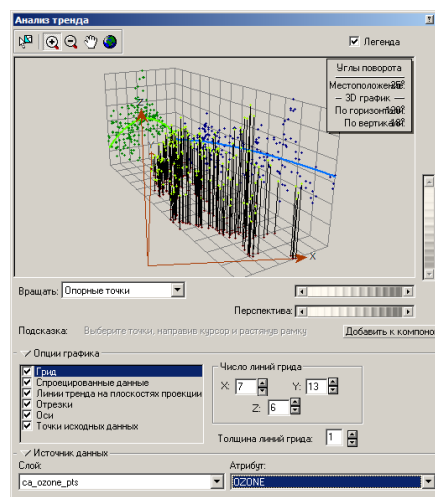


Выбранный набор данных

Выбранный атрибут

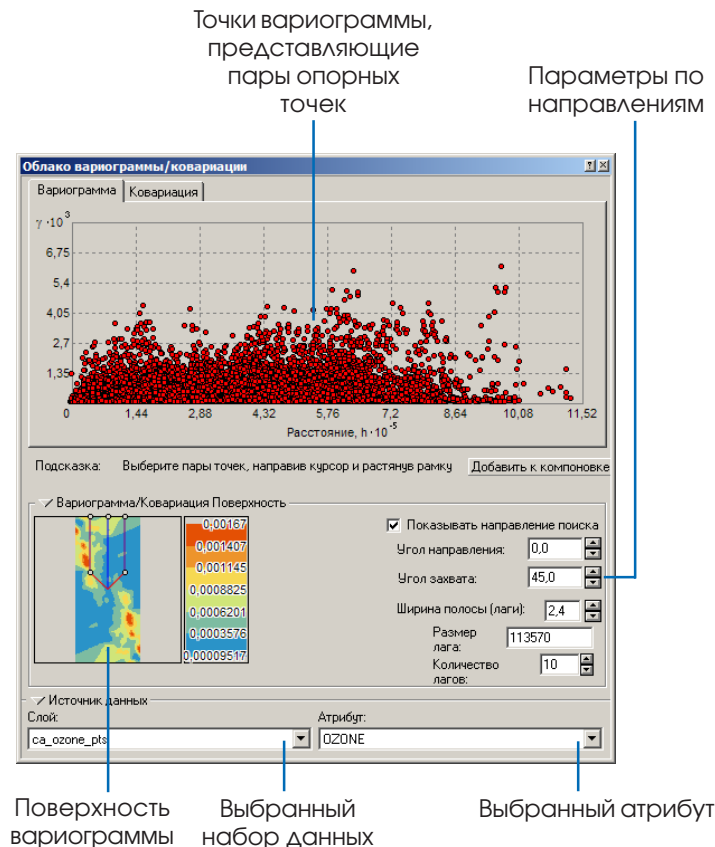
Инструмент Анализ тренда позволяет увидеть данные в трехмерном изображении. Координаты опорных точек наносятся в плоскости  $x, y$ . В каждой опорной точке высотой отрезка показано значение точки (по оси  $z$ ). Уникальная особенность инструмента Анализа тренда состоит в том, что значения точек затем

проецируются на плоскости  $x, z$  и  $y, z$ , образуя точечные графики. Эти графики могут рассматриваться как боковые проекции трехмерных данных. Затем к точечным графикам на плоскостях проекций подбираются полиномы, аппроксимирующие их расположение. Дополнительная особенность состоит в том, что вы можете вращать данные, чтобы выделить в значениях тренды по направлениям. Существует целый ряд других функциональных особенностей, которые позволяют вам вращать и менять перспективу целого изображения, менять размер и цвет точек и линий, удалять плоскости и точки, и выбирать степень полинома, который будет использован для аппроксимации точечных графиков. На диаграмме, приведенной внизу, данные развернуты под углом 25 градусов по часовой стрелке, а для аппроксимации точечного графика используется полином второй степени. На задней плоскости виден квадратичный тренд (показан зеленым цветом): кривая начинается с низких значений, затем значения растут, и потом снова падают. Тренд на правой плоскости (показан голубой линией) является более линейным и более постепенным.

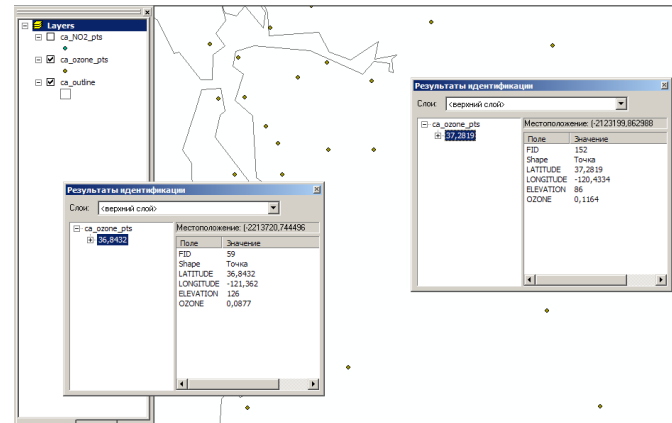


## Облако вариограммы/ковариации

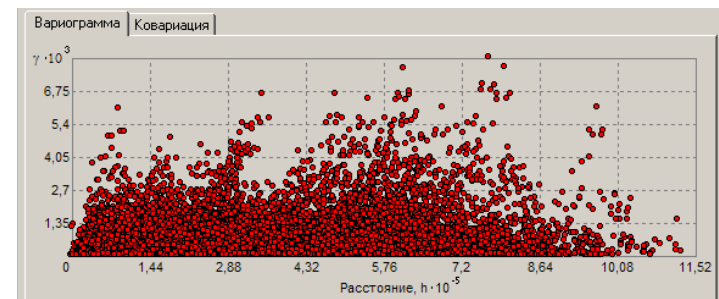
Облако вариограммы/ковариации - это эмпирическая вариограмма (на которой показаны значения половины квадрата разности между измеренными величинами) и ковариация для всех пар опорных точек из набора данных; на облаке вариограммы эти значения показаны как функция расстояния между двумя точками.



Пусть  $z(s_i)$  обозначает значение набора данных в  $i$ -той точке.



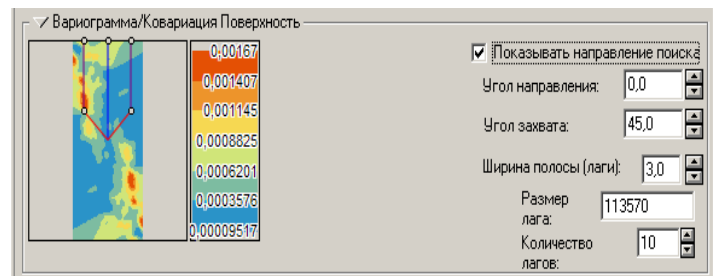
Тогда значение эмпирической вариограммы для пары  $(i, j)$  равно  $0.5 \cdot (z(s_i) - z(s_j))^2$ , а эмпирическая ковариация -  $(z(s_i) - \bar{z})(z(s_j) - \bar{z})$ , где  $\bar{z}$  - среднее значение данных. Облако вариограммы/ковариации может быть использовано для изучения локальных характеристик пространственной автокорреляции в наборе данных и определения выпадающих значений в выборке. Облако вариограммы/ковариации выглядит следующим образом:



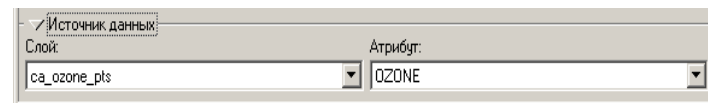
где каждая красная точка - это эмпирическая вариограмма (половина квадрата разницы значений, нанесенная на график как функция расстояния между точками) для всех пар опорных точек из набора данных. Вы можете выделить точки с помощью инструмента выбора и увидеть связанные пары точек в ArcMap.

Вы также можете воспользоваться поверхностью вариограммы с возможностями поиска по направлению. Значения на облаке вариограммы распределяются по бинам, основанным на направлениях и расстоянии между парами опорных точек. Эти значения бинов усредняются и сглаживаются, а затем по ним строится поверхность вариограммы. Справа дана цветовая шкала и значения на границах перехода между цветовыми обозначениями. Область отображения поверхности вариограммы контролируется размером лага и количеством лагов. (см. Главу 3, 'Принципы геостатистического анализа', для дополнительной информации о поверхности вариограммы, бинах и лаге вариограммы.)

Вы можете просматривать поднаборы значений на облаке вариограммы, выбрав направление поиска, а затем щелкнуть на указатель направления, если вы хотите изменить его размер или ориентацию. Вы также можете щелкнуть по серой стрелке в верхней части панели, чтобы временно скрыть эту часть инструмента.



Вы выбираете набор данных и атрибут с использованием следующей панели:

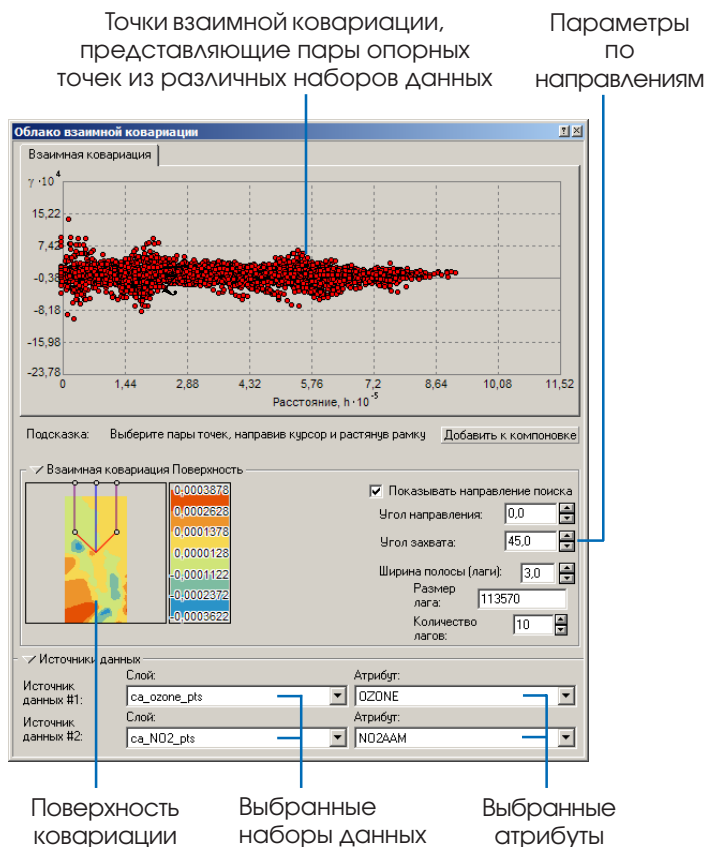


где OZONE - поле таблицы атрибутов, в котором хранятся значения концентрации озона. Вы можете щелкнуть по серой стрелке вверху панели, чтобы временно спрятать эту часть инструмента.

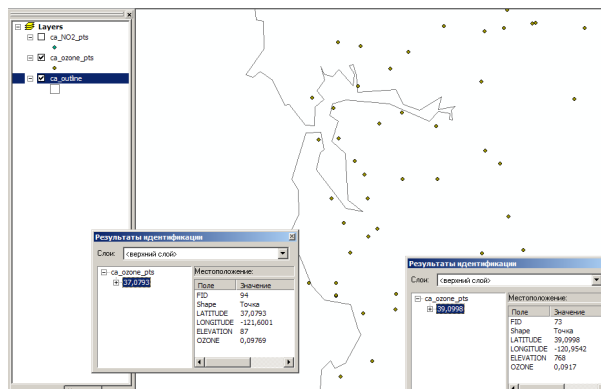


## Облако взаимной ковариации

Облако взаимной ковариации показывает эмпирическую взаимную ковариацию для всех пар точек между двумя наборами данных; точки на графике даны как функция расстояния между двумя точками.

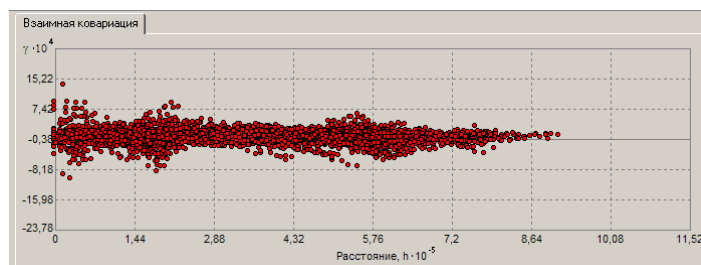


Пусть  $z(s_i)$  обозначает значение  $i$ -ой точки в первом наборе данных, а  $y(t_j)$  - значение  $j$ -той точки во втором наборе данных.



Тогда значение эмпирической взаимной ковариации для пары  $(i, j)$  находится по формуле  $(z(s_i) - \bar{z})(y(t_j) - \bar{y})$ ,

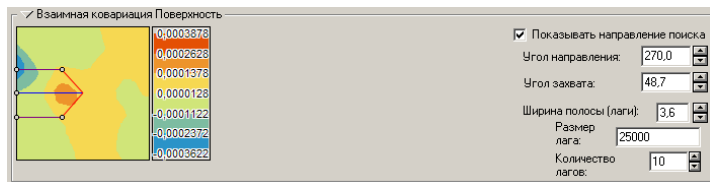
где  $\bar{z}$  и  $\bar{y}$  - средние значения для первого и второго наборов данных, соответственно. Облако взаимной ковариации может быть использовано для локальных характеристик пространственной автокорреляции между двумя наборами данных; по облаку взаимной ковариации могут быть определены также пространственные сдвиги в корреляции между двумя наборами данных. Облако взаимной ковариации выглядит следующим образом:



где каждая красная точка соответствует эмпирической взаимной ковариации между парами опорных точек в каждом наборе данных. Вы можете выбрать точки и просмотреть их связанные пары в ArcMap. (Чтобы различать, какие пары точек относятся к каждому из наборов данных, в диалоге свойств каждого набора данных задайте свой цвет выбора.)

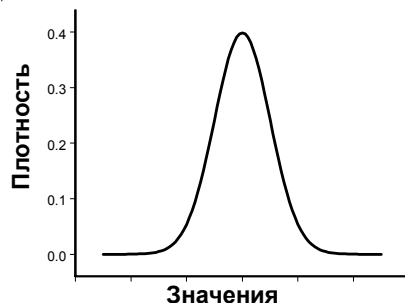
Также вы можете работать с поверхностью ковариации и воспользоваться возможностью поиска по направлениям. Значения на облаке взаимной ковариации сгруппированы в бины, основанные на сходном направлении и одинаковом расстоянии между парами опорных точек. Затем эти значения бинов усредняются и сглаживаются для построения поверхности взаимной ковариации. Справа дана цветовая шкала, на которой приведены значения, соответствующие цветовой границе перехода между биномы. Область отображения поверхности взаимной ковариации контролируется размером лага и количеством лагов. (См. Главу 3, 'Принципы геостатистического анализа', для дополнительной информации о поверхности вариограммы, бинах и лагах.)

Вы можете просмотреть значения поднаборов данных на облаке взаимной ковариации, отметив галочкой опцию Показывать направление поиска, а затем, щелкнув по указателю направления, менять его размер или ориентацию. Вы можете также щелкнуть по серой стрелке в верхней части панели, чтобы временно скрыть эту часть инструмента.



# Изучение распределения данных

Определенные методы кригинга работают лучше, если данные подчиняются закону нормального распределения, функция плотности которого выглядит следующим образом:

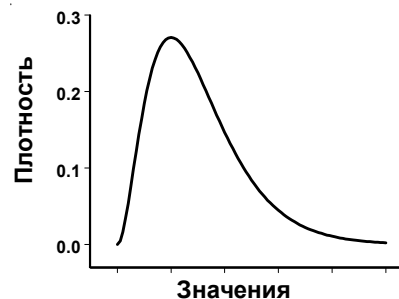


В частности, карты квантилей и вероятности, построенные с использованием методов ординарного, простого и универсального кригинга, используют данные, которые подчиняются закону нормального распределения.

Как уже упоминалось в Главе 3, метод кригинга основывается также на предположении о стационарности. Это предположение, в частности, требует соблюдения условия того, что все распределения значений данных имеют одинаковую изменчивость. Часто в природе мы наблюдаем, что с увеличением значений, увеличивается также их изменчивость. Чтобы привести данные к нормальному распределению и выполнить условие их равной изменчивости, вы можете выполнить различные преобразования наборов данных.

Инструменты гистограммы и нормального графика КК позволяют использовать несколько методов преобразований, включая метод Вох—Сох (‘рубка рулевого’ - также известно как степенное преобразование), логарифмический и арксинуса (для дополнительной информации см. Главу 7, ‘Использование аналитических инструментов при построении поверхностей’). Преобразование по методу Вох—Сох определяется выражением  $Y(s) = (Z(s)^l - 1) / (l + 1)$  для  $l \neq 0$ . Этот метод преобразования следует использовать в том случае, если ваши данные состоят из подсчетов встречаемости какого-либо

явления. Для таких типов данных дисперсия часто определяется по среднему значению. Это означает, что если на части изучаемой территории встречаемость маленькая, изменчивость в этом локальном районе будет меньше, чем изменчивость в другом районе, где встречаемость больше. В этом случае, известно, что если вы сначала извлечете корень квадратный из всех значений ваших данных, это поможет сделать дисперсии более постоянными для всей изучаемой территории, и также часто позволяет привести данные к нормальному распределению. Преобразование с использованием квадратного корня применимо, когда  $l = 0.5$ . Логарифмическое преобразование обычно рассматривается, как частный случай преобразований по методу Вох—Сох, когда  $l = 0$ ,  $Y(s) = \ln(Z(s))$  для  $Z(s) > 0$ , и ‘ln’ - натуральный логарифм. Логарифмическое преобразование часто используется в тех случаях, когда данные имеют распределение с положительной асимметрией, и в данных есть несколько очень больших значений. Вы можете локализовать эти большие значения на изучаемой территории, а логарифмическое преобразование поможет сделать дисперсии более постоянными и нормализовать ваши данные. Распределение с положительной асимметрией выглядит следующим образом:



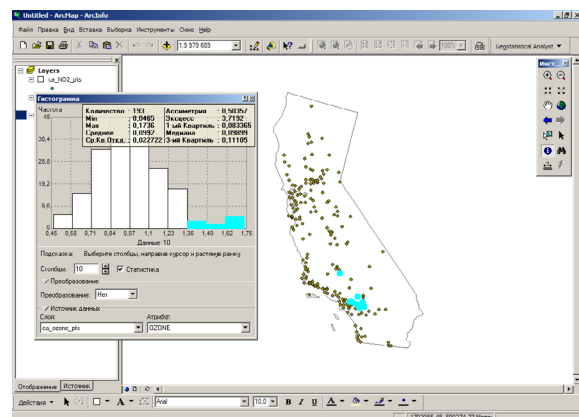
Преобразование по методу арксинуса имеет выражение  $Y(s) = \sin^{-1}(Z(s))$ , если значения  $Z(s)$  находятся в интервале от 0 до 1. Преобразование по методу арксинуса может быть использовано для данных, выраженных в относительных единицах (или в процентах). Часто, когда данные выражены пропорцией, дисперсия будет наи-

меньшей для значений, близких к 0 и 1, и наибольшей для значений, близких к 0.5. Преобразование по методу арксинуса поможет сделать дисперсии более постоянными для всей изучаемой территории и часто приводит к нормализации распределения данных.

Чтобы понять, какие преобразования данных необходимо выполнить для того, чтобы привести данные к нормальному распределению, вы можете воспользоваться инструментами гистограммы и нормального графика КК. Это же преобразование (по методу арксинуса), скорее всего, также приведет к выравниванию дисперсий, что поможет выполнить условие стационарности данных.

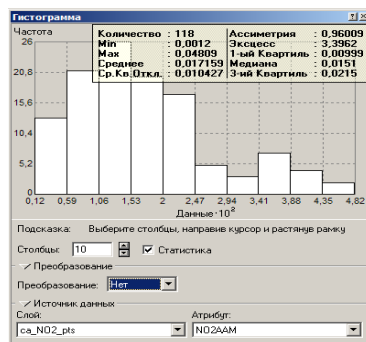
## Использование инструмента гистограммы для изучения распределений

С помощью инструмента гистограммы вы можете изучить форму распределения визуально. Просмотрев статистику, в частности значения среднего и медианы, вы можете определить центр распределения. Обратите внимание на рисунке внизу, что форма гистограммы похожа на колокол, и поскольку значения среднего и медианы очень похожи, это распределение близко к нормальному. Вы можете также выделить цветом экстремальные значения в хвосте гистограммы и посмотреть, как они пространственно распределены в ландшафте. На

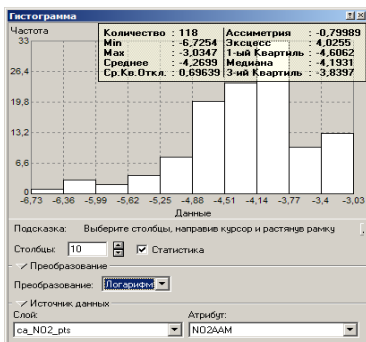


нижнем рисунке, на котором показаны данные по концентрации озона, наиболее высокие значения расположены, как и ожидалось в городских районах.

Если ваши данные имеют сильную асимметрию, вы можете проверить результат применения преобразований к вашим данным. На верхней диаграмме показано асимметричное распределение до применения преобразований.



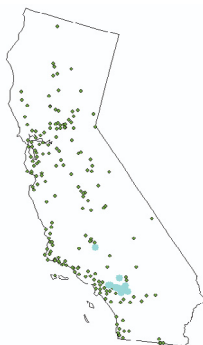
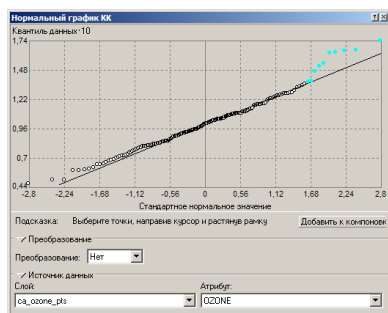
К асимметричным данным применено логарифмическое преобразование, которое привело данные к распределению, близкому к нормальному.



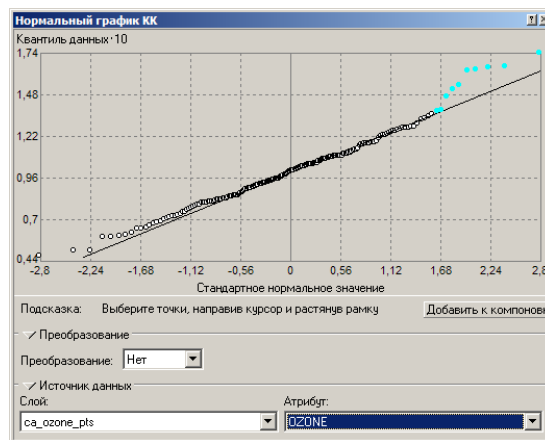
## Изучение распределений с помощью графиков Квантиль-квантиль

Для двух идентичных распределений общий график КК (квантиль-квантиль) будет прямой линией. Таким образом, сравнение этой линии с точками на нормальном графике КК дает представление об одномерной нормальности. Если данные асимметричны (т.е. далеки от нормального распределения), точки будут отклоняться от прямой линии.

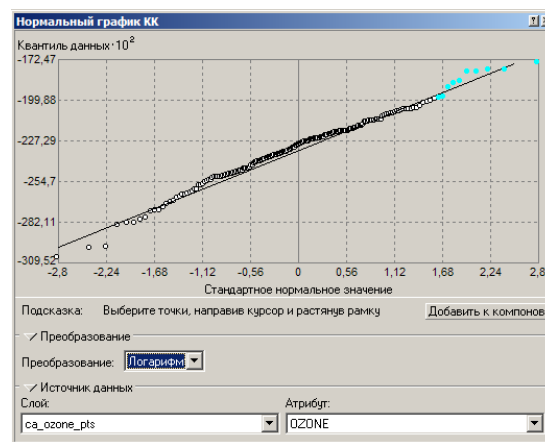
На нижней диаграмме квантили стандартного нормального распределения нанесены на нормальный график КК по оси x, а квантили набора данных - по оси y. Вы можете видеть, что график близок к прямой линии. Главное отклонение от этой линии происходит в области высоких значений концентрации озона. Инструмент нормального графика КК позволяет вам выбрать точки, которые не попадают на прямую линию. Выбранные точки выделяются цветом в виде данных ArcMap. (Здесь видно, что они сконцентрированы на небольшом участке рядом с Лос-Анджелесом.)



Тот же набор данных, что мы трансформировали в предыдущем примере с гистограммой, также был преобразован с помощью инструмента нормального графика КК (рисунки справа). Обратите внимание, что на первом графике точки сильно отклонялись от прямой линии.



Однако, если вы посмотрите на следующий рисунок, вы увидите, что после логарифмического преобразования набора данных, точки лежат ближе к прямой линии.



## Изучение распределения ВАШИХ ДАННЫХ

Инструменты ESDA помогают вам в изучении распределения ваших данных.

Вы хотите знать, подчиняются ли ваши данные нормальному распределению (кривая в форме колокола). Среднее и медиана при нормальном распределении будут иметь близкие значения, асимметрия будет близка к нулю, а эксцесс должен иметь значение, примерно равное 3.

Если данные имеют значительную асимметрию, вы можете попробовать преобразовать их с тем, чтобы привести к нормальному распределению. При построении поверхности преобразование данных надо выполнять с осторожностью, поскольку полученные интерполированные значения будут подвергнуты обратному преобразованию, и это обратное преобразование даст приблизительно несмещенные значения интерполяции с примерной стандартной ошибкой кригинга.

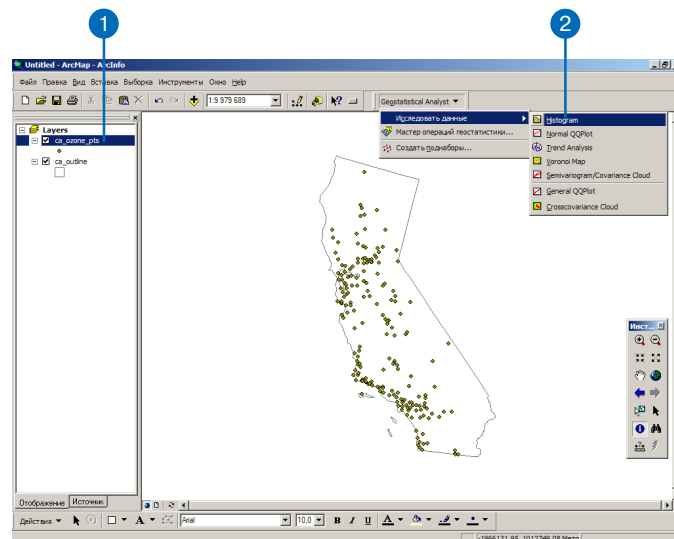
### Подсказка

#### График КК

Если данные подчиняются нормальному распределению, точки будут расположены близко к прямой линии.

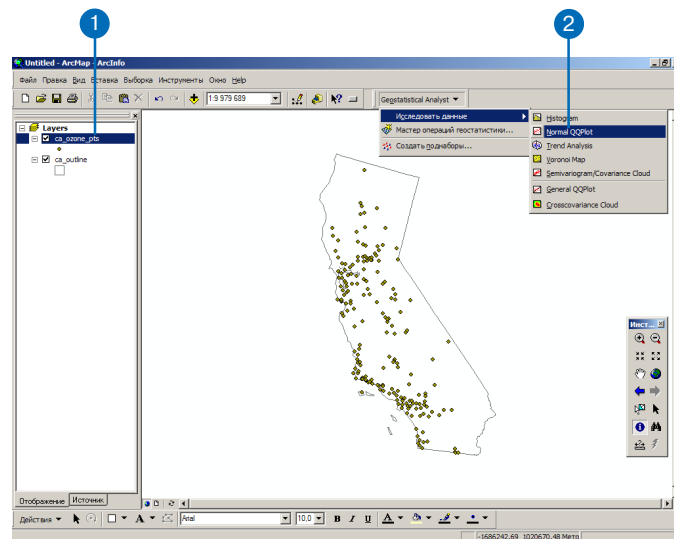
## Изучение распределения с помощью инструмента Гистограмма

1. В таблице содержания ArcMap выберите точечный слой, данные которого вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Гистограмма.



## Изучение распределения с помощью графика КК

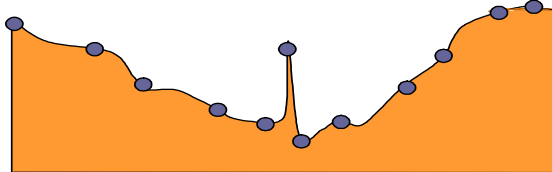
1. В таблице содержания ArcMap выберите точечный слой, данные которого вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Нормальный график КК.



## Поиск глобальных и локальных выпадающих значений в наборе данных

Глобальное выпадающее значение - это очень высокое или очень низкое (относительно других точек в наборе данных) измеренное значение в опорной точке. Например, если 99 из 100 точек имеют значение в интервале между 300 и 400, но сотая точка имеет значение, равное 750, эта сотая точка может рассматриваться как глобальное выпадающее значение.

Локальное выпадающее значение - это измеренное значение в опорной точке, которое попадает в обычный интервал значений всего набора данных, но если вы посмотрите на точки, расположенные в окрестности этой точки, то по сравнению с ними, это значение необычно высокое или низкое. Например, на диаграмме внизу показан разрез долины в ландшафте. Видно, что в центре долины есть одна точка, высота которой слишком велика относительно соседних точек, но при этом значение не выпадает из набора данных в целом.



Выпадающие значения важно определять по двум причинам: они могут либо указывать на ненормальность явления, либо на то, что была допущена ошибка при измерении или записи значения этой точки.

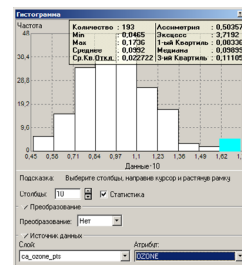
Если выпадающее значение соответствует действительной ненормальности явления, это может стать самой существенной точкой исследования и важным фактором в понимании изучаемого явления. Например, опорная точка, расположенная на рудной жиле какого-либо минерала, может иметь резко выделяющееся значение, и именно это место будет представлять наибольший интерес для горнодобывающей компании.

Если выпадающие значения вызваны ошибками при вводе данных, и эти значения опорных точек точно являются неправильными,

они должны быть либо исправлены, либо удалены из набора данных до построения поверхности. Выпадающие значения могут исказить поверхность проинтерполированных значений, а также оказать отрицательное влияние на моделирование вариограммы и на использование области поиска соседства.

### Поиск выпадающих значений по гистограмме

Инструмент гистограммы позволяет вам выбирать точки на хвостах распределения. Выбранные точки отображаются в виде данных ArcMap. Если экстремальные значения соответствуют отдельным опорным точкам (т.е., окруженным точками с очень разными значениями), может потребоваться дополнительное изучение данных, и, при необходимости, удаление точки с экстремальным значением.



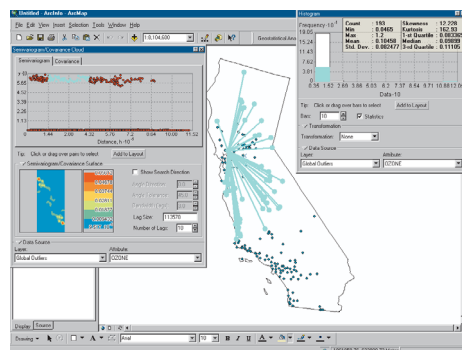
В приведенном выше примере, высокие значения концентрации озона не являются экстремальными или выпадающими, и, следовательно, нет необходимости удалять их из набора данных.

### Определение выпадающих значений по облаку вариограммы/ковариации

Если в вашем наборе данных присутствует глобальное выпадающее, нехарактерно высокое для набора данных значение, все пары точек, образующиеся с участием этого экстремального значения, будут иметь высокие значения и на облаке вариограммы, независимо от их удаленности друг от друга. Это можно увидеть по облаку вариограммы, приведенному на рисунке на следующей странице. Обратите внима-

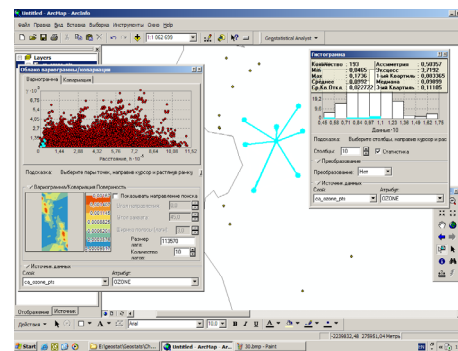


ние на два основных слоя вариограммы. Если вы выделите точки в верхнем слое, как показано на рисунке, в виде данных ArcMap вы увидите, что все эти высокие значения соответствуют парам с одной точкой - точкой, имеющей экстремально высокое значение. Таким образом, верхний слой точек был образован всеми парами опорных точек с единственной точкой, имеющей экстремально высокое значение, а нижний слой соответствует парам, которые образуют остальные опорные точки. Когда вы смотрите на гистограмму, которая также приведена на рисунке внизу, вы видите одно высокое значение в правом хвосте гистограммы, которое также указывает на глобальное выпадающее значение. Возможно, была допущена ошибка при вводе данных, и в таком случае, следует исправить значение или удалить точку из набора данных.



Когда в данных существует локальное выпадающее значение, значение этой точки не будет выходить за пределы значений всего распределения, но будет иметь нехарактерную величину по сравнению с соседними точками. На рисунке справа вы видите группу пар точек, расположенных близко к друг другу, но имеющих высокое значение вариограммы (они расположены в начале оси x, что означает, что они находятся на местности рядом, и высоко по оси y, что означает, что они имеют высокие значения вариограммы). Если вы выберете эти точки, вы увидите, что все они образуют пару с одной опорной точкой. Если вы посмотрите на гистограмму, вы увидите, что данные

подчиняются нормальному закону распределения, и на гистограмме нет необычной точки. Обсуждаемая точка выделена цветом в низком (правом) хвосте гистограммы и образует пары с более высокими соседними значениями (см. выделенные на гистограмме значения). Эта точка может быть локальным выпадающим значением. Необходимо дальнейшее исследование данных.



## Поиск выпадающих значений по карте Вороного

Для определения возможных экстремальных значений данных могут быть применены карты Вороного, построенные с использованием методов энтропии и кластерного анализа.

Значения энтропии показывают степень различия между соседними ячейками. В природе следует ожидать, что объекты, расположенные ближе друг к другу будут более похожи, чем удаленные друг от друга. Следовательно, локальные выпадающие значения могут быть определены с помощью областей с высокими значениями энтропии.

Метод кластерного анализа определяет ячейки, которые отличаются от своих соседей. Следует предположить, что значение, относящееся к ячейке, будет походить хотя бы на одно значение соседней ячейки. Таким образом, и этот инструмент может быть использован для определения локальных выпадающих значений.

## Определение глобальных и локальных выпадающих значений

Для определения глобального выпадающего значения, обратите внимание на необычно высокие или низкие значения на гистограмме и две ясно различимых горизонтальных группы точек на облаке вариограммы. Локальные выпадающие значения можно определить по высоким значениям вариограммы, относящимся к одной точке в начале оси x на облаке вариограммы.

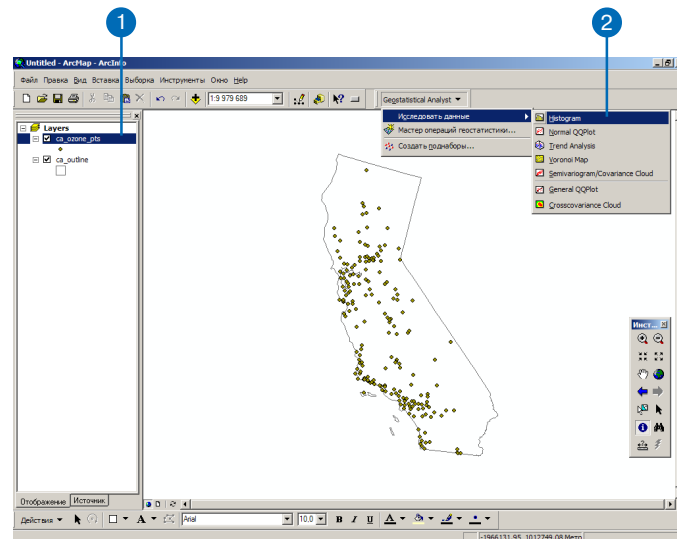
И глобальные, и локальные выпадающие значения могут оказывать отрицательное воздействие на поверхность интерполированных значений, путем влияния на модель вариограммы и на искомые значения.

### См. также

*См. Главу 3, 'Принципы геостатистического анализа', для дополнительной информации о выпадающих (экстремальных) значениях.*

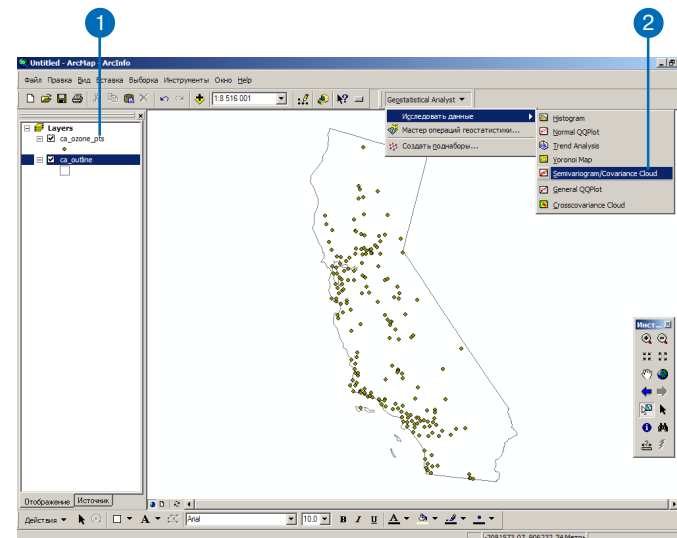
## Определение глобальных выпадающих значений с использованием инструмента Гистограмма

1. В таблице содержания ArcMap выберите слой точечных или полигональных объектов, которые вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Гистограмма.



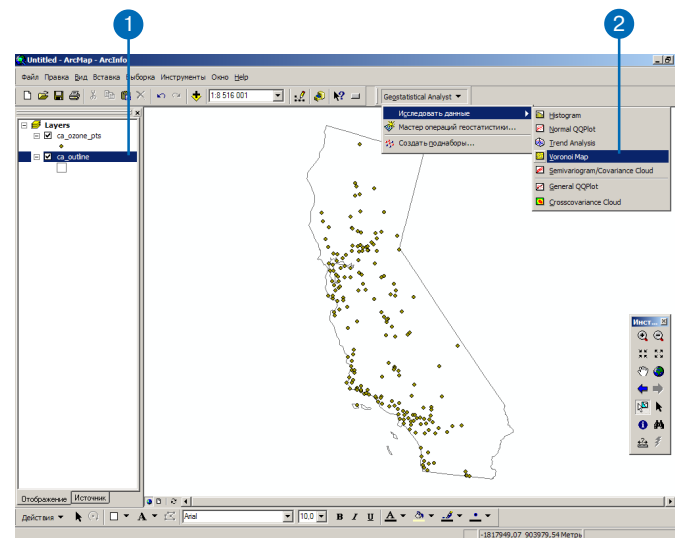
## Определение глобальных выпадающих значений по облаку вариограммы/ковариации

1. В таблице содержания ArcMap выберите слой точечных или полигональных объектов, которые вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Облако вариограммы/ковариации.



## Обнаружение локальных выпадающих значений с использованием карты Вороного

1. В таблице содержания ArcMap выберите слой точечных или полигональных объектов, которые вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Карта Вороного.



## Определение глобальных трендов

Поверхность может состоять из двух главных компонентов: фиксированного глобального тренда и случайной вариации на микроуровне. Глобальный тренд часто рассматривается как структура с фиксированным средним. Случайная вариация на микроуровне (иногда называемая случайной ошибкой) может быть смоделирована для двух составляющих: пространственной автокорреляции и эффекта самородка.

Если вы пришли к выводу, что в ваших данных присутствует глобальный тренд, то вы должны решить, как описать его с помощью модели. Используйте ли вы для построения поверхности детерминистский метод или геостатистический, как правило, зависит от целей ваших исследований. Если вы хотите моделировать только глобальный тренд и построить сглаженную поверхность, для создания результирующей поверхности вы можете воспользоваться интерполяцией по методу глобального полинома или локальных полиномов (см. Главу 3, ‘Принципы геостатистического анализа’, и Главу 5, ‘Детерминистские методы интерполяции пространственных данных’). Однако, вы можете учесть тренд при использовании геостатистического метода, вычтете его, а затем моделировать оставшуюся часть как случайную вариацию на микроуровне. Главная причина для вычитания тренда в геостатистике - удовлетворение предположения о стационарности (см. Главу 3, ‘Принципы геостатистического анализа’).

Если вы вычтете тренд при использовании геостатистического метода, вы будете моделировать случайную вариацию на микроуровне для остатков значений. Для того, чтобы вы получили обоснованные значения интерполяции, тренд будет снова автоматически добавлен в ваши данные.

Если вы раскладываете данные на тренд и вариацию на микроуровне, вы предполагаете, что тренд является фиксированным, а вариация на микроуровне - случайной. В данном случае понятие “случайный” не означает “непредсказуемый”; скорее, такой случайный процесс регулируется правилами теории вероят-

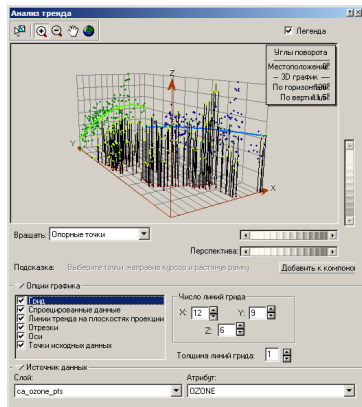
ности, которые включают правила зависимости от соседних значений и носят название автокорреляции. Результирующая поверхность - это сумма фиксированной и случайной поверхностей. То есть, следует подумать об использовании двух слоев: того, который никогда не меняется, в то время как другой меняется случайным образом. Например, предположим, что вы изучаете биомассу. Если вам необходимо вернуться в прошлое, скажем на 1000 лет назад, и устремиться в будущее, в наши дни, та часть поверхности величины биомассы, которая соответствует глобальному тренду, останется неизменной. Однако, та часть поверхности величины биомассы, которая соответствует вариации на микроуровне, изменится. Неизменяемый глобальный тренд может быть отнесен за счет фиксированных эффектов, таких как топография. Вариация на микроуровне может быть вызвана менее постоянными явлениями, которые невозможно проследить в столь длительный период времени, такими как количество осадков, поэтому предполагается, что они являются случайными и коррелирующими.

Если вы сможете определить тренд и оценить его количественно, вы получите более глубокое представление о своих данных, и, таким образом, сможете принять более правильные решения по работе с ними. Если вы вычтете тренд, вы сможете более аккуратно смоделировать случайную вариацию на микроуровне, поскольку глобальный тренд не будет оказывать влияние на пространственный анализ ваших данных.

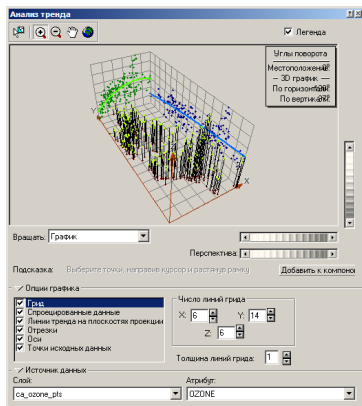
### Изучение глобального тренда с помощью инструмента анализа тренда

Инструмент анализа тренда поднимает точки над изображением интересующего вас участка до высоты значений изучаемого атрибута на трехмерной диаграмме исследуемой территории. Точки затем проецируются в двух направлениях (по умолчанию, север и восток) на две плоскости, перпендикулярные плоскости карты. Кривая полинома подбирается для каждого спроецированного изображения. Поверхность всей карты может

вращаться в любом направлении, что также приводит к изменению направлений, соответствующих плоскостям проекции. Если кривая, проходящая через спроецированные точки, близка к прямой линии (как голубая линия на правой плоскости проекции на рисунке на следующей странице), в данных тренд отсутствует.

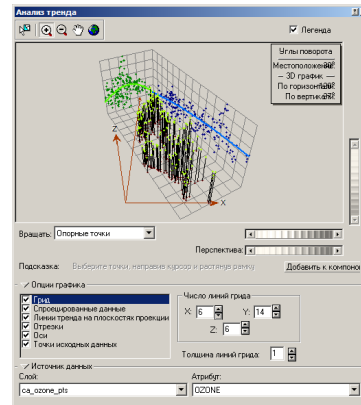


Если кривая может быть описана при помощи полинома, как, например, выгнутая вверх кривая на левой задней плоскости проекции, показанная на диаграмме зеленым цветом, можно



предположить, что в ваших данных существует глобальный тренд.

Если вы повернете опорные точки на 30 градусов, как показано на диаграмме внизу, тренд становится более явным и выглядит, как сильно выгнутая U-образная кривая. Для описания данных может быть подобран полином второго порядка (см. Главу 3, 'Принципы геостатистического анализа', и Главу 5, 'Детерминистские методы интерполяции пространственных данных'). С помощью точной настройки, возможной при использовании инструмента анализа тренда, может быть определено истинное направление тренда. В нашем примере, наиболее сильно тренд проявляется в направлении с юго-востока на северо-запад.



## Определение глобальных трендов

Для определения глобального тренда в ваших данных, необходимо выявить линию, отличающуюся от прямой, на плоскости проекции в инструменте анализа тренда.

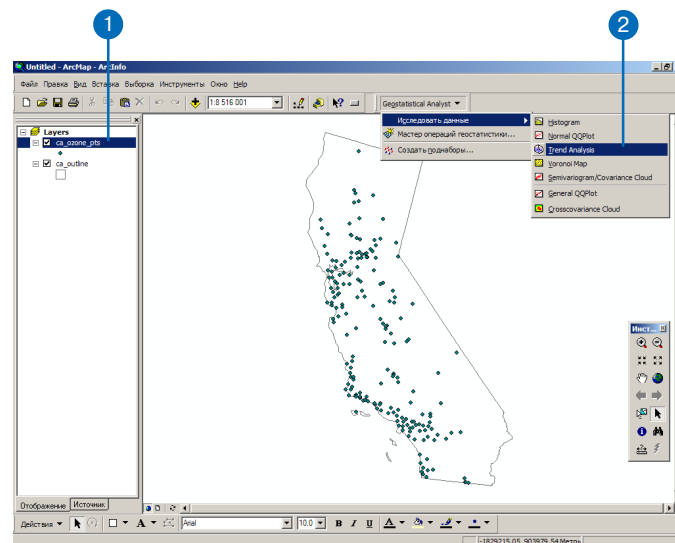
Если в ваших данных присутствует глобальный тренд, вы можете построить поверхность с использованием одного из детерминистских методов интерполяции (например, глобального или локального полинома), либо вычестть тренд из ваших данных перед моделированием вариограммы/ковариации для кригинга.

### См. также

См. Главу 5, 'Детерминистские методы интерполяции пространственных данных', для дополнительной информации о детерминистских методах интерполяции и Главу 7, 'Использование аналитических инструментов при построении поверхностей', для дополнительной информации о трендах.

## Определение глобальных трендов с помощью инструмента Анализ тренда

1. В таблице содержания ArcMap выберите слой точечных или полигональных объектов, которые вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Анализ Тренда.



# Изучение пространственной автокорреляции и вариации по направлениям

Исследуя данные, вы получаете представление о пространственной корреляции между измеренными значениями. Такое понимание может быть использовано для принятия более обоснованных решений при выборе моделей при выполнении пространственной интерполяции.

## Пространственная автокорреляция

Вы можете исследовать пространственную корреляцию, которая существует в ваших данных, путем изучения различных пар опорных точек. Облако вариограммы образуется при нанесении на график расстояний между двумя точками и величины половины квадрата разности между значениями в этих точках. Расстояние между точками откладывается по оси  $x$ , а половина квадрата разности значений точек - по оси  $y$ . Каждая точка на вариограмме соответствует паре опорных точек, а не индивидуальной точке на карте.

Если данные являются пространственно зависимыми, пары точек, расположенных близко друг от друга (в начале оси  $x$ ) должны иметь меньшие значения квадрата разности (располагаться в нижней части оси  $y$ ). По мере того, как точки удаляются друг от друга (перемещение вправо по оси  $x$ ), в целом, величина квадрата разности должна расти (перемещение вверх по оси  $y$ ). Часто существует некое расстояние, за пределами которого квадрат разности остается примерно постоянным. Считается, что точки, удаленные друг от друга на расстояние, большее этого некоего расстояния, не коррелируют друг с другом.

Основным допущением для геостатистических методов является предположение, что любые две пары точек, расположенных на одинаковом расстоянии и в одинаковом направлении друг от друга, должны иметь похожие значения квадрата разности. Такая взаимосвязь носит название стационарности (см. Главу 3, 'Принципы геостатистического анализа', и Главу 7, 'Использование аналитических инструментов при построении поверхностей').

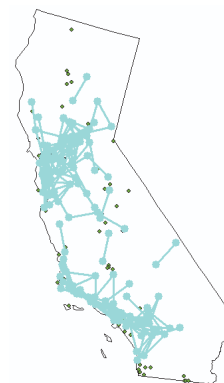
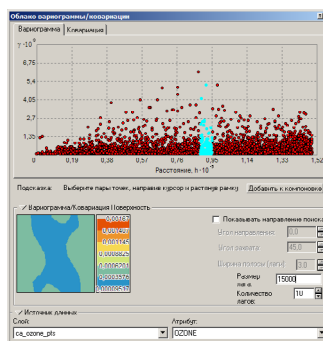
Пространственная корреляция может зависеть только от расстояния между двумя точками, что носит название изотропии. Одна-

ко, возможно, что то же самое значение корреляции могут иметь точки, удаленные друг от друга на другое расстояние или точки, расположенные по-другому относительно друг друга в пространстве. Однако в действительности, может случиться так, что объекты, удаленные друг от друга на значительное расстояние в одном направлении, будут больше похожи между собой, чем объекты, расположенные ближе друг к другу в другом направлении. Такое влияние по направлениям видно на вариограммах и диаграммах ковариации и носит название анизотропии.

Важно исследовать анизотропию, поскольку если вы обнаружите присутствие влияния по направлениям в корреляции, вы можете учесть их в моделях вариограммы и ковариации. Это, в свою очередь, влияет на геостатистический метод интерполяции.

## Исследование пространственной структуры с помощью инструмента Облако Вариограммы/ковариации

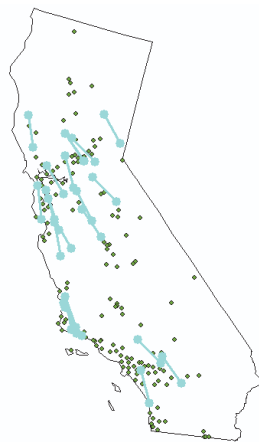
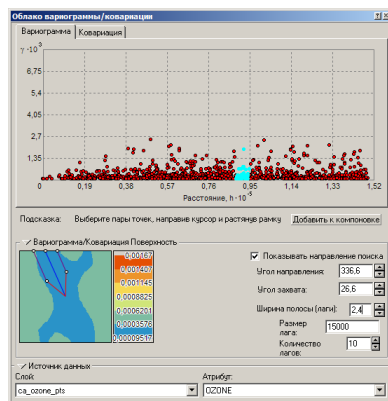
Инструмент Облако вариограммы/ковариации может быть использован для изучения корреляции ваших данных. Рассмотрим набор данных со значениями концентрации озона. На следующем рисунке обратите внимание, что вы можете выбрать все пары точек, которые находятся на определенном расстоянии друг от друга, выделив их на вариограмме.



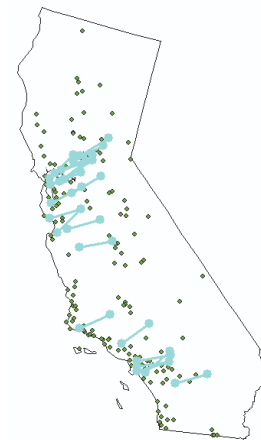
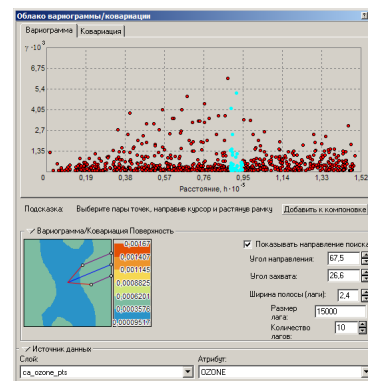


## Определение влияния по направлениям с помощью инструмента облака Вариограммы/Ковариации

В предыдущем примере вы использовали инструмент Облако вариограммы/ковариации для определения общей корреляции в данных. Однако, взглянув на поверхность вариограммы, можно предположить, что в значениях вариограммы могут быть различия в зависимости от направления. Если вы выберете опцию Показывать направление поиска и зададите угол и ширину полосы, как показано на следующем рисунке, вы обнаружите, что связанные точки имеют очень похожие значения, поскольку значения вариограммы являются относительно низкими.



Теперь, если вы измените направление связей, как показано на следующем рисунке, вы можете увидеть, что некоторые связанные точки имеют значения, которые сильно отличаются друг от друга, что приводит к высоким значениям вариограммы. Это означает, что точки, отстоящие друг от друга на расстояние около  $0.9 \times 10^5$  метров в направлении под углом 70/250 градусов в среднем больше различаются между собой, чем точки в направлении под углом 160/340 градусов. Напомним, что свойство, при котором вариация меняется в одном направлении быстрее, чем в другом, носит название анизотропии. При интерполяции поверхности с использованием Мастера операций геостатистики модуля Geostatistical Analyst, для учета анизотропии вы можете использовать модели вариограммы.



## Изучение пространственной структуры и вариации по направлениям

Изучение пространственной структуры позволяет выявить пространственную автокорреляцию опорных точек и определить наличие влияний по направлениям.

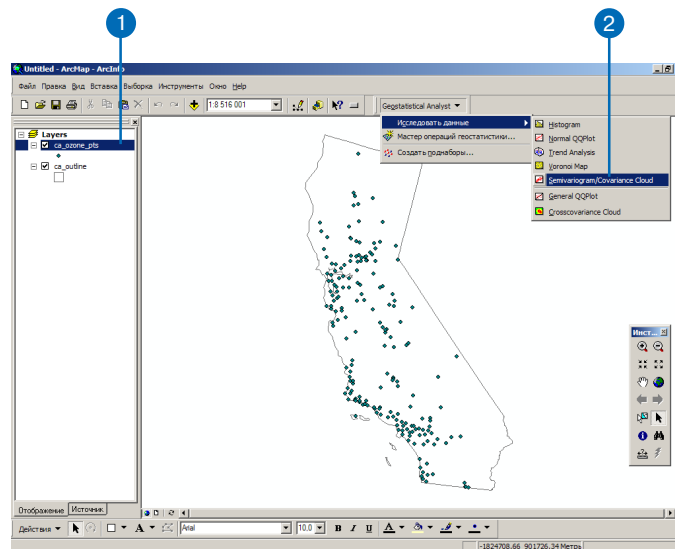
Выберите пары точек, расположенных близко друг к другу (слева по оси x на вариограмме), которые должны быть в наибольшей степени схожи (значения которых находятся в нижней части оси y). При удалении точек друг от друга (движение вправо по оси x), вариация должна возрастать (значения находятся выше по оси y). Если пары точек на вариограмме образуют прямую горизонтальную линию, в данных может отсутствовать пространственная корреляция, поэтому она может не учитываться при построении поверхности.

### См. также

См. Главу 7, 'Использование аналитических инструментов при построении поверхностей', для дополнительной информации о моделировании вариограммы и трендах по направлениям.

## Изучение пространственной структуры

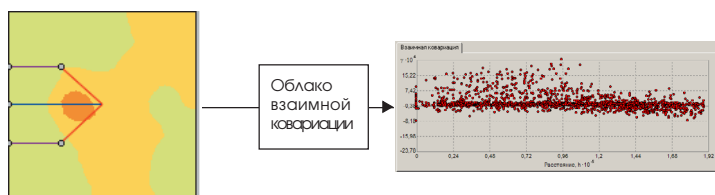
1. В таблице содержания ArcMap выберите слой точечных или полигональных объектов, которые вы хотите изучить.
2. На панели инструментов Geostatistical Analyst выберите опцию Исследовать данные, затем строку Облако данных, затем строку Облако вариограммы/ковариации.



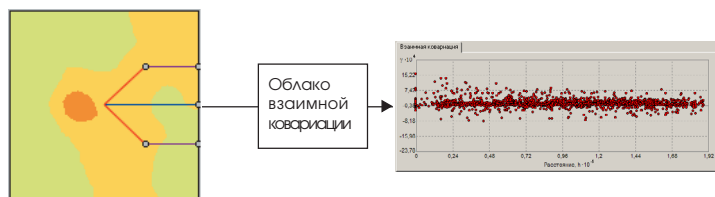
# Изучение ковариации между несколькими наборами данных

## Изучение ковариации между несколькими наборами данных

Инструмент Облако взаимной ковариации может быть использован для определения взаимной корреляции между двумя наборами данных. Рассмотрим значения концентрации озона (первый набор данных - Ozone) и двуокиси азота  $\text{NO}_2$  (второй набор данных  $\text{NO}_2$ ). Обратите внимание, что взаимная корреляция между значениями концентрации  $\text{NO}_2$  и озона выглядит асимметричной. Темно-красный участок показывает, что самые высокие значения концентрации характерны для тех районов, где значения  $\text{NO}_2$  получены со смещением к западу от значений концентрации озона. Инструмент поиска по направлению поможет определить причину этого явления. Когда он ориентирован на запад, мы получаем следующую картину:



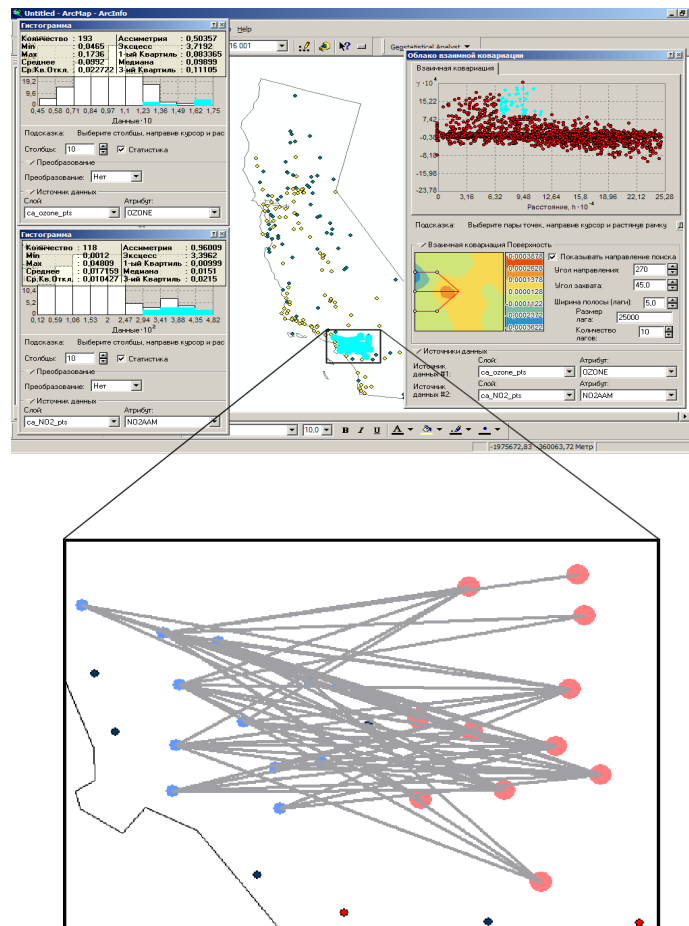
а когда он направлен на восток, следующую:



Очевидно, что значения ковариации выше, когда инструмент направления поиска ориентирован на запад. Теперь вы можете изучить, какие пары точек дают высокие значения взаимной ковариации. Для этого воспользуемся инструментами Облако взаимной ковариации и Гистограмма (рисунок на стр. 110).

Если инструмент направления поиска указывает на запад, и вы выберете на графике несколько точек с высокими значениями взаимной ковариации, вы увидите, что большинство пар данных расположены в районе Лос-Анджелеса. Вы также можете заметить, что значения концентрации  $\text{NO}_2$  смещены к западу от значений концентраций озона. По гистограммам можно определить, что высокие значения ковариации для всех пар данных характерны для тех точек, где и значения концентрации  $\text{NO}_2$  (голубые столбики на гистограмме  $\text{NO}_2$ ), и значения концентрации озона (оранжевые столбики на гистограмме Ozone) выше своих соответствующих средних. Итак, вы определили, что асимметрия взаимной ковариации является результатом смещения высоких значений концентрации  $\text{NO}_2$  к западу от высоких значений концентрации озона в районе Лос-Анджелеса. Обратите внимание, что вы можете также получить высокие значения взаимной ковариации, когда выбранные пары из обоих наборов данных имеют значения ниже соответствующих средних для каждого из набора данных. В действительности, в целом следует ожидать, что высокие значения взаимной ковариации будут характерны для пар точек, имеющих значения как выше, так и ниже их соответствующих средних, и такие пары могут встречаться на различных участках изучаемой территории. Исследовав данные, вы определили, что значения взаимной ковариации в районе Лос-Анджелеса отличаются от значений на остальной территории штата (см. рисунок на стр. 110). Воспользовавшись этой информацией, вы можете предположить, что причиной результатов, полученных с помощью инструмента исследования взаимной ковариации, может быть непостоянное среднее в данных, и попробовать вычестить тренды из данных со значениями концентрации  $\text{NO}_2$  и озона, либо разделить изучаемую территорию на районы и выполнить моделирование с помощью кригинга и кокринга в пределах этих районов.

## Изучение корреляции между двумя наборами данных



Образование пар между точками с высокой взаимной корреляцией между значениями концентрации озона и двуокиси азота для района Лос-Анджелеса

## Изучение пространственной ковариации между несколькими наборами данных

Инструмент позволяет изучить взаимную ковариацию двух наборов данных.

Проверьте, является ли поверхность ковариации симметричной, и воспользуйтесь инструментом направления поиска, чтобы определить, аналогично ли облако взаимной ковариации выглядит для всех направлений.

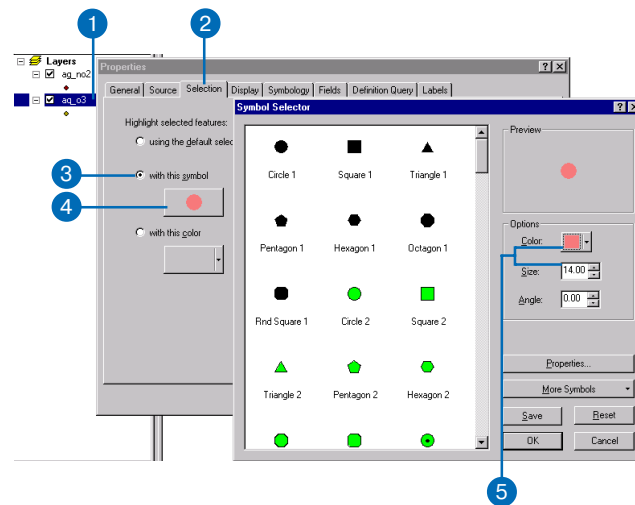
Если вы обнаружите, что для значений двух наборов данных характерно пространственное смещение, или необычно высокие значения взаимной ковариации, вы можете исследовать, где это происходит. Если вы заметите, что необычные значения взаимной ковариации характерны для отдельных точек или ограниченных участков изучаемой территории, возможно, вы захотите что-либо предпринять, например, обратить внимание на определенные значения в данных, вычесть тренд из ваших данных или разделить данные на участки.

## Изучение пространственной ковариации с помощью инструмента Облако взаимной ковариации

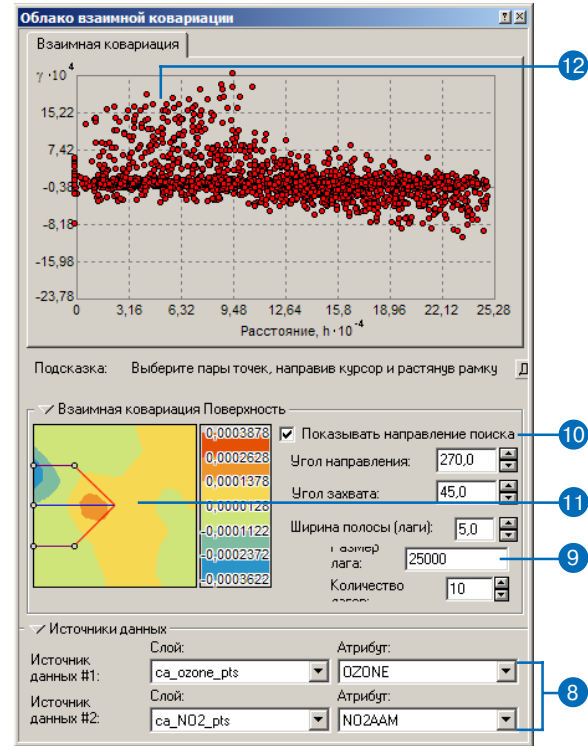
1. В таблице содержания ArcMap нажмите правой клавишей мыши на слое точечных пространственных объектов, который будет первым слоем, участвующим в анализе взаимной ковариации, и выберите Свойства.
2. Откройте закладку Выборка.
3. Отметьте опцию выделения объектов с помощью данного символа.
4. Щелкните по кнопке с изображением символа.
5. Выберите цвет и размер символа, который вы будете использовать для выделенных объектов.

Повторите шаги 1–5 для второго слоя, который будет использоваться в анализе взаимной ковариации, но выберите символ другого цвета и размера.

6. Выделите эти два слоя в таблице содержания ArcMap, удерживая клавишу Ctrl.
7. На панели Geostatistical Analyst выберите опцию Исследовать данные, а затем строку Облако взаимной ковариации. ►



8. Выберите соответствующие атрибуты для каждого слоя из списка Атрибут.
9. Введите размер лага и количество лагов.
10. Отметьте опцию Показывать направление поиска.
11. Удерживайте курсор мыши на центральной синей линии на поверхности ковариации и поворачивайте стрелку направления поиска до тех пор, пока она не будет соответствовать углу, при котором вы предполагаете наличие смещения; в нашем примере это угол 270 градусов (он дан в окне "угол направления").
12. Выберите несколько точек на графике ковариации, удерживая левую клавишу мыши и растягивая инструмент выбора так, чтобы захватить несколько точек. Просмотрите, какие точки выбраны на карте в ArcMap.





# Детерминистские методы интерполяции пространственных данных

## 5

### В ЭТОЙ ГЛАВЕ

- Как работает интерполяция по методу взвешенных расстояний
- Построение поверхности с использованием интерполяции по методу взвешенных расстояний (IDW)
- Как работает интерполяция по методу глобального полинома
- Создание карты с использованием метода глобального полинома
- Как работает интерполяция по методу локальных полиномов
- Создание карты с использованием метода локальных полиномов
- Как работает интерполяция с использованием радиальных базисных функций
- Создание карты с использованием интерполяции на основе радиальных базисных функций

Существует две основные группы методов интерполяции: детерминистские методы и геостатистические. Детерминистские методы интерполяции строят поверхность по опорным точкам, основываясь либо на степени схожести точек выборки (как например, метод взвешенных расстояний), либо на степени сглаживания (как например, радиальные базисные функции). Геостатистические методы интерполяции (например, кригинг) используют статистические свойства опорных точек. Геостатистические методы количественно определяют пространственную корреляцию между опорными точками и учитывают расположение опорных точек в пространстве вокруг искомой точки. Геостатистические методы будут обсуждаться в Главе 6, 'Построение поверхности с использованием методов геостатистики'.

Детерминистские методы интерполяции могут быть разделены на две группы: глобальные и локальные. Глобальные методы вычисляют искомые значения с использованием всего набора данных. Локальные методы используют для вычисления искомого значения только опорные точки, расположенные в окрестностях искомой, и относятся только к небольшим участкам изучаемой территории. Модуль Geostatistical Analyst в качестве глобального интерполятора использует метод глобального полинома, а в качестве локального - методы взвешенных расстояний, локальных полиномов и радиальных базисных функций.

Поверхность, построенная с использованием детерминистских методов, может как проходить, так и не проходить через опорные точки. Метод интерполяции, который дает в опорной точке значение, равное измеренному, носит название жесткого интерполятора. Нежесткий интерполятор в опорной точке дает значение, отличное от измеренного (то есть, аппроксимирует значение в опорной точке). Последнее позволяет избежать острых пиков или впадин на результирующей поверхности. Метод взвешенных расстояний и радиальные базисные функции являются жесткими интерполяторами, в то время как глобальные и локальные полиномы - нежесткими интерполяторами.



# Как работает интерполяция по методу взвешенных расстояний

Интерполяция по методу взвешенных расстояний (IDW) использует предположение, что объекты, расположенные ближе к другу в большей степени похожи, чем удаленные друг от друга. Чтобы найти значение в какой-либо точке, метод IDW использует опорные точки, находящиеся в окрестностях искомой. Эти опорные точки будут оказывать большее влияние на интерполируемое значение, чем те, которые от нее удалены на значительное расстояние. Таким образом, метод IDW предполагает, что каждая опорная точка оказывает локальное влияние, которое уменьшается с расстоянием. Точкам, находящимся в окрестностях искомой, присваиваются весовые значения большие, чем удаленным от нее точкам. Отсюда и пошло название метода: метод (обратных) взвешенных расстояний.

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

$\hat{Z}(s_0)$  - искомое значение для точки  $s_0$ .

$N$  число опорных точек, находящихся в окрестности искомой точки и используемых в вычислениях.

$\lambda_i$  - веса, присвоенные каждой опорной точке, из числа тех, которые будут использованы в вычислениях. Эти веса уменьшаются с расстоянием.

$Z(s_i)$  измеренное значение в точке  $s_i$ .

Формула метода выглядит следующим образом:

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p} \quad \sum_{i=1}^N \lambda_i = 1,$$

где:

Веса определяются по следующей формуле:

С увеличением расстояния вес уменьшается за счет коэффициента  $p$ .

Величина  $d_{i0}$  - это расстояние между искомой точкой,

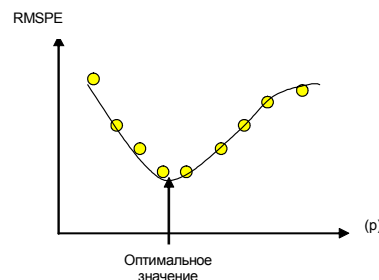
$s_0$ , и  $i$ -той опорной точкой,  $s_i$ .

Параметр степени  $p$  влияет на присвоение весов опорным точкам; это означает, что по мере того, как увеличивается расстояние между опорными точками и искомой точкой, влияние (или вес), которое опорная точка будет оказывать на искомую точку, уменьшается по экспоненте.

Сумма весов опорных точек, которые будут использованы при выполнении интерполяции, должна быть равна 1.

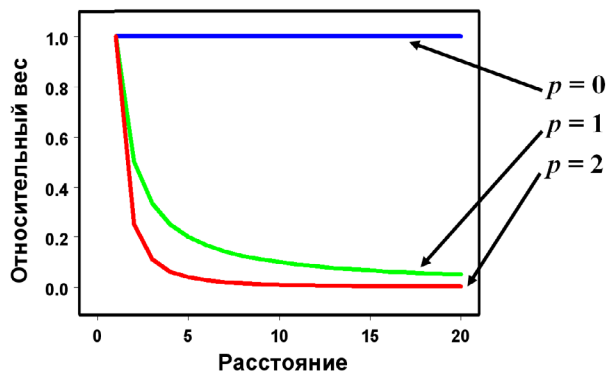
## Степенная функция

Оптимальное значение  $p$  определяется путем минимизации среднеквадратичной ошибки вычислений (RMSPE). Значение среднеквадратичной ошибки является статистической величиной и рассчитывается при перекрестной проверке (см. Главу 7, 'Использование аналитических инструментов при построении поверхностей'). При перекрестной проверке каждая опорная точка исключается из вычислений и сравнивается с проинтерполированным значением для этого местоположения. Среднеквадратичная ошибка RMSPE - это суммарная статистическая величина, количественно определяющая ошибку интерполируемой поверхности. Модуль Geostatistical Analyst подставляет несколько вариантов значения степени в формулу метода взвешенных расстояний (IDW), чтобы определить при каком значении степени среднеквадратичная ошибка минимальна. На графике внизу



показано, как модуль Geostatistical Analyst вычисляет оптимальную степень. Значение RMSPE наносится на график относительно различных степеней, использованных для одного и того же набора данных. Через точки проводится кривая (описываемая локальным полиномом второй степени), и по этой кривой определяется оптимальное значение степени, при котором среднеквадратичная ошибка минимальна.

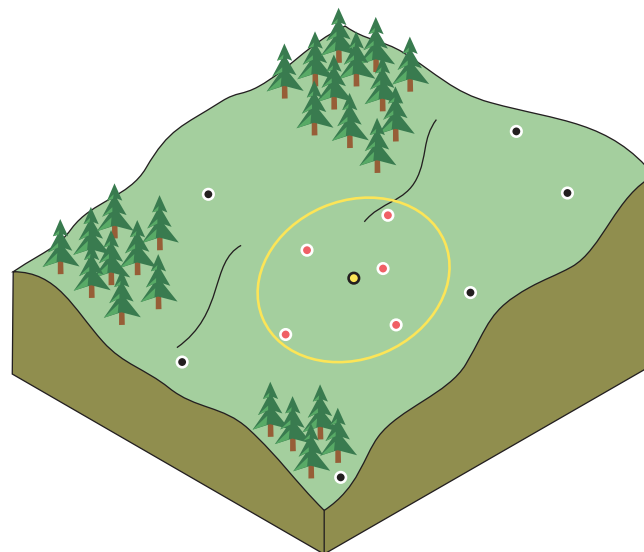
Веса обратно пропорциональны расстоянию, возведенному в степень  $p$ . В результате, по мере увеличения расстояния, веса быстро уменьшаются. Насколько быстро это происходит, зависит от значения  $p$ . Если  $p = 0$ , с увеличением расстояния веса не уменьшаются, и поскольку каждый вес  $I_i$  будет иметь одно и то же значение, искомым результатом будет представлять собой среднее из всех значений опорных точек. По мере возрастания степени  $p$ , веса для удаленных точек быстро уменьшаются, что видно на диаграмме, приведенной внизу. Если значение  $p$  очень велико, только точки, находящиеся в непосредственной близости от искомой, будут влиять на полученное значение.



В модуле Geostatistical Analyst используются функции со степенью выше 1. При значении  $p = 2$  метод носит название интерполяции по методу квадратичных взвешенных расстояний (inverse distance squared weighted).

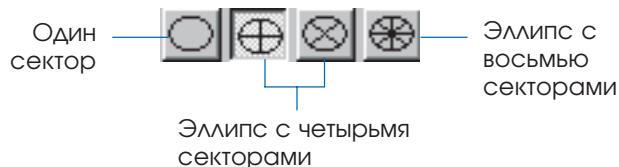
## Поиск соседей

Поскольку объекты, расположенные ближе друг к другу более похожи, чем удаленные друг от друга, при удалении точек значения опорных точек будут иметь все меньшую взаимосвязь со значением искомой точки. Чтобы ускорить процесс вычисления, мы можем принять вес значительно удаленных точек, с небольшим влиянием на искомую точку, за ноль. В результате, обычно ограничивают количество опорных точек, которые будут использованы при расчете искомого значения, путем определения области поиска соседства. Заданная форма, в пределах которой выбираются соседние точки, ограничивает дальность и направление поиска опорных точек, которые будут использованы при выполнении интерполяции. На следующем рисунке показано, какие пять опорных точек (соседей) будут участвовать при вычислении значения искомой точки, обозначенной желтым цветом.

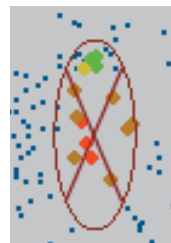


Форма области соседства зависит от исходных данных и поверхности, которую вы хотите построить. Если определение весов не зависит от влияния по направлениям, вы должны учитывать точки равномерно во всех направлениях. Для этого область соседства должна иметь форму круга. Однако, если для ваших данных характерно влияние по направлениям, такое, как, например, преобладающий ветер, вы можете использовать для определения области соседства эллипс, большая ось которого направлена параллельно ветру. Регулирование области поиска соседей для такого направленного влияния обоснованно, поскольку вы знаете, что точки, расположенные в направлении по ветру, даже удаленные друг от друга, будут больше похожи, чем точки, расположенные перпендикулярно к преобладающему направлению ветра.

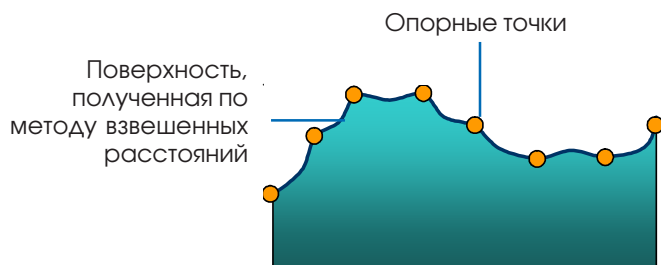
После того, как определена форма области поиска соседей, вы можете также ограничить, какие точки, попадающие в эту область, должны быть использованы. Вы можете определить максимальное и минимальное количество используемых точек, а также вы можете разделить область поиска соседей на сектора. Если вы разделите эту область на сектора, ограничения по максимальному и минимальному количеству точек будут применены к каждому сектору. Существуют различные способы деления на сектора (см. рисунок внизу).



Точки, выделенные в виде данных диалога Поиск соседей, показывают опорные точки с весами, которые будут использованы для поиска значения искомой точки в центре эллипса. Соседи попадают в показанный эллипс. В данном примере, двум точкам (красным) в западном секторе и одной в южном секторе будут присвоены веса более 10 процентов. Вес точки (желтой), расположенной в северном секторе, будет от 3 до 5 процентов.



Поверхность, построенная по методу взвешенных расстояний (IDW), зависит от выбора степени ( $p$ ) и способа поиска соседей. Метод взвешенных расстояний - это жесткий интерполятор, при котором максимальные и минимальные значения на проинтерполированной поверхности (см. рисунок) могут иметь только опорные точки. Результирующая поверхность чувствительна к кластеризации и присутствию в данных экстремальных значений. Метод взвешенных расстояний предполагает, что поверхность была получена с использованием локальной вариации, которая может быть учтена с помощью определения области поиска соседей.



## Построение поверхности с использованием метода взвешенных расстояний (IDW)

Метод IDW предполагает, что поверхность получена на основе локальной вариации. Он лучше работает, если опорные точки равномерно распределены по территории и если они не объединены в кластеры. При выполнении интерполяции по методу IDW важны следующие параметры: условия поиска соседей, степень ' $p$ ', и анизотропия, если она есть (см. Главу 3, 'Принципы геостатистического анализа').

### Подсказка

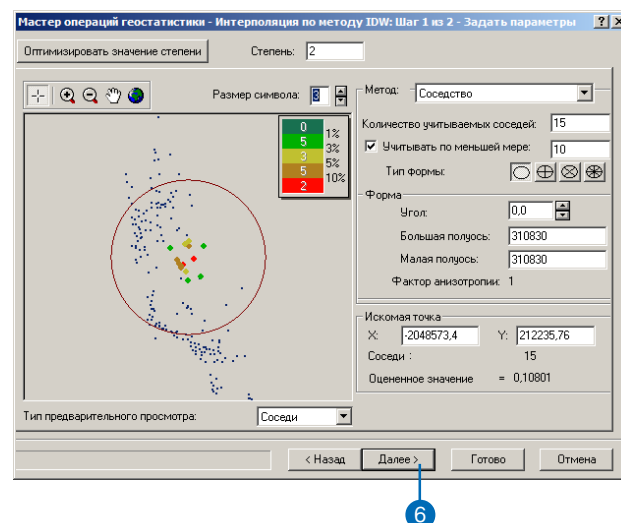
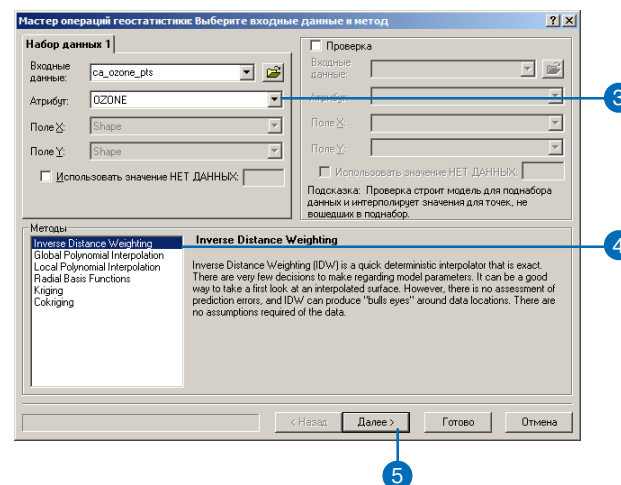
**Использование файла базы данных вместо точечного слоя**  
Вместо слоя ArcMap может быть использован файл базы данных. Для этого в диалоге Выберите исходные данные и метод следует нажать кнопку Пролистать и выбрать требуемый файл базы данных.

### См. также

Для дополнительной информации о параметрах, устанавливаемых в диалоге Поиск соседства и изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Построение карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, который будет использован для интерполяции по методу взвешенных расстояний (IDW).
2. Запустите модуль Geostatistical Analyst.
3. В диалоге Выберите входные данные и метод в окне Атрибут определите показатель, на основе значений которого будет выполняться интерполяция по методу взвешенных расстояний.
4. Выберите Метод взвешенных расстояний.
5. Нажмите Далее.
6. В диалоге Задать параметры установите требуемые параметры и нажмите Далее.
7. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
8. В диалоге Информация о результирующем слое нажмите ОК.



## Подсказка

### Создание учебных и тестовых наборов данных

При выполнении проверки используются два набора данных: учебный и тестовый. Учебный набор данных содержит значения опорных точек, которые будут использованы для создания модели интерполяции. Тестовый набор данных нужен для подтверждения достоверности полученных данных. Учебный набор данных вводится как Набор данных 1, а тестовый - как Набор данных для проверки. См. раздел 'Выполнение проверки для геостатистического слоя, созданного из поднабора данных', в Главе 9, для дополнительной информации о создании поднаборов данных.

## Подсказка

### Использование проверки

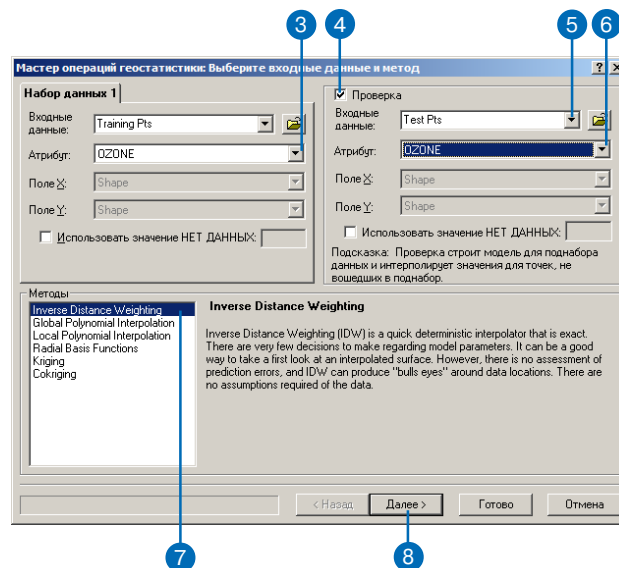
Убедитесь, что учебный набор данных содержит достаточное для точного представления поверхности количество опорных точек. Если учебный набор данных слишком мал, экстремальные (выпадающие) значения данных могут исказить параметры модели и полученные результаты.

## См. также

Для получения дополнительной информации об определении параметров для диалога Поиск соседства и изучения диалога Перекрестная проверка обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с использованием проверки

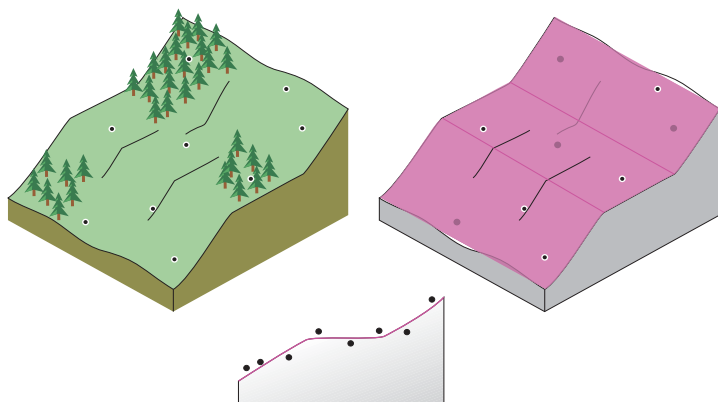
1. В таблице содержания ArcMap выберите точечный слой, который будет использован для интерполяции по методу взвешенных расстояний.
2. Запустите модуль Geostatistical Analyst.
3. В диалоге Выберите входные данные и метод определите показатель, на основе значений которого будет выполняться интерполяция по методу взвешенных расстояний.
4. Отметьте галочкой опцию Проверка.
5. В меню Входные данные выберите точечный слой или пролистайте папки, чтобы найти требуемый файл.
6. В диалоге Выберите входные данные и метод в окне Атрибут определите показатель, на основе значений которого будет выполняться проверка интерполяции по методу взвешенных расстояний.
7. Выберите Метод взвешенных расстояний.
8. Нажмите Далее.
9. В диалоге Задать параметры установите требуемые параметры и нажмите Далее.
10. Изучите результаты в диалогах Перекрестная проверка и Проверка и нажмите Готово.
11. В диалоге Информация о результирующем слое нажмите ОК.



## Как работает интерполяция по методу глобального полинома

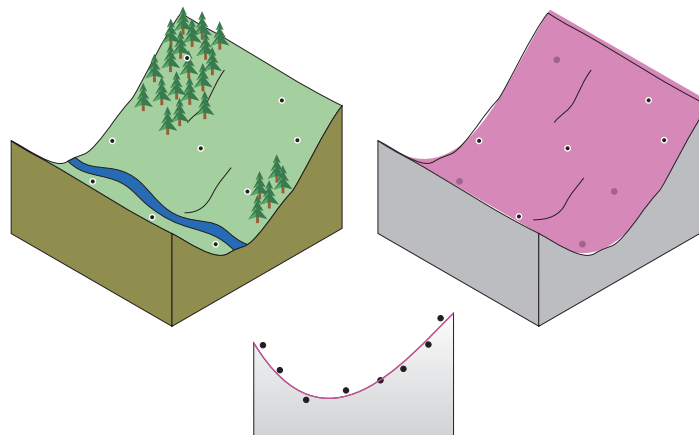
Интерполяция по методу глобального полинома подбирает сглаженную поверхность, построенную по опорным точкам с помощью математической функции (полинома). Поверхность, построенная по методу глобального полинома, меняется постепенно и грубо передает общий характер данных.

Поверхность, построенную с помощью интерполяции по методу глобального полинома, можно представить как лист бумаги, размещенный так, чтобы он прошел через точки, поднятые до значения высоты. На нижнем рисунке это проиллюстрировано для



набора опорных точек, отобранных на пологом склоне (лист бумаги показан лиловым цветом).

Но плоский лист бумаги не может точно описать ландшафт, в котором есть долина. Однако, если вы один раз перегнете кусок бумаги, вы сможете гораздо лучше подогнать его под форму поверхности. Добавление определенного выражения в математическую формулу дает аналогичный результат - перегиб плоскости. Ровная плоскость (на листе бумаги нет перегибов) может быть описана полиномом первой степени (линейной функцией). Один перегиб плоскости будет соответствовать полиному



второго порядка (квадратичная функция), два перегиба - полиному третьего порядка (кубическая функция), и так далее, вплоть до 10, максимально возможной для модуля Geostatistical Analyst степени полинома. На следующем рисунке показано, как можно описать долину с помощью полинома второго порядка.

Лист бумаги редко будет проходить непосредственно через опорные точки, что делает метод интерполяции по методу глобального полинома нежестким интерполятором (метод глобального полинома аппроксимирует значения в опорных точках). Некоторые точки будут расположены над листом бумаги, некоторые - ниже. Однако, если вы сложите значения превышений значений опорных точек над плоскостью и недобора значений опорных точек до значений на плоскости, эти две суммы должны быть равны. Поверхность, показанная лиловым цветом, получена при подборе плоскости по методу наименьших квадратов. Результирующая поверхность минимизирует сумму квадратов разности между действительными значениями опорных точек и их значениями на плоскости.



## Когда использовать интерполяцию по методу глобального полинома

Результатом интерполяции по методу глобального полинома является сглаженная математическая поверхность, которая отражает постепенные тренды в поверхности изучаемой территории.

Глобальная интерполяция используется для:

1. Подбора поверхности для опорных точек в том случае, если поверхность медленно меняется от участка к участку на изучаемой территории (например, загрязнение над промышленными территориями).
2. Изучения и/или удаления эффектов длительных или глобальных трендов. В таком случае метод часто называют методом анализа тренда поверхности.

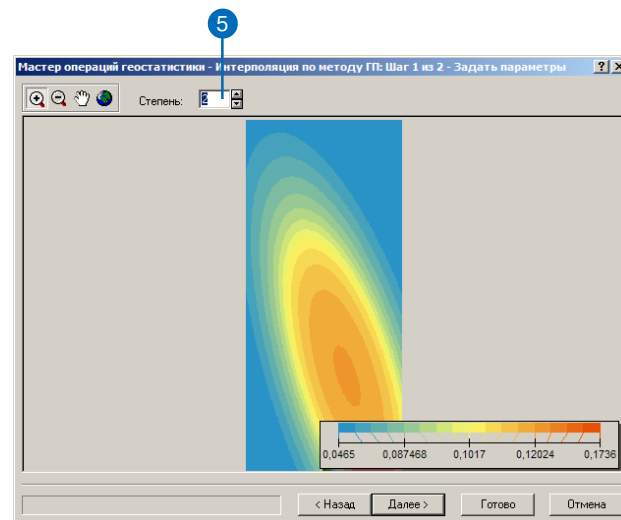
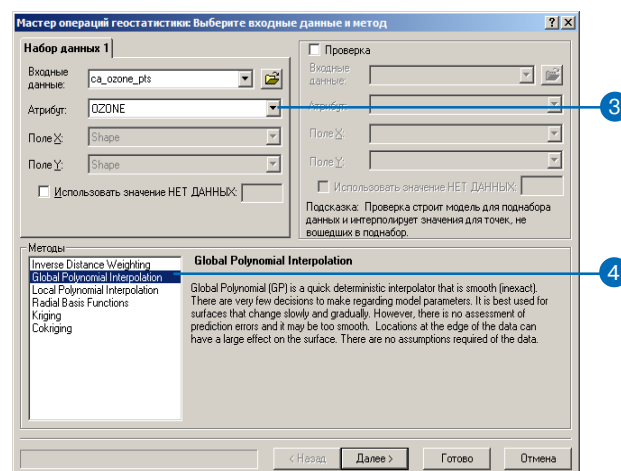
Интерполяция по методу глобального полинома строит медленно изменяющиеся поверхности с использованием полиномов маленьких степеней, которые могут описывать некоторые физические процессы (например, загрязнение и направление ветра). Следует заметить, однако, что чем сложнее полином, тем труднее связать с ним некий физический процесс. Кроме того, вычисленные поверхности очень чувствительны к экстремальным значениям (очень низким или очень высоким, особенно по краям изучаемой территории).

## Создание карты с использованием метода глобального полинома

Для построения карты проинтерполированных значений используйте метод глобального полинома. Этот метод дает лучшие результаты, если атрибут в пределах изучаемой территории меняется медленно, а также, если вы применяете его для изучения эффектов глобальных трендов или анализа поверхности тренда. Поверхность, построенная по методу глобального полинома, чувствительна к экстремальным значениям, особенно по краям изучаемой территории. Моделируемый атрибут в пределах изучаемой территории должен меняться медленно.

## Создание карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, который будет использован для интерполяции по методу глобального полинома.
2. Запустите модуль Geostatistical Analyst.
3. В диалоге Выберите входные данные и метод в окне Атрибут определите показатель, на основе значений которого будет выполняться интерполяция по методу глобального полинома.
4. Выберите интерполяцию по методу глобального полинома и нажмите Далее.
5. В диалоге Задать параметры для интерполяции по методу глобального полинома определите порядок полинома и нажмите Далее.
6. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
7. В диалоге Информация о результирующем слое нажмите ОК.



## Как работает интерполяция по методу локальных полиномов

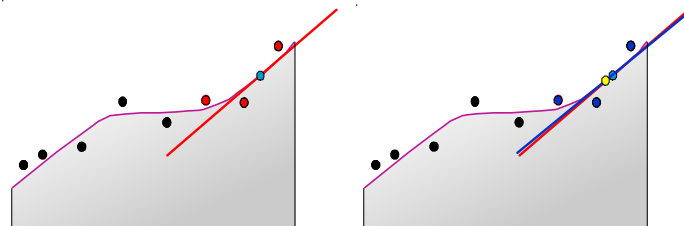
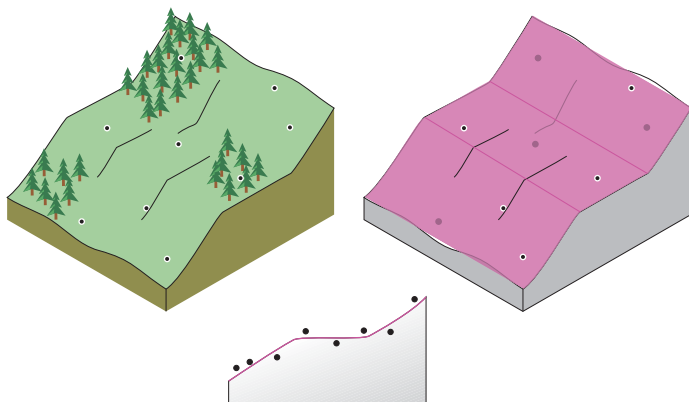
Интерполяция по методу глобального полинома применяет полином ко всей поверхности. Интерполяция по методу локальных полиномов использует несколько полиномов, каждый из которых подбирается для отдельного участка (участки граничат между собой и перекрываются). Поиск соседей может быть определен с использованием диалога Поиск соседства (см. раздел об интерполяции по методу взвешенных расстояний в этой главе и Главу 7, 'Использование аналитических инструментов при построении поверхностей'). В диалоге могут быть определены форма области поиска, максимальное и минимальное количество используемых точек и конфигурация секторов поиска. Помимо этого, можно задать ширину полосы поиска наряду с параметром степени, который, в зависимости от расстояния, будет уменьшать веса опорных точек, попадающих в область соседства. Таким образом, интерполяция по методу локального полинома позволяет строить поверхности, больше учитывающие локальную вариацию.

Глобальный полином первого порядка позволяет провести через опорные точки ровную поверхность; глобальный полином второго порядка описывает поверхность с перегибом, что позволяет использовать его для участков с долиной; глобальный поли-

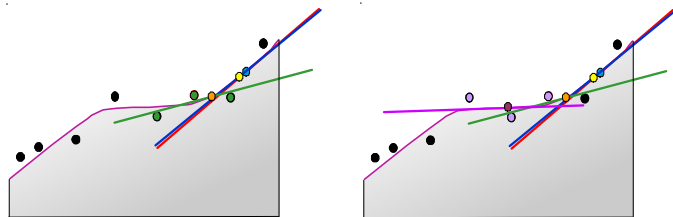
ном третьего порядка допускает наличие двух перегибов в поверхности; и т.д. Однако, в тех случаях, когда у поверхности другая форма, как в нашем примере, когда сначала мы видим склон, затем поверхность выравнивается, а затем снова образуется склон, единый глобальный полином не сможет достаточно хорошо описать форму поверхности. Более точно отразить характер поверхности смогут несколько плоскостей, построенных с использованием полиномов (см. нижний рисунок).

Интерполяция по методу локальных полиномов подбирает полином определенной степени (например, нулевой, первой, второй и третьей), используя точки только из заданной области соседства. Соседние области перекрываются, и значение, используемое для каждой искомой точки - это значение выбранного полинома в центре области соседства.

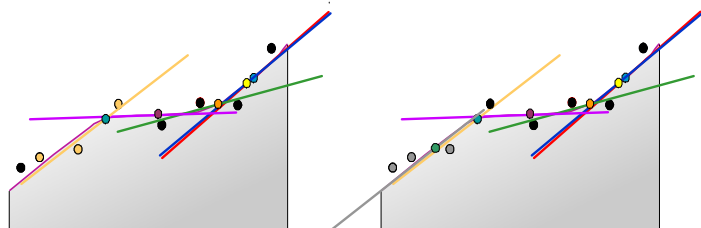
На нижнем рисунке приведен профиль для опорных точек со значениями высот (поперечный разрез). На левом рисунке, показаны три соседние точки (обозначены красным цветом), использованные для расчета полинома первой степени и линия полинома (красная линия), по которой получено значение искомой точки, обозначенной голубым цветом. Значение второй точки (обозначенной на правом рисунке желтым цветом) вычислено с использованием другого полинома первой степени. Точка расположена очень близко к первой точке, и в вычислениях были использованы те же самые опорные точки; но присвоенные им веса немного отличались друг от друга, поэтому и подобранный полином (голубая линия) несколько отличается от первого.



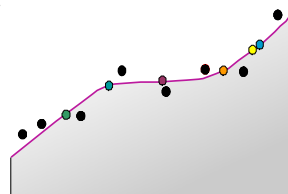
Этот процесс повторяется, при этом центр смещается в последующую искомую точку, и для определения значений этих точек подбираются локальные полиномы. На двух нижних рисунках показаны еще две искомые точки, вычисленные в процессе построения результирующей поверхности. Значение оранжевой точки получено по подобранному полиному, показанному зеленой линией, на основе значений зеленых опорных точек. Значение коричневой точки получено с использованием полинома, показанного сиреневым цветом.



На двух следующих рисунках отображены еще два подобранных полинома (желтая и серая линия) для двух искомых точек (бирюзовая и зеленая точки).



Этот процесс повторяется для всех точек. На нижнем рисунке вы можете видеть, как строится поверхность (малиновая линия) для опорных точек.



Модель оптимизируется путем повторяющейся перекрестной проверки результирующих поверхностей, рассчитанных с использованием различных параметров. Оптимальный параметр выбирается таким образом, чтобы минимизировать среднеквадратичную ошибку подобно тому, как это делается при выборе параметра степени ' $p$ ' при использовании интерполяции по методу взвешенных расстояний (IDW).

## Когда использовать локальную интерполяцию

Интерполяция по методу глобального полинома может быть использована для построения гладких поверхностей и для определения трендов длительного воздействия в наборе данных. Однако, в науках о Земле изучаемая переменная, как правило, изменчива в короткий период времени и, вместе с тем, имеет долговременный тренд. Когда в наборе данных проявляется кратковременная вариация (или вариация на микроуровне), карты, построенные при помощи метода локальной интерполяции, могут помочь выявить эту кратковременную вариацию.

Интерполяция с использованием локальных полиномов чувствительна к расстоянию до соседних точек. По этой причине, вы можете предварительно просмотреть данные перед тем, как создать результирующий слой.

Как и в случае с интерполяцией по методу взвешенных расстояний, вы можете определить модель, которая учитывает анизотропию (обратитесь к разделу этой главы, посвященному методу взвешенных расстояний).

## Создание карты с использованием интерполяции по методу локальных полиномов

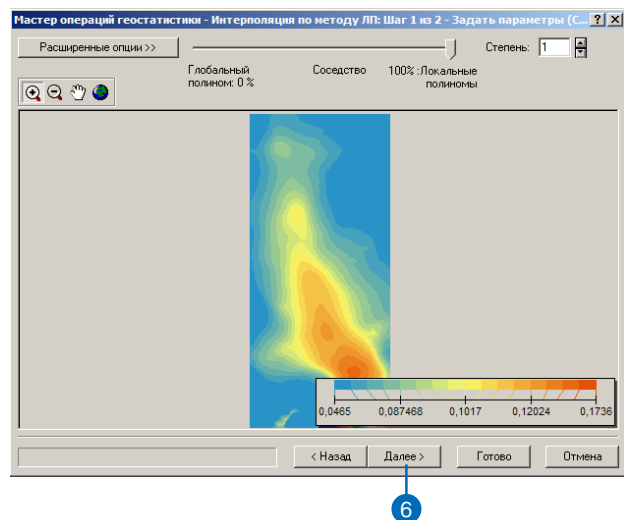
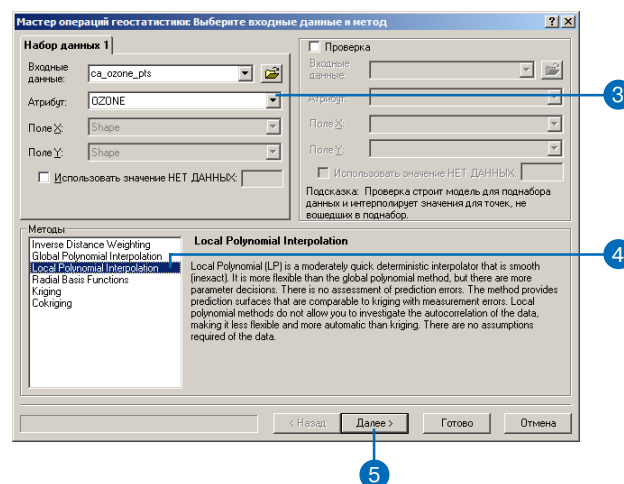
Метод интерполяции с использованием локальных полиномов не является жестким интерполятором (аппроксимирует значения в опорных точках). С его помощью строится сглаженная поверхность. Лучше ее применять, когда данным присуща кратковременная вариация (вариация на микроуровне).

### См. также

*Для дополнительной информации об определении параметров в диалоге Поиск соседства и изучения диалога Перекрестной проверки обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.*

## Создание карты проинтерполированных значений

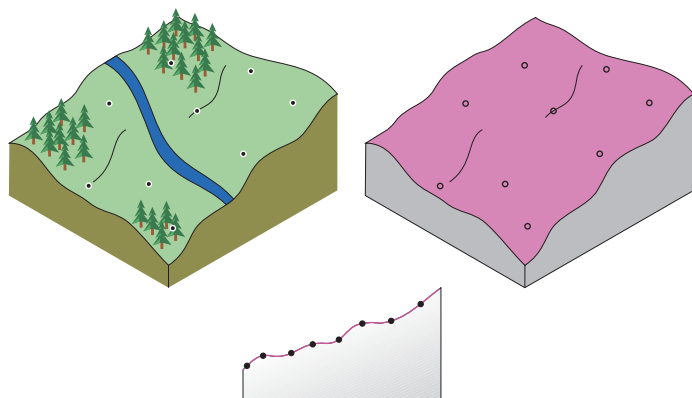
1. В таблице содержания ArcMap выберите точечный слой, который будет использован для интерполяции по Методу локальных полиномов.
2. Запустите модуль Geostatistical Analyst.
3. В диалоге Выберите входные данные и метод в окне Атрибут определите показатель, на основе значений которого будет выполняться интерполяция по методу локальных полиномов.
4. Выберите метод интерполяции с использованием локальных полиномов.
5. Нажмите Далее.
6. В диалоге Задать параметры определите параметры, которые вы будете использовать при выполнении интерполяции по методу локальных полиномов и нажмите Далее.
7. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
8. В диалоге Информация о результирующем слое нажмите ОК.



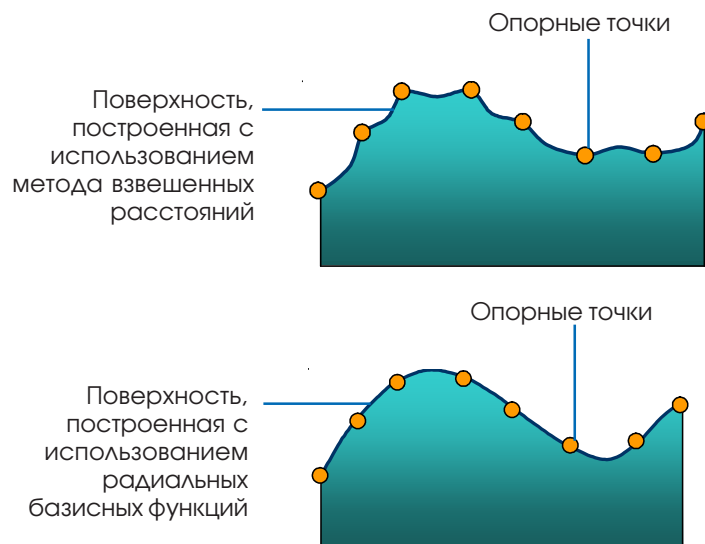
## Как работает интерполяция с использованием радиальных базисных функций

Радиальные базисные функции (Radial basis functions - RBF) - это целый ряд жестких методов интерполяции; то есть, поверхность, построенная с использованием этих функций, будет проходить через все опорные точки. Существует пять различных видов функций: плоский сплайн, сплайн с натяжением, полностью регуляризованный сплайн, функция мультиквадриков, и обратный мультиквадрик. Каждая радиальная функция имеет различную форму и результаты для различных поверхностей интерполяции. Методы RBF - форма искусственных нейронных сетей.

Методы RBF концептуально похожи на метод “резинового листа”, когда лист проходит через все опорные точки, и при этом минимизируется общая кривизна поверхности. Выбранная базисная функция определяет, как резиновый лист пройдет через опорные точки. На диаграмме внизу наглядно показано, как поверхность, построенная с использованием радиальной базисной функции RBF, проходит через опорные точки с высотами. На профиле обратите внимание, что поверхность проходит через значения опорных точек.



Будучи жесткими интерполяторами, методы RBF отличаются от интерполяторов, использующих глобальные и локальные полиномы, поскольку эти два метода являются нежесткими интерполяторами и не предполагают прохождения поверхности через опорные точки (аппроксимируют значения в опорных точках). При сравнении методов с использованием радиальных базисных функций и метода взвешенных расстояний, другого жесткого интерполятора, следует отметить, что метод IDW никогда не даст значений, которые будут выше максимальных или ниже минимальных значений опорных точек (см. приведенные ниже профили).



В отличие от метода взвешенных расстояний, функции RBF могут давать значения выше максимальных и ниже минимальных измеренных значений (см. нижний разрез на рисунке на стр.126).

Оптимальные параметры функций определяются так же, как и для метода взвешенной интерполяции и локальных полиномов - с использованием перекрестной проверки (см. раздел этой главы, посвященный методу взвешенных расстояний).

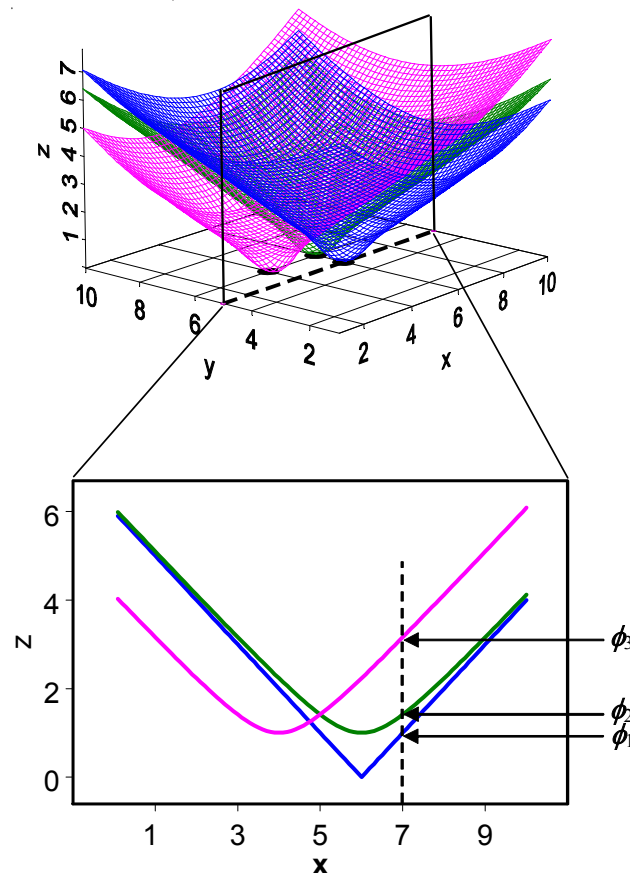
### Когда использовать радиальные базисные функции

Радиальные базисные функции используются для построения сглаженных поверхностей для большого количества опорных точек. Функции дают хорошие результаты для плавно меняющихся поверхностей, таких как рельеф.

Эти методы не подходят в тех случаях, когда на поверхности происходит резкое изменение значений на коротком расстоянии по горизонтали и/или в тех случаях, когда вы предполагаете, что в исходных данных могут быть ошибки или неточности.

### Теоретические основы использования радиальных базисных функций

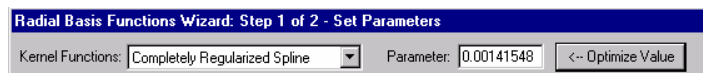
В модуле Geostatistical Analyst радиальные базисные функции формируются над каждой опорной точкой. РБФ - это функция, которая меняется с расстоянием. Например, на следующем рисунке показаны три точки, и для каждой из них функция РБФ показана своим цветом.





В данном примере, радиальная базисная функция - просто расстояние от каждой точки, поэтому над каждой точкой она образует перевернутый конус. Если вы посмотрите на сечение плоскости  $x, z$  для значения  $y = 5$  (второй рисунок на предыдущей странице), вы увидите разрез каждой из приведенных РБФ. Теперь предположим, что вам надо найти значение функции для точки  $y = 5, x = 7$ . Значение каждой из рассматриваемых функций RBF в искомой точке может быть определено по графику, показанному на рисунке, (значения обозначены буквами  $f_1, f_2$  и  $f_3$ ) и зависит от расстояния до каждой из точек. Интерполятор образуется путем нахождения взвешенного среднего  $w_1 f_1 + w_2 f_2 + w_3 f_3 + \dots$ . Вопрос заключается в том, как определить эти веса? Ведь вы совсем не использовали значения данных! Веса  $w_1, w_2, w_3$ , и так далее, должны удовлетворять следующему условию: если искомая точка будет помещена в точку с измеренным значением, значение данных будет проинтерполировано точно. Это приводит к образованию системы из  $N$  уравнений с  $N$  неизвестными, для которой могут быть найдены однозначные решения. Таким образом, поверхность проходит через опорные точки, то есть интерполятор является жестким. Функция RBF, приведенная выше, является особым случаем мультиквадриков. В модуле Geostatistical Analyst можно также использовать другие функции РБФ, такие как полностью регуляризованный сплайн, плоский сплайн, сплайн с натяжением, и обратные мультиквадрики. Часто разница между ними невелика, но у вас могут быть причины для выбора одной из них, либо вы можете попробовать использовать несколько функций, а затем для выбора окончательной применить перекрестную проверку. Каждая функция РБФ имеет параметр, который контролирует “сглаживание” поверхности.

Для всех методов, за исключением обратных мультиквадриков, чем выше значение параметра, тем выше сглаживание поверхности; обратное утверждение верно для функции обратных мультиквадриков.



## Создание карты с использованием интерполяции на основе радиальных базисных функций

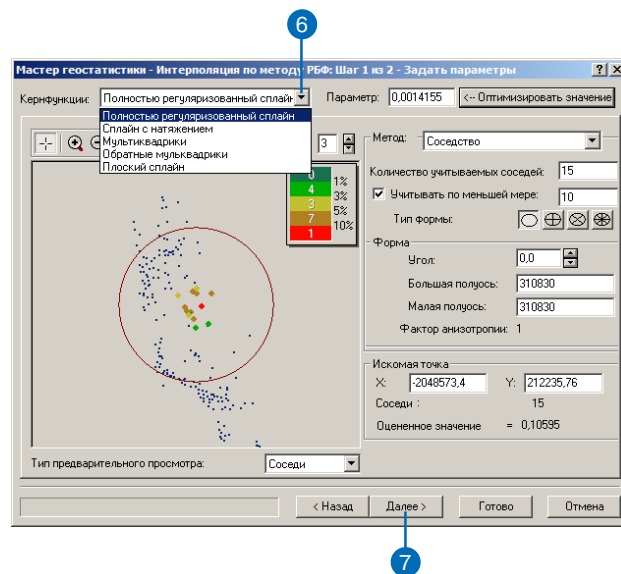
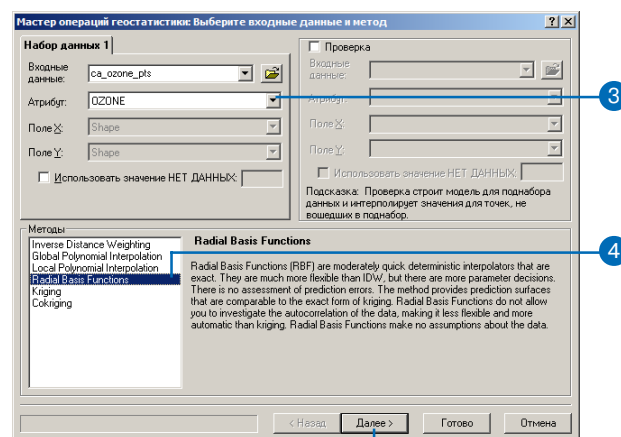
Функции РБФ - жесткие интерполаторы, которые строят гладкие поверхности. Они дают хорошие результаты для плавно меняющихся поверхностей. Поскольку интерполаторы являются жесткими, функции РБФ могут быть чувствительны к экстремальным (выпадающим) значениям.

### См. также

Для дополнительной информации об определении параметров в диалоге Поиск соседства и изучения диалога Перекрестной проверки обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с использованием функций РБФ

1. В таблице содержания ArcMap выберите точечный слой, который будет использоваться для интерполяции с использованием радиальных базисных функций.
2. Запустите модуль Geostatistical Analyst.
3. В диалоге Выберите входные данные и метод выберите показатель, на основе значений которого будет выполняться интерполяция с использованием радиальных базисных функций.
4. Выберите метод Радиальные базисные функции.
5. Нажмите Далее.
6. В диалоге Задать параметры для интерполяции с использованием РБФ в открывающемся списке Кернфункции выберите нужную радиальную функцию.
7. В диалоге Задать параметры задайте необходимые параметры для выбранной РБФ и нажмите Далее.
8. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
9. В диалоге Информация о результирующем слое нажмите ОК.





# Построение поверхности с использованием методов геостатистики

## В ЭТОЙ ГЛАВЕ

- Что такое геостатистические методы интерполяции?
- Изучение моделей кригинга
- Изучение результирующих поверхностей
- Создание карты с использованием параметров, предложенных по умолчанию
- Изучение преобразований и трендов
- Картографирование с применением методов:
- Ординарного кригинга
- Простого кригинга
- Универсального кригинга
- Индикаторного кригинга
- Вероятностного кригинга
- Дизъюнктивного кригинга
- Кокригинга

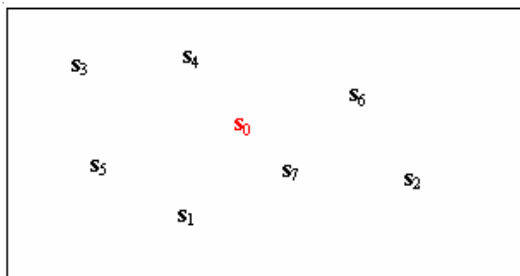
В предыдущей главе вы познакомились с детерминистскими методами интерполяции. Детерминистские методы используют взаимное расположение опорных точек для построения поверхности (Метод взвешенных расстояний) или на основе значений опорных точек подбирают математическую функцию, описывающую поверхность (глобальный и локальный полиномы и радиальные базисные функции). В этой главе будет дан обзор различных геостатистических методов интерполяции. Как следует из их названия, геостатистические методы строят поверхность с учетом статистических свойств используемых данных. Поскольку геостатистика основывается на статистике, эти методы позволяют строить не только поверхности интерполируемых значений, но также поверхности ошибок и неопределенности интерполяции, что помогает вам оценить ее качество.

Многие методы связаны с геостатистикой, но все они относятся к семейству кригинга. В модуле Geostatistical Analyst возможно использование ординарного, простого, универсального, вероятностного, индикаторного и дизъюнктивного кригинга, наряду с дополняющим их кокригингом. Эти методы кригинга не только строят поверхности интерполируемых значений и ошибок, но они могут быть также использованы для создания карт вероятности и квантилей.

Кригинг разделен на две отдельные задачи: количественная оценка пространственной структуры данных и вычисление искомых значений. Количественная оценка структуры, известная как вариография, - это подбор пространственно-зависимой модели для ваших данных. Чтобы найти неизвестное значение в определенной точке, кригинг воспользуется подобранной при вариографии моделью, взаимным расположением (конфигурацией) пространственных данных и значениями опорных точек, находящихся в окрестностях искомой точки. Модуль Geostatistical Analyst предоставляет множество инструментов, которые помогают вам в определении требуемых параметров. Кроме того, вы можете воспользоваться параметрами, предлагаемыми по умолчанию.

## Что такое геостатистические методы интерполяции?

Геостатистика, в своем начальном понимании, означала статистику “о Земле”, используемую в географии и геологии. Сейчас геостатистика широко используется во многих областях и образует ветвь пространственной статистики. Изначально, в пространственной статистике, понятие геостатистики было синонимом кригинга, являющегося статистической версией интерполяции. В настоящее время определение расширено и включает в себя не только метод кригинга, но и многие другие методы интерполяции, включая детерминистские методы, рассмотренные в Главе 5, ‘Детерминистские методы интерполяции пространственных данных’. Модуль Geostatistical Analyst - это реализация такого расширенного определения геостатистики. Одна из существенных особенностей геостатистики состоит в том, что изучаемое явление имеет значение (не обязательно измеренное) в любой точке изучаемой территории, например, количество нитратов в почве или концентрация озона в атмосфере, т.е. является непрерывным. Важно уметь определить типы данных, которые могут быть соответствующим образом проанализированы с использованием методов геостатистики. Предположим, что показанный внизу прямоугольник - изучаемая нами территория. Точки на ней обозначены буквами  $s_i$ , а номер каждой конкретной точки - нижним индексом  $i$ .



Предположим, что вы отобрали данные в точках с  $s_1$  по  $s_7$ , и хотите найти значение точки  $s_0$ , показанной красным цветом. Это пример интерполяции. Кригинг предполагает, что вы мо-

жете поместить точку  $s_0$  в любой точке изучаемой территории, и в этой точке  $s_0$  данные имеют какое-то действительное значение. Например, если данные включают значения содержания нитратов  $s_1, \dots, s_7$ , то в точке  $s_0$  содержание нитратов имеет какое-то значение, которое вы не измерили, но хотите вычислить. Обратите внимание, что данные отбираются так, как будто вы измеряете точечные события, в то время как содержание нитратов имеет площадное распространение, то есть такие данные являются пространственно непрерывными.

В статистике эти значения часто характеризуются, как относящиеся к одному из следующих типов:

- Непрерывное; любое число, например, -1.4789, 10965.6891, и т. п.
- Целое, например, ... -2, -1, 0, 1, 2, ...
- Ранжированное качественное значение; например, худший, средний, лучший
- Неранжированное качественное значение; например, лес, сельскохозяйственные земли, городская застройка
- Бинарное; например, 0 или 1

Слово “непрерывное” может вызвать в данном случае некоторую путаницу. Если данные пространственно непрерывны и имеют непрерывное значение при многомерном нормальном распределении, а также если вам известна корреляция многомерного распределения, в таком случае, кригинг является оптимальным интерполятором. Однако, если учесть, что различные формы кригинга разрабатывались так, чтобы вместить все перечисленные выше типы данных, кригинг - метод аппроксимации, который хорошо работает на практике.

## Изучение различных моделей кригинга

Методы кригинга полагаются на математические и статистические модели. Учет вероятности в статистической модели отличает методы кригинга от детерминистских методов, описанных в Главе 5, ‘Детерминистские методы интерполяции пространственных данных’. При кригинге вы связываете некую вероятность с выполняемой вами интерполяцией; это означает, что значения не могут быть получены по статистической модели абсолютно точно. Рассмотрим пример с измеренными значениями содержания нитратов в почве. Очевидно, что даже при наличии большой выборки, вы не сможете вычислить точное значение содержания нитратов в какой-нибудь конкретной точке, в которой измерения не проводились. Следовательно, вы можете только попытаться проинтерполировать ее значение, и при этом оценить ошибку интерполяции.

Методы кригинга основываются на понятии корреляции. Корреляцию часто определяют как тенденцию двух типов переменных к взаимозависимости. Например, уровень цен на фондовой бирже имеет тенденцию к позитивным изменениям при низком спросе, поэтому говорят, что они обратно коррелируют. Кроме того, можно утверждать, что уровень цен на бирже имеет положительную корреляцию, что означает, что он коррелирует сам с собой. На фондовом рынке два значения цены будут иметь тенденцию к совпадению, если они относятся к двум датам, следующим одна за другой, в отличие от дат, которые разделяет год. Величина, при которой корреляция исчезает, может быть выражена как функция расстояния.

На следующем рисунке корреляция показана как функция расстояния. Это определяющая характеристика геостатистики. В классической статистике предполагается, что наблюдения являются независимыми; следовательно наблюдения не коррелируют между собой. В геостатистике, информация о положении точек наблюдения в пространстве позволяет вам вычислить расстояния между точками наблюдения и смоделировать корреляцию как функцию расстояния.



Также обратите внимание, что в целом, цены фондового рынка растут, и такая тенденция носит название “тренда”. Для геостатистических данных, вы оперируете теми же терминами, которые могут быть выражены простой математической формулой,

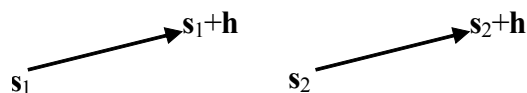
$$Z(s) = \mu(s) + \varepsilon(s),$$

где  $Z(s)$  - интересующая нас переменная, разложенная на детерминистский тренд  $\mu(s)$ , и случайные, коррелирующие ошибки  $\varepsilon(s)$ . Символ  $s$  просто указывает на положение точки; то есть обладает координатами  $x$ - (долгота) и  $y$ - (широта). Различные варианты этой формулы образуют основу для различных типов кригинга, поэтому она стоит усилий, потраченных на ознакомление с ней.

Не имеет значения, насколько сложным в модели является тренд, - составляющая  $\mu(s)$  все равно не сможет дать точных проинтерполированных значений. В этом случае, делаются некоторые допущения относительно ошибки  $\varepsilon(s)$ ; а именно, предполагается, что среднее значение ошибки будет равно 0 и корреляция между  $\varepsilon(s)$  и  $\varepsilon(s+h)$  не зависит от действительного положения точки  $s$ , а только от взаимного расположения двух точек, или расстояния  $h$ . Это необходимо



для оценки функции корреляции. Например, для примера на следующем рисунке, предполагается, что случайные ошибки для пар точек, соединенных стрелками, будут одинаково коррелировать.



Далее, изучите тренд. Он может быть простой константой; то есть,  $m(s) = m$  для всех точек, и если  $m$  известно, это и является моделью, на которой основан ординарный кригинг. Он также может быть представлен линейной функцией самих пространственных координат, например,

$$m(s) = b_0 + b_1x + b_2y + b_3x^2 + b_4y^2 + b_5xy,$$

где поверхность тренда представлена полиномом второй степени и является линейной регрессией пространственных координат  $x$  и  $y$ . Тренды, которые варьируют, и для которых коэффициенты регрессии неизвестны, образуют модели универсального кригинга. Если же полностью известен тренд (т.е., известны все параметры и ковариаты), независимо от того, является он постоянным или нет, он образует модель для простого кригинга.

Теперь рассмотрим левую часть выражения,  $Z(s) = m(s) + e(s)$ . Вы можете выполнить преобразования значения  $Z(s)$ . Например, вы можете изменить его на индикаторную переменную, то есть значениям  $Z(s)$  будет присвоен 0, если они ниже некоторой величины (например, 0.12 ppm для концентрации озона) или 1, если они превышают какое-либо значение. Затем, вы можете спрогнозировать вероятность того, что значения  $Z(s)$  - выше порогового, и такая интерполяция будет носить название индикаторного кригинга. Вы можете также выполнить другие преобразования для значений  $Z(s)$ , назвав их функцией

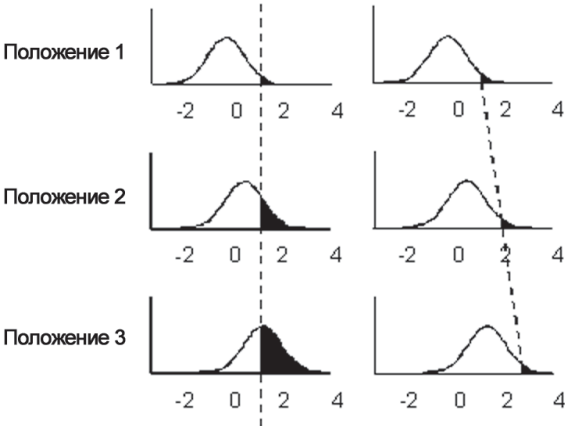
$f_i(Z(s_i))$  для  $i$ -той переменной. Из функций переменных вы можете образовать интерполяторы; например, если вы хотите найти значение показателя в точке  $s_0$ , вы формируете интерpolator дизъюнктивного кригинга из значений функции  $g(Z(s_0))$  с использованием данных функции  $f_i(Z(s_i))$ . В модуле Geostatistical Analyst, функция  $g$  - либо индикаторное преобразование, либо отсутствие преобразования.

И наконец, рассмотрим случай, когда у вас есть больше одного типа переменных, и вы формируете модели  $Z_j(s) = m_j(s) + e_j(s)$  для  $j$ -того типа переменной. Вы можете учесть различные тренды для каждой переменной, и помимо этого, автокорреляцию для ошибок  $e_j(s)$ ; существует также взаимная корреляция между ошибками  $e_j(s)$  и  $e_k(s)$  для двух типов переменных. Например, вы можете учесть взаимную корреляцию между двумя переменными, такими как концентрация озона и определенного вещества, и они необязательно должны быть измерены в одних и тех же точках. Модели, основанные на более чем одной переменной, образуют базу для кокригинга. Вы можете создать индикаторную переменную для значений  $Z(s)$  и, если вы будете вычислять искомое значение с использованием исходных непретворенных данных  $Z(s)$  в модели кокригинга, вы получите вероятностный кригинг. Если изначально у вас есть более одной переменной, для которой вы хотите интерполировать поверхность, вы можете рассматривать ординарный кокригинг, универсальный кокригинг, простой кокригинг, индикаторный кокригинг, вероятностный кокригинг и дизъюнктивный кокригинг, как многовариантные расширения различных типов кригинга, описанных ранее.



# Изучение типов результирующих поверхностей

Кригинг и кокригинг - методы интерполяции, и их цель - построить поверхность предполагаемых (проинтерполированных) значений. Вы можете также захотеть получить ответ на вопрос: “Насколько точны полученные значения?” Можно создать три типа карт предполагаемых значений, и два из них имеют связанные с ними стандартные ошибки. На предыдущих страницах, методы кригинга были сгруппированы по типу моделей, которые они используют; в этом разделе мы приводим их классификацию по целям использования. Рассмотрим следующий рисунок. Предполагается, что искомые значения в трех точках подчиняются нормальному закону распределения.



В таком случае, искомое значение будет находиться в центре каждой кривой, и можно построить карту проинтерполированных значений для всей поверхности. Рассмотрим три левых рисунка. Вероятность того, что значение в искомой точке превысит пороговое значение, к примеру, равное 1, равна площади под кривой справа от пунктирной линии. Распределение проинтерполированных значений меняется с добавлением каждой точки. Таким образом, сохраняя пороговое значение постоянным,

можно построить карту вероятности для всей поверхности. Рассмотрим три правых рисунка. Квантиль с пятью процентами вероятности в правой части распределения будет равен значению, полученному при пересечении пунктирной линией оси x. Снова заметим, что распределение проинтерполированных значений меняется с каждой точкой. Таким образом, сохраняя значение вероятности постоянным, для всей поверхности можно построить карту квантилей. Для карт проинтерполированных значений и для карт вероятности могут быть созданы карты стандартных ошибок. В таблице даны различные методы и результирующие карты, которые могут быть построены с их помощью, а также основные ограничения для каждого из методов.

Кригинг и Кокригинг	Расчет	Стандартная ошибка расчета	Карта Квантилей	Карта вероятностей	Стандартная ошибка показателей
Ординарный	√	√	√ *	√ *	
Универсальный	√	√	√ *	√ *	
Простой	√	√	√ *	√ *	
Показатель				√	√
Вероятность				√	√
Дизьюктивный	√ +	√ +		√ +	√ +

\* Требуется многомерного нормального распределения

+ Требуется парного двумерного нормального распределения

## Создание карты по методу кригинга с использованием параметров, предложенных по умолчанию

Результирующая поверхность создается с использованием метода ординарного кригинга на основе параметров, предлагаемых по умолчанию.

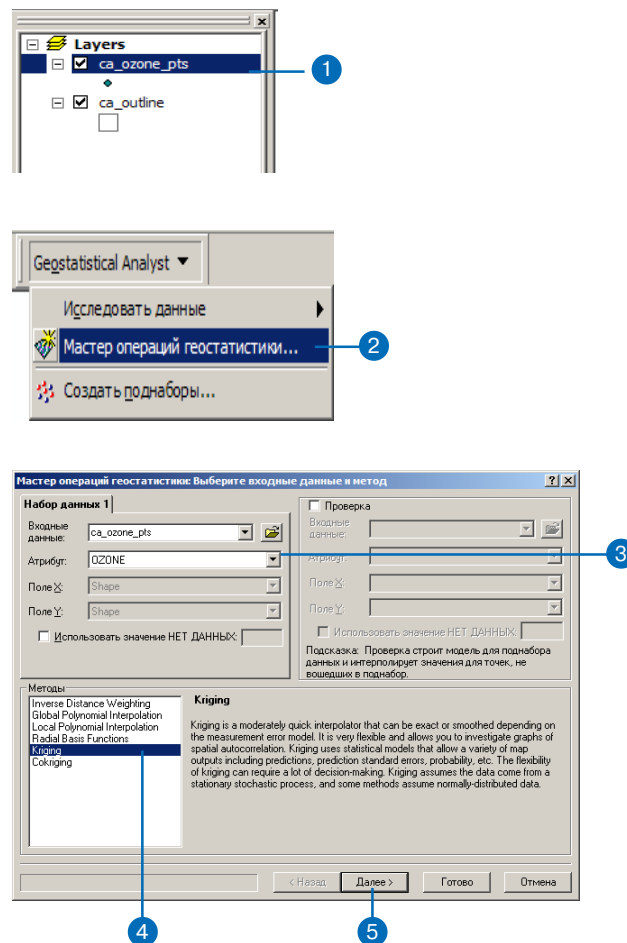
Воспользуйтесь этим методом, если вы незнакомы с геостатистикой и многими параметрами, указанными в диалогах мастера операции; либо если вы хотите визуально изучить данные, представив их в картографическом виде; и наконец, если вы хотите построить предварительную поверхность, которая позволит определить, как настройка параметров может повлиять на результирующую поверхность. Данные в опорных точках должны относиться к явлению, имеющему непрерывное распределение в пространстве.

### Подсказка

**Использование кнопки Готово**  
После того, как вы определили данные и метод в первом диалоге, вы можете нажать кнопку Готово. Модуль построит поверхность с параметрами, предложенными для данного метода по умолчанию.

## Использование параметров, предлагаемых по умолчанию

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить кригинг.
2. Запустите Мастер операций геостатистики.
3. В окне Атрибут выберите показатель, для которого вы хотите выполнить интерполяцию по методу кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Во всех последующих диалогах нажимайте Далее.
7. В диалоге Перекрестная проверка нажмите Готово.
8. В диалоге с информацией о результирующем слое нажмите ОК.



# Изучение преобразований и трендов

Кригинг, как интерполятор, не выдвигает к данным требования нормальности распределения. Однако, как вы видели в предыдущем разделе, подчинение данных закону нормального распределения обязательно для создания карт вероятности и карт квантилей для ординарного, простого и универсального кригинга. Если рассматривать только интерполяторы, которые образованы на основе взвешенных средних, кригинг - лучший интерполятор, отвечающий условию несмещенности оценки, независимо от того, имеют ли ваши данные нормальное распределение. Однако, если данные подчиняются закону нормального распределения, кригинг является лучшим интерполятором среди всех несмещенных интерполяторов, а не только тех, которые основываются на использовании средневзвешенных значений. Кригинг также использует предположение, что все случайные ошибки обладают свойством стационарности второго порядка, что предполагает, что случайные ошибки имеют нулевое среднее значение, и что две случайные ошибки зависят только от расстояния, которое их разделяет, и направления вектора, соединяющего эти точки, а не от их точного местоположения. Преобразования и исключение тренда из данных может помочь подтвердить предположение о нормальности и стационарности. Интерполяция с использованием ординарного, простого и универсального кригинга для общих преобразований по методу Вох—Сох и с использованием арксинуса, носит название трансгауссова кригинга. Логарифмическое преобразование - это особый случай преобразования по методу Вох—Сох, оно обладает специальными свойствами при интерполяции и носит название логарифмически нормального (или логнормального) кригинга. Здесь показаны опции преобразований и анализа тренда, которые доступны для каждого метода кригинга. В таблицах показано также, что выполняется в первую очередь: вычитание тренда или преобразование в том случае, если выбраны обе опции. Подробности о преобразованиях и трендах приведены в Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

Преобразование и тренд для первой переменной:

Тип кригинга	BAL	NST	Тренд
OK	yes (1st if TR)	no	TR (2nd if BAL)
SK	yes	yes	no
UK	yes (1st if T)	no	T (2nd if BAL)
IK	no	no	no
PK*	no	no	no
DK	yes (1st if TR)	yes (2nd if TR)	TR (1st if NST, 2nd if BAL)

Преобразование и тренд для второй переменной (кокригинг):

Тип кригинга	BAL	NST	Тренд
OK	yes (1st if TR)	no	TR (2nd if BAL)
SK	yes	yes	no
UK	yes (1st if T)	no	T (2nd if BAL)
IK	no	no	no
PK	yes (1st if TR)	no	TR (2nd if BAL)
DK	yes (1st if TR)	yes (2nd if TR)	TR (1st if NST, 2nd if BAL)

## Определения

Тренд: фиксированное воздействие, составленное из пространственных координат, использованных в линейной модели  
Первая (первичная) переменная: переменная, для которой вычисляются значения с использованием кригинга или кокригинга  
Вторые (вторичные) переменные: сопеременные (значения которых не вычисляются) при использовании кокригинга

## Сокращения

BAL—преобразования по методу Вох—Сох, арксинуса и логарифмические  
NST—преобразование по методу нормальных меток  
TR—вычитание тренда, или внешний тренд  
T—тренд, или внутренний тренд  
SV—вторичная переменная, или значения ковариат для кокригинга

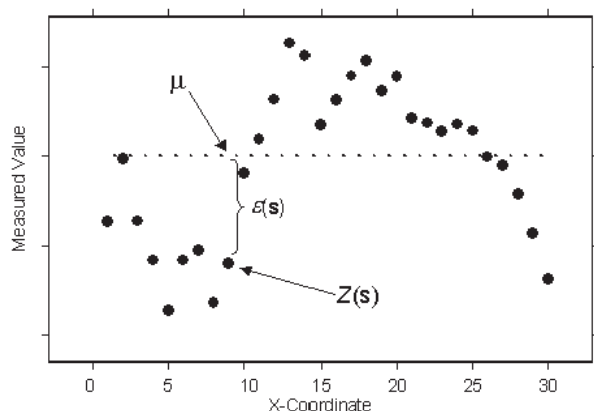
\*Примечание: Для РК (вероятностного кригинга), первичные переменные состоят из индикаторов исходной переменной —эта исходная переменная затем рассматривается как вторая переменная для кокригинга.

## Что такое ординарный кригинг?

Ординарный кригинг использует модель,

$$Z(s) = \mu + \varepsilon(s),$$

где  $m$  - неизвестная постоянная. Один из основных моментов, касающихся ординарного кригинга, - является ли предположение о постоянном среднем оправданным. Иногда существуют веские научные причины для того, чтобы отбросить это предположение. Однако, являясь простым методом интерполяции, он обладает удивительной гибкостью. Следующий рисунок - пример для одного пространственного измерения.



По рисунку можно предположить, что эти данные представляют собой отметки высот, полученные по линии, пересекающей долину и гору. Также можно предположить, что данные имеют больший разброс слева и более сглажены справа. На самом деле, эти данные были получены с использованием модели ординарного кригинга с постоянным средним  $m$ . Истинное, но неизвестное среднее показано пунктирной линией. Таким образом, ординарный кригинг может быть использован для данных, в которых, возможно, присутствует тренд. Нет способа решить, основываясь только на данных, является ли изучаемый участок результатом только автокорреляции (между ошибками  $\varepsilon(s)$  с кон-

стантой  $m$ ) или трендом (когда значение  $m(s)$  меняется с каждой точкой  $s$ ). Часто это решение зависит от решаемой научной задачи.

Ординарный кригинг может использовать либо вариограмму, либо ковариационные функции (которые являются математическим описанием автокорреляции), он может использовать преобразования и вычитание тренда, и он может также допустить наличие ошибок в измерениях; см. Главу 7, 'Использование аналитических инструментов при построении поверхностей', для более подробной информации.

## Создание карты с использованием ординарного кригинга

Используйте ординарный кригинг для создания карт проинтерполированных значений, квантилей, вероятности и карт стандартных ошибок. Он учит неизвестное постоянное значение. Данные, полученные в опорных точках, должны относиться к непрерывному явлению.

### Подсказка

#### Важные параметры

*Соответствующее преобразование, возможное вычитание тренда из поверхности, модели вариограммы/ковариации и область поиска соседства.*

### Подсказка

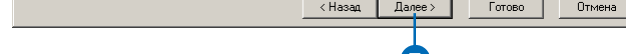
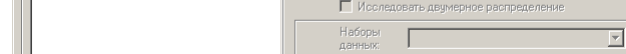
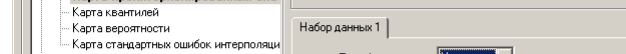
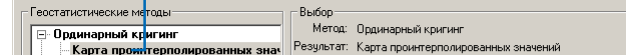
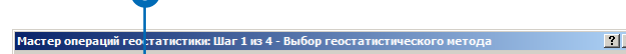
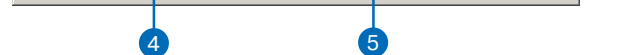
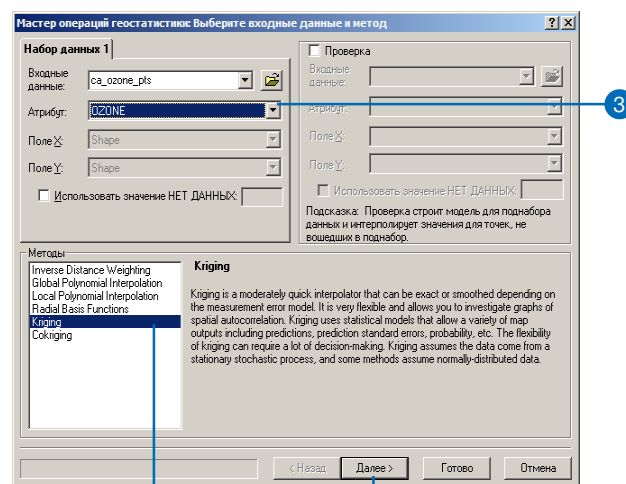
**Использование файла базы данных вместо точечного слоя**  
В диалоге Выберите входные данные и метод вместо слоя ArcMap может быть выбран файл базы данных. Для этого надо нажать кнопку Пролить и выбрать соответствующий файл из базы данных.

### См. также

Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. В окне Атрибут выберите показатель, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Из меню Ординарный кригинг выберите опцию Интерполяция.
7. Нажмите Далее.
8. В диалоге Моделирование вариограммы/ковариации задайте требуемые параметры и нажмите Далее.
9. В диалоге Поиск соседства задайте требуемые параметры поиска и нажмите Далее.
10. В диалоге Перекрестная проверка изучите полученные результаты и нажмите Готово.
11. В диалоге с информацией о результирующем слое нажмите ОК.



## Подсказка

### Создание учебного и тестового наборов данных

При выполнении проверки, используются два набора данных: учебный набор данных и тестовый набор данных. Учебный набор данных содержит опорные точки, по которым будет выполняться интерполяция. Тестовый набор данных будет использован для проверки вычислений. Учебный набор данных вводится как Набор данных I, а тестовый набор данных - как Проверочный набор данных. См. раздел 'Выполнение проверки для геостатистического слоя, созданного из поднабора данных' в Главе 9 для получения дополнительной информации о создании поднаборов данных.

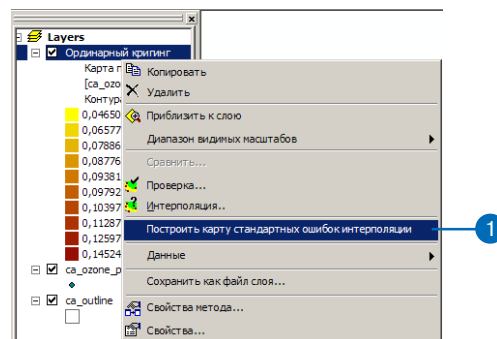
## Подсказка

### Использование проверки

Убедитесь, что учебный набор данных содержит достаточное для точного представления поверхности количество опорных точек. Если учебный набор слишком мал, экстремальные значения данных могут сместить параметры модели и исказить полученные результаты.

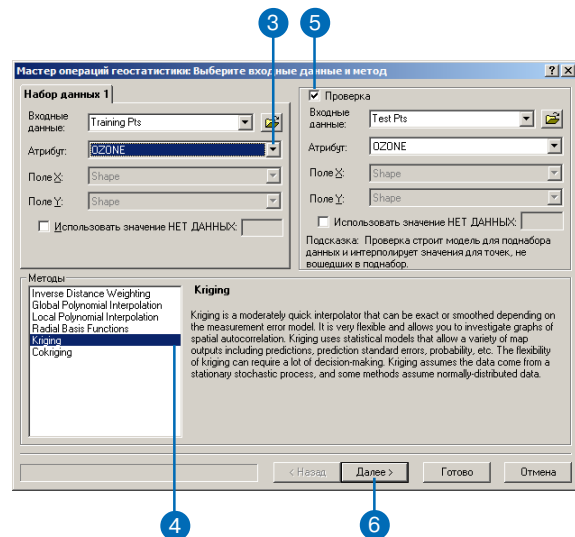
## Создание карты стандартных ошибок интерполяции

1. В таблице содержания ArcMap нажмите правую клавишу мыши на поверхности проинтерполированных значений, построенной с использованием ординарного кригинга, и из контекстного меню выберите опцию Построить карту стандартных ошибок интерполяции.



## Создание карты проинтерполированных значений с использованием проверки

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. В окне Атрибут выберите показатель, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
4. Из списка методов выберите Кригинг.
5. Отметьте галочкой опцию Проверка и задайте набор данных и показатель, которые будут использоваться при проверке.
6. Нажмите Далее.
7. Повторите шаги с 6 по 10 упражнения 'Создание карты проинтерполированных значений' с предыдущей страницы, изучите полученные результаты в диалоге Проверка и затем нажмите Готово.



## Подсказка

### Применение инструментов ESDA для определения параметров модели

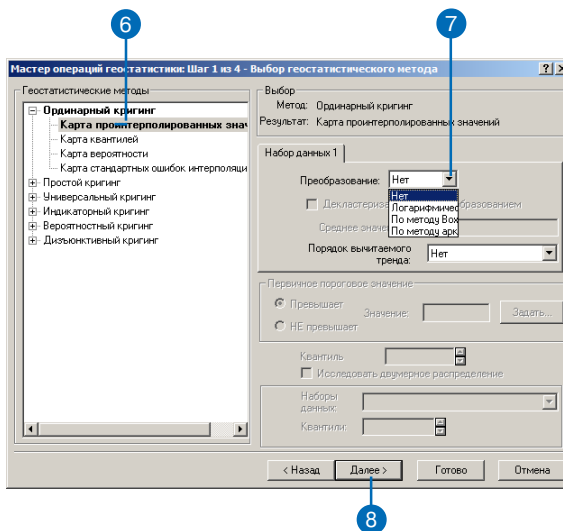
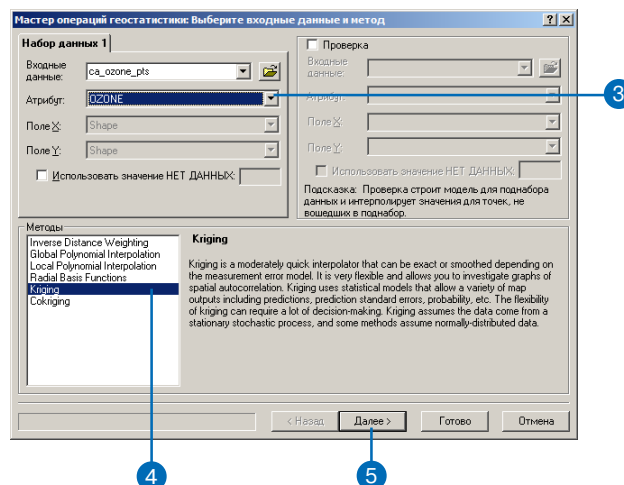
Воспользуйтесь инструментами исследовательского анализа пространственных данных (ESDA) для принятия решений по преобразованиям, вычитанию тренда и наличию эффекта выпадающих (экстремальных) значений в моделях вариограммы/ковариации. Подтвердите свое решение с использованием проверки и перекрестной проверки.

## См. также

Для дополнительной информации о преобразованиях и определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с применением преобразований

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите атрибут, для которого вы хотите выполнить ординарный кригинг.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Ординарный кригинг и выберите опцию Карта проинтерполированных значений.
7. Из меню Преобразование выберите требуемый метод преобразования.
8. Нажмите Далее.
9. Выполните шаги с 9 по 12 упражнения 'Создание карты проинтерполированных значений с использованием метода вычитания тренда', которое дано на следующей странице.





## Подсказка

### Вычитание тренда

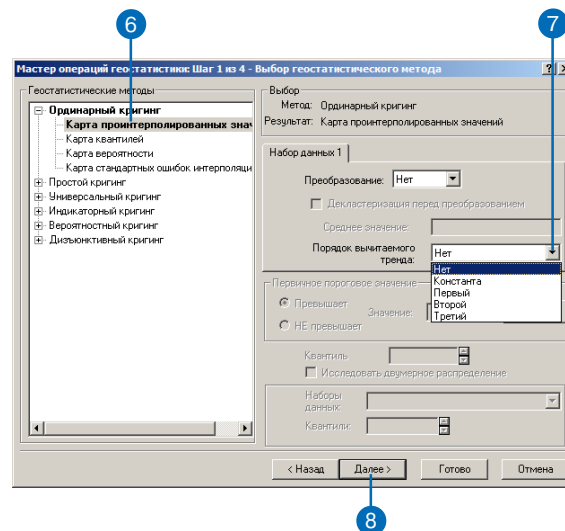
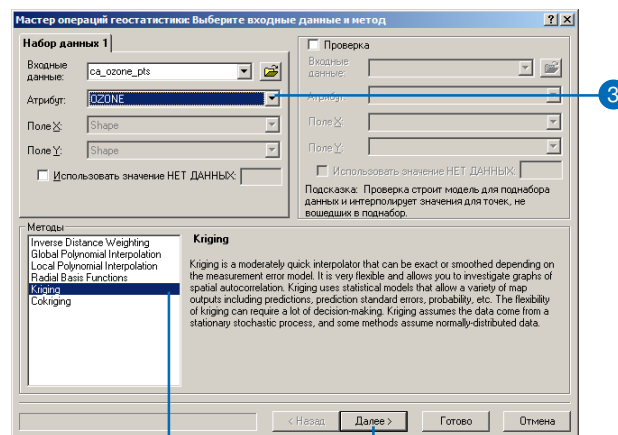
После того, как значение тренда будет вычтено из исходных данных, ординарный кригинг будет выполнен для значений остатков.

## См. также

Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с использованием метода вычитания тренда

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. В окне Атрибут выберите показатель, для которого вы хотите выполнить интерполяцию по методу ординарного кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Из списка Ординарный кригинг выберите опцию Карта проинтерполированных значений.
7. В окне Порядок вычитаемого тренда выберите необходимую опцию.
8. Нажмите Далее.
9. В диалоге Вычитание тренда задайте необходимые параметры и нажмите Далее.
10. Задайте необходимые параметры в диалоге Моделирование вариограммы/ковариации и нажмите Далее.
11. Задайте необходимые параметры в диалоге Поиск соседства и нажмите Далее.
12. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
13. В диалоге с информацией о результирующем слое нажмите ОК.

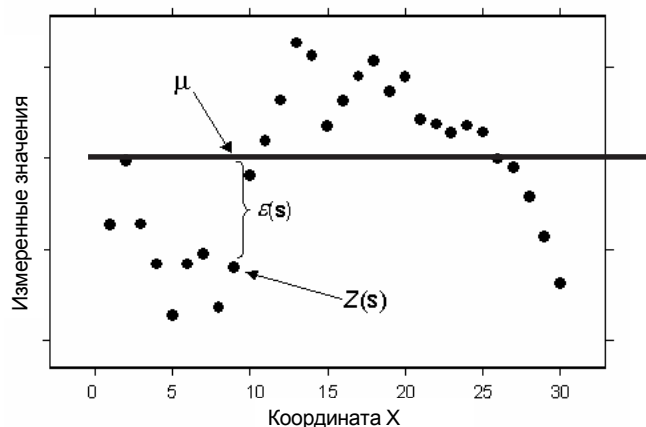


## Что такое простой кригинг?

Простой кригинг использует модель,

$$Z(s) = m + e(s)$$

где  $m$  - известная константа. Например, на следующем рисунке, который использует тот же набор данных, что и примеры для



пояснения понятий ординарного и универсального кригинга, точки наблюдений даны черными кружочками. Известная константа — сплошная жирная линия — равна  $m$ . (Сравните с ординарным кригингом.) Для простого кригинга, поскольку предполагается, что вы точно знаете значение  $m$ , точно известны также значения  $e(s)$  для ваших точек. При ординарном кригинге вы оцениваете значение  $m$ , следовательно, вы также оцениваете значение  $e(s)$ . Если вам известно точное значение  $e(s)$ , вы сможете оценить автокорреляцию и сделать это лучше, чем если бы вы оценивали значения  $e(s)$ . Предположение, что вам будет точно известно значение  $m$ , часто является нереалистичным. Однако, иногда имеет смысл предположить, что модель, имеющая физический смысл, дает известный тренд. Тогда вы можете

взять разницу между значениями этой модели и измеренными значениями, которая носит название остатков, и применить метод простого кригинга к этим значениям остатков, приняв за известное, что тренд в этих остатках равен нулю.

Простой кригинг может использовать либо вариограммы, либо ковариационные функции (являющиеся математическим способом определения автокорреляции), он может использовать преобразования, и он может также допустить наличие ошибок в измерениях; см. Главу 7, 'Использование аналитических инструментов при построении поверхностей', для более подробной информации.

## Создание карты с использованием простого кригинга

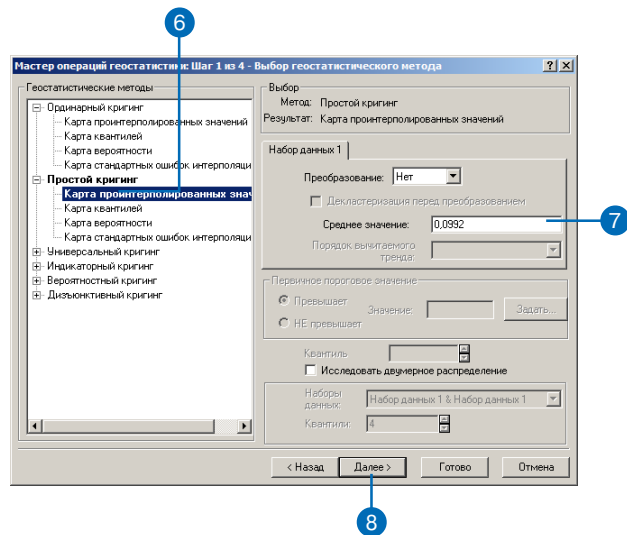
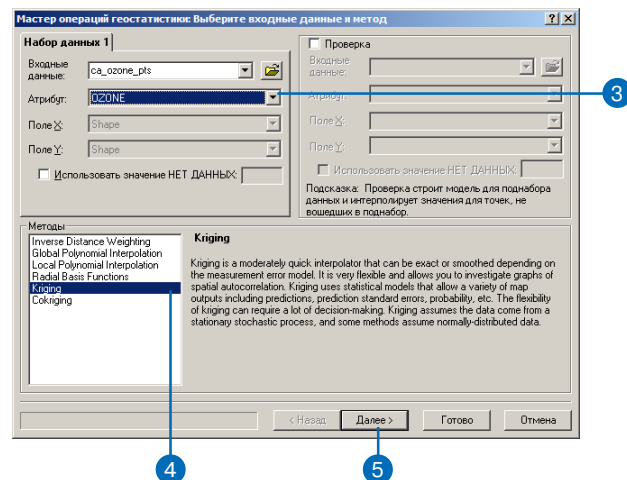
Используйте простой кригинг для создания карт проинтерполированных значений, квантилей, вероятности и карт стандартных ошибок. Он учит известное постоянное значение. Данные, полученные в опорных точках, должны относиться к непрерывному явлению.

### См. также

Для дополнительной информации о преобразованиях и определении параметров в диалогах *Моделирование вариограммы/ковариации* и *Поиск соседства*, а также для изучения диалога *Перекрестная проверка*, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Простой кригинг и выберите опцию Карта проинтерполированных значений.
7. Задайте Среднее значение.
8. Нажмите Далее.
9. Задайте необходимые параметры в диалоге Моделирование вариограммы/ковариации и нажмите Далее.
10. Задайте необходимые параметры в диалоге Поиск соседства и нажмите Далее.
11. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
12. В диалоге с информацией о результирующем слое нажмите ОК.



## Подсказка

### Проверка на двумерное распределение

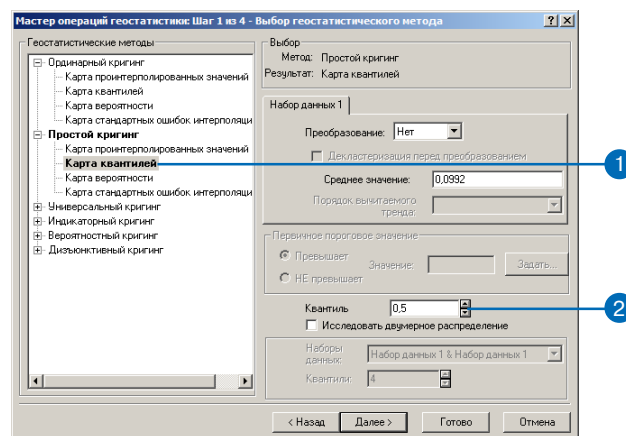
Проверяйте свои данные на двумерную нормальность. См. Главу 7, 'Использование аналитических инструментов при построении поверхностей'.

## См. также

Для дополнительной информации о преобразованиях и определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

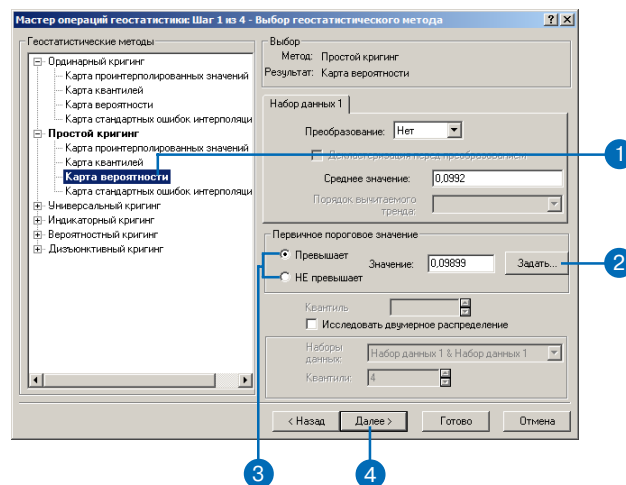
## Создание карты квантилей

1. Повторите шаги, описанные в упражнении 'Создание карты проинтерполированных значений' с предыдущей страницы, но в шаге 6, выберите опцию Карта квантилей, а не Карта проинтерполированных значений.
2. В окне Квантиль с помощью стрелок поменяйте значение уровня квантилей.
3. Повторите шаги с 7 по 12, описанные в упражнении 'Создание карты проинтерполированных значений' с предыдущей страницы.



## Создание карты вероятности

1. Повторите шаги, описанные в упражнении 'Создание карты проинтерполированных значений' с предыдущей страницы, за исключением того, что в шаге 6, выберите опцию Карта вероятности, а не Карта проинтерполированных значений.
2. В диалоге Первичное пороговое значение введите пороговое значение или используйте для этого кнопку Задать...
3. Отметьте опцию Превышает или Не превышает.
4. Нажмите Далее.
5. Повторите шаги с 7 по 12, описанные в упражнении 'Создание карты проинтерполированных значений' с предыдущей страницы.



## Подсказка

### Создание учебного и тестового наборов данных

При выполнении проверки, используется два набора данных: учебный набор данных и тестовый набор данных. Учебный набор данных содержит опорные точки, по которым будет выполняться интерполяция. Тестовый набор данных будет использован для проверки вычислений. Учебный набор данных вводится как Набор данных 1, а тестовый набор данных - как Проверочный набор данных. См. раздел 'Выполнение проверки для геостатистического слоя, созданного из поднабора данных' в Главе 9, для получения дополнительной информации о создании поднаборов данных.

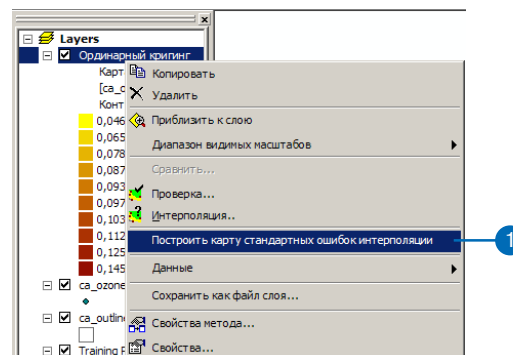
## Подсказка

### Использование проверки

Убедитесь, что учебный набор данных содержит достаточное для точного представления поверхности количество опорных точек. Если учебный набор слишком мал, экстремальные значения данных могут сместить параметры модели и исказить полученные результаты.

## Создание карты стандартной ошибки интерполяции

1. В таблице содержания ArcMap нажмите правую клавишу мыши на созданной с использованием простого кригинга поверхности проинтерполированных значений и из контекстного меню выберите опцию Построить карту стандартных ошибок интерполяции.



## Подсказка

### Инструменты исследовательского анализа пространственных данных ESDA

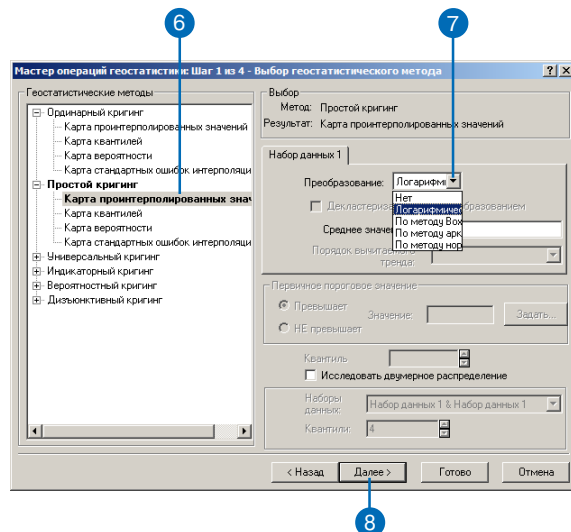
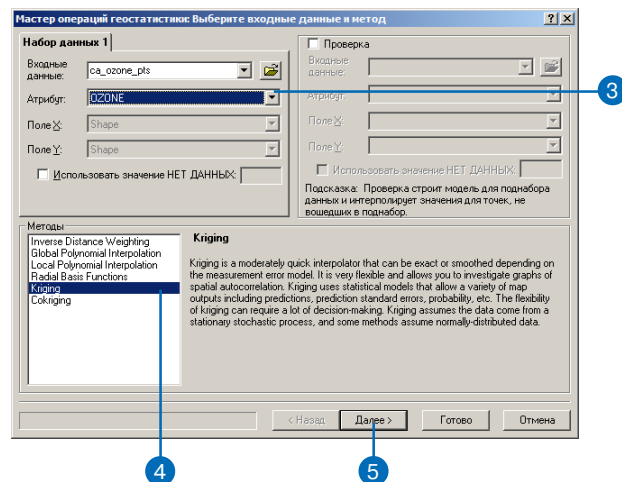
Воспользуйтесь инструментами исследовательского анализа пространственных данных (ESDA) для принятия решений по преобразованиям, и влиянию экстремальных значений на модель вариограммы/ковариации. Подтвердите свое решение с использованием проверки и перекрестной проверки.

## См. также

Для дополнительной информации о преобразованиях и определении параметров в диалогах *Моделирование вариограммы/ковариации* и *Поиск соседства*, а также для изучения диалога *Перекрестная проверка*, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с применением преобразований

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Простой кригинг и выберите опцию Карта проинтерполированных значений.
7. В окне Преобразование выберите требуемый метод преобразования.
8. Нажмите Далее.
9. Повторите шаги с 9 по 12 упражнения 'Создание карты проинтерполированных значений с использованием метода вычитания тренда', приведенного ранее в этой главе.



## Подсказка

### Использование декластеризации

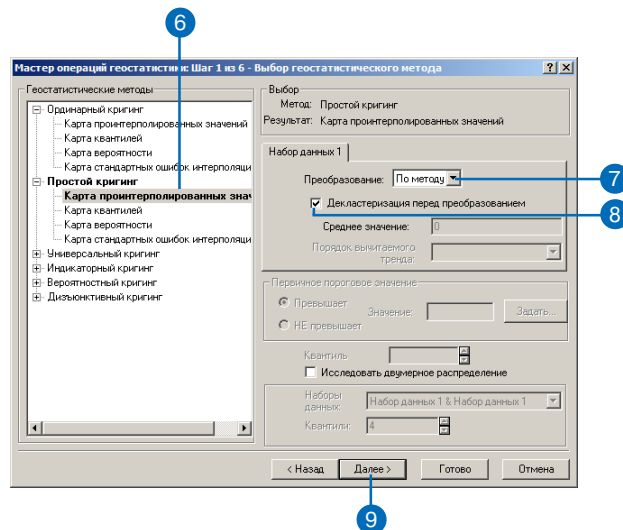
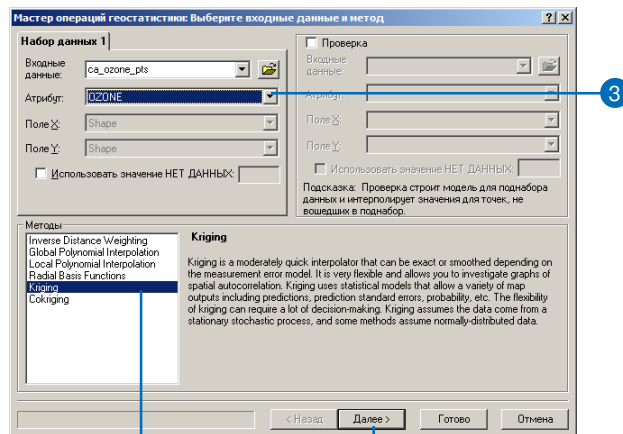
Если вы используете преобразование по методу нормальных меток, и ваши данные были изначально отобраны с различной густотой для различных участков территории, попробуйте выполнить декластеризацию данных. См. Главу 7, 'Использование аналитических инструментов при построении поверхностей'.

## См. также

Для дополнительной информации о преобразованиях и определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с применением преобразований с декластеризацией

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Простой кригинг и выберите опцию Карта проинтерполированных значений.
7. В окне Преобразование выберите опцию По методу нормальных меток.
8. Отметьте галочкой опцию Декластеризация перед преобразованием.
9. Нажмите Далее.
10. В диалоге Декластеризация задайте требуемые параметры и нажмите Далее.
11. Задайте требуемые параметры в диалоге Преобразование по методу нормальных меток и нажмите Далее.
12. Повторите шаги с 9 по 12 упражнения 'Создание карты проинтерполированных значений', приведенного ранее в этой главе.





## Подсказка

### Двумерное распределение

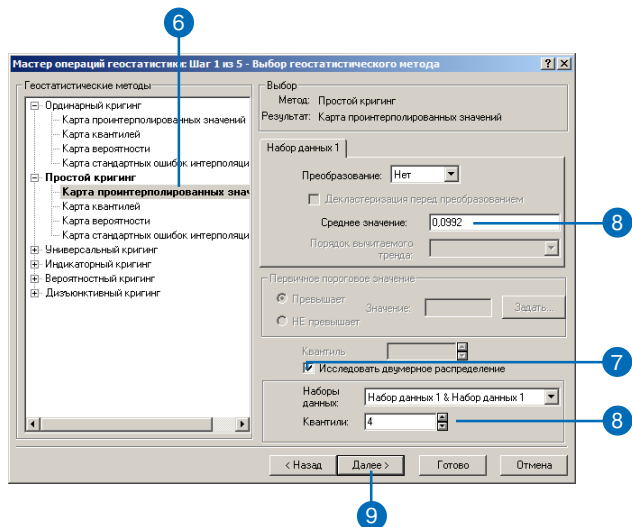
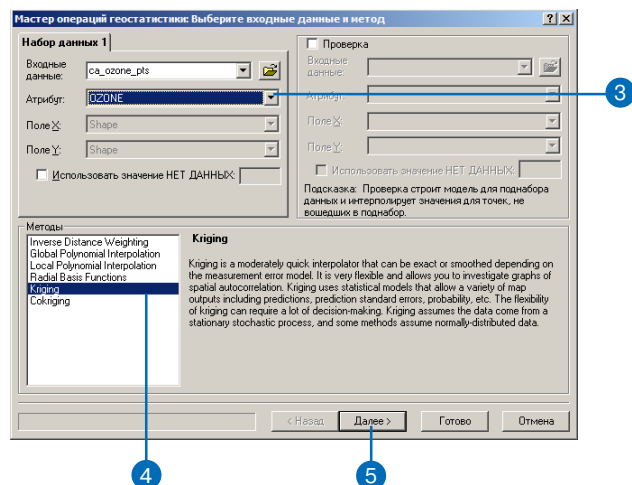
Проверяйте свои данные на двумерную нормальность. См. Главу 7, 'Использование аналитических инструментов при построении поверхностей'.

## См. также

Для дополнительной информации об определении параметров в диалогах Вычитание тренда, Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Изучение двумерного распределения при создании карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу простого кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Простой кригинг и выберите опцию Карта проинтерполированных значений.
7. Отметьте галочкой опцию Исследовать двумерное распределение.
8. Задайте Среднее значение и количество проверяемых квантилей.
9. Нажмите Далее.
10. Задайте необходимые параметры в диалоге Моделирование вариограммы/ковариации и нажмите Далее.
11. Изучите диалог Исследовать двумерное распределение и нажмите Далее.
12. Повторите шаги с 10 по 12 упражнения 'Создание карты проинтерполированных значений', приведенного ранее в этой главе.

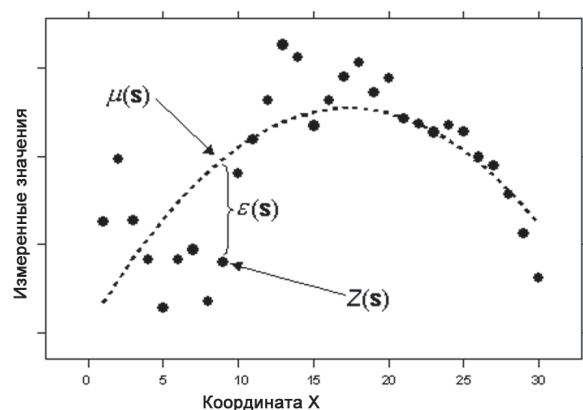


# Что такое универсальный кригинг?

Универсальный кригинг предполагает модель,

$$Z(s) = m(s) + e(s),$$

где  $m(s)$  - некая детерминистская функция. Например, на следующем рисунке, данные для которого совпадают с использованными в примере для ординарного кригинга, опорные точки даны черными кружочками.



Тренд представлен полиномом второго порядка (длинная пунктирная линия) - и равен  $m(s)$ . Если вы вычтете полином второго порядка из ваших данных, вы получите ошибки,  $e(s)$ , которые считаются случайными. Среднее значение всех  $e(s)$  равно 0. Автокорреляция моделируется из случайных ошибок  $e(s)$ . Рисунок вверху выглядит как пример полиномиальной регрессии из любого базового курса статистики. В действительности, он показывает, что представляет собой универсальный кригинг. Вы выполняете регрессию для пространственных координат, как для объяснительных переменных. Однако, вместо того, чтобы рассматривать ошибки  $e(s)$  как независимые величины, вы моделируете их как автокоррелирующие. Повторим совет, дан-

ный нами при рассмотрении ординарного кригинга: нет способа принять решение о правильном разделении значения на составляющие, основываясь только на одних данных.

Универсальный кригинг может использовать либо вариограммы, либо ковариационные функции (являющиеся математическим способом определения автокорреляции), он может использовать преобразования для данных, из которых должны быть вычтены тренды, и он может также допустить наличие ошибок в измерениях. См. Главу 7, 'Использование аналитических инструментов при построении поверхностей', для более подробной информации.

## Создание карты с использованием универсального кригинга

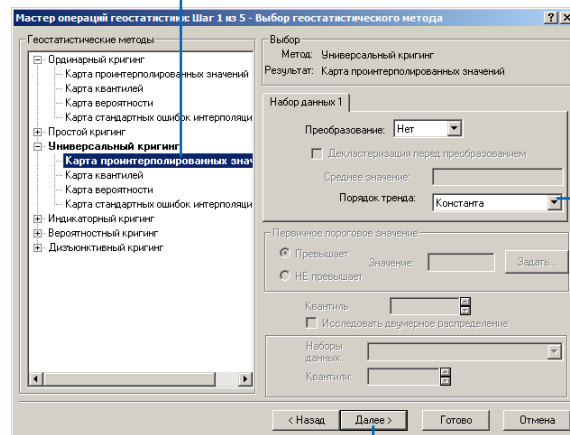
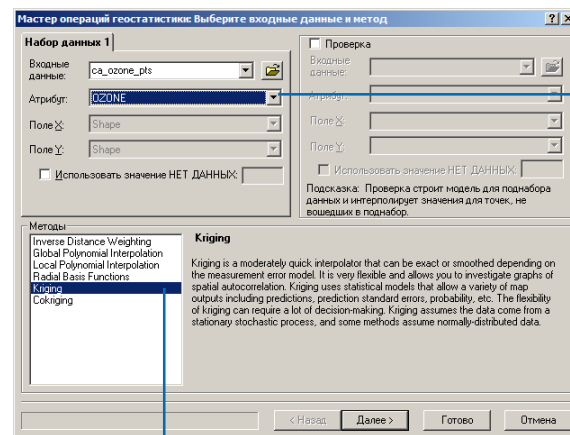
Используйте универсальный кригинг для создания карт проинтерполированных значений, квантилей, вероятности и карт стандартных ошибок интерполяции. Он учит среднее значение тренда. Данные, полученные в опорных точках, должны относиться к явлению, имеющему непрерывное распространение.

### См. также

Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу универсального кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите поле атрибута, для которого вы хотите выполнить универсальный кригинг.
4. Из списка методов выберите Кригинг и нажмите Далее.
5. Откройте список для строки Универсальный кригинг и выберите опцию Карта проинтерполированных значений.
6. В окне Порядок тренда выберите требуемый порядок полинома для тренда.
7. Нажмите Далее.
8. Задайте требуемые параметры в диалоге Вычитание тренда и нажмите Далее.
9. Задайте необходимые параметры в диалоге Моделирование вариограммы/ковариации и нажмите Далее.
10. Задайте необходимые параметры в диалоге Поиск соседства и нажмите Далее.
11. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
12. В диалоге с информацией о результирующем слое нажмите ОК.



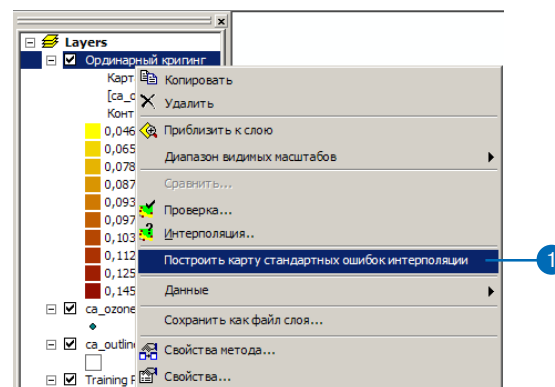
## Подсказка

### Применение инструментов ESDA для определения преобразований

Воспользуйтесь инструментами исследовательского анализа пространственных данных (ESDA) для принятия решений по преобразованиям, вычитанию тренда и наличию эффекта выпадающих (экстремальных) значений в моделях вариограммы/ковариации. Подтвердите свое решение с использованием проверки и перекрестной проверки.

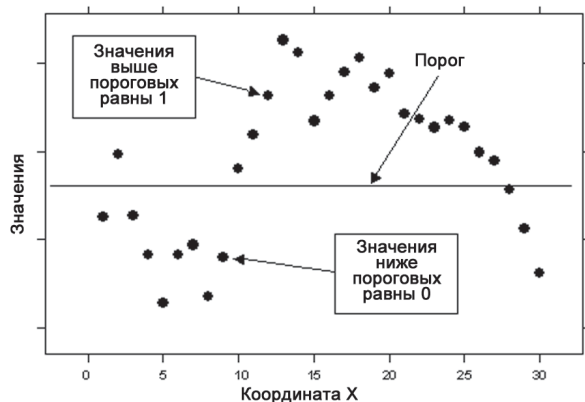
## Создание карты стандартных ошибок интерполяции

1. В таблице содержания ArcMap нажмите правую клавишу мыши на поверхности проинтерполированных значений, построенной с использованием универсального кригинга, и из контекстного меню выберите опцию Построить карту стандартных ошибок интерполяции.



## Что такое пороговые значения?

Переменная, которая имеет непрерывные значения, может быть преобразована в бинарную (0 или 1) переменную путем выбора некоего порогового (критического) значения. В модуле Geostatistical Analyst, если значения в опорных точках выше порогового, им присваивается 1, если значения в опорных точках ниже порогового, им присваивается 0.

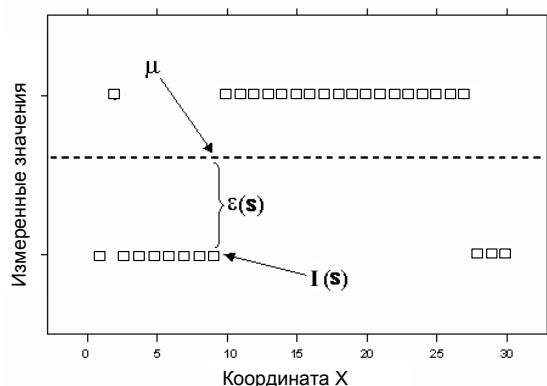


## Что такое индикаторный кригинг?

Индикаторный кригинг предполагает модель,

$$I(s) = m + e(s),$$

где  $m$  - неизвестная константа, а  $e(s)$  - бинарная переменная. Бинарные данные могут быть созданы для непрерывных данных с использованием порогового (критического) значения, либо значения в опорных точках могут изначально, при выполнении наблюдений, фиксироваться как 0 или 1. Например, у вас может быть набор данных, который содержит информацию о том, расположена точка в лесу или нет, и бинарная переменная в данном случае будет означать принадлежность к определенному классу. Используя бинарные переменные, индикаторный кригинг действует так же, как и ординарный кригинг. Например, на следующем рисунке, для которого использованы те же данные, что и для ординарного, универсального и простого кригинга, данные были преобразованы в бинарные с использованием порогового значения, аналогично тому, как было описано в предыдущем разделе 'Что такое пороговые значения?'.



Бинарные данные в опорных точках показаны прямоугольниками. Неизвестное среднее для всех индикаторных перемен-

ных показано пунктирной линией, и равно  $m$ . (Сравните с ординарным кригингом.) Как и для ординарного кригинга, вы предполагаете, что значения  $e(s)$  автокоррелируют. Обратите внимание, что поскольку индикаторные переменные равны 0 или 1, значения, полученные в результате интерполяции по методу индикаторного кригинга, будут находиться в диапазоне между 0 и 1 и могут быть интерпретированы, как вероятности того, что переменная будет равна 1 или попадет в класс, обозначенный как 1. Если для создания индикаторной переменной использовалось пороговое значение, то на карте с результатами интерполяции будут показаны вероятности того, что пороговое значение будет превышено (или наоборот, искомые значения будут ниже порогового).

Возможно также создание нескольких индикаторных переменных для одного и того же набора данных путем выбора нескольких пороговых значений. В таком случае, первое пороговое значение образует первичную индикаторную переменную, а другие индикаторные переменные используются в кокригинге как вторичные переменные.

Индикаторный кригинг может использовать либо вариограммы, либо ковариационные функции (являющиеся математическим способом определения автокорреляции); см. Главу 7, 'Использование аналитических инструментов для построения поверхностей', для более подробной информации.

## Создание карты с использованием индикаторного кригинга

Используйте индикаторный кригинг для создания карт вероятности и карт стандартных ошибок индикаторов. Индикаторный кригинг учитывает неизвестное постоянное среднее. Данные, полученные в опорных точках должны относиться к непрерывному явлению.

### Подсказка

#### Важные параметры

*Заданное пороговое значение (которое определяет, каким интерполированным значениям будет присвоен "0", а каким "1"), модели ковариации/вариограммы, и область поиска соседства.*

### Подсказка

#### Выбор порогового значения

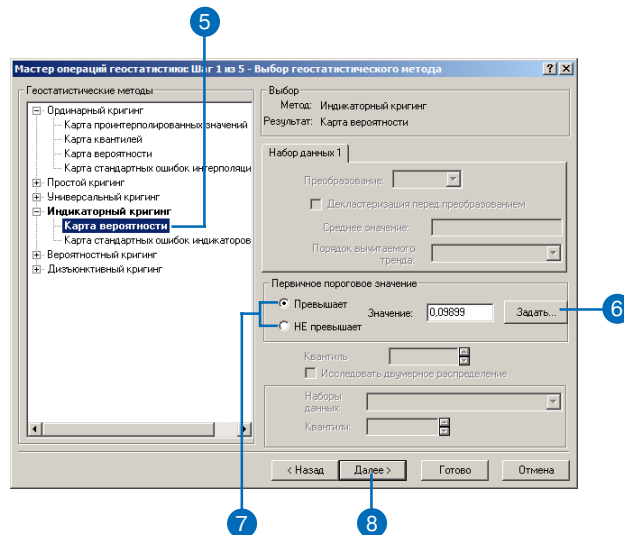
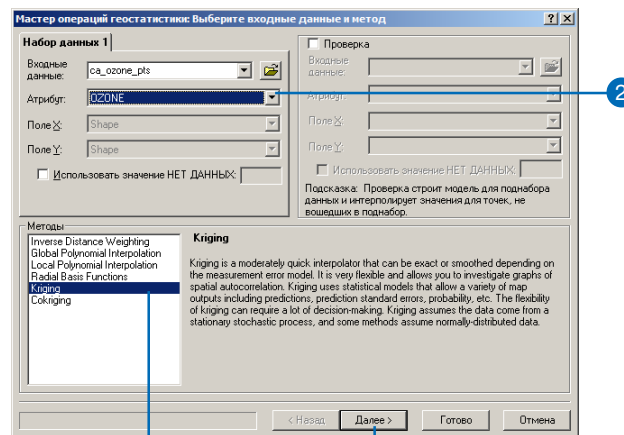
*Оценка вариограммы/ковариации затруднительна, когда индикаторные переменные изначально все равны нулю или единице. Выбирайте пороговое значение, которое позволит создать набор индикаторных значений, в котором будут и нули, и единицы.*

### См. также

*Для дополнительной информации о преобразованиях и определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.*

## Создание карты вероятности

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу индикаторного кригинга. Запустите модуль Geostatistical Analyst.
2. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу индикаторного кригинга.
3. Из списка методов выберите Кригинг.
4. Нажмите Далее.
5. Откройте список для строки Индикаторный кригинг и выберите опцию Карта вероятности.
6. В диалоге Первичное пороговое значение введите пороговое значение или используйте для этого кнопку Задать...
7. Отметьте опцию Превышает или Не превышает.
8. Нажмите Далее.
9. В диалоге Дополнительные отсекатели определите дополнительные отсекатели.
10. Задайте необходимые параметры в диалогах Моделирование вариограммы/ковариации и Поиск соседства и в каждом из них нажмите Далее.
11. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
12. В диалоге с информацией о результирующем слое нажмите ОК.





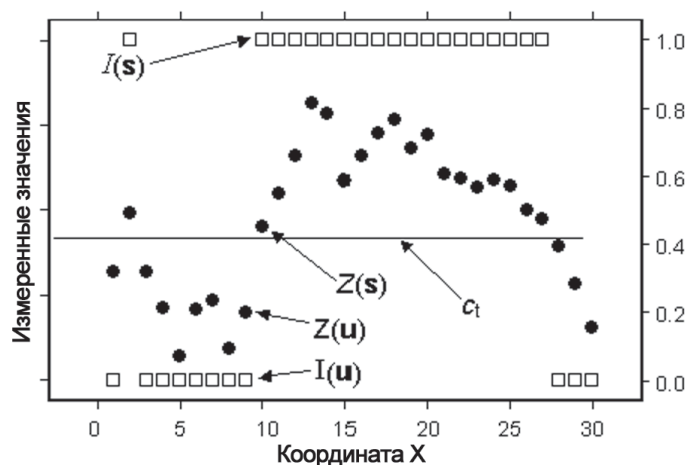
## Что такое вероятностный кригинг?

Вероятностный кригинг предполагает модель,

$$I(s) = I(Z(s) > c_1) = m_1 + e_1(s)$$

$$Z(s) = m_2 + e_2(s),$$

где  $m_1$  и  $m_2$  - неизвестные константы, а  $I(s)$  - бинарная переменная, созданная с использованием порогового индикатора  $I(Z(s) > c_1)$ . Обратите внимание, что теперь существует два типа случайных ошибок,  $e_1(s)$  и  $e_2(s)$ , и, следовательно, существует автокорреляция для каждой из них и взаимная корреляция между ними. Вероятностный кригинг пытается делать то же самое, что и индикаторный кригинг, но для того, чтобы выполнить работу лучше, он использует кокригинг. Например, на следующем рисунке, для которого использованы те же данные, что и для ординарного, универсального, простого и индикаторного кригинга, обратите внимание на точку  $Z(u=9)$ , индикаторная переменная которой равна  $I(u) = 0$ , и  $Z(s=10)$ , индикаторная переменная которой равна  $I(s) = 1$ .



Если бы вы захотели вычислить значение посередине между ними (для координаты  $x=9.5$ ), тогда использование только индикаторного кригинга дало бы значение около 0.5. Однако, вы можете заметить, что значение  $Z(s)$  едва превышает пороговое значение, а  $Z(u)$  - гораздо меньше него. Следовательно, у вас есть причины предположить, что индикаторное значение в точке 9.5 должно быть меньше 0.5. Вероятностный кригинг пытается учитывать дополнительную информацию в исходных данных, помимо бинарной переменной. Однако, это имеет свою цену. Вам необходимо выполнить гораздо больше оценок, включая оценку автокорреляции для каждой переменной, наряду с их взаимной корреляцией. Каждый раз, когда вы оцениваете неизвестные параметры автокорреляции, вы вносите большую неопределенность, поэтому, возможно, вероятностный кригинг не стоит затрачиваемых на него дополнительных усилий.

Вероятностный кригинг может использовать либо вариограммы, либо ковариационные функции (являющиеся математическим способом определения автокорреляции), в также взаимную ковариацию (являющуюся математическим способом определения взаимной корреляции); см. Главу 7, 'Использование аналитических инструментов для построения поверхностей', для более подробной информации.

## Создание карты с использованием вероятностного кригинга

Используйте вероятностный кригинг для создания карт вероятности и карт стандартных ошибок индикаторов. Данные, полученные в опорных точках, должны относиться к непрерывному явлению.

### Подсказка

#### Важные параметры

*Заданное пороговое значение (которое определяет, каким вычисленным значениям будет присвоен “0”, а каким “1”), модели ковариации/вариограммы, и область поиска соседства, используемая в модели.*

### Подсказка

#### Выбор порогового значения

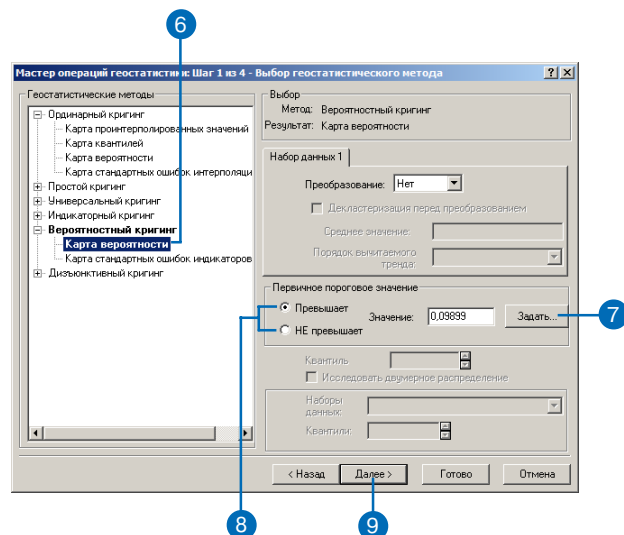
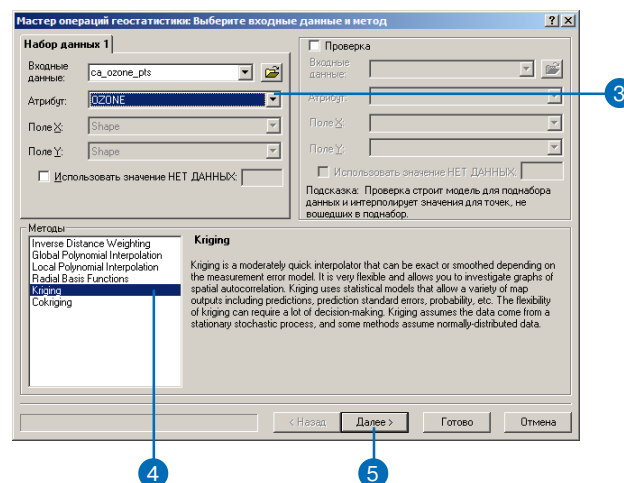
*Оценка вариограммы/ковариации затруднительна, когда индикаторные переменные изначально все равны нулю или единице. Если возможно, выбирайте пороговое значение, которое позволит создать набор индикаторных значений, в котором будут и нули, и единицы.*

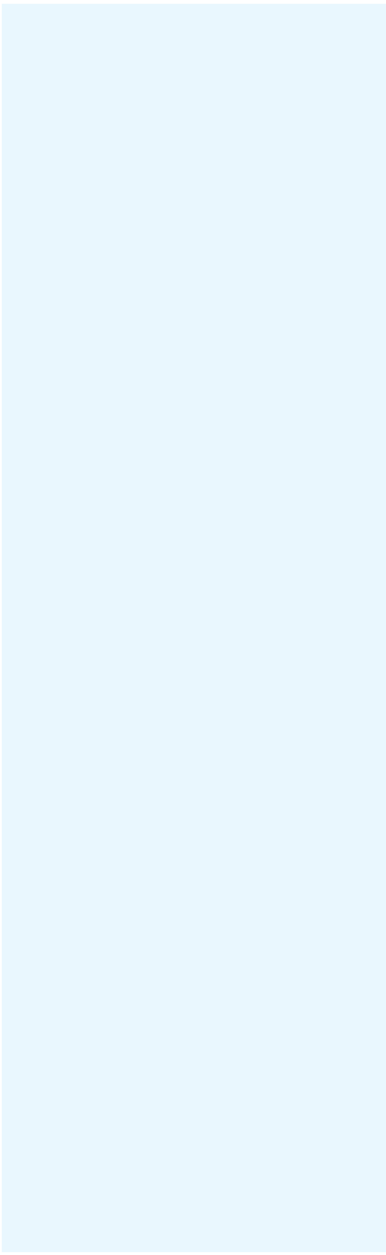
### См. также

*Для дополнительной информации о преобразованиях, вычислениях тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, ‘Использование аналитических инструментов при построении поверхностей’.*

## Создание карты вероятности

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу вероятностного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите поле атрибута, для которого вы хотите выполнить вероятностный кригинг.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Вероятностный кригинг и выберите опцию Карта вероятности.
7. В диалоге Первичное пороговое значение введите пороговое значение или используйте для этого кнопку Задать...
8. Отметьте опцию Превышает или Не превышает.
9. Нажмите Далее.
10. Задайте необходимые параметры в диалогах Моделирование вариограммы/ковариации и Поиск соседства и в каждом из них нажмите Далее.



- 
12. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
  13. В диалоге с информацией о результирующем слое нажмите ОК.

## Что такое дизъюнктивный кригинг?

Дизъюнктивный кригинг предполагает модель,

$$f(Z(s)) = m_1 + e(s)$$

где  $m_1$  - неизвестная константа и  $f(Z(s))$  - некая произвольная функция  $Z(s)$ . Обратите внимание, что вы можете записать выражение  $f(Z(s)) = I(Z(s) > c_1)$ , поэтому индикаторный кригинг является частным случаем дизъюнктивного кригинга. В модуле Geostatistical Analyst с помощью дизъюнктивного кригинга вы можете вычислить либо само проинтерполированное значение, либо его индикаторное значение.

Сравним ординарный и индикаторный кригинг с дизъюнктивным. Ординарный кригинг использует линейную комбинацию данных, и окончательная формула интерполятора выглядит следующим образом:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i),$$

где  $\hat{Z}(s_0)$  - проинтерполированное значение,  $\{Z(s_i)\}$  - исходные данные, а  $\{\lambda_i\}$  - веса кригинга. Задача ординарного кригинга - найти оптимальные веса,  $\{\lambda_i\}$ . Индикаторный кригинг образует интерполятор,

$$\hat{f}(Z(s_0) > c_t) = \sum_{i=1}^n \lambda_i I(Z(s_i) > c_t),$$

и, также пытается найти оптимальные веса,  $\{\lambda_i\}$ . Однако возможно, вам удастся найти большее количество общих функций для данных, которые помогут вам вычислить некую функцию переменной в искомой точке.

Дизъюнктивный кригинг генерализирует индикаторный кригинг, чтобы сформировать интерполятор,

$$\hat{g}(Z(s_0)) = \sum_{i=1}^n f_i(Z(s_i)),$$

В модуле Geostatistical Analyst доступные функции  $g(Z(s_0))$  - это просто сами значения  $Z(s_0)$  и  $I(Z(s_0) > c_t)$ . В целом, дизъюнктивный кригинг пытается сделать больше, чем ординарный кригинг. Поскольку награда может быть большей, то и затраты соответственно возрастают. Дизъюнктивный кригинг предполагает наличие двумерного нормального распределения данных (см. Главу 7, 'Использование аналитических инструментов при построении поверхностей') и аппроксимаций функций  $f_i(Z(s_i))$ ; эти ограничения трудно проверяемы, и поэтому решения являются математически и вычислительно сложными.

Дизъюнктивный кригинг может использовать либо вариограмму, либо ковариационные функции (которые являются математическим описанием автокорреляции), и он также может использовать преобразования данных и вычитание тренда; см. Главу 7, 'Использование аналитических инструментов при построении поверхностей', для более подробной информации.

## Создание карты с использованием дизъюнктивного кригинга

Используйте дизъюнктивный кригинг для создания карт проинтерполированных значений, вероятности и карт стандартных ошибок интерполяции. Данные, полученные в опорных точках, должны относиться к непрерывному явлению и подчиняться закону двумерного нормального распределения.

### Подсказка

#### Важные параметры

*Соответствующее преобразование и вычитание тренда, модели вариограммы/ковариации и поиск соседства.*

### Подсказка

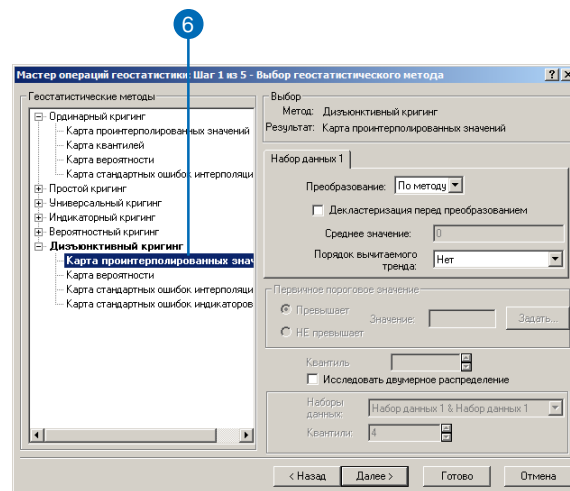
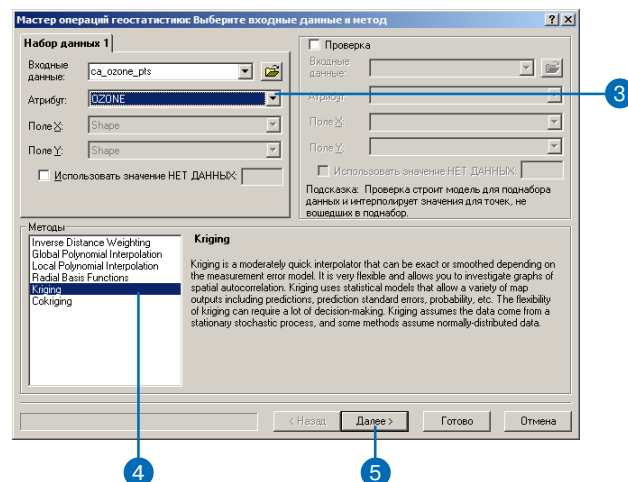
**Использование файла базы данных вместо точечного слоя**  
В диалоге Выберите входные данные и метод вместо слоя ArcMap может быть выбран файл базы данных. Для этого надо нажать кнопку Пролить и выбрать соответствующий файл из базы данных.

### См. также

Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу дизъюнктивного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. В окне Атрибут выберите показатель, для которого вы хотите выполнить интерполяцию по методу дизъюнктивного кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Дизъюнктивный кригинг и выберите опцию Карта проинтерполированных значений. Задайте Среднее значение. Помимо этого, задайте преобразование по методу нормальных меток и нажмите Далее.
7. Если указано преобразование по методу нормальных меток, определите требуемые параметры в диалоге Преобразование по методу нормальных меток и нажмите Далее.
8. В диалоге Моделирование вариограммы/ковариации задайте требуемые параметры и нажмите Далее.
9. В диалоге Поиск соседства задайте требуемые параметры поиска и нажмите Далее.
10. В диалоге Перекрестная проверка изучите полученные результаты и нажмите Готово.
11. В диалоге с информацией о результирующем слое нажмите OK.



## Подсказка

### Карта стандартных ошибок интерполяции

Карта стандартных ошибок количественно определяет неопределенность вычислений. Если данные подчиняются закону нормального распределения, истинное значение будет находиться в интервале, определяемом вычисленным значением  $\pm$  двукратная стандартная ошибка интерполяции в примерно 95 процентах случаев.

## См. также

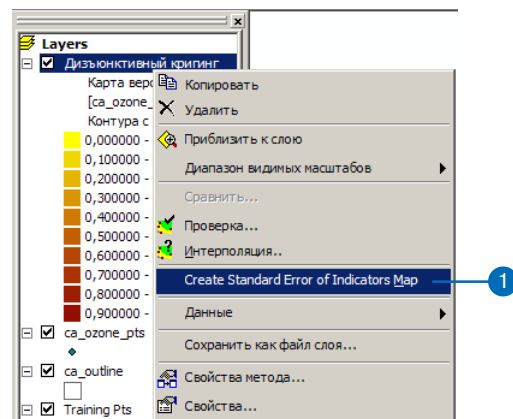
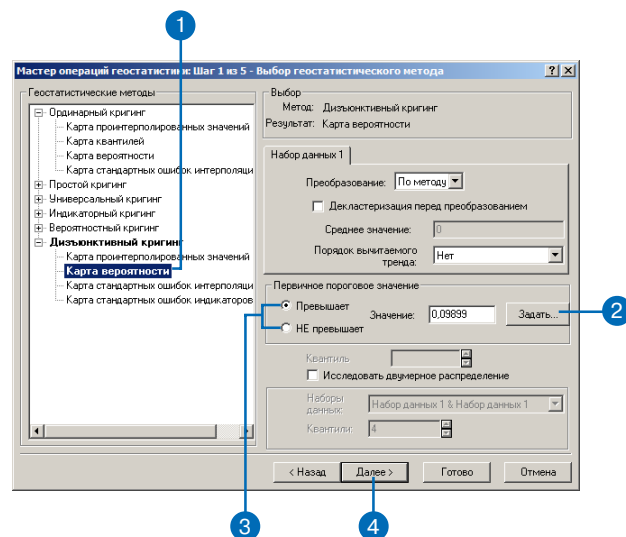
Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты вероятности

1. Повторите шаги упражнения 'Создание карты проинтерполированных значений', приведенного на предыдущей странице, за исключением того, что в шаге 6 вместо опции Карта проинтерполированных значений выберите опцию Карта вероятности.
2. В диалоге Первичное пороговое значение введите пороговое значение или используйте для этого кнопку Задать...
3. Отметьте опцию Превышает или Не превышает.
4. Определите Среднее значение и задайте преобразование по методу нормальных меток. Нажмите Далее.
5. Повторите шаги с 7 по 11 упражнения 'Создание карты проинтерполированных значений', приведенного ранее в этой главе.

## Создание карты стандартных ошибок интерполяции

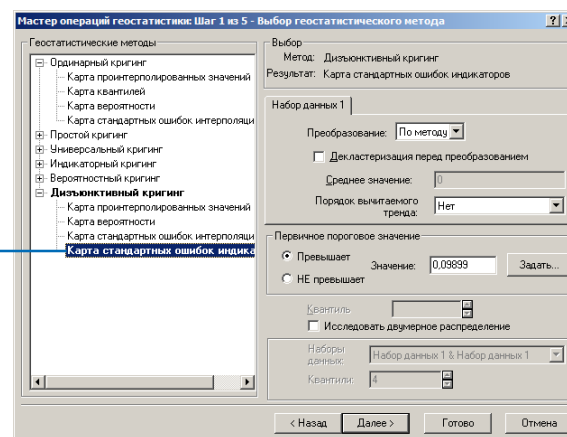
1. В таблице содержания ArcMap нажмите правую клавишу мыши на поверхности проинтерполированных значений, построенной с использованием дизъюнктивного кригинга, и из контекстного меню выберите опцию Построить карту стандартных ошибок интерполяции.



## Создание карты стандартной ошибки индикаторов

1. Повторите шаги упражнения ‘Создание карты вероятности’, приведенного ранее в этой главе, за исключением того, что в шаге 6 вместо опции Карта проинтерполированных значений выберите опцию Карта стандартной ошибки индикаторов.

1





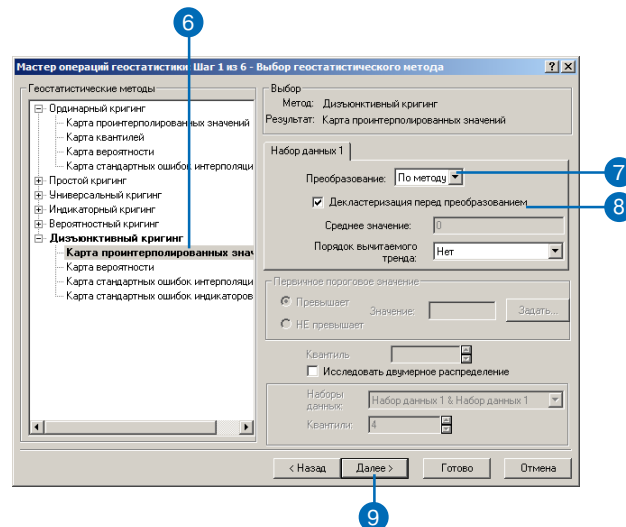
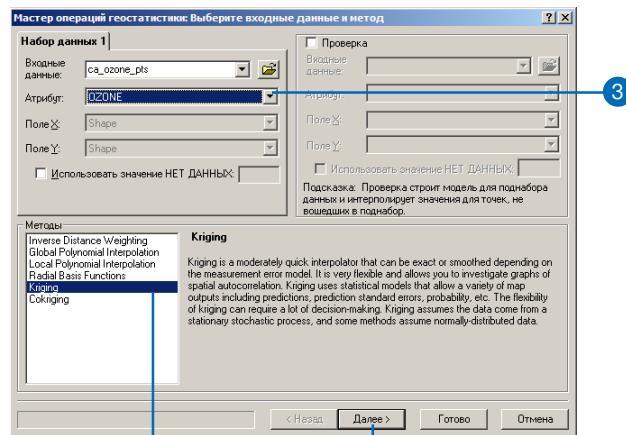
## Подсказка

### Использование декластеризации

Если вы используете преобразование по методу нормальных меток, и ваши данные были изначально отобраны с различной плотностью для различных участков территории, попробуйте выполнить декластеризацию данных. См. Главу 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений с применением декластеризации

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу дизъюнктивного кригинга.
2. Запустите модуль Geostatistical Analyst.
3. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу дизъюнктивного кригинга.
4. Из списка методов выберите Кригинг.
5. Нажмите Далее.
6. Откройте список для строки Дизъюнктивный кригинг и выберите опцию Карта проинтерполированных значений.
7. В окне Преобразование выберите опцию По методу нормальных меток.
8. Отметьте галочкой опцию Декластеризация перед преобразованием.
9. Нажмите Далее.
10. В диалоге Декластеризация задайте требуемые параметры и нажмите Далее.
11. Задайте требуемые параметры в диалоге Преобразование по методу нормальных меток и нажмите Далее.
12. Повторите шаги с 8 по 11 упражнения 'Создание карты проинтерполированных значений', приведенного ранее в этой главе.



## Подсказка

### Изучение двумерного распределения

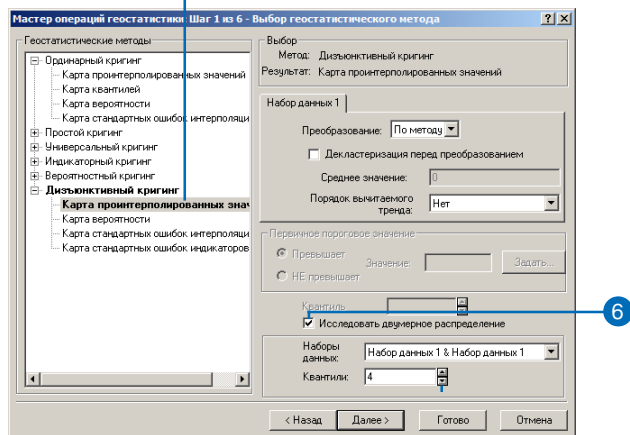
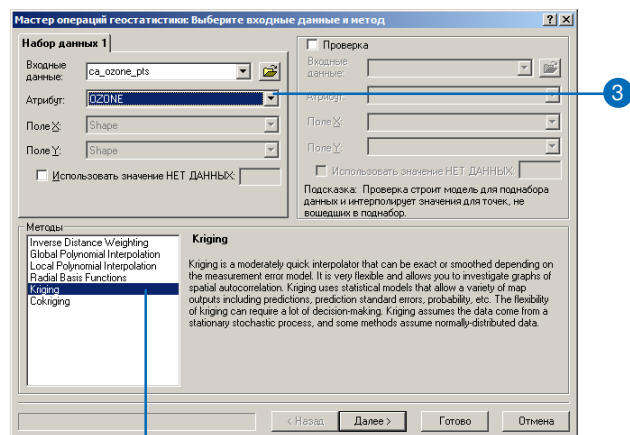
Проверяйте свои данные на двумерную нормальность. См. Главу 7, 'Использование аналитических инструментов при построении поверхностей'.

## См. также

Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Изучение двумерного распределения при создании карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить интерполяцию по методу дизъюнктивного кригинга.
2. На панели Geostatistical Analyst выберите опцию Мастер операций геостатистики.
3. Выберите атрибут, для которого вы хотите выполнить интерполяцию по методу дизъюнктивного кригинга.
4. Из списка методов выберите Кригинг. Нажмите Далее.
5. Откройте список для строки Дизъюнктивный кригинг и выберите опцию Карта проинтерполированных значений.
6. Отметьте галочкой опцию Исследовать двумерное распределение и задайте Среднее значение или Преобразование по методу нормальных меток.
7. Задайте необходимые параметры в диалоге Моделирование вариограммы/ковариации и нажмите Далее.
8. Изучите диалог Исследовать двумерное распределение и нажмите Далее.
9. Повторите шаги с 9 по 11 упражнения 'Создание карты проинтерполированных значений', приведенного ранее в этой главе.



## Что такое кокригинг?

Кокригинг использует информацию для нескольких типов переменных. Для нахождения более точных искоемых значений применяется первая рассматриваемая переменная -  $Z_1$ , наряду с автокорреляцией для значений  $Z_1$  и взаимной корреляцией между значениями  $Z_1$  и всеми другими типами переменных. Кокригинг обращается к информации, заключенной в других переменных, чтобы помочь в нахождении искоемых значений, но столь сложный процесс имеет свою цену. Кокригинг требует намного большего количества оценок, которые включают как оценку автокорреляции для каждой переменной, так и взаимной корреляции для всех переменных. Теоретически, вы не можете ухудшить значения, полученные с помощью кригинга, поскольку если в данных нет взаимной корреляции между переменными, вы можете вернуться к автокорреляции между значениями  $Z_1$ . Однако, каждый раз, когда вы оцениваете неизвестные параметры автокорреляции, вы вводите большую неопределенность, поэтому цели повышения точности могут не оправдать затрачиваемых на них дополнительных усилий.

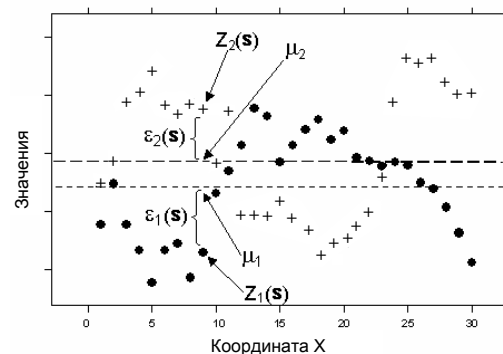
Ординарный кокригинг предполагает модели,

$$Z_1(s) = m_1 + e_1(s)$$

$$Z_2(s) = m_2 + e_2(s),$$

где  $m_1$  и  $m_2$  - неизвестные константы. Обратите внимание, что теперь у вас есть два типа случайных ошибок,  $e_1(s)$  и  $e_2(s)$ , и следовательно, для каждой из них существует автокорреляция и взаимная корреляция между ними. Ординарный кокригинг пытается вычислить значения  $Z_1(s_0)$ , так же, как и ординарный кригинг, но в своей попытке выполнить работу лучше, он использует информацию ковариаты  $\{Z_2(s)\}$ . Например, на рисунке показаны те же данные, что были использованы для примера с ординарным кригингом, но здесь в них добавлена вторая переменная.

Обратите внимание, что значения  $Z_1$  и  $Z_2$  автокоррелируют. Также заметьте, что когда  $Z_1$  ниже среднего  $m_1$ ,  $Z_2$  часто превышает свое среднее  $m_2$ , и наоборот. Таким образом,  $Z_1$  и  $Z_2$  обладают отрицательной взаимной корреляцией. В этом примере, каждая точка  $s$  имеет значения  $Z_1(s)$  и  $Z_2(s)$ ; однако это условие является необязательным, и часто каждому типу переменных может соответствовать свой набор данных. Главная интересующая нас переменная - это  $Z_1$ , и для



более точного вычисления искоемых значений учитывается и автокорреляция, и взаимная корреляция.

Другие методы кокригинга, включая методы универсального, простого, индикаторного, вероятностного и дизъюнктивного кокригинга, - это все обобщение соответствующих методов кригинга для случаев, когда вы работаете с несколькими наборами данных. Например, индикаторный кокригинг может быть осуществлен путем использования нескольких пороговых значений для ваших данных. А затем, для того, чтобы вычислить пороговое значение, интересующее вас в первую очередь, можно воспользоваться бинарными данными. При таком способе, индикаторный кокригинг будет аналогичен вероятностному кригингу, но он может быть более чувствителен к экстремальным значениям и другим непонятным (с возможными ошибками) данным.

Кокригинг может использовать либо вариограммы, либо ковариационные функции (которые являются математическим описанием автокорреляции) и взаимную ковариацию (которая является математической формой выражения взаимной корреляции), он может использовать преобразования и вычитание тренда, и он может также допустить наличие ошибок в измерениях в тех же ситуациях, что и различные методы кригинга (ординарный, простой и универсальный кригинг); см. Главу 7, 'Использование аналитических инструментов при построении поверхностей', для более подробной информации.

## Создание карты с использованием кокригинга

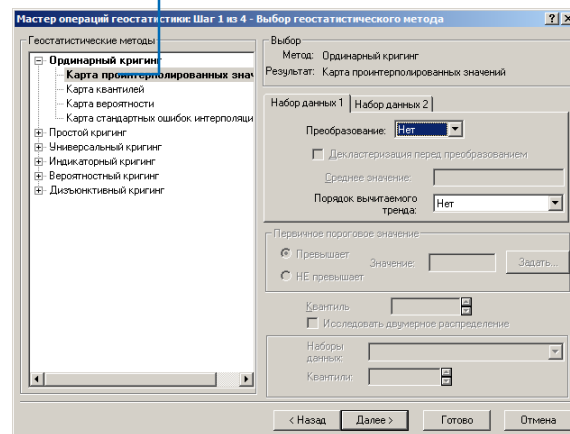
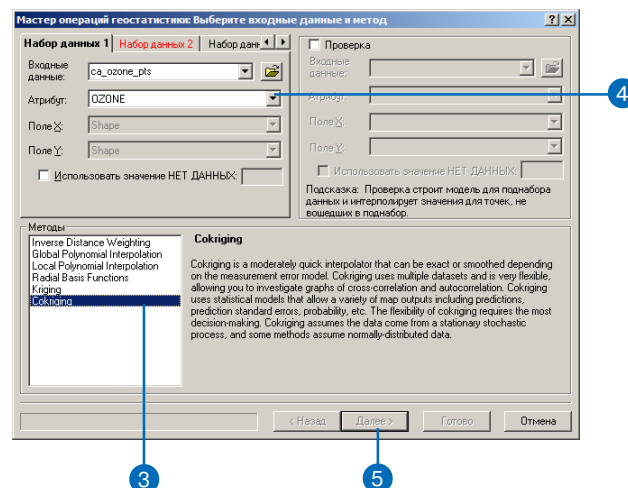
Используйте метод кокригинга для создания карт проинтерполированных значений, вероятности, квантилей, стандартной ошибки индикаторов и стандартной ошибки интерполяции, при сохранении тех же условий, что и для каждого из методов кригинга из обсуждавшихся в этой главе.

### См. также

Для дополнительной информации о преобразованиях, вычитании тренда, определении параметров в диалогах Моделирование вариограммы/ковариации и Поиск соседства, а также для изучения диалога Перекрестная проверка, обратитесь к Главе 7, 'Использование аналитических инструментов при построении поверхностей'.

## Создание карты проинтерполированных значений

1. В таблице содержания ArcMap выберите точечный слой, для которого вы хотите выполнить кокригинг.
2. На панели Geostatistical Analyst выберите опцию Мастер операций геостатистики.
3. Из списка методов выберите Кокригинг.
4. Выберите атрибут, для которого вы хотите выполнить кокригинг для всех наборов данных (переходите от закладки одного набора данных к закладке другого набора данных, чтобы задать параметры).
5. Нажмите Далее.
6. Выберите требуемый тип кокригинга и тип результирующей карты в диалоге Выбор геостатистического метода. Нажмите Далее.
7. В диалоге Вычитание тренда определите параметры, если для всех наборов данных был задан порядок полинома для вычитаемого тренда. Нажмите Далее.
8. В диалоге Моделирование вариограммы/ковариации задайте требуемые параметры для всех наборов данных. Нажмите Далее.
9. В диалоге Поиск соседства задайте требуемые параметры поиска и нажмите Далее.
10. В диалоге Перекрестная проверка изучите полученные результаты и нажмите Готово.
11. В диалоге с информацией о результирующем слое нажмите ОК.





# Использование аналитических инструментов при построении поверхностей

# 7

## В ЭТОЙ ГЛАВЕ

- Исследование пространственной структуры: вариография
- Определение области поиска соседства
- Выполнение перекрестной проверки для оценки выбранных параметров
- Оценка протокола решений с использованием проверки
- Сравнение моделей
- Моделирование распределений и определение преобразований
- Проверка данных на соответствие двумерному нормальному распределению
- Применение декластеризации для определения выборки
- Вычитание трендов из данных

При построении поверхности вы проходите через многие стадии. На каждой из этих стадий вы определяете целый ряд параметров. Модуль Geostatistical Analyst предоставляет серии диалогов-подсказок, содержащих аналитические инструменты, которые призваны помочь вам в определении этих параметров. Некоторые из этих диалогов и инструментов, например, определение области поиска опорных точек для вычислений, перекрестная проверка и проверка, применимы практически к всем методам интерполяции. Другие, например, моделирование вариограмм, преобразования, вычитание тренда, декластеризация и проверка данных на соответствие двумерному нормальному распределению, применяются только для геостатистических методов (кригинга и кокригинга).

В каждый диалог-помощник включены серии задач, которые могут быть выполнены с использованием этих инструментов. В этой главе обсуждаются понятия, используемые для наиболее часто выполняемых задач, а также стадии их выполнения. Необходимо ли использовать эти инструменты все, или следует применить только некоторые из них, зависит от ваших данных. Для всех параметров модуль Geostatistical Analyst предоставляет надежные значения по умолчанию, некоторые из которых рассчитываются на основе ваших конкретных данных. Таким образом, вы можете получить расширенное представление об изучаемом вами явлении, отличное от первоначального и основывающееся на данных, полученных с помощью инструментов исследовательского анализа модуля Geostatistical Analyst. Эти представления вы можете использовать для уточнения параметров, что поможет вам построить более точную поверхность.

# Исследование пространственной структуры: вариография

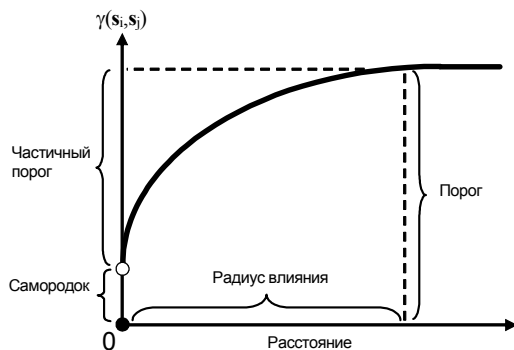
## Вариограммы и ковариационные функции

Функции вариограммы и ковариации (известные также как структурные функции) количественно характеризуют предположение о том, что чем ближе точки расположены друг к другу, тем они более похожи. И вариограмма, и ковариация определяют меру статистической корреляции как функцию расстояния. Вариограмма определяется следующим образом,

$$g(s_i, s_j) = S \text{ var}(Z(s_i) - Z(s_j)),$$

где *var* - дисперсия.

Если две точки  $s_i$  и  $s_j$  расположены близко друг к другу, и эта близость определяется расстоянием  $d(s_i, s_j)$ , ожидается, что эти точки будут похожи, и разность их значений,  $Z(s_i) - Z(s_j)$ , будет маленькой. По мере удаления друг от друга точек  $s_i$  и  $s_j$ , они становятся все менее похожи друг на друга, и следовательно разность их значений,  $Z(s_i) - Z(s_j)$ , будет расти. Это можно увидеть на следующем рисунке, который показывает анатомию типичной вариограммы.



Обратите внимание, что дисперсия разности возрастает с расстоянием, поэтому вариограмму можно рассматривать как функцию различия. Существует несколько терминов, часто ассоциирующихся с этой функцией, и они также используются в модуле Geostatistical Analyst.

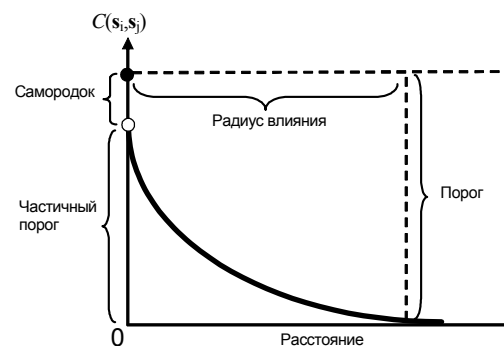
Значение (высота графика), начиная с которого вариограмма выравнивается, носит название порога. Оно часто состоит из двух частей: разрыва в начале графика, который носит название эффекта самородка, и частичного порога, которые вместе и образуют порог. Эффект самородка далее может быть разделен на ошибку измерений и вариацию на микроуровне, и в силу того, что любой из этих компонентов может быть равен нулю, эффект самородка может быть отнесен целиком за счет одного из них. Расстояние, при котором вариограмма выравнивается до значения порога, носит название радиуса влияния.

Ковариационная функция определяется следующим образом:

$$C(s_i, s_j) = \text{cov}(Z(s_i), Z(s_j)),$$

где *cov* - ковариация.

Ковариация - это масштабируемая версия корреляции. Следовательно, когда две точки,  $s_i$  и  $s_j$ , расположены поблизости, и предполагается, что они имеют сходные значения, значения их ковариации (корреляции) будут большими. По мере того, как  $s_i$  и  $s_j$  удаляются друг от друга, они становятся все менее похожи, и их ковариация стремится к нулю. Это можно увидеть на следующем рисунке, который показывает анатомию типичной ковариационной функции.



Обратите внимание, что функция ковариации уменьшается с расстоянием, поэтому ее можно рассматривать как функцию схождения.

Между вариограммой и ковариационной функцией существует связь, которая может быть выражена, как

$$g(s_i, s_j) = \text{порог} - C(s_i, s_j),$$

и эту связь можно увидеть по рисункам. Поскольку эти функции равноценны, в модуле Geostatistical Analyst при выполнении интерполяции вы можете использовать любую из них. (Напомним, что все вариограммы в модуле Geostatistical Analyst имеют значение порога.)

Вариограммы и ковариации не могут быть просто какими-то функциями. Для того, чтобы интерполяция имела неотрицательные стандартные ошибки кригинга, только некоторые функции могут быть использованы в качестве ковариационной или функции вариограммы. Модуль Geostatistical Analyst предлагает несколько функций, которые вы можете применить для анализа данных. Вы можете также воспользоваться моделями, составленными из нескольких—такая конструкция часто дает более достоверные модели (в модуле Geostatistical Analyst возможно составить комбинацию из не более, чем четырех функций). В некоторых случаях бывает так, что вариограммы есть, а ковариационных функций нет. Например, существует линейная вариограмма, но у нее нет порога, и, следовательно, нет соответствующей ковариационной функции. В модуле Geostatistical Analyst используются только модели с порогами. Не существует правил, которые без труда и быстро позволили бы вам выбрать “лучшую” модель вариограммы. Чаще всего вы подбираете модель для эмпирической вариограммы или ковариационной функции визуально. В качестве помощников вы можете использовать проверку или перекрестную проверку.

## Что такое ошибка измерений?

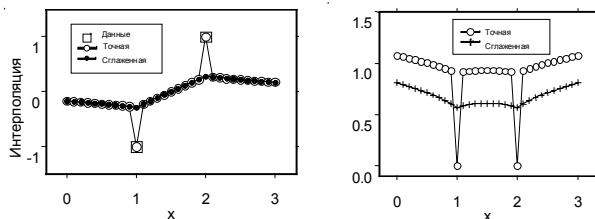
Три метода кригинга—ординарный, простой и универсальный—позволяют использовать модели ошибок измерения. Ошибка измерения может быть получена, когда в одной и той же точке выполнено несколько наблюдений, и их значения отличаются друг от друга. Например, вы можете получить образец в опорной точке с пробами грунта или воздуха, а затем разделить этот образец на несколько проб. Потребность в этом может возникнуть, если используемый вами инструмент для выполнения анализа дает некую погрешность. Другой пример: вы можете направить почву, разделив пробу, на анализ в различные лаборатории. В некоторых случаях, для прибора могут быть задокументированы отклонения в точности измерений. В таком случае, вы можете просто включить известные отклонения в измерениях в вашу модель. Модель ошибки измерения выглядит следующим образом:

$$Z(s) = m(s) + e(s) + d(s),$$

где  $d(s)$  - где ошибка измерений,  $m(s)$  и  $e(s)$  - аналогичны значениям, используемым в модели кригинга, приведенной в Главе 6. В этой модели, эффект самородка состоит из дисперсии  $e(s)$  (которая носит название вариации на микроуровне) и дисперсии  $d(s)$  (которая носит название ошибки измерений). В модуле Geostatistical Analyst вы можете определить, в какой пропорции в эффекте самородка присутствуют вариация на микроуровне и вариация, вызванная ошибкой измерений; либо вы можете использовать модуль Geostatistical Analyst для оценки ошибки измерений, если у вас есть множественные измерения для каждой точки; либо вы можете ввести известное значение погрешности (вариации), вызванной измерениями. По умолчанию предлагается отсутствие (ноль) погрешности, вызванной измерениями. Когда в данных нет ошибки измерений, кригинг является жестким интерполятором, что означает, что если выполнить интерполяцию для точки, в которой были отобраны данные, вычисленное значение совпадет с измеренным. Однако,



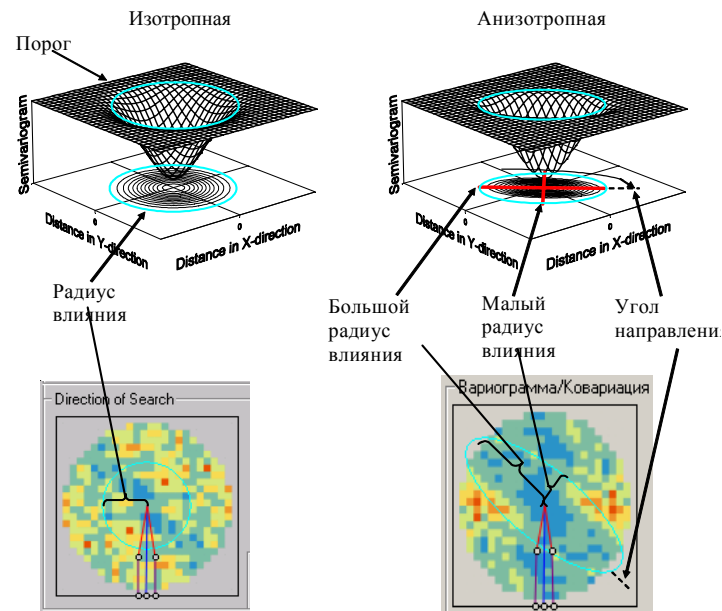
если существуют ошибки измерений, вы должны учитывать, что проинтерполированное значение будет фильтрованным,  $m(s_0) + e(s_0)$ , с невыраженной ошибкой измерения. В точках, где отбирались данные, фильтрованное значение не равно измеренному значению. Эффект от выбора моделей ошибки измерений состоит в том, что ваша окончательная карта может быть более сглаженной и иметь меньшие стандартные ошибки, чем при использовании версии жесткого кригинга. Это проиллюстрировано с помощью примера на нижнем рисунке; жесткий кригинг и сглаженный кригинг показаны для случая, когда есть только две опорные точки (1 и 2) со значениями -1 и 1 для модели, построенной без учета изменчивости, вызванной измерениями, и для модели, в которой эффект самородка полностью состоит из вариации, вызванной измерениями.



## Анизотропия: Вариограммы и ковариационные функции по направлениям

Поскольку вы работаете в двумерном пространстве, вы можете предположить, что вариограмма и ковариационная функция меняются не только с расстоянием, но и с направлением. Такое свойство носит название анизотропии. Рассмотрим две точки,  $s_i$  и  $s_j$ , и вектор между ними, который обозначим как  $s_i - s_j$ . Этот вектор будет иметь значение расстояния не только по оси  $x$ , но и по оси  $y$ . Помимо этого, вы можете рассматривать вектор в полярных координатах, то есть учитывать его длину и угол направления. Здесь анизотропия описана для вариограммы; очевидно, что те же самые соображения применимы и для ковари-

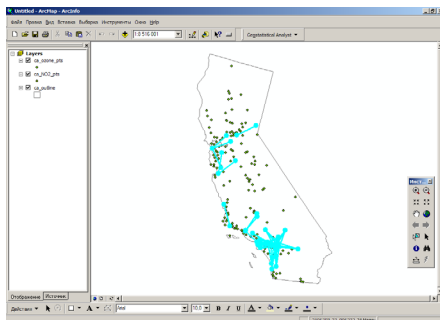
ационной функции. Вариограмма, отображенная на двумерной плоскости с осями координат, выглядит следующим образом:



Изотропная модель выглядит одинаково во всех направлениях, тогда как анизотропная модель достигает порога быстрее в одних направлениях, чем в других. Длина большой (длинной) оси, при которой достигается порог, носит название большого радиуса влияния, а длина короткой оси - малого радиуса влияния; также вам известен угол поворота отрезка, образующего большой радиус влияния. В модуле Geostatistical Analyst, контур области влияния показан голубой линией (окружность или контур эллипса) на поверхности эмпирической вариограммы.

## Эмпирические вариограммы и ковариационные функции

Вариограмма и функции ковариации - это теоретические величины, которые вы не можете наблюдать, поэтому оцениваете их, исходя из ваших данных, с использованием так называемых эмпирических вариограмм и эмпирических ковариационных функций. Часто, вы можете постигнуть величины, изучив то, каким образом они оцениваются. Сначала взгляните на эмпирическую вариограмму. Предположим, вы возьмете все пары точек, расположенных друг от друга на примерно одинаковом расстоянии и примерно в одном и том же направлении, как и пары точек, соединенных на следующем рисунке зелеными отрезками.



Для всех пар точек  $s_i$  и  $s_j$ , соединенных отрезками, вычислим

$$\text{среднее}[(z(s_i) - z(s_j))^2],$$

где  $z(s_i)$  - измеренное значение в точке  $s_i$ . Если все пары точек  $s_i$  и  $s_j$  расположены близко друг к другу, предполагается, что значения  $z(s_i)$  и  $z(s_j)$  будут иметь сходные значения, следовательно, когда вы найдете их разности и вычислите их квадрат, среднее значение квадратов разностей должно быть маленьким. По мере удаления  $s_i$  и  $s_j$ , предполагается, что их значения будут сильнее отличаться друг от друга, следовательно, когда вы найдете их разности и вычислите их квадрат, среднее значение квадратов разностей должно стать больше.

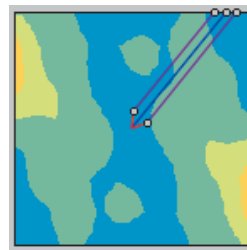
Посмотрите на ковариационную функцию. Для всех пар точек  $s_i$  и  $s_j$ , соединенных отрезками, вы вычисляете

$$\text{среднее}[(z(s_i) - \bar{z})(z(s_j) - \bar{z})],$$

где  $z(s_i)$  - измеренное значение в точке  $s_i$ , а  $\bar{z}$  - среднее из всех значений опорных точек. Если все пары  $s_i$  и  $s_j$  расположены близко друг к другу, предполагается, что либо оба значения  $z(s_i)$  и  $z(s_j)$  будут выше среднего  $\bar{z}$ , либо оба будут ниже среднего. В любом случае, результат выражения будет положительным, поэтому при нахождении среднего значения всех произведений вы получите положительное значение. Если же  $s_i$  и  $s_j$  удалены друг от друга, ожидается, что примерно в половине случаев произведения будут отрицательными, а в половине случаев - положительными, следовательно, их среднее будет стремиться к нулю.

В модуле Geostatistical Analyst средние значения, вычисленные по формуле, приведенной выше, для всех пар точек, удаленных на одинаковое расстояние и в одинаковом направлении, наносятся на поверхность вариограммы или ковариации. Например, здесь приведена поверхность эмпирической вариограммы.

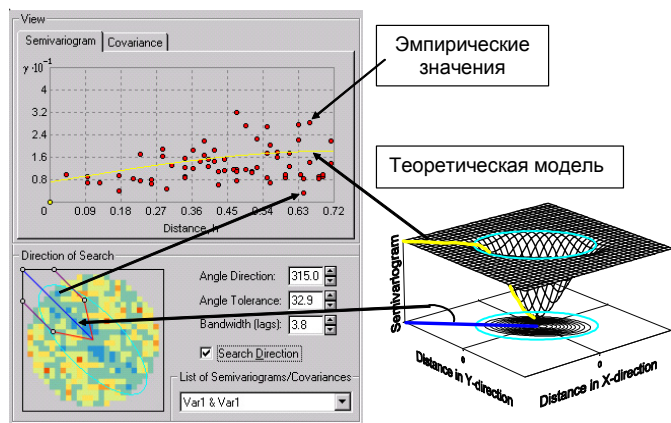
Размер ячеек носит название размера лага, а количество ячеек носит название количества лагов, и их можно задать в модуле Geostatistical



Analyst. Количество лагов в данном примере равно 12, и подсчитывается как число соседних ячеек по прямой горизонтальной или вертикальной линии от центра до края изображения поверхности.

## Использование эмпирических данных для оценки теоретических моделей

Теперь для оценки теоретических моделей, которые и будут в действительности использованы для интерполяторов по методу кригинга и вычисления стандартных ошибок, вам необходимо воспользоваться вариограммами и ковариационными функциями. На следующем рисунке показаны и оцененная теоретическая модель и эмпирические значения.

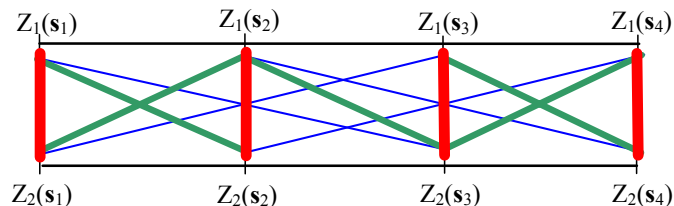


## Модели взаимной ковариации

Если у вас есть несколько наборов данных, и вы хотите использовать кокригинг, тогда вам необходимо разработать модели взаимной ковариации. Поскольку у вас есть несколько наборов данных, вы храните последовательность (ряд) переменных с индексами, где  $Z_k(s_j)$  показывает случайную переменную из набора данных  $k$  в точке  $s_j$ . Тогда функция взаимной ковариации между типами данных  $k$  и  $m$  будет определяться как,

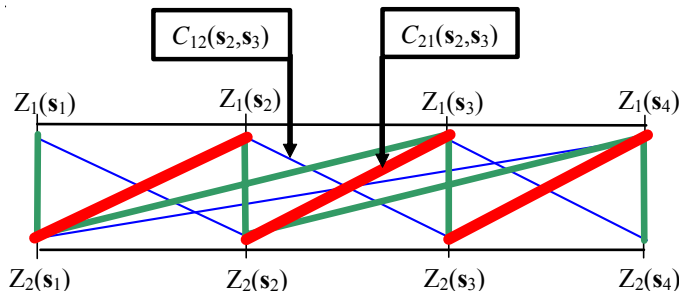
$$C_{km}(s_i, s_j) = \text{cov}(Z_k(s_i), Z_m(s_j)).$$

Перед нами трудно уловимый и часто сбивающий с толку факт:  $C_{km}(s_i, s_j)$  может быть асимметричной. В целом,  $C_{km}(s_i, s_j) \neq C_{mk}(s_i, s_j)$  (обратите внимание на переменную мест в индексах). Чтобы понять, почему это происходит, обратимся к следующему примеру.



Предположим, что у вас есть данные, расположенные в одном измерении, вдоль линии, и они выглядят следующим образом:

В этом примере, переменные для типа 1 и 2 расположены вдоль линии равномерно, на одинаковом расстоянии; толстой красной линией показано самое высокое значение взаимной ковариации, зеленой линией - меньшие значения взаимной ковариации, а тонкой синей линией - наименьшие значения взаимной ковариации; если линия отсутствует, это означает нулевую взаимную ковариацию. На рисунке видно, что  $Z_1(s_i)$  и  $Z_2(s_j)$  имеют максимальное значение взаимной ковариации, когда  $s_i = s_j$ , и это значение уменьшается по мере удаления  $s_i$  и  $s_j$  друг от друга. В этом примере,  $C_{km}(s_i, s_j) = C_{mk}(s_i, s_j)$ . Однако, взаимная ковариация может быть “смещена”:



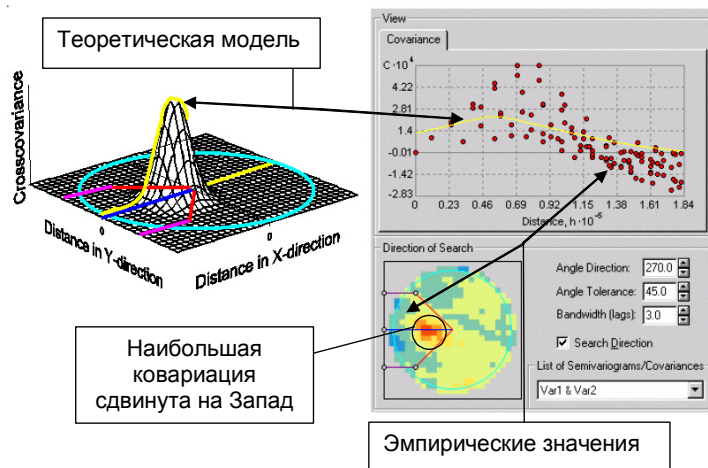
Теперь, к примеру, обратите внимание, что  $C_{12}(s_2, s_3)$  имеет минимальное значение взаимной ковариации (тонкая синяя линия), в то время, как  $C_{21}(s_2, s_3)$  имеет максимальное значение взаимной ковариации (толстая красная линия), поэтому здесь  $C_{km}(s_i, s_j) \neq C_{mk}(s_i, s_j)$ . Относительно  $Z_1$ , взаимные ковариации  $Z_2$  были смещены на -1. Если вы выберете опцию отображения параметров смещения, модуль Geostatistical Analyst оценит любое смещение во взаимной ковариации между двумя наборами данных в двух измерениях.

Эмпирические взаимные ковариации рассчитываются по формуле:

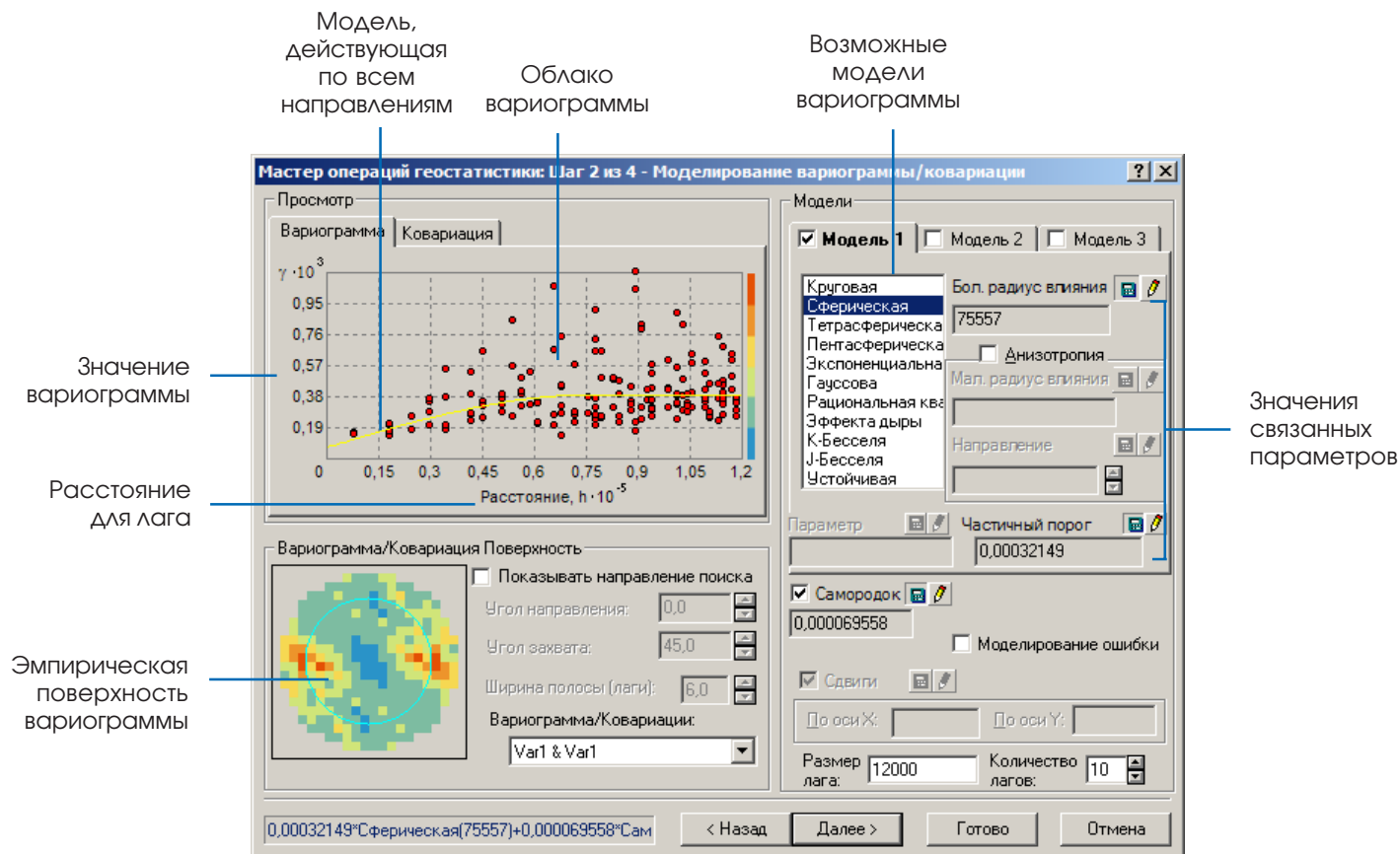
$$\text{среднее}[(z_1(s_i) - \bar{z}_1)(z_2(s_j) - \bar{z}_2)],$$

где  $Z_k(s_i)$  - измеренное значение для набора данных  $k$  в точке  $s_i$ ,  $\bar{z}_k$  - среднее для набора данных  $k$ ; среднее взято для всех  $s_i$  и  $s_j$ , удаленных друг от друга в определенном направлении на определенное расстояние. Что касается вариограмм, модуль Geostatistical Analyst показывает и эмпирическую, и подобранную модель для взаимной ковариации.

Выбор различных моделей ковариации, использование составных моделей ковариации, и выбор анизотропии, приведет к изменению теоретической модели. Вы можете сделать предварительный выбор модели, визуально определив, насколько хорошо она описывает эмпирические значения. Изменение размера лага и количества лагов и добавление смещений изменит эмпирическую поверхность ковариации, что приведет к соответствующим изменениям в теоретической модели. Модуль Geostatistical Analyst вычисляет значения по умолчанию. Однако, вы не должны ограничивать себя рамками этих параметров и воспользоваться возможностями по подбору различных значений. Для оценки достоверности результатов выполните также проверку и перекрестную проверку моделей и попробуйте найти научное обоснование для выбора наиболее подходящей модели.



## Диалог моделирования вариограммы/ковариации





## Подсказка

## Автокорреляция ваших данных

Рассмотрите положение точек вариограммы относительно желтой линии (модели). Если точки расположены близко к линии в одном направлении и разбросаны в другом, в ваших данных может быть автокорреляция в одном из направлений.

### Подсказка

## Это только визуальное исследование

Исследование направленной автокорреляции с использованием инструмента направления поиска наряду с указателем направления на поверхности вариограммы является исключительно визуальным (вокруг подобранной изотропной модели). Чтобы подобрать модель с учетом автокорреляции по направлению, отметьте галочкой опцию Анизотропия.

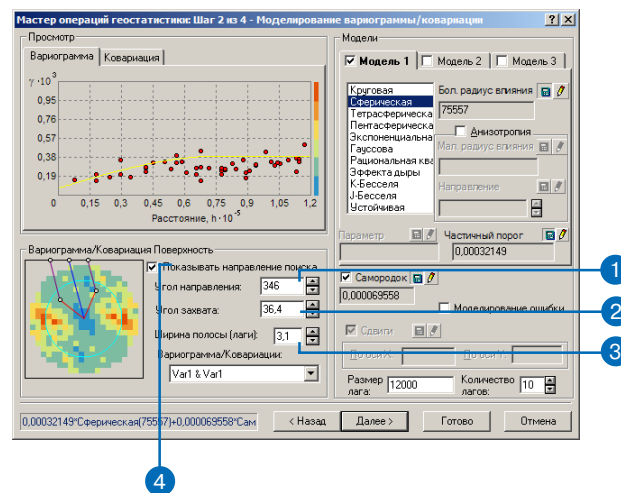
## Изучение автокорреляции по направлениям

1. Отметьте галочкой опцию Показывать направление поиска.
2. В окне Угол направления прокрутите значения стрелочками вверх-вниз, чтобы определить требуемый угол направления.

Или на карте вариограммы, выбрав мышью и удерживая среднюю синюю линию, меняйте направление указателя.

3. В окне Угол захвата прокрутите значения вверх-вниз, чтобы подобрать угловой допуск.  
Или на карте вариограммы, выберите и удерживайте любую из двух красных линий и растяните угол поиска по требуемой величины.
4. Поменяйте с помощью стрелок ширину полосы.

Или на карте вариограммы выберите и удерживайте любую из двух фиолетовых линий, ограничивающих указатель направления поиска. Увеличьте или уменьшите ширину полосы поиска.





## Подсказка

### Выбор опции анизотропии

Обратите внимание, что если выбрана опция анизотропии, вместо одной желтой линии будет отображаться несколько линий. Желтые линии показывают модели вариограммы для нескольких различных направлений. Модели (желтые линии) - теоретические "наилучшим образом подобранные" модели эмпирической вариограммы. Модуль Geostatistical Analyst автоматически вычисляет оптимальные параметры (например, большой радиус влияния, и угол направления) для учета анизотропного влияния.

## Подсказка

### Проверка вариограммы после того, как выбрана опция анизотропии

После выбора опции Анизотропия отметьте галочкой опцию Показывать направление поиска. Если модель подобрана правильно, желтая линия будет меняться по мере того, как в зависимости от угла направления поиска, меняется график эмпирической вариограммы.

## Подсказка

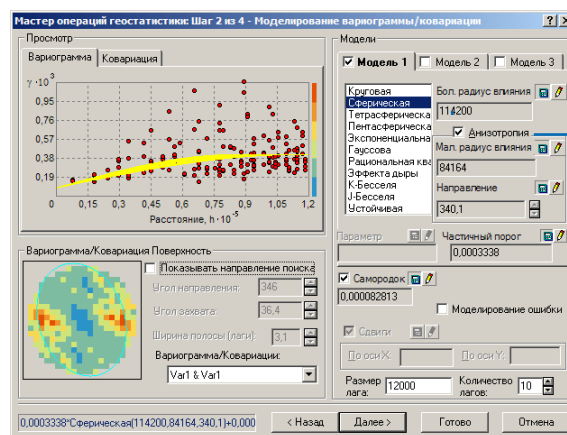
### Вычисление оптимальных параметров

Иконка с изображением карандаша, расположенная над окошками для ввода значений большого и малого радиусов влияния и угла направления, позволяет вам редактировать вводимые значения. Если вы нажмете на иконку калькулятора слева от карандаша, будут вычислены оптимальные значения, которые отобразятся в соответствующих окошках вводимых параметров.

## Моделирование анизотропии

1. Отметьте галочкой опцию Анизотропия в диалоге Моделирование вариограммы/ковариации.

Обратите внимание, что когда выбрана опция анизотропии, на графике вместо одной отображается несколько желтых линий.

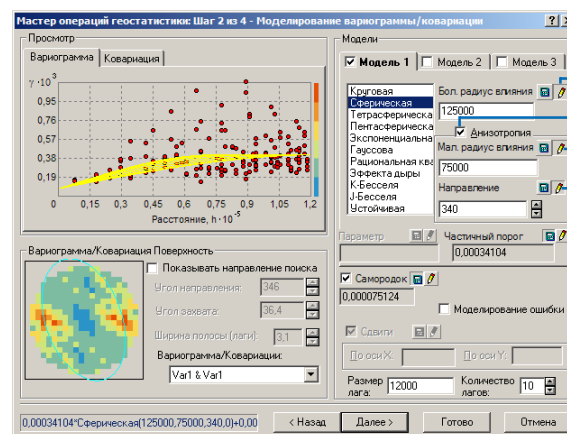


## Изменение параметров анизотропии

1. Отметьте опцию Анизотропия в диалоге Моделирование вариограммы/ковариации.

Примечание: Активируются опции Малый радиус влияния и Направление.

2. Чтобы изменить большой радиус влияния, нажмите на иконку с изображением карандаша над окошком, используемым для ввода значения большого радиуса влияния (что сделает окно ввода активным) и введите требуемое значение.
3. Чтобы изменить малый радиус влияния, нажмите на иконку с изображением карандаша, расположенную над окошком для ввода значения малого радиуса влияния (что позволит активировать окно ввода) и введите требуемое значение.



4. Чтобы изменить угол направления, нажмите на иконку с изображением карандаша, расположенную над окошком для ввода значения угла направления (что позволит активировать окно ввода) и введите требуемое значение.

## Подсказка

### Выбор размера лага

Правило большого пальца состоит в том, что размер лага, умноженный на количество лагов, должен быть меньше, чем половина максимального расстояния между точками для вашего набора данных.

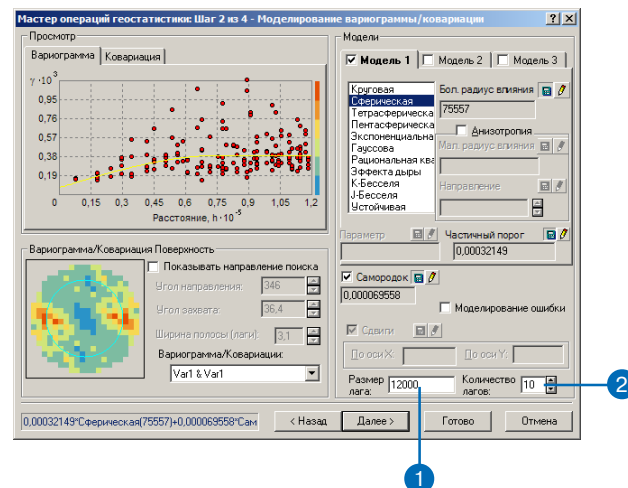
## Подсказка

### Вычисление оптимальных параметров

Иконка с изображением карандаша, расположенная над окошками для ввода значений частичного порога и самородка позволяет вам редактировать вводимые значения. Если вы нажмете на иконку калькулятора слева от карандаша, будут вычислены оптимальные значения, которые отобразятся в соответствующих окошках вводимых параметров.

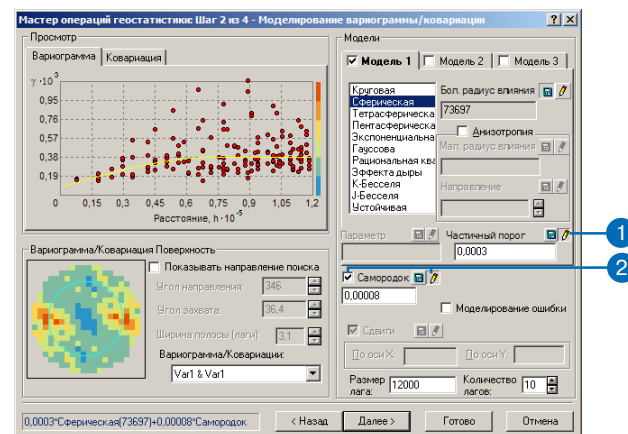
## Изменение размера лага и количества лагов

1. В диалоге Моделирование вариограммы/ковариации впечатайте требуемое значение размера лага.
2. Поменяйте значение количества лагов, уменьшая или увеличивая его с помощью прокрутки, или введите необходимое значение в окне Количество лагов.



## Изменение значений частичного порога и самородка

1. В диалоге Моделирование вариограммы/ковариации нажмите иконку с изображением карандаша, расположенную над окошком для ввода значения частичного порога (которая активирует окно ввода) и введите необходимое значение.
2. Отметьте опцию Самородок, нажмите иконку с изображением карандаша, расположенную над окошком для ввода значения самородка (которая активирует окно ввода) и введите необходимое значение.



## Подсказка

### Множественные измерения

Если у вас есть множественные измерения для опорных точек, позвольте модулю Geostatistical Analyst вычислить для вас ошибку измерений. Для этого отметьте опцию ОИ (Ошибка измерений) и нажмите кнопку Оценить.

## Подсказка

### Инструментальная ошибка

Для многих научных приборов может быть известна точность измерений. Основываясь на этой информации, вы можете ввести эти известные значения ошибки измерений.

## Подсказка

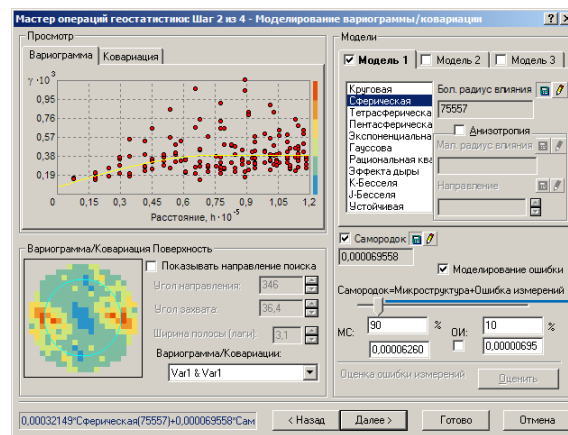
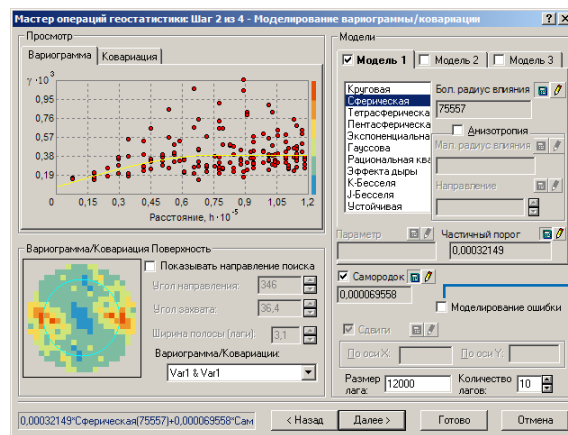
### Нулевая ошибка измерений

Если ошибка измерений равна нулю, вы получите “жесткий” кригинг.

## Работа с ошибкой измерений

1. Отметьте опцию Моделирование ошибки.
2. Перемещайте движок, чтобы определить процентное соотношение вариации на микроуровне (или микроструктуры) и ошибки измерений для значения эффекта самородка.

Или введите проценты или значения в окошках для ввода данных.



## Определение размера области поиска соседства

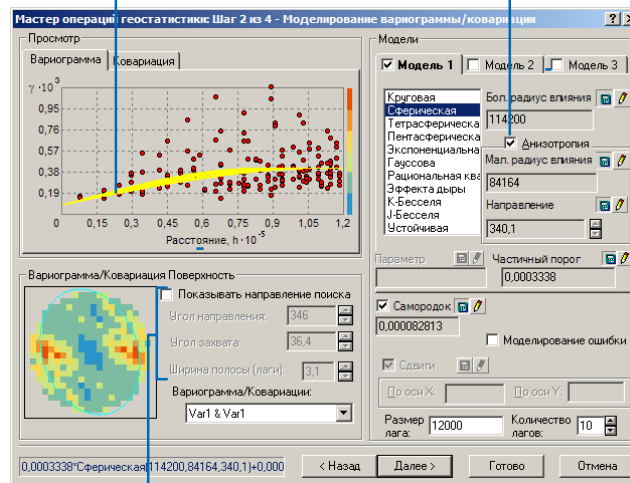
Объекты, расположенные ближе друг к другу, должны быть более похожими — по мере удаления опорных точек от точки с неизвестным значением, они могут оказаться не слишком полезными для вычисления значения в искомой точке. При некотором расстоянии опорные точки не будут коррелировать с искомой точкой, и возможно, что они будут даже расположены в области, которая сильно отличается от той, где находится искомая точка. Поэтому, общепринятым является определение области поиска соседства, ограничивающей количество и расположение точек, которые будут использованы при интерполяции. Существует два механизма контроля используемых точек, один из которых определяет форму области соседства, а другой устанавливает ограничения на выбор точек, попадающих в эту область.

Форма области будет диктоваться исходными данными. Например, если ваши данные отобраны равномерно, и нет автокорреляции, связанной с направлением (данные изотропны), вы захотите включить в вычисления точки, расположенные одинаково во всех направлениях от искомой. Для этого, вы, возможно, захотите определить область поиска соседства в форме круга. Однако, если вам известно, что существует автокорреляция, связанная с направлением (анизотропия), к примеру, вызванная ветровым переносом загрязнителей, при интерполяции поверхности в качестве области поиска соседства вы можете использовать эллипс, большая ось которого параллельна направлению ветра. Вы поступите таким образом, поскольку вам известно, что точки, расположенные по направлению ветра от искомой, даже на значительных расстояниях будут больше похожи на точку, для которой мы интерполируем значение, чем те, которые расположены в направлении, перпендикулярном направлению ветра.

Форма области поиска соседства должна определяться на основе понимания пространственного расположения опорных точек и пространственной автокорреляции набора данных. Изучение пространственного расположения опорных точек и автокорреляции выполняется при помощи инструментов ESDA (Исследовательского анализа пространственных данных), включенных в Мастер операций геостатистики. Например, на изображении диалога, приве-

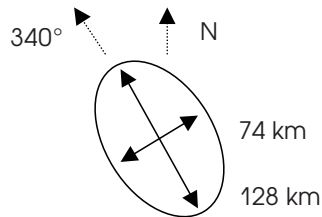
денном на этой странице, вы можете видеть, что автокорреляция, связанная с направлением (т.е., анизотропия), влияет на форму вариограммы. Кривая вариограммы медленно растет в направлении северо-северо-запад (примерно  $340^\circ$ ) и достигает значения, определяющего радиус влияния, примерно на расстоянии 114 км (самая нижняя желтая линия на вариограмме). В западно-юго-западном направлении, кривая вариограммы растет быстрее (самая верхняя желтая линия). Радиус влияния в этом направлении составляет примерно 84 км. Поскольку точки, отстоящие друг от друга на расстояние, превышающее радиус влияния, не коррелируют, информация может быть использована для определения стратегии поиска.

Отметьте опцию Анизотропия, чтобы показать конверт вариограммы (т.е., то как кривая модели вариограммы меняется в зависимости от направления)

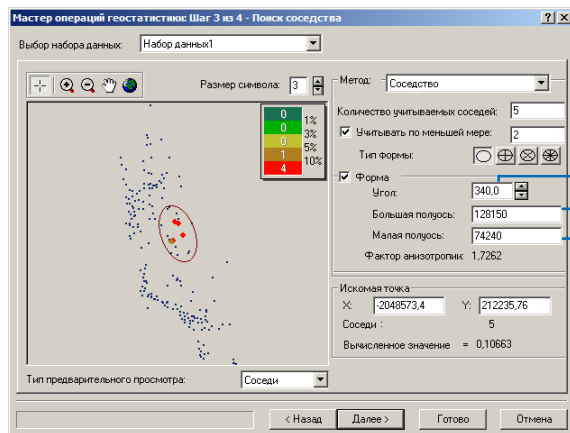
Конверт  
вариограммы

Используйте инструмент направления поиска для определения значения радиуса влияния для конкретного направления

Область поиска соседства в данном примере может быть определена как эллипс, с размерами большой и малой полуосей 128 и 74 км, соответственно, и углом поворота, равным  $340^\circ$ .



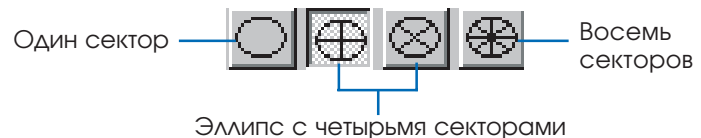
Диалог Поиск соседства позволяет вам определить размеры большой и малой полуосей и направление большой полуоси. Окружность будет иметь одинаковые значения размера большой и малой полуоси.



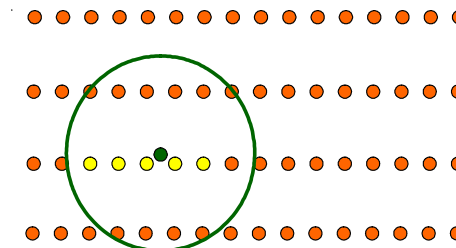
Угол оси

Размер оси

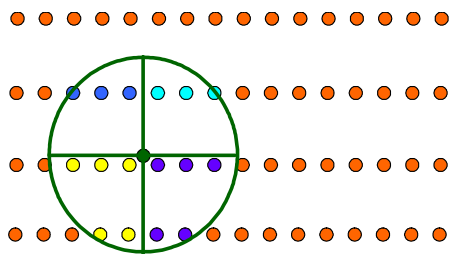
После того, как определена форма области, с помощью второго механизма, контролирующего поиск соседства, можно задать ограничения на выбор точек внутри области. Во-первых, необходимо установить число точек, попадающих в область соседства, которые вы будете использовать при выполнении интерполяции. Модуль Geostatistical Analyst позволяет вам выбрать желательное и минимальное количество используемых точек. Во-вторых, во избежание смещения в определенном направлении, окружность или эллипс могут быть поделены на сектора, для каждого из которых будет выбираться одинаковое количество точек.



Особенно полезно это делать, когда опорные точки располагаются в узлах регулярной сетки. На диаграмме внизу показан случай, при котором точки, расположенные ближе всего к искомой, относятся к одному и тому же поперечному разрезу. Ограничения, заданные областью поиска в форме круга, состоят в том, что значение искомой точки будет определяться по пяти соседним точкам. Неизвестное значение показано зеленым цветом, а пять ближайших соседних точек, - желтым цветом. Другие точки внутри окружности не включаются в вычисления, поскольку расположены дальше от искомой точки.

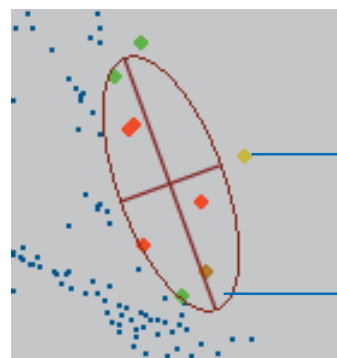


Интерполяция может быть более точной, если в вычислениях будут использованы точки и из других рядов. На диаграмме внизу показана область соседства с четырьмя секторами, в которую должно быть включено по меньшей мере три точки из каждого сектора, но не больше пяти. Для вычисления неизвестного значения в искомой точке (которая показана зеленым цветом) будет использовано всего 16 точек. Точки каждого сектора, включаемые в интерполяцию, показаны по секторам разными цветами.



Количество используемых точек и секторов должно определяться объективно и основываться на пространственном распределении опорных точек. С одной стороны, вам необходимо иметь достаточное количество точек, чтобы получить достоверный результат для любой искомой точки; с другой стороны, вы не захотите включать в вычисления точки, расположенные слишком далеко от искомой. Например, если вы выберете максимальное количество точек, равное пяти, а минимальное - равное двум (т.е., в вычислениях должно участвовать по крайней мере две точки) и четыре сектора, общее количество используемых точек будет равно 20. Если в любом из секторов нет минимально необходимого количества точек, программа выбирает ближайшие возможные точки за пределами этого сектора. Однако, поиск вне пределов сектора может быть лимитирован областью, образующейся при продлении линий, ограничивающих сектор, на бесконечное расстояние. Если в любом из расширенных секторов поиска нет пяти точек, общее количество точек будет меньше 20.

Цветом обозначены абсолютные значения весов точек, выделенных в окне вида данных, которые будут присвоены каждой точке при вычислении искомого значения (точка соответствует центру окружности или эллипса). В приведенном примере, искомая точка находится на пересечении секторов, и четыре точки (показанные красным цветом) будут иметь веса более чем 10 процентов. Чем больше абсолютное значение веса, тем большее влияние точка будет иметь при интерполяции значения искомой точки.



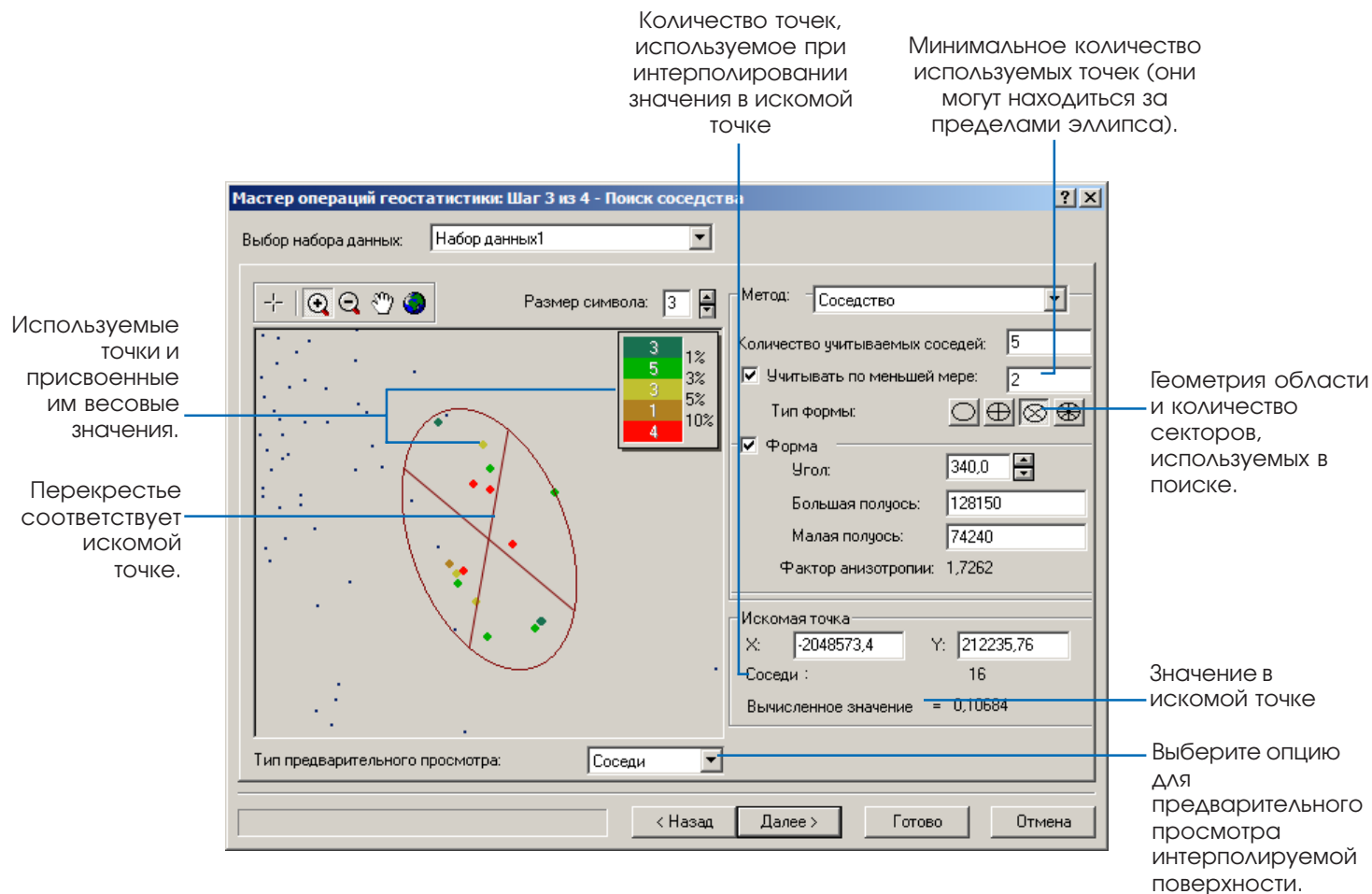
Если в область соседства не попадает достаточное количество точек, программа выбирает ближайшие возможные точки (точка, показанная на рисунке желтым цветом будет включена в верхний правый сектор).

В этот сектор попадают только две точки.

Помимо использования вариограммы, существует еще несколько средств для определения формы области соседства и ограничений на использование опорных точек. Изучение пространственного распределения опорных точек в ArcMap может также помочь в определении соседства. Например, опорные точки могут быть отобраны вдоль линий, образующих грид с точками, которые в направлении восток-запад расположены ближе друг к другу, чем в направлении север-юг. В этой ситуации, большая полуось соответствующего эллипса области поиска может быть направлена с юга на север. Размеры осей эллипса могут быть определены таким образом, чтобы примерно равное количество точек попадало в границы этого эллипса в обоих направлениях.



## Диалог Поиск соседства



## Определение размера области поиска соседства

Размер области поиска соседства определяет форму области и ограничение на использование точек, попадающих в эту область и применяемых при интерполировании неизвестного значения в искомой точке.

Вы задаете параметры соседства путем определения положения точек в окне вида данных и использования начальных значений, полученных при работе с инструментами ESDA и моделировании вариограммы/ковариации.

### Подсказка

#### Оценка соседства

Влияние использования области соседства может быть оценено с использованием инструментов перекрестной проверки и проверки, которые включены в модуль *Geostatistical Analyst*. При необходимости, область поиска соседства может быть переопределена, и на ее основе построена новая поверхность.

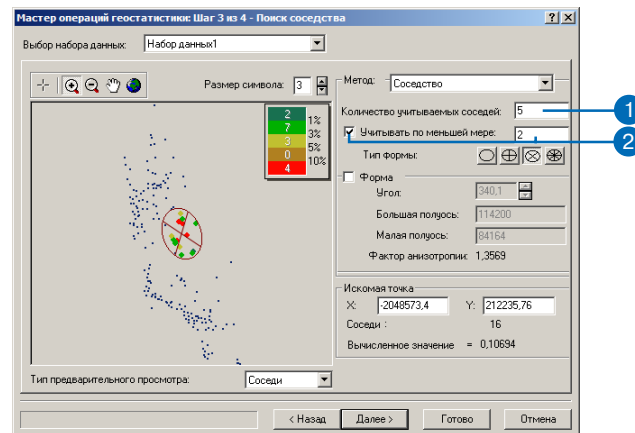
## Изменение числа точек, включаемых в область соседства

1. В диалоге Поиск соседства при помощи стрелочек поменяйте количество учитываемых соседей (включаемых в поиск соседних точек).

Или введите требуемое значение.

2. Чтобы определить минимальное количество точек, включаемых в область соседства, отметьте опцию Учитывать по меньшей мере и с помощью стрелочек определите требуемое значение.

Или введите значение в поле ввода.



## Подсказка

### Весовые значения

Веса отображаются абсолютными значениями, то есть значения -6 процентов и 6 процентов будут “равны”.

## Подсказка

Значение малой полуоси может быть задано большим, чем значение большой полуоси.

## Изменение формы области соседства

1. В диалоге Поиск соседства выберите иконку с изображением нужного эллипса, чтобы изменить форму области соседства, предложенную по умолчанию.

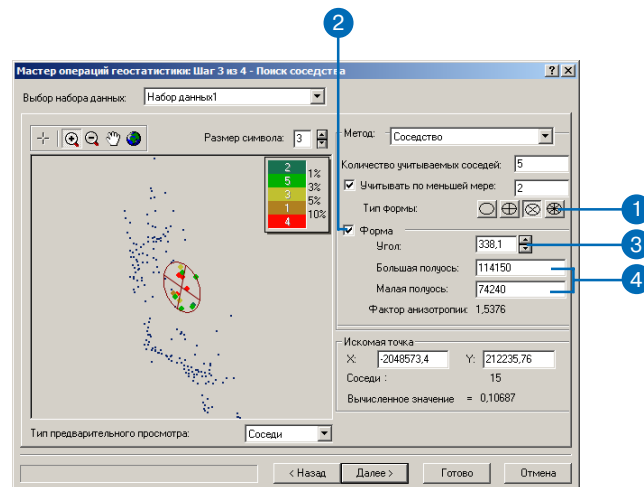
2. Отметьте опцию Форма.

Параметры диалога станут активными.

3. С помощью стрелочек задайте значение угла, чтобы изменить угол направления большой оси эллипса.

4. В окошках для ввода данных напечатайте требуемые размеры большой и малой полуосей, что приведет к изменению формы эллипса.

В окне диалога отразятся внесенные вами изменения.



## Подсказка

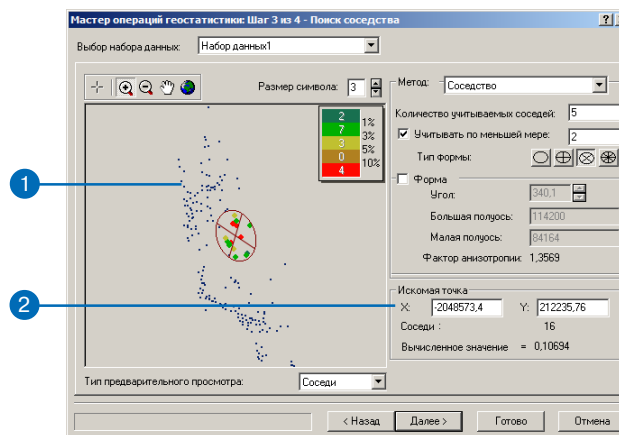
### Проинтерполированное значение

В диалоге Поиск соседства найдите рамку Искомая точка; значения полей X и Y соответствуют координатам искомой точки на карте (которая также является центром эллипса в окне отображения данных).

## Определение проинтерполированного значения в определенной точке

1. Число соседей, используемых для интерполяции, и проинтерполированное значение показаны ниже полей для ввода данных (в нижней правой части диалога). Чтобы определить новую искомую точку, нажмите на нее в области отображения данных диалога.
2. Или введите значения координат X и Y.

Проинтерполированное значение и количество соседей будут тут же обновлены в соответствии с вновь выбранной точкой.

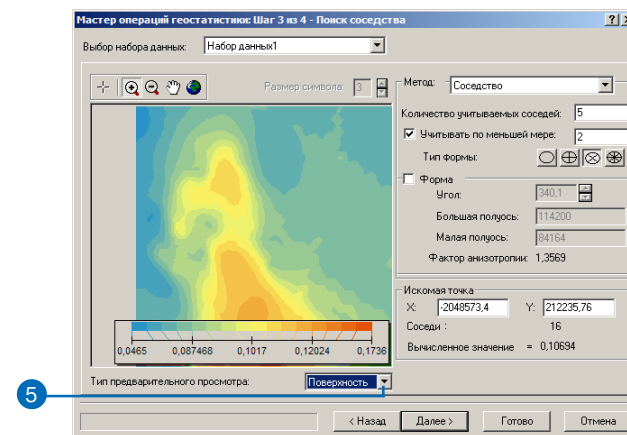
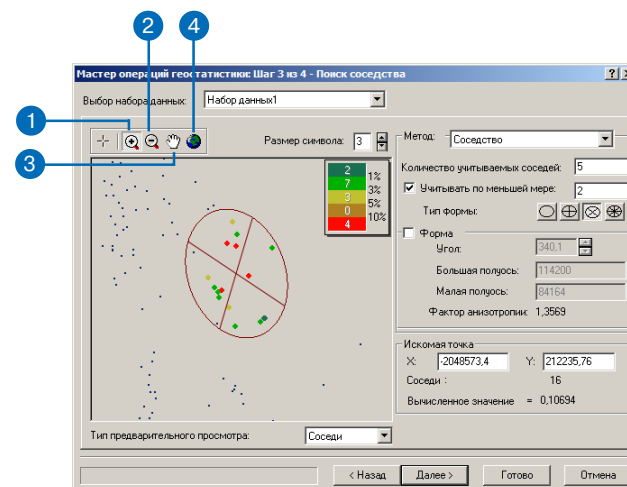


## Изменение вида карты

1. Чтобы увеличить масштаб отображения карты в диалоге Поиск соседства нажмите кнопку Увеличить, затем растяните прямоугольник на ту область, изображение которой вы хотите увеличить.
2. Чтобы уменьшить изображение, нажмите на кнопку Уменьшить.
3. Чтобы двигаться по изображению карты, нажмите кнопку Переместить, расположите курсор на изображении карты и, удерживая левую клавишу мыши, меняйте положение курсора.

Карта будет смещаться в соответствии с движением курсора.

4. Чтобы отобразить карту для всей исследуемой территории, нажмите кнопку Полный экстен.
5. Чтобы осуществить предварительный просмотр результирующей поверхности, выберите в меню Тип предварительного просмотра опцию Поверхность. Чтобы вернуться к предыдущему виду, выберите опцию Соседи.



## Выполнение проверки и перекрестной проверки

Перед построением окончательной поверхности, вы должны составить себе некое представление о том, насколько хорошо подобранная вами модель интерполирует значения в искомым точках. Перекрестная проверка и проверка помогают вам принять обоснованное решение о том, какая из моделей наилучшим образом выполняет интерполяцию значений. Вычисление статистики служит диагностикой, которая показывает, являются ли модель и связанные с ней параметры приемлемыми.

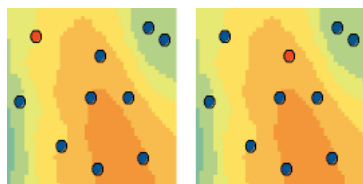
Перекрестная проверка и проверка убирают из набора данных одну или несколько опорных точек, а затем для этой точки вычисляют значение. Таким способом, вы можете сравнить проинтерполированное и измеренное значение и на основе этого получить полезную информацию о модели кригинга (например, параметрах вариограммы и поиске соседей). Далее мы рассмотрим разницу между проверкой и перекрестной проверкой.

Перекрестная проверка использует все данные для оценки модели автокорреляции. Затем она последовательно, по одной, убирает опорную точку и интерполирует для этого местоположения соответствующее значение. Например, на диаграмме внизу показаны 10 опорных точек, распределенных случайным образом. Перекрестная проверка пропускает точку (показанную красным цветом) и интерполирует значение показателя для этого местоположения с использованием остающихся девяти точек (показанных синим цветом). Вычисленное и действительное значение пропущенной опорной точки сравниваются. Эта процедура повторяется для второй точки, и так далее. Для всех точек, перекрестная проверка сравнивает измеренные и проинтерполированные значения. По сути, перекрестная проверка несколько “мошенничает”, используя все данные для оценки модели автокорреляции. После завершения перекрестной проверки, некоторые из опорных точек могут быть

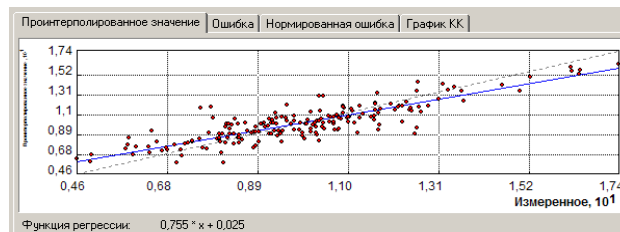
“отодвинуты в сторону” как необычные, требующие повторного подбора модели автокорреляции.

Проверка сначала исключает из вычислений часть данных — назовем ее “тестовый набор данных” — а затем использует оставшиеся данные — назовем их “учебный набор данных” — для создания моделей тренда и автокорреляции, которые будут использованы в интерполяции. В модуле Geostatistical Analyst вы создаете тестовый и учебный набор данных с использованием инструмента Создать поднаборы данных. Однако, типы графиков и суммарная статистика, используемые для сравнения проинтерполированных и истинных значений, аналогичны и для перекрестной, и для простой проверки. Проверка создает модель только для поднабора данных, поэтому она напрямую не проверяет вашу окончательную модель, которая должна использовать полный набор опорных точек. Скорее, проверка вычисляет, является ли правильным “протокол” решений, например, выбор модели вариограммы, выбор размера лага, выбор условий поиска соседей, и т.д. Если протокол решений хорошо работает для набора данных, используемого при проверке, вы можете с уверенностью предположить, что он будет также хорошо работать и для всего набора данных.

В модуле Geostatistical Analyst есть несколько графиков и данных суммарной статистики, позволяющих сравнить измеренные и проинтерполированные значения. Начнем с графика рассеивания, на котором по одной оси нанесены проинтерполированные значения, а по другой — измеренные значения. Можно ожидать, что точки будут лежать вблизи линии 1:1 (черная пунктирная линия внизу). Однако, наклон обычно меньше единицы. Это является свойством кригинга, который имеет тенденцию к завышению маленьких значений и занижению высоких значений (см. нижний рисунок).

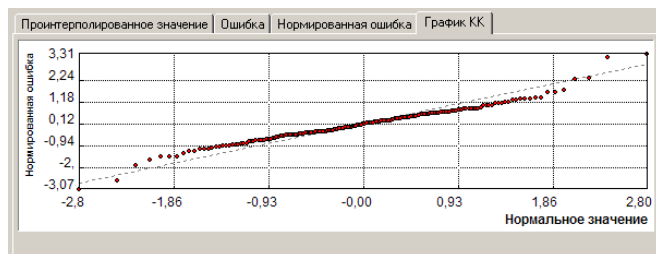


И так далее для  
всех точек



Подобранная модель, проходящая через точки, показана синим цветом, а уравнение этой прямой приведено под графиком. График ошибок аналогичен графику интерполяции, за исключением того, что здесь истинные значения вычтены из проинтерполированных значений. Для графика нормированных ошибок, измеренные значения вычтены из проинтерполированных, а затем поделены на оцененные стандартные ошибки кригинга. Все эти три графика позволяют увидеть, насколько хорошо кригинг выполняет интерполяцию. Если все данные были независимыми (нет автокорреляции), все проинтерполированные значения будут одними и теми же (каждое проинтерполированное значение будет средним из измеренных), следовательно, голубая линия будет горизонтальной. При наличии автокорреляции и хорошей модели кригинга, голубая линия должна располагаться близко к линии 1:1 (черной пунктирной линии). Вы также можете видеть точки, рассеянные около линии (некоторые показаны на предыдущем рисунке на стр. 172). Чем ближе точки расположены к линии 1:1, тем лучше. Окончательный график является графиком КК (Квантиль-квантиль). Он показывает квантили разности между проинтерполированными и измеренными значениями, деленными на оцененные стандартные ошибки кригинга и соответствующие квантили из стандартного нормального распределения. Если ошибки интерполяции (отклонения от истинных значений) подчиняются нормальному распределению, точки должны лежать примерно вдоль пунктирной линии. При этом вы с уверенностью можете пользоваться методами, основывающимися на нормальности распределения (например, для создания карты квантилей с использованием ординарного кригинга).

См. Главу 4, 'Исследовательский анализ пространственных данных', для более подробной информации о графиках КК.



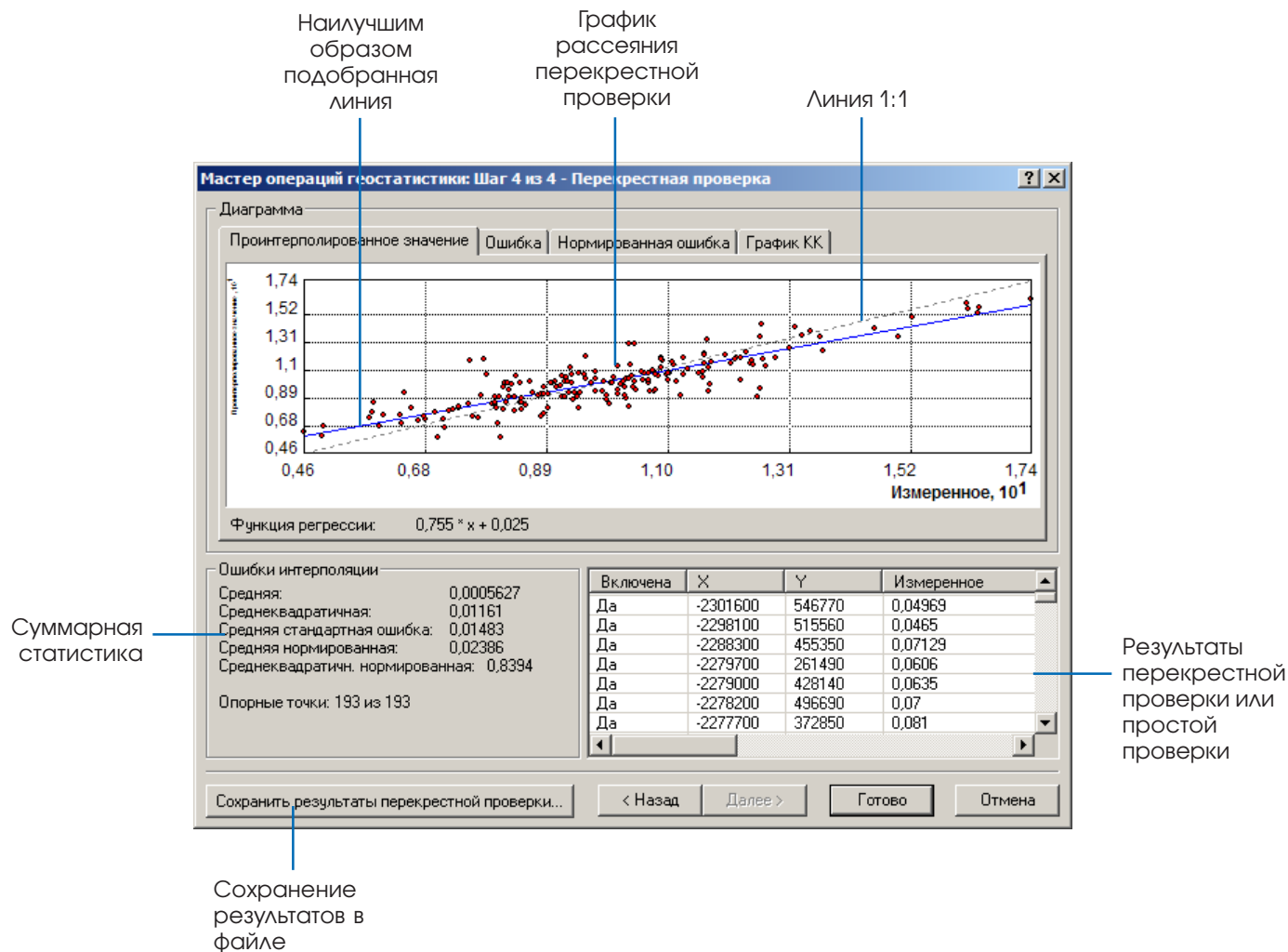
И наконец, в левом нижнем углу дана суммарная статистика по ошибкам интерполяции с использованием кригинга. Вы воспользуетесь этими данными для диагностики по трем основным моментам:

1. Вы захотите, чтобы проинтерполированные вами значения оказались несмещенными (центрированными по измеренным значениям). Если ошибки вычислений не смещены, средняя ошибка интерполяции должна быть примерно равной нулю. Однако, это значение зависит от масштаба данных, поэтому чтобы их стандартизировать, используют нормированные ошибки интерполяции, которые получаются при делении ошибок интерполяции на их стандартные ошибки. Среднее значение нормированных ошибок интерполяции должно быть примерно равно нулю.
2. Вы бы хотели, чтобы проинтерполированные вами значения были как можно ближе к измеренным значениям. Среднеквадратичные ошибки интерполяции вычисляются как корень квадратный из среднего квадратов расстояний (размеров зеленых отрезков), показанных на диаграмме интерполяции, приведенной выше. Чем короче зеленые отрезки, тем ближе проинтерполированные значения к своим истинным значениям, и тем меньше среднеквадратичные ошибки интерполяции. Эта статистика может быть использована для сравнения различных моделей путем визуального анализа близости проинтерполированных и измеренных значений. Чем меньше среднеквадратичная ошибка интерполяции, тем лучше.
3. Вы захотите, чтобы ваша оценка неопределенности, или стандартные ошибки интерполяции, были достоверными. Каждая из моделей кригинга дает оцененные стандартные ошибки интерполяции. Помимо выполнения интерполяции, мы оцениваем отклонение проинтерполированных значений от измеренных. Важно получить корректные отклонения. Например, в ординарном кригинге (предполагающем, что отклонения подчиняются закону нормального распределения), карты квантилей и вероятности зависят от стандартных ошибок кригинга в такой же степени, как и сами проинтерполированные значения. Если средние стандартные ошибки близки к среднеквадратичной ошибке интерполяции, тогда вы сможете корректно оценить неопределенность при интерполировании значений.



Если средние стандартные ошибки больше, чем среднеквадратичные ошибки интерполяции, вы завышаете оценку отклонений в ваших вычислениях; если средние стандартные ошибки меньше, чем среднеквадратичные ошибки интерполяции, вы занижаете оценку отклонений в ваших вычислениях. Другой способ - разделить каждую ошибку вычислений на оцененную стандартную ошибку интерполяции. В среднем, они должны быть похожи, и следовательно, среднеквадратичные нормированные ошибки должны быть близки к единице, если стандартные ошибки интерполяции достоверны. Если среднеквадратичные нормированные ошибки больше 1, вы занижаете оценку неопределенности проинтерполированных значений; если же среднеквадратичные нормированные ошибки меньше 1, вы завышаете оценку неопределенности проинтерполированных значений.

## Диалог Cross Validation and Validation (Перекрестная проверка и проверка)



## Выполнение перекрестной проверки для оценки выбранных параметров

Перекрестная проверка позволяет вам определить “насколько хороша” ваша модель. Ваша цель - получить нормированное среднее ошибки вычислений, примерно равное 0, небольшие среднеквадратичные ошибки интерполяции, среднюю стандартную ошибку близкую к среднеквадратичным ошибкам интерполяции, и нормированные среднеквадратичные ошибки интерполяции, примерно равные 1.

Точки должны быть расположены как можно ближе к серой пунктирной линии. Обратите внимание на точки, которые сильно отклоняются от линии.

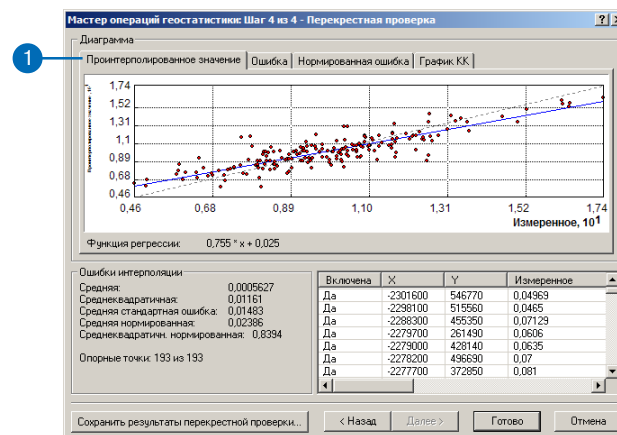
### Подсказка

#### Выбор точек

Когда в таблице результатов выбрана строка, точка на графике выделяется цветом. Для сортировки значений выберите столбец. Это поможет найти определенные точки на графике.

## Изучение полученной модели

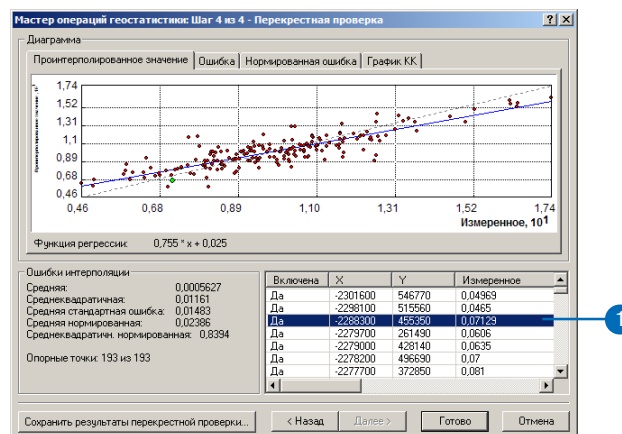
1. В диалоге Перекрестная проверка выберите одну из вкладок Проинтерполированные значения, Ошибка, Нормированная ошибка, или График КК в зависимости от метода, который вы хотите использовать для просмотра результатов.



## Выбор определенной точки

1. В диалоге Перекрестная проверка в таблице, данной в правой нижней части диалога, выберите строку, соответствующую интересующей вас точке.

Когда строка выбрана, точка выделяется цветом на диаграмме в верхней части диалога.



## Подсказка

### Просмотр всех строк и столбцов

Используйте линейки прокрутки для просмотра всех строк и столбцов таблицы.

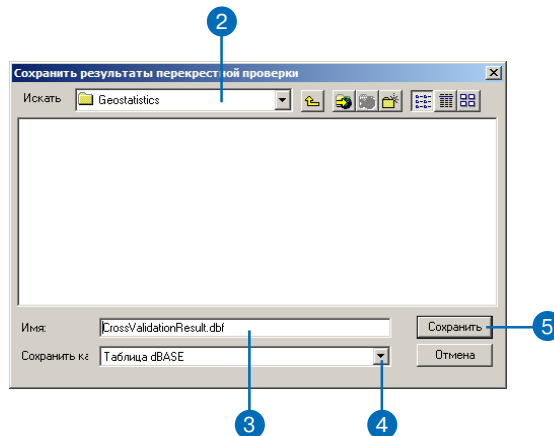
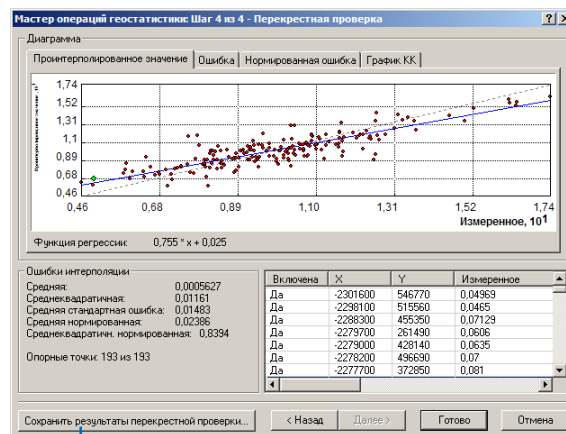
## Подсказка

### Просмотр сохраненной таблицы

Чтобы просмотреть сохраненную в файле таблицу, выберите Добавить на стандартной панели инструментов, перейдите к базе данных, дважды щелкните по иконке, а затем добавьте таблицы. В таблице содержания ArcMap выберите таблицу и, нажав правую клавишу мыши, перейдите к опции Открыть.

## Сохранение статистики перекрестной проверки в файл

1. Нажмите кнопку Сохранить результаты перекрестной проверки.
2. Перейдите к папке на диске, в которой вы хотите сохранить набор данных.
3. Наберите название набора данных.
4. Выберите тип набора данных.
5. Нажмите Сохранить.



## Оценка протокола решений с использованием проверки

Проверка позволяет оценить полученные вами результаты интерполяции.

Ваша цель - получить нормированное среднее ошибок вычислений, примерно равное 0, небольшие среднеквадратичные ошибки вычислений, средние стандартные ошибки около среднеквадратичных ошибок вычислений, и нормированные среднеквадратичные ошибки интерполяции, примерно равные 1.

Точки должны быть расположены как можно ближе к серой пунктирной линии. Обратите внимание на точки, которые сильно отклоняются от линии.

### Подсказка

#### Разделение набора данных на тестовый и учебный поднаборы

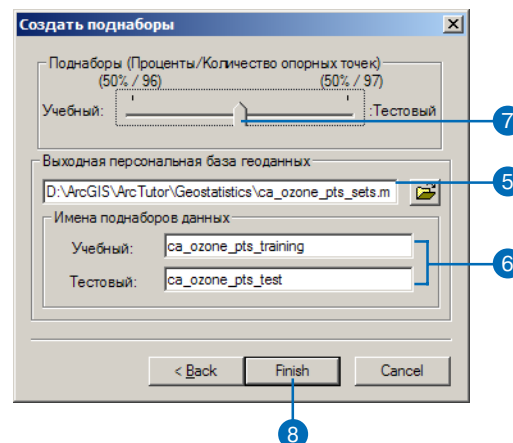
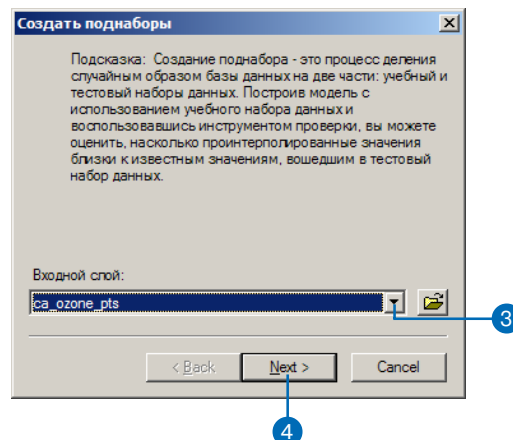
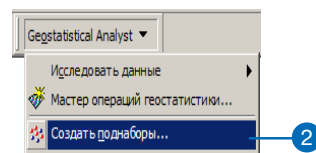
Установите движок посередине, чтобы разбить данные примерно поровну между тестовым и учебным набором.

### См. также

См. раздел 'Выполнение перекрестной проверки и проверки для оценки выбранных параметров' (предваряющего данные упражнения) для сравнения понятий проверки и перекрестной проверки.

## Создание поднаборов данных для использования их в проверке модели

1. Добавьте набор данных, для которого вы хотите создать поднаборы, в ArcMap.
2. Выберите опцию Создать поднаборы данных.
3. Из списка выберите набор данных, для которого вы хотите создать поднабор.
4. Нажмите Далее.
5. Необязательный шаг: измените местоположение результирующего файла базы геоданных на диске и/или его название.
6. Необязательный шаг: измените названия поднаборов, предложенные по умолчанию.
7. Удерживая движок, передвиньте его в требуемое положение.
8. Нажмите Готово.



## Подсказка

### Просмотр всех строк и столбцов

Используйте линейки прокрутки для просмотра всех строк и столбцов таблицы.

## Подсказка

### Выделение значений

Когда в таблице результатов выбрана строка, точка на графике выделяется цветом. Для сортировки значений выберите столбец. Это поможет найти определенные точки на графике.

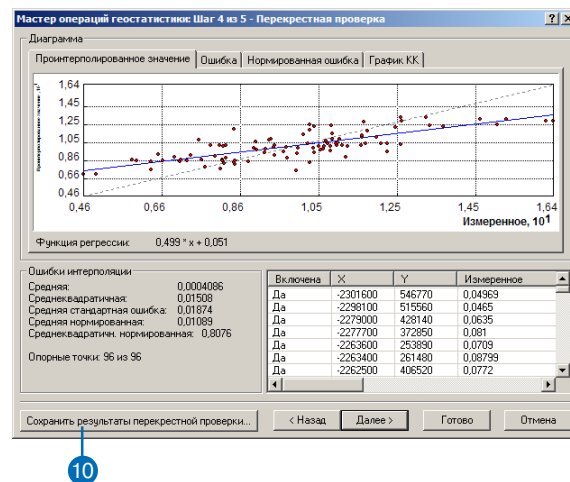
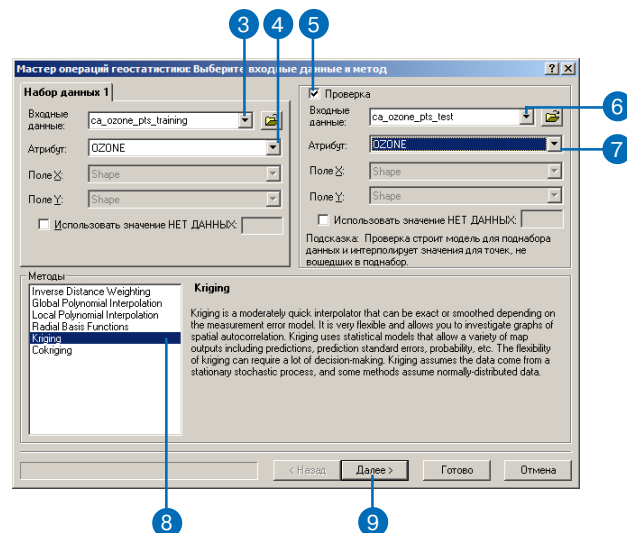
## Подсказка

### Просмотр сохраненной таблицы

Чтобы просмотреть сохраненную в файле таблицу, выберите кнопку Добавить на стандартной панели инструментов, перейдите к базе данных, дважды щелкните по иконке, а затем добавьте таблицу. В таблице содержания ArcMap выберите таблицу и, нажав правую клавишу мыши, перейдите к опции Открыть.

## Выполнение проверки

1. Нажмите на кнопку Добавить данные и перейдите к тестовому и учебному набору данных (созданных при выполнении предыдущего упражнения). Нажмите Добавить.
2. Запустите Мастер операций геостатистики.
3. В окне Входные данные выберите учебный набор данных (созданный в предыдущем упражнении).
4. В окне Атрибут выберите показатель, значения которого вы хотите использовать в интерполяции.
5. Отметьте опцию Проверка.
6. В окне Входные данные выберите тестовый набор данных (созданный в предыдущем упражнении).
7. В окне Атрибут выберите тот же самый атрибут, который вы выбирали для учебного набора данных.
8. Выберите метод, который вы хотите использовать.
9. Нажимайте Далее в этом и во всех последующих диалогах до тех пор, пока вы не дойдете до диалога Проверка.
10. Необязательный шаг: сохраните таблицу с результатами проверки в базе данных.



## Сравнение моделей

Сравнение помогает вам определить, насколько модель, которая использовалась для создания геостатистического слоя, лучше, чем какая-либо другая модель. Два геостатистических слоя, которые вы сравниваете, могут быть созданы с использованием двух различных моделей (например, по методу взвешенных расстояний IDW и ординарного кригинга), либо они могут быть созданы с использованием одной и той же модели, но с различными параметрами. В первом случае, вы сравниваете, какой из методов лучше подходит для использования с вашими данными, во втором - вы изучаете влияние различных исходных параметров на модель при построении результирующей поверхности.

Диалог Сравнение использует статистику перекрестной проверки аналогично тому, как было рассказано в предыдущем разделе. Однако, он позволяет вам изучить статистику и графики для различных моделей совместно, в одном окне. В целом, лучшая модель - это та модель, нормированная средняя ошибка которой близка к нулю, и которая характеризуется наименьшей среднеквадратичной ошибкой интерполяции, имеет значение средней стандартной ошибки, близкое к среднеквадратичной ошибке вычислений, и нормированную среднеквадратичную ошибку, близкую к единице.

Как правило, перед тем, как одна из поверхностей будет признана “лучшей” и будет либо сама использоваться как окончательная, либо войдет в другую, “большую” модель, предназначенную для решения существующей проблемы (например, модель условий строительства новых домов), создается несколько поверхностей. Вы можете последовательно сравнивать каждую из поверхностей с другой, исключая “худшую” из двух сравниваемых, до тех пор, пока не останутся две “лучшие” поверхности, и вы не перейдете к сравнению их друг с другом. Вы можете прийти к заключению, что для данного конкретного анализа, лучшая из двух окончательных поверхностей и является “лучшей” из возможных.

### Что следует учитывать при сравнении методов и моделей

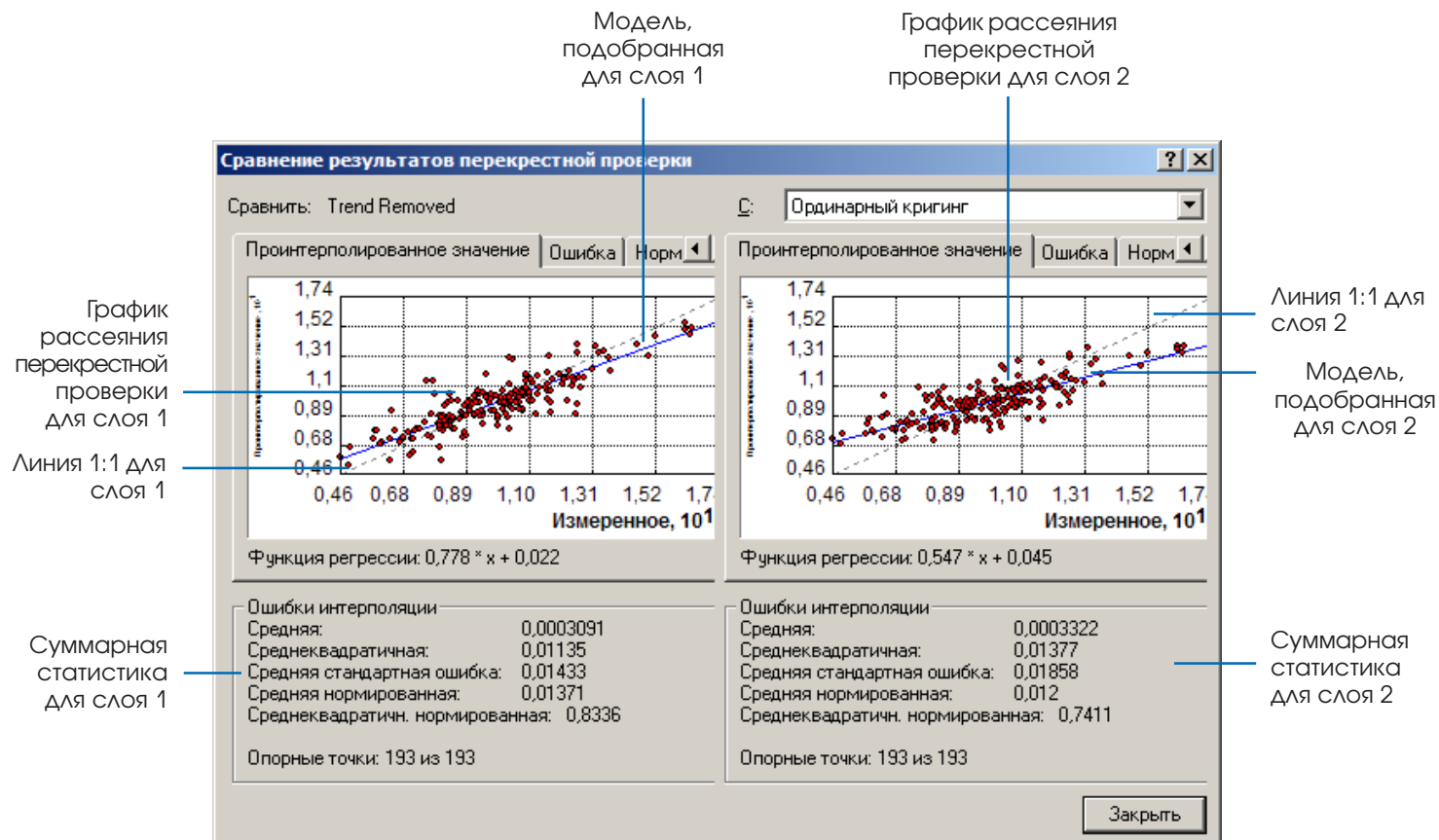
Существует два момента, на которые следует обратить внимание при сравнении результатов различных методов и/или моделей: один из них - оптимальность, второй - достоверность.

Например, для определенной модели среднеквадратичная ошибка интерполяции может быть очень маленькой. Исходя из этого, вы можете прийти к заключению, что это и есть “оптимальная” модель. Однако, при сравнении с другой моделью, может оказаться, что среднеквадратичная ошибка интерполяции ближе к средней оцененной стандартной ошибке интерполяции. Это более достоверная модель, потому что когда вы интерполируете значение в точке, не имея данных, у вас есть только оцененные стандартные ошибки, которые и используются для оценки неопределенности ваших вычислений. Когда средние оцененные стандартные ошибки интерполяции близки к ее среднеквадратичным ошибкам, полученным в результате перекрестной проверки, вы уверены, что стандартные ошибки интерполяции являются удовлетворительными.

Наряду со статистикой, приведенной в диалоге Сравнение, вы должны также использовать первичную информацию, имеющуюся в вашем наборе данных и извлеченную вами с помощью инструментов ESDA (Исследовательского анализа пространственных данных) при оценке того, какая из моделей является “лучшей”. Обратитесь к предыдущему разделу по перекрестной проверке и проверке для более подробной информации о том, как была получена статистика, и как она должна использоваться в анализе.



## Диалог Сравнение результатов перекрестной проверки



## Сравнение моделей

При сравнении моделей вы должны искать ту, нормированное среднее которой ближе к нулю, а также имеющую наименьшую среднеквадратичную ошибку интерполяции, среднюю стандартную ошибку вычислений, ближайшую к среднеквадратичной ошибке интерполяции, и нормированную среднеквадратичную ошибку, значение которой ближе всего к единице.

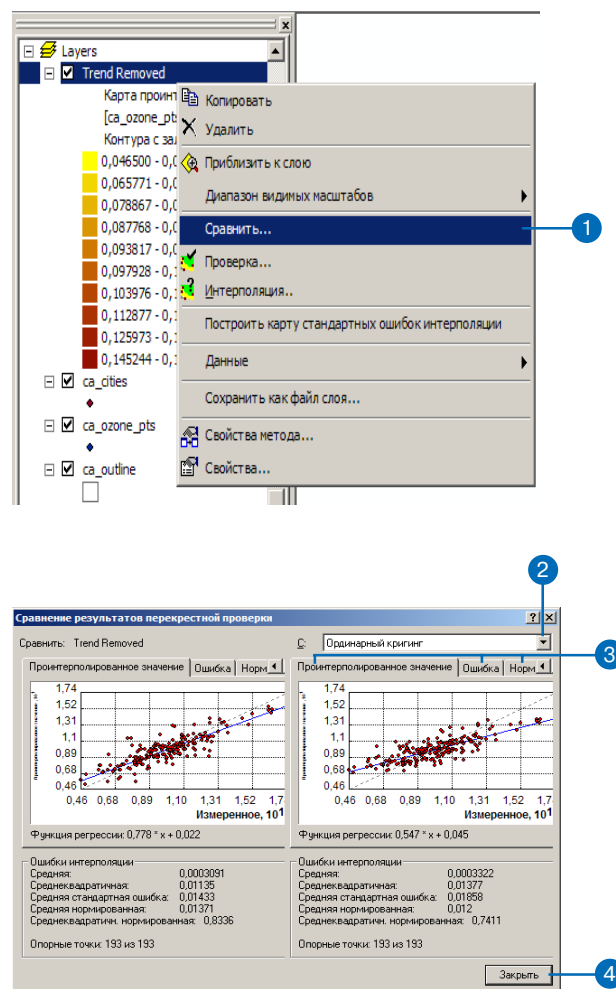
Чтобы иметь возможность сравнить модели, у вас должно быть два геостатистических слоя (созданных с использованием модуля Geostatistical Analyst). Эти два слоя могут быть созданы с использованием различных методов интерполяции (например, метода взвешенных расстояний - IDW и ординарного кригинга), либо с использованием одного и того же метода, но различных параметров.

**См. также**

*См. раздел 'Выполнение перекрестной проверки и проверки для оценки выбранных параметров' (предваряющего данные упражнения) для сравнения понятий проверки и перекрестной проверки.*

## Выполнение сравнения

1. Выберите в таблице содержания ArcMap один из слоев, который вы будете сравнивать, нажмите правую клавишу мыши и выберите опцию Сравнить.
2. В меню C: выберите второй слой, с которым вы будете сравнивать первый.
3. Выберите различные закладки для просмотра некоторых результатов сравнения.
4. Нажмите Заккрыть, чтобы закрыть диалог Сравнение результатов перекрестной проверки.



### Преобразования по методу Box–Cox, арксинуса и логарифмические

Модуль Geostatistical Analyst дает возможность использовать несколько методов преобразований, включая метод Box–Cox (также известный как степенные преобразования), логарифмические преобразования, и преобразования по методу арксинуса. Предположим, что у вас есть данные наблюдений в опорных точках  $Z(s)$ , и есть некое преобразование  $Y(s) = t(Z(s))$ . Как правило, вы хотите найти преобразование, которое приведет к тому, что значения  $Y(s)$  будут подчиняться закону нормального распределения. Часто результатом преобразований является постоянная дисперсия данных для всей изучаемой территории. Теперь рассмотрим каждое преобразование.

Преобразование по методу Box–Cox выглядит следующим образом:

$$Y(s) = (Z(s)^\lambda - 1) / \lambda$$

для  $\lambda \neq 0$ . Например, предположим, что ваши данные состоят из подсчетов встречаемости какого-либо явления. Для этих типов данных, дисперсия часто соотносится со средним значением. Это означает, что если для части изучаемой территории количество подсчетов мало, изменчивость в этом районе будет меньше, чем изменчивость в другом районе, в котором количество подсчетов больше. В этом случае, преобразование, использующее квадратный корень, поможет вам сделать дисперсии более постоянными для всей изучаемой территории и, кроме этого, оно часто приводит данные к нормальному распределению. Преобразование по методу квадратного корня является частным случаем преобразования Box–Cox ( $\lambda = 0.5$ ).

Логарифмическое преобразование, по сути, является также частным случаем преобразования по методу Box–Cox при  $\lambda = 0$ ; формула преобразования имеет следующий вид:

$$Y(s) = \ln(Z(s))$$

где  $Z(s) > 0$ , а  $\ln$  - натуральный логарифм. Следствием логарифмического преобразования является метод интерполяции, известный как логнормальный кригинг, тогда как для всех остальных значений  $\lambda$ , соответствующий метод интерполяции известен как трансгауссов кригинг. Логарифмическое преобразование часто используется в тех случаях, когда данные имеют асимметричное распределение, и в них есть несколько очень больших значений. Эти большие значения могут быть локализованы на изучаемой территории, и логарифмическое преобразование поможет вам сделать дисперсии более постоянными и нормализовать ваши данные.

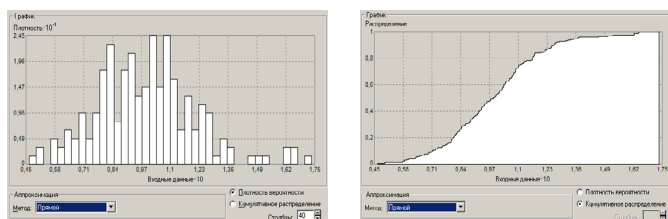
Преобразование по методу арксинуса выглядит следующим образом:

$$Y(s) = \sin^{-1}(Z(s))$$

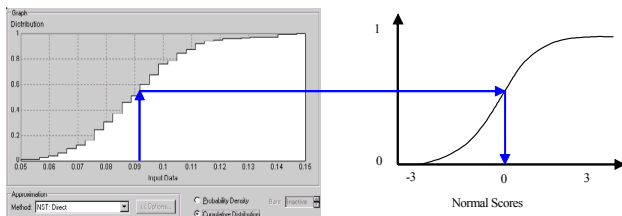
где  $Z(s)$  лежат в интервале между 0 и 1. Преобразование по методу арксинуса может быть использовано для относительных данных или данных, выраженных в процентах. Часто, когда данные относительные, дисперсия наименьшая для значений близких к 0 и 1 и наибольшая, когда значения близки к 0.5. Тогда преобразование по методу арксинуса часто приводит к тому, что дисперсии постоянны для всей изучаемой территории и, кроме того, часто его результатом является нормальное распределение данных.

## Преобразование по методу нормальных меток

Преобразование по методу нормальных меток ранжирует значения вашего набора данных от самых маленьких до самых больших и сопоставляет эти классы с соответствующими классами графика нормального распределения. Затем преобразование осуществляется следующим образом: значения берутся из соответствующего класса нормального распределения. Это можно видеть на следующих диаграммах. На первой показана гистограмма, которая обычно строится при проведении исследовательского анализа данных. По-другому данные можно отобразить, воспользовавшись графиком кумулятивного распределения, который показан на правом рисунке.

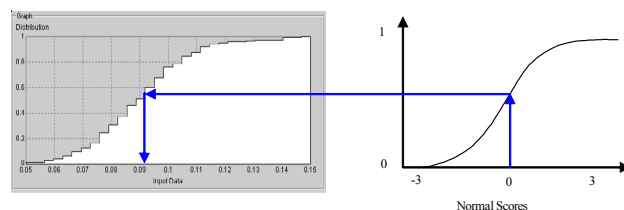


Чтобы получить преобразование по методу нормальных меток, воспользуйтесь кумулятивным распределением. Возьмите кумулятивное распределение наблюдений и сопоставьте его с кумулятивным распределением стандартного нормального распределения.



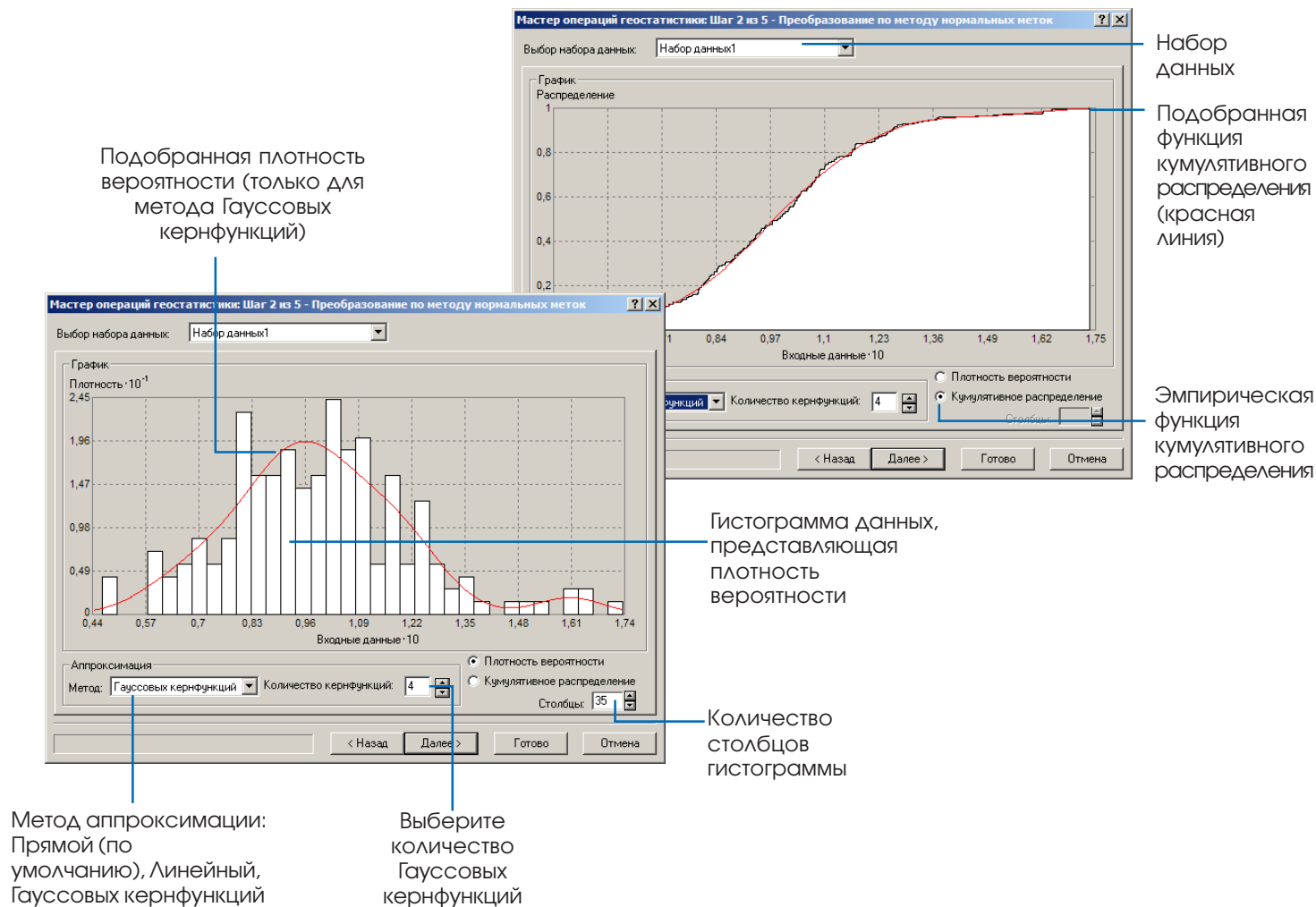
деления. Это хорошо видно при рассмотрении процедуры сопоставления графиков, показанной на рисунке.

В данном примере, значение исходных данных, равное 0.09 по методу нормальных меток преобразуется в значение, примерно равное нулю. В модуле Geostatistical Analyst можно использовать три метода аппроксимации: прямой, линейный и Гауссовых керн-функций. Прямой метод использует наблюдаемое кумулятивное распределение; линейный метод подбирает линии между каждым “шагом” кумулятивного распределения; и метод Гауссовых керн-функций аппроксимирует вероятностное распределение путем подбора линейной комбинации из кумулятивных распределений плотности. После выполнения интерполяции в преобразованном масштабе, необходимо выполнить обратное преобразование, чтобы привести вычисленные значения к исходному масштабу. Например, если вы используете прямую аппроксимацию нормального распределения, обратное преобразование будет выглядеть следующим образом:



Выбор метода аппроксимации зависит от допущений, сделанных вами, и от степени сглаживания аппроксимации. Прямой метод - наименее сглаженный и характеризуется наименьшим количеством допущений; линейный метод занимает промежуточное положение, а метод Гауссовых керн-функций обладает самым сглаженным обратным преобразованием, но наряду с этим и самые серьезные допущения (например, то, что распределение данных может быть аппроксимировано с помощью сочетания ограниченного количества нормальных распределений).

## Диалог Преобразования по методу нормальных меток



## Сравнение преобразований по методу нормальных меток с другими преобразованиями

Преобразования по методу нормальных меток могут быть сопоставлены с преобразованиями по методу Box—Cox, арксинуса, и логарифмическими преобразованиями. Наиболее существенная разница между ними состоит в том, что функция преобразований NST меняет каждый определенный набор данных, в то время как преобразование BAL не делает этого (например, функция логарифмического преобразования - всегда натуральный логарифм). Однако цель преобразований по методу NST - сделать случайные ошибки всей совокупности (а не только опорных точек) нормально распределенными. Таким образом, важно, что кумулятивное распределение опорных точек отражает истинное кумулятивное распределение всей совокупности данных. С точки зрения классической статистики, преобразование NST, примененное к непространственным данным, может рассматриваться как метод, основанный на ранжировании. Однако, метод нормальных меток может быть полезен в геостатистике, потому что, когда данные зависимы, с использованием метода нормальных меток легче определить и смоделировать автокорреляцию. По этой причине, преобразование по методу нормальных меток должно осуществляться после вычитания тренда, поскольку ковариация и вариограммы рассчитываются для остатков (значений, оставшихся после вычитания тренда). Сопоставьте это с логарифмическим преобразованием, когда исключаются любые связи между дисперсией и трендом. Следовательно, после логарифмических преобразований вычитаются тренд и модель автокорреляции (необязательно). Часто вследствие этого появляется примерно нормальное распределение остатков, но в целом это не является специфической целью логарифмических преобразований, в отличие от преобразований по методу нормальных меток.

## Использование преобразований (логарифмического, по методу Box–Cox и арксинуса)

Использование преобразований позволяет сделать дисперсии постоянными для всей изучаемой территории и привести данные ближе к нормальному распределению.

Воспользуйтесь гистограммой и нормальными графиками КК в инструментах ESDA, чтобы, испытав различные методы преобразований, получить нормальное распределение данных.

Некоторые методы геостатистики решающим образом зависят от данных, подчиняющихся нормальному распределению, — например, дизъюнктивный кригинг, карты квантилей и вероятности для ординарного, простого и универсального кригинга. Поэтому выполнение преобразований может привести данные к распределению, которое ближе к нормальному, чем у исходных данных.

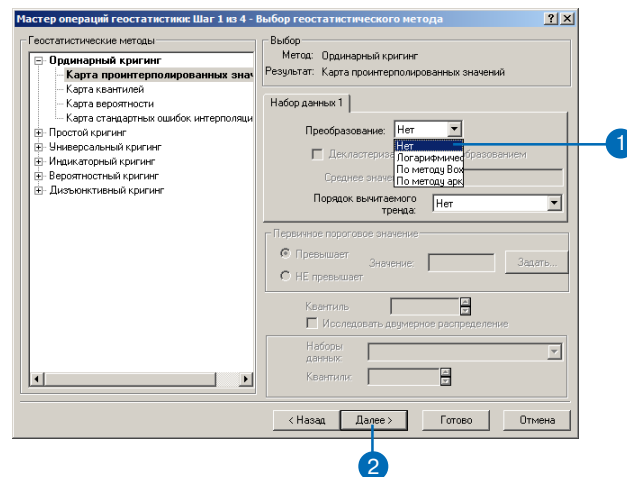
### Подсказка

#### Когда вы можете выполнять преобразования

Опция выполнения преобразований может быть использована, когда в диалоге Выберите входные данные и метод вы выбираете методы кригинга и кокригинга.

## Использование преобразований

1. В диалоге Выбор геостатистического метода выберите нужный метод преобразований из меню Преобразование.
2. Нажмите Далее.
3. Переходите от диалога к диалогу для того, чтобы построить поверхность.





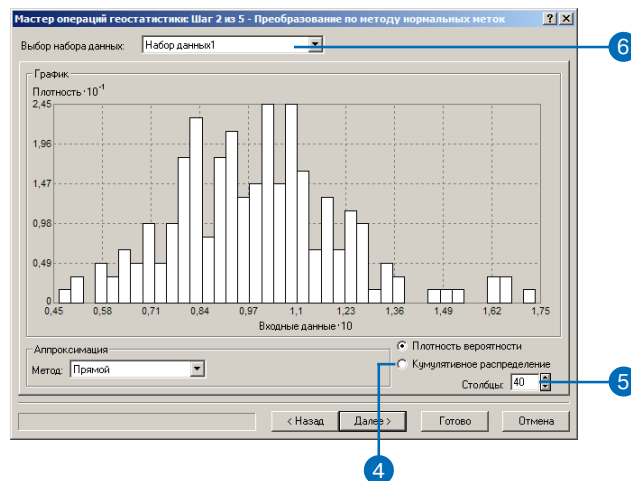
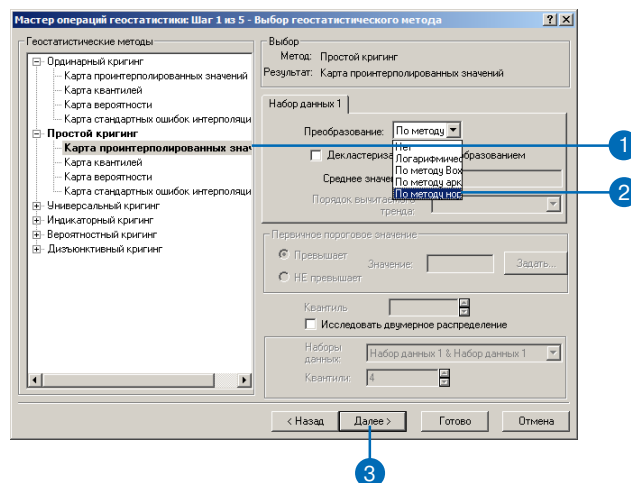
## Использование преобразований по методу нормальных меток

По методу нормальных меток данные будут преобразованы и приведены к одномерному нормальному распределению, что позволит применить к ним методы простого и дизъюнктивного кригинга.

Желательно сравнить подобранную модель с эмпирической функцией кумулятивного распределения для каждого из трех методов аппроксимации, используемых в преобразованиях по методу нормальных меток.

## Моделирование распределений

1. Для исследования распределения выберите интерполяцию по методу простого или дизъюнктивного кригинга.
2. В диалоге Выбор геостатистического метода, из меню Преобразование выберите опцию По методу нормальных меток.
3. Нажмите Далее.
4. Или, чтобы поменять способ отображения графика, выберите опцию Кумулятивное распределение.
5. Или, задайте количество столбцов, которое должно быть отображено на гистограмме.
6. В окне Выбор набора данных выберите другой набор данных (только для кокригинга, когда у вас есть два или большее количество наборов данных).



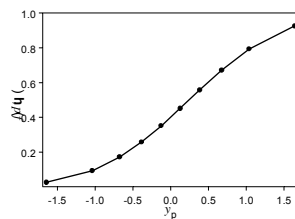
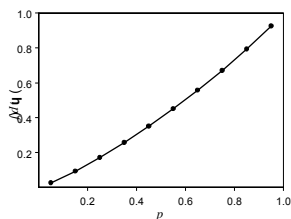
# Проверка на двумерное нормальное распределение

Дизъюнктивный кригинг требует соблюдения условия подчинения данных двумерному нормальному распределению. Помимо этого, для того, чтобы создать карты вероятности и квантилей, предположим, что данные подчиняются полному многомерному нормальному распределению. Чтобы проверить данные на одномерное нормальное распределение, вы можете воспользоваться нормальными графиками КК. Теперь проверим данные на двумерное нормальное распределение. (Ни одна из этих проверок не гарантирует, что данные подчиняются полному многомерному нормальному распределению, но иногда разумно выдвинуть это предположение, основываясь на использовании этих инструментов диагностики.) Рассмотрим следующее вероятностное утверждение:

$$f(p, h) = \text{Prob}[Z(s) \leq z_p, Z(s + h) \leq z_p]$$

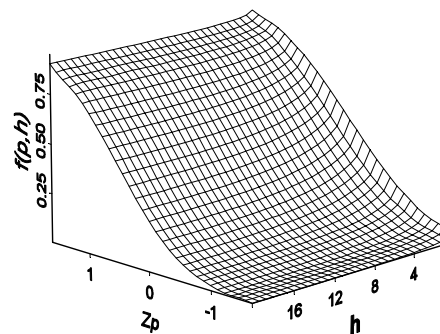
где  $z_p$  - стандартный нормальный квантиль для некоей вероятности  $p$ . Например, известны значения стандартных нормальных квантилей: при  $p = 0.975$ ,  $z_p = 1.96$ ; при  $p = 0.5$ ,  $z_p = 0$ ; а при  $p = 0.025$ ,  $z_p = -1.96$ . Выражение вероятности, приведенное выше, берет значение переменной  $Z$  в точке  $s$  и другой переменной  $Z$  в некоей точке  $s + h$ , и дает значение вероятности того, что *оба* этих значения меньше  $z_p$ . Эта формула вероятности является функцией  $f(p, h)$ , зависящей от  $p$  (и, соответственно,  $z_p$ ) и  $h$ . Функция будет также зависеть от степени автокорреляции между  $Z(s)$  и  $Z(s + h)$ .

Теперь предположим, что  $Z(s)$  и  $Z(s + h)$  подчиняются двумерному нормальному распределению. Если значение автокорреляции известно, для функции  $f(p, h)$  существуют формулы. Предположим, что  $h$  - константа, и только значение  $p$  меняется. Тогда можно предположить, что функция будет выглядеть следующим образом:



Правый график похож на график кумулятивного распределения вероятности. Теперь предположим, что значение  $p$  остается постоянным, и функция  $f(p, h)$  меняется в зависимости от  $h$ .

Предположим, что значение  $h$  очень мало. В этом случае, вероятность  $\text{Prob}[Z(s) \leq z_p, Z(s + h) \leq z_p]$  очень близка к вероятности  $\text{Prob}[Z(s) \leq z_p] = p$ . Далее, предположим, что значение  $h$  очень велико. В этом случае, вероятность  $\text{Prob}[Z(s) \leq z_p, Z(s + h) \leq z_p]$  очень близка к вероятности  $\text{Prob}[Z(s) \leq z_p] \cdot \text{Prob}[Z(s + h) \leq z_p] = p^2$  (потому что значения  $Z(s)$  и  $Z(s + h)$  практически независимы). Таким образом, для фиксированного значения  $p$ , следует ожидать, что значения функции  $f(p, h)$  будут меняться от  $p$  до  $p^2$ . Теперь рассмотрим выражение  $f(p, h)$  как функцию обоих параметров  $p$  и расстояния  $h$ , вы можете наблюдать нечто похожее на следующем рисунке.



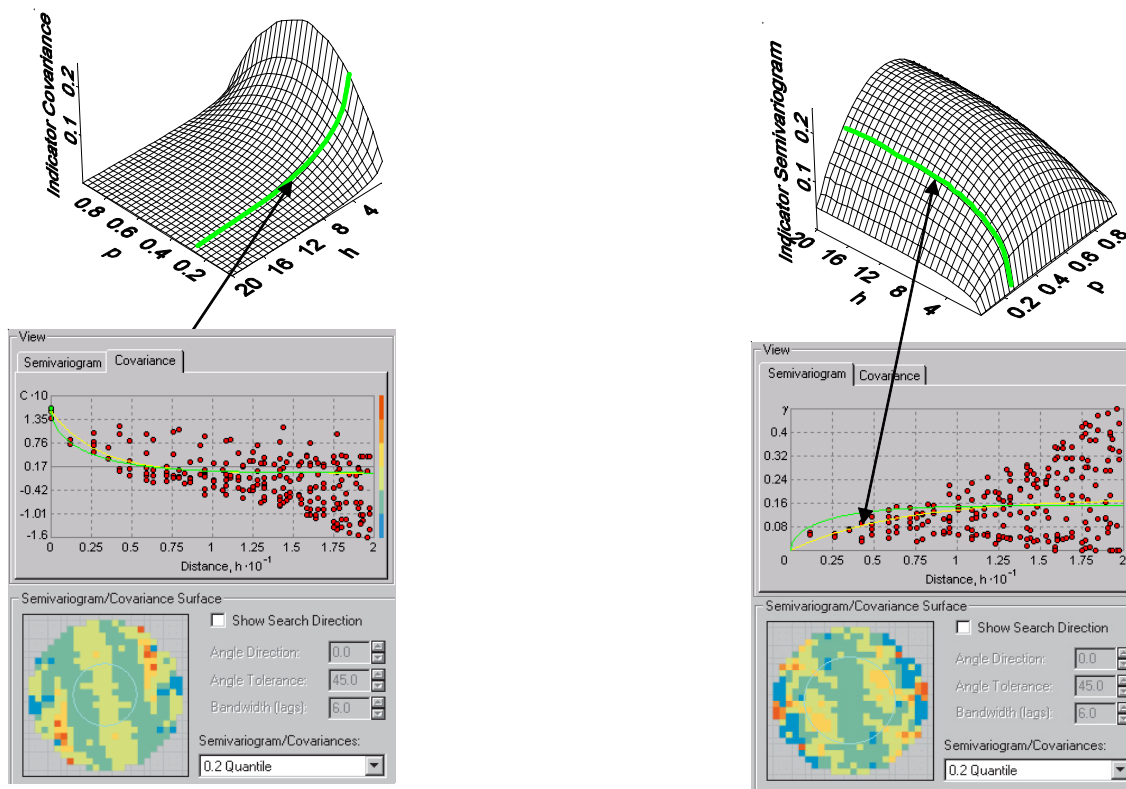
Эта функция может быть конвертирована в вариограммы и ковариационные функции для индикаторов. Если вероятность  $\text{Prob}[Z(s) \leq z_p, Z(s + h) \leq z_p] = E[I(Z(s) \leq z_p) I(Z(s + h) \leq z_p)]$ , где  $I(\text{утверждение})$  - индикаторная функция, которая равна 1, если *утверждение* истинно, или 0, если *утверждение* ложно, тогда ковариационная функция для индикаторов с фиксированным значением  $p$  выглядит следующим образом:

$$C_1(h; p) = f(p, h) - p^2$$

а вариограмма для индикаторов с фиксированным  $p$ :

$$g_1(h; p) = p - f(p, h)$$

Следовательно, вы можете оценить вариограмму и ковариационную функцию индикаторов исходных данных и использовать эту оценку для получения ожидаемых вариограмм и ковариационных функций индикаторов для различных значений  $p$ . Например, они могут выглядеть следующим образом:



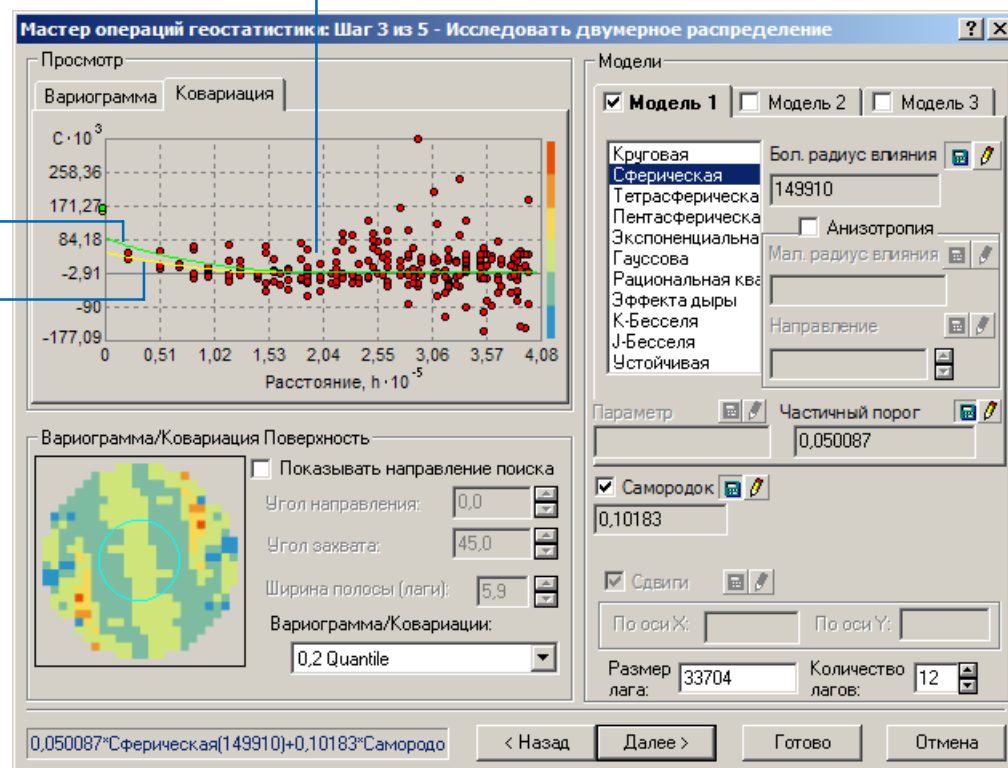
Красные точки на верхнем рисунке (в диалоге Моделирование вариограммы/ковариации) - это значения эмпирической ковариации и вариограммы для индикаторных переменных. Зеленая линия - это кривая, построенная теоретическим способом, для индикаторной вариограммы или индикаторной ковариации, предполагающая, что используемые данные подчиняются двумерному нормальному распределению, а желтая линия - модель, подобранная для исходных индикаторных данных. Таким образом, зеленая линия и желтая линия должны совпасть, если данные имеют двумерное нормальное распределение.

## Диалог Исследовать двумерное распределение

Теоретическая кривая  
ковариации  
индикаторов,  
предполагающая, что  
используемые данные  
подчиняются  
двумерному  
нормальному  
распределению  
(зеленая линия)

Подобранная  
ковариационная  
функция для  
эмпирической  
ковариации  
индикаторов  
(желтая линия)

Эмпирическая  
ковариация для  
индикаторов



## Проверка двумерного распределения

Это средство проверить, подчиняются ли данные двумерному нормальному распределению. Желтая и зеленая линии должны располагаться близко друг к другу, если данные подчиняются закону двумерного нормального распределения.

Чтобы можно было применить дизъюнктивный кригинг, необходимо, чтобы используемые данные имели двумерное нормальное распределение. Создание карт вероятности и квантилей с использованием метода простого кригинга требует условия соответствия данных многомерному нормальному распределению, и проверка двумерного распределения может помочь подтвердить предположение о многомерном нормальном распределении.

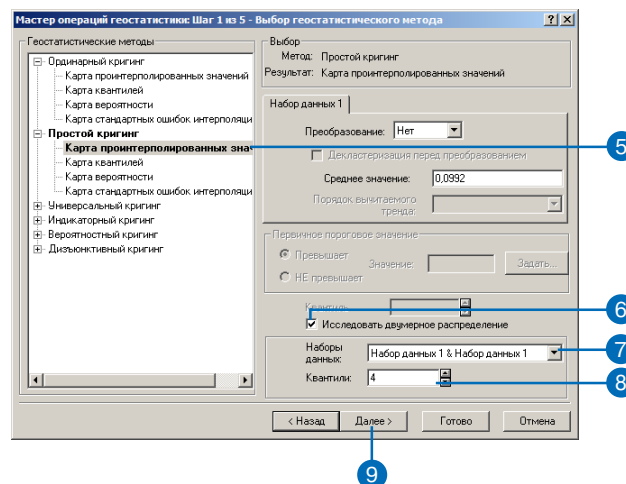
### Подсказка

**Методы, которые позволяют проверить двумерное распределение**

*В диалоге Выбор геостатистического метода выберите метод кригинга или кокригинга, а затем метод простого кригинга/коккригинга или дизъюнктивного кригинга/коккригинга.*

## Проверка двумерного распределения

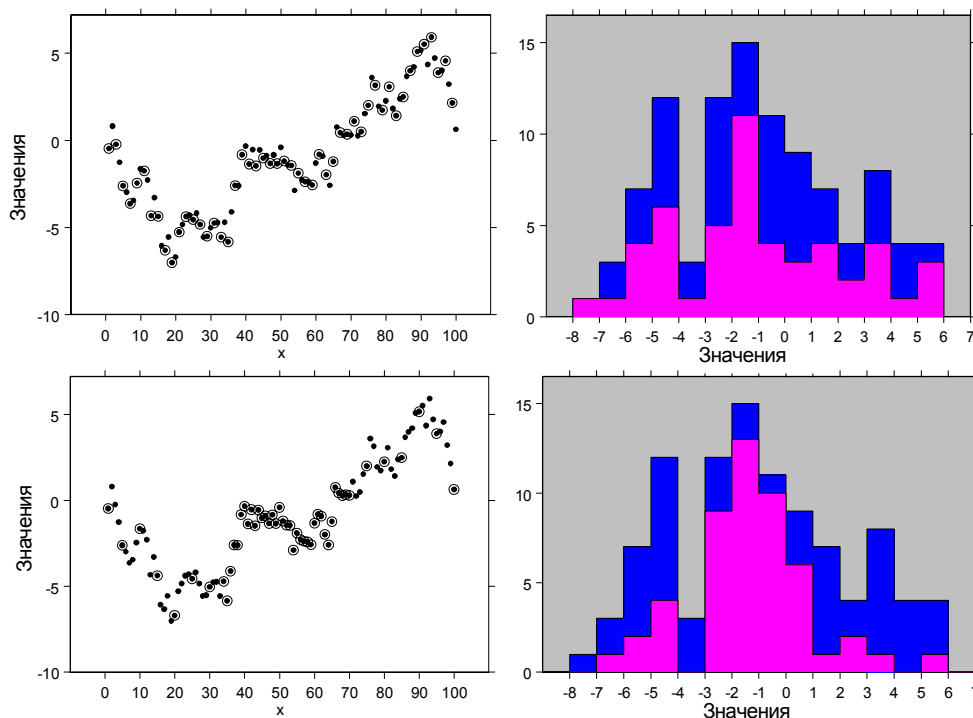
1. На панели инструментов ArcMap нажмите кнопку Добавить данные и добавьте слой, данные которого вы хотите проверить на двумерное распределение.
2. Запустите Мастер операции геостатистики.
3. Выберите метод кригинга или кокригинга.
4. Нажмите Далее в диалоге Выберите входные данные и метод.
5. Выберите метод простого кригинга/коккригинга или дизъюнктивного кригинга/коккригинга.
6. Отметьте опцию Исследовать двумерное распределение и выберите опцию Преобразование по методу нормальных меток.
7. Выберите комбинацию наборов данных, которую вы хотите использовать (только для кокригинга).
8. Наберите значение количества проверяемых квантилей.
9. Нажмите Далее.
10. Выберите Метод аппроксимации, задайте параметры, и отметьте либо опцию Плотность вероятности либо Кумулятивное распределение в диалоге Преобразование по методу нормальных меток. Нажмите Далее. ►



11. Задайте необходимые параметры в диалоге Моделирование вариограммы/ковариации и нажмите Далее.
12. Изучите двумерное распределение в диалоге Исследование двумерного распределения. Нажмите Далее.

## Применение декластеризации для данных, отобранных с различной плотностью

Очень часто опорные точки не распределены в пространстве равномерно или случайным образом. По разным причинам, данные могут отбираться в одних районах с большей плотностью, чем в других. Как вы убедились в предыдущем разделе, для правильного применения преобразований по методу нормальных меток важно, чтобы гистограмма (и кумулятивное распределение) выборки правильно отражала гистограмму для всей совокупности данных. Если данные неравномерно распределены в пространстве и при этом пространственно автокоррелируют, результирующая гистограмма выборки может не соответствовать гистограмме для данных всей совокупности. На следующих рисунках приведен пример, иллюстрирующий это утверждение.

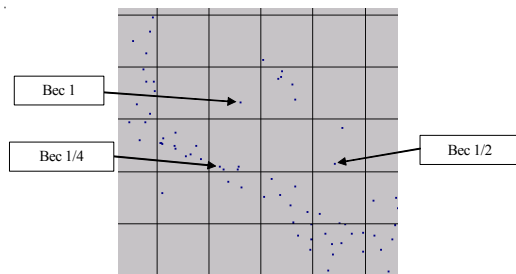


На левом верхнем рисунке, вся совокупность значений в 100 точках вдоль линии показана черными точками. Эти значения были получены с использованием пространственно коррелированного процесса с постоянным средним значением и сильной позитивной корреляцией. Выборка - это все другие точки, начиная с первой, показанные маленькими окружностями. Справа показаны гистограммы: синим цветом - гистограмма совокупности, малиновым цветом - гистограмма выборки.



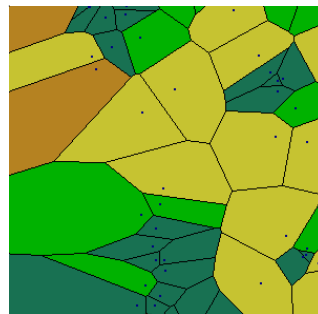
Поскольку выборка равна примерно половине всей совокупности, столбики гистограммы выборки должны быть равны половине высоты столбиков гистограммы совокупности, с некоторыми отклонениями. На левом нижнем рисунке на стр.211 выборка сделана с различной густотой: до 34-ой точки отобрана каждая пятая, затем до 70-ой отобраны все точки, а затем до конца - снова каждая пятая. В конечном итоге выборка снова включает половину точек популяции. Выборка с различной густотой до середины опорных точек приводит к тому, что выборка содержит высокую долю значений, близких к среднему, и, следовательно, столбики гистограммы примерно равны столбикам гистограммы для всей популяции для значений от -3 до 1. Вследствие этого, на гистограмме выборки плохо представлены очень низкие и очень высокие значения.

Одно из решений, которое может быть использовано для выборки с различной густотой, состоит во взвешивании данных; при этом опорным точкам, расположенным на участках с высокой густотой точек, присваиваются меньшие веса (что приведет к уменьшению высоты столбиков гистограммы выборки для значений от -3 до 1 для данных, отобранных с различной густотой, из предыдущего примера), а точкам, расположенным по территории разреженно, присваиваются большие веса (что приведет к увеличению высоты столбиков гистограммы для низких и высоких значений данных). Модуль Geostatistical Analyst позволяет использовать два метода. Метод, предложенный по умолчанию, - это декластеризация по ячейкам). При использовании этого метода, прямоугольные ячейки образуют грид, совмещаемый с опорными точками, и каждой точке ячейки присваивается вес, обратно пропорциональный количеству точек, попадающих в эту ячейку. Пример приведен на следующем рисунке.



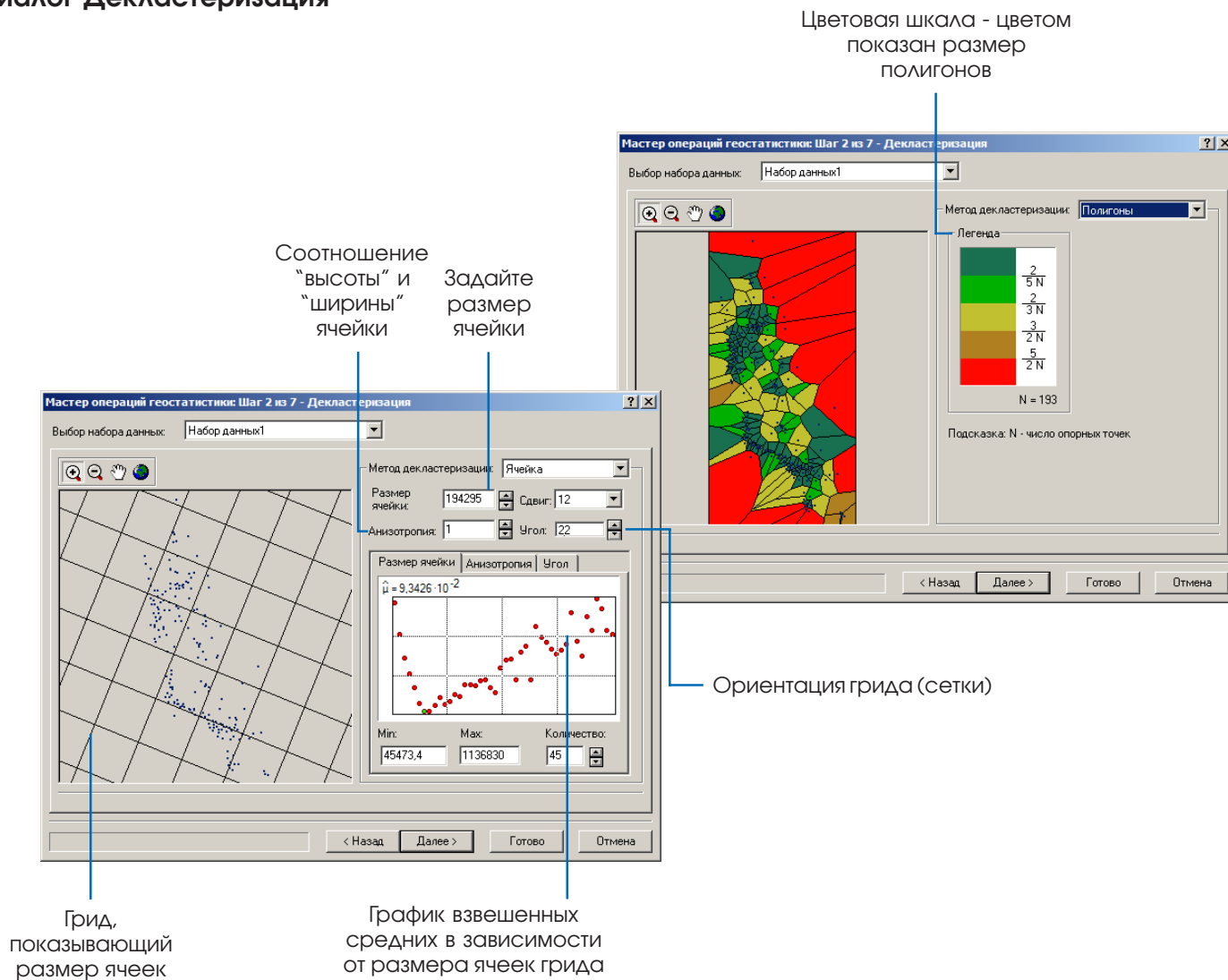
Все, что остается сделать вам, - это выбрать размер грида. Можно принять несколько схем, и для того, чтобы выбрать какую из этих схем лучше применить, вы можете обратиться к справочной литературе. Один из инструментов модуля Geostatistical Analyst позволяет построить график, на котором показано взвешенное среднее значение для всех данных при различных размерах ячеек. Предлагается выбрать размер ячейки, соответствующий минимальному взвешенному среднему, если данные были отобраны с различной густотой в районах с высокими значениями, и, напротив, выбрать размер ячейки, соответствующий максимальному взвешенному среднему, если данные были отобраны с различной густотой в районах с низкими значениями.

Другая схема использует полигональный метод, при котором для каждой опорной точки полигон определяется таким образом, чтобы все точки, попадающие внутрь этого полигона, были бы ближе к этой точке, чем к любой другой опорной точке (см. рисунок).



Опорные точки показаны маленькими точками, а вокруг них нарисованы полигоны, цвет которых соответствует размеру полигонов. Идея состоит в том, чтобы присвоить каждой опорной точке вес, пропорциональный площади, которую он "образует". Проблема этого метода состоит в том, что для краевых точек трудно корректно определить значения весов. Краевым точкам часто могут быть присвоены завышенные значения весов, так как полигон простирается до границы области исследования. В модуле Geostatistical Analyst, граница - это прямоугольник, что часто приводит к тому, что краевым точкам присваиваются высокие значения.

## Диалог Декластеризация



# Декластеризация данных, отображенных с различной густотой

Существует два способа выполнить декластеризацию ваших данных: методом подбора грида и методом полигонов Вороного.

Выборка должна быть сформирована таким образом, чтобы быть репрезентативной для всей поверхности. Однако, часто выборка осуществляется в местах с наибольшей концентрацией точек, что приводит к искажению отображения поверхности. Декластеризация учитывает неравномерное расположение опорных точек путем их соответствующего взвешивания, что приводит к возможности более точного построения поверхности.

## Подсказка

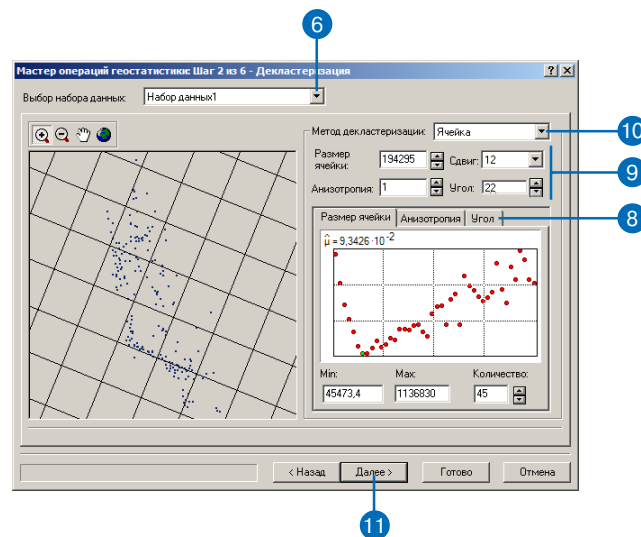
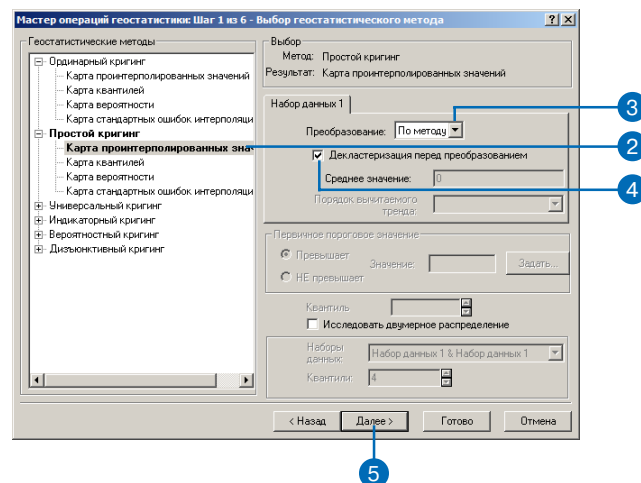
### Использование декластеризации

Декластеризация может быть применена только в том случае, если используется преобразование по методу нормальных меток.

Чтобы иметь возможность воспользоваться преобразованием по методу нормальных меток, выберите методы вероятностного, простого или дизъюнктивного кригинга/кокринга.

## Выполнение декластеризации по методу ячейки

1. В диалоге Выберите входные данные и метод выберите кригинг или кокринг.
2. Выберите один из методов кригинга или кокринга: вероятностный, дизъюнктивный, или простой.
3. В диалоге Преобразование выберите преобразование по методу нормальных меток.
4. Выберите опцию Декластеризация перед преобразованием.
5. Нажмите Далее.
6. В меню Выбор набора данных определите набор данных, который вы хотите отобразить (только для кокринга).
7. Задайте необходимые параметры.
8. Выбирайте различные закладки для того, чтобы просмотреть диаграммы, отражающие изменение выборки в зависимости от Размера ячейки, Анизотропии и Угла наклона ячейки.
9. Измените размер ячейки, анизотропию, смещение и угол наклона ячейки, чтобы определить экстремум графика.
10. Или из меню Метод декластеризации выберите Полигоны, чтобы отобразить полигоны, по которым будет выполняться декластеризация.
11. Нажмите Далее. ►



12. Из меню Аппроксимация выберите метод аппроксимации, определите параметры, и отметьте опцию Плотность вероятности или Кумулятивное распределение в диалоге Преобразование по методу нормальных меток. Нажмите Далее.
13. В диалоге Моделирование вариограммы/ковариации задайте требуемые параметры и нажмите Далее.
14. Определите параметры в диалоге Поиск соседства и нажмите Далее.
15. Изучите результаты в диалоге Перекрестная проверка и нажмите Готово.
16. В диалоге Информация о результирующем слое нажмите ОК.

## Вычитание трендов из данных

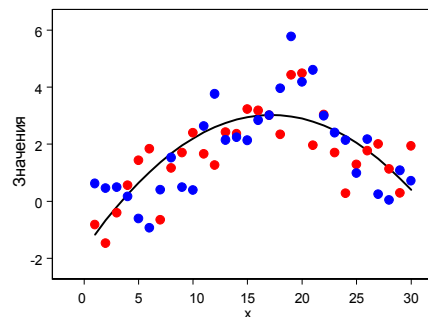
Иногда вам может понадобиться вычесть тренд поверхности из данных и применить кригинг или кокригинг для данных с вычтенным трендом, которые носят название остатков. Рассмотрим аддитивную модель,

$$Z(s) = m(s) + e(s)$$

где  $m(s)$  - некая детерминистская поверхность (которая носит название тренда) и  $e(s)$  - пространственно коррелированная ошибка. Теоретически, тренд зафиксирован, что означает, что если вы будете моделировать данные снова и снова, тренд меняться не будет. Однако, вы увидите флуктуации в моделируемых поверхностях, что вызвано наличием автокоррелированных случайных ошибок. Обычно тренд в пространстве меняется постепенно, в то время как случайные ошибки меняются быстрее. Метеорологическим примером тренда может быть наблюдаемое вами (и известное из теории) постепенное изменение температуры в зависимости от широты. Однако, метеорологические наблюдения для каждого конкретного дня отражают локальные отклонения, которые возникают вследствие движения фронтальных масс, различий в подстилающей поверхности, облачности, и т.п. и являются труднопредсказуемыми, следовательно, локальные отклонения моделируются как автокоррелирующие.

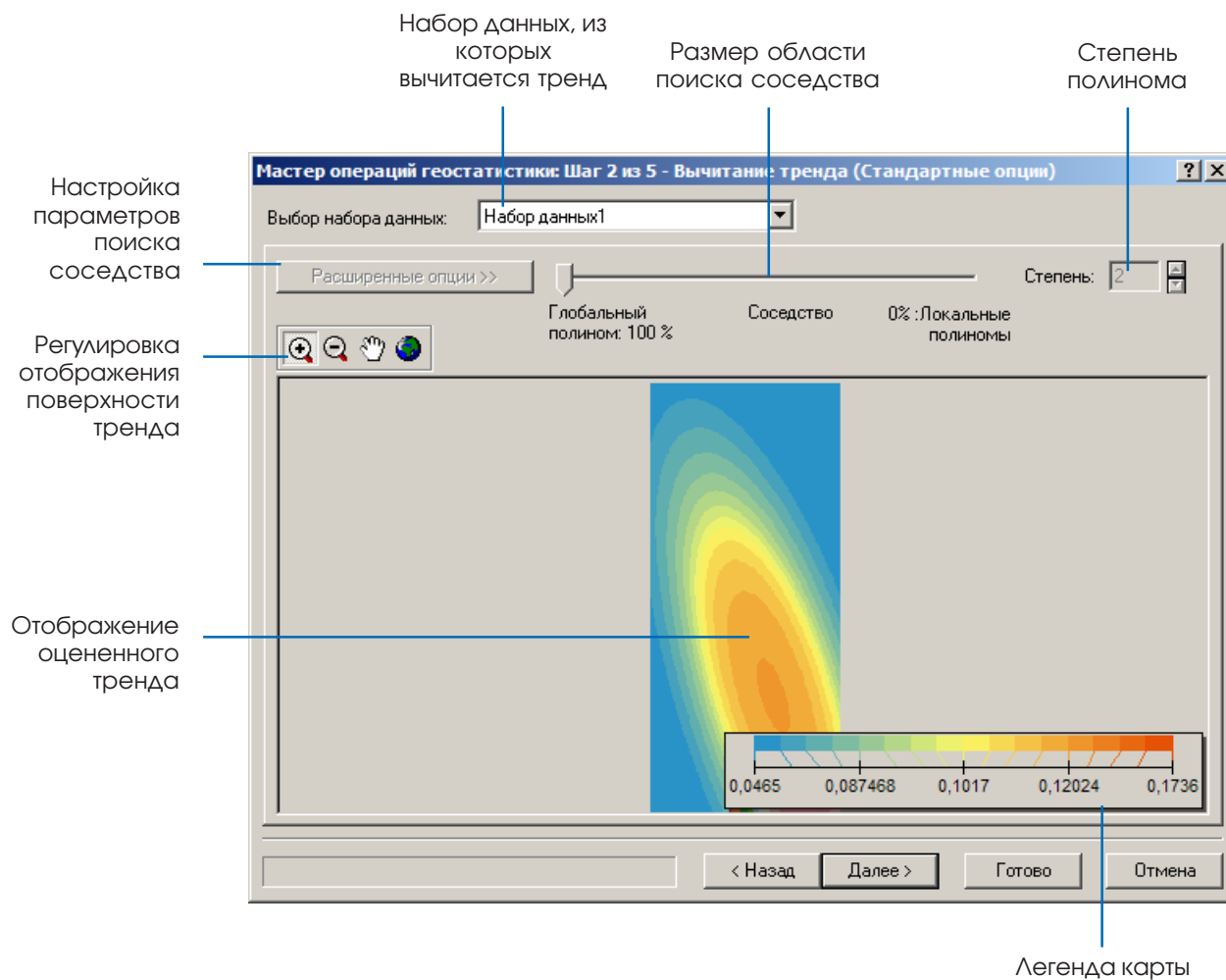
К сожалению, нет никакого магического способа однозначно разложить данные на тренд и случайные ошибки. Изложенные далее соображения могут помочь вам в этом процессе. На следующем рисунке показаны данные, полученные при помощи двух моделей. Одна из них - ординарный кригинг, где  $Z(s) = m + e(s)$  и ошибки  $e(s)$  коррелируют. Процесс имеет среднее  $m = 0$  и экспоненциальную вариограмму. Другой набор данных был смоделирован с использованием универсального кригинга:  $m(s) = b_0 + b_1x(s) + b_2x^2(s)$  (сплошная линия на рисунке); но ошибки были независимы, со средним значением, равным 0, и дисперсией, равной 1.

Как видите, трудно определить, какие точки соответствуют какой модели (синие точки соответствуют значениям, полученным при использовании ординарного кригинга, а красные точки - универсального кригинга с независимыми ошибками). Пространственная автокорреляция может позволить строить различные поверхности интерполяции, и этот пример демонстрирует, что



часто может быть трудно выбрать между моделями, основанными только на данных. В целом, вы должны придерживаться ординарного кригинга до тех пор, пока у вас не возникнут серьезные причины для вычитания поверхности тренда. Это вызвано тем, что лучше использовать простые модели. Если вы вычитаете поверхность тренда, параметров, которые вам необходимо оценить, больше. Двумерная поверхность второй степени к отсекающему параметру добавляет еще пять параметров, требующих оценки. Чем большее количество параметров оценивается, тем менее точной становится модель. Однако, в некоторых случаях пространственные координаты служат как заменитель для некоего известного тренда в данных. Например, урожайность может меняться в зависимости от широты—причиной этого являются не сами координаты, а такие показатели, меняющиеся с широтой, как температура, влажность, осадки и т.д. В данном случае, вычитание поверхностей тренда может иметь смысл. Повторим снова, что чем проще поверхность (полином первой или второй степени), тем лучше. В модуле Geostatistical Analyst возможно также применение локального полиномиального сглаживания как опции вычитания тренда. Существует реальная опасность слишком точного подбора данных при использовании трендов и сохранении слишком маленькой дисперсии в отклонениях, что приводит к ошибкам и некорректному учету неопределенности интерполяции. При применении моделей, учитывающих тренд, всегда проверяйте их с использованием перекрестной проверки и особенно проверки.

## Диалог Вычитание тренда (Стандартная опция)



## Удаление глобальных и локальных трендов из данных: вычитание тренда

Чтобы определить присутствует ли в ваших данных глобальный тренд, воспользуйтесь инструментом ESDA, который носит название Анализ тренда (см. Главу 4, 'Исследовательский анализ пространственных данных').

По существу, вы раскладываете свои данные на детерминистскую составляющую - тренд, и случайную автокоррелированную составляющую. После вычитания тренда, кригинг будет выполнен для остатков. Перед выполнением окончательной интерполяции, тренд снова прибавляется к результирующей поверхности.

### Подсказка

#### Степень трендов

Для определения поверхности тренда используйте глобальные полиномы низких степеней, если только у вас нет серьезных оснований поступить по-другому.

## Интерактивная оценка тренда

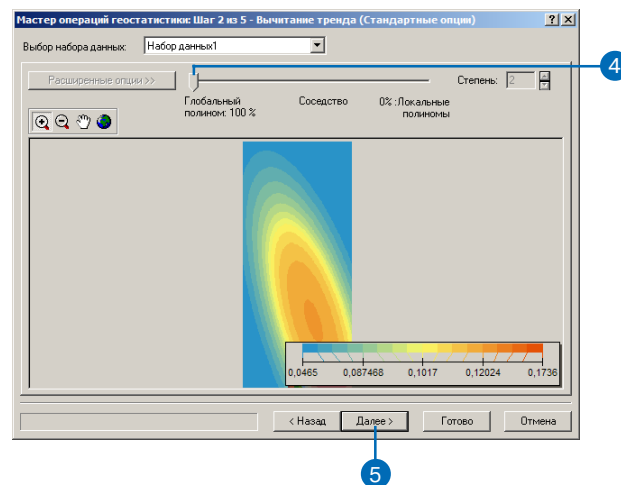
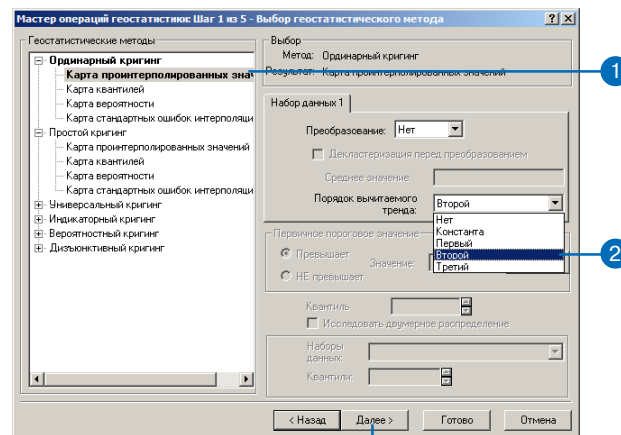
1. В диалоге Выбор геостатистического метода выберите ординарный, универсальный кригинг или дизъюнктивный кригинг и нужный тип результирующей поверхности в диалоге Геостатистические методы.
2. В меню Порядок вычитаемого тренда задайте степень полинома.
3. Нажмите Далее.

Примечание: Если вы зададите какую-либо степень полинома (все опции, кроме Нет), после нажатия кнопки Далее появится диалог Вычитание тренда.

4. В диалоге Вычитание тренда, перемещая движок между двумя крайними положениями, вы определяете размер окна, позволяющего перейти от глобального к локальному полиному.

Вы можете также задать параметры поиска соседства, нажав на кнопку Расширенные опции.

5. Нажмите Далее.





# Отображение геостатистических слоев и управление ими

## 8

### В ЭТОЙ ГЛАВЕ

- Что такое геостатистический слой?
- Добавление слоев
- Работа со слоями на карте
- Управление слоями
- Просмотр геостатистических слоев в ArcCatalog
- Представление геостатистического слоя
- Изменение символики геостатистического слоя
- Классификация данных
- Определение масштаба, при котором будет отображаться геостатистический слой
- Экстраполяция значений точек, находящихся за пределами исследуемой территории
- Сохранение и экспорт геостатистических слоев

ArcMap и модуль Geostatistical Analyst предлагают множество инструментов, с помощью которых вы можете отображать данные и управлять ими. Воспользовавшись инструментами отображения, вы можете получить прекрасную карту и, проанализировав ее, глубже понять свои данные с тем, чтобы иметь возможность принимать более эффективные решения. Исследование данных особенно важно при использовании модуля Geostatistical Analyst, поскольку через понимание вы приходите к созданию более правильных моделей и построению более точных поверхностей.

Хотя инструменты управления и не помогают напрямую при построении поверхностей в модуле Geostatistical Analyst, эти инструменты незаменимы для организации вашей работы с картой и могут помочь вам в процессе обдумывания решений.

Многие из инструментов отображения данных и управления ими, используемые для работы с любым из слоев ArcMap, применимы также для работы с геостатистическим слоем. В этой главе, мы рассмотрим только основные инструменты работы со слоями данных, наиболее часто используемые в модуле Geostatistical Analyst, и те инструменты, которые используются только для работы с геостатистическими слоями. Обратитесь, пожалуйста, к книге *“ArcMap. Руководство пользователя”* для информации о других инструментах, применимых ко всем типам слоев данных, включая геостатистические слои.

## Что такое геостатистический слой?

В ArcMap географические данные организованы по слоям. Для представления различных данных существуют различные типы слоев. Слой пространственных объектов ArcMap может содержать полигоны, соответствующие типам почв, точки, определяющие объем биомассы, величина которого измерена в определенных местах на поверхности Земли, или линии, которыми показаны, к примеру, тропы. Растровый слой может представлять собой аэрофотографию или грид расстояний от дорог. Другие типы слоев - это слои TIN для трехмерных поверхностей, слой CAD для хранения листов карты CAD и геостатистический слой, в котором хранятся результаты анализа, проведенного с помощью модуля Geostatistical Analyst.

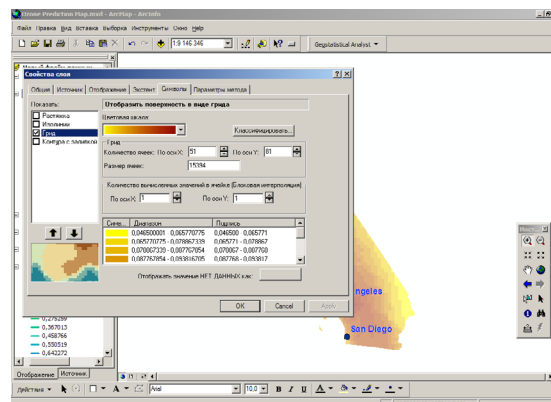
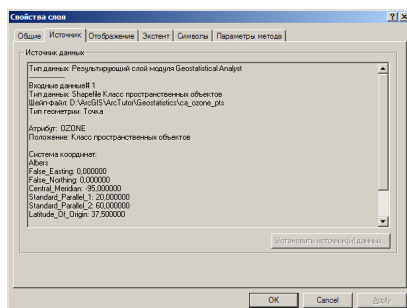
Функциональность геостатистического слоя аналогична функциональности любого слоя ArcMap. Вы можете добавлять его в ArcMap, удалять его, отображать и менять символику. Однако, геостатистический слой отличается от других слоев способом создания и хранения. Геостатистический слой может быть создан только при использовании модуля Geostatistical Analyst. Большинство типов слоев ArcMap хранят ссылку на источник данных, символику отображения слоя и другие характеристики. Геостатистический слой хранит источник данных, с использованием которых он был создан (обычно точечный слой пространственных объектов), символику и другие характеристики, наряду с параметрами модели интерполяции. В диалоге Свойств

ва геостатистического слоя вы можете просмотреть данные об исходном источнике данных и о параметрах модели.

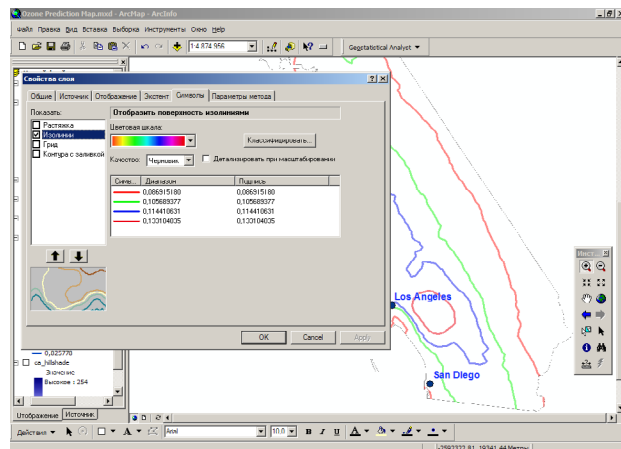
Помимо определения источника исходных точек и параметров модели, вы можете также получить общую информацию о слое в закладке Общие, просмотреть и изменить область отображения карты в закладке Экстент, изменить символогию в закладке Символы, установить степень прозрачности слоя и определить необходимость отображения всплывающих подсказок в закладке Отображение.

Геостатистический слой можно отобразить четырьмя следующими способами: контурами с заливкой, изолиниями, гридом или растяжкой цвета. Вы можете комбинировать несколько способов при отображении одного слоя для достижения различных эффектов. Полный диапазон символов и контролируемых параметров существует для каждого из возможных форматов.

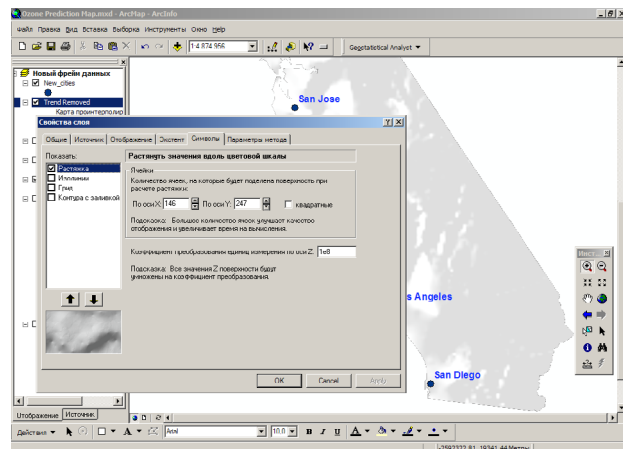
## Контура с заливкой



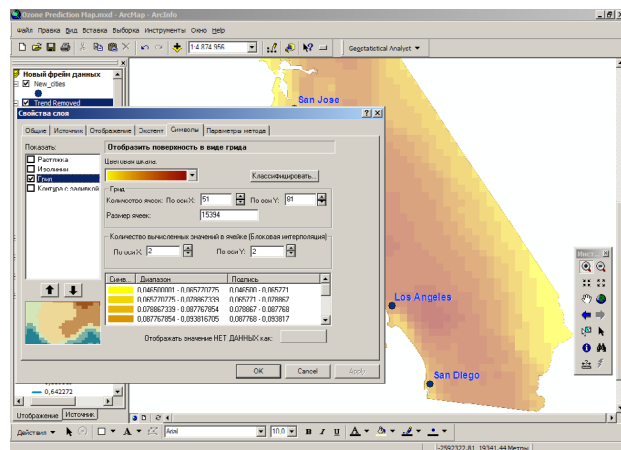
## Изолинии



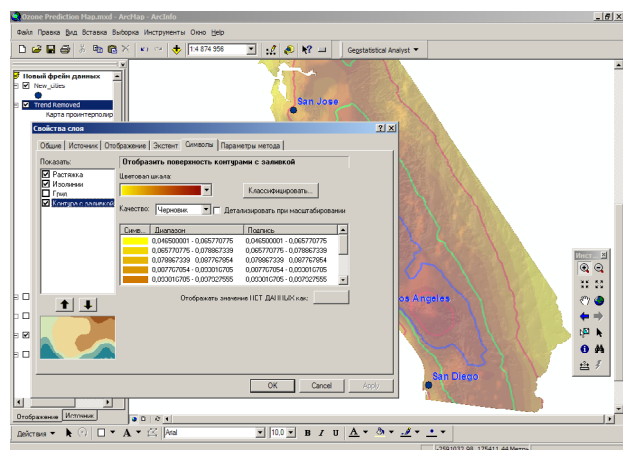
## Растяжка



## Грид



## Сочетание изолиний, контуров с заливкой и растяжки



## Добавление слоев

В ArcMap географические данные организованы по слоям. Существует несколько различных типов слоев, но при использовании модуля Geostatistical Analyst наибольший интерес представляют точечный, полигональный и растровый.

Слой пространственных объектов может содержать полигональные пространственные объекты, такие как типы почв или землепользования, линейные пространственные объекты, такие как дорожная сеть, или точечные пространственные объекты, такие как места преступлений или точки отбора проб на загрязняющие вещества. В модуле Geostatistical Analyst в качестве исходного чаще всего используется точечный слой. Геостатистический слой представляет собой поверхность, созданную в результате анализа, выполненного с использованием модуля Geostatistical Analyst. Растровый слой может содержать изображение, полученное со спутника, отсканированный снимок или изображение типов леса при помощи грида. Часто, для последующего анализа, геостатистический слой конвертируется в растровый слой.

### Подсказка

#### Отображение “спрятанных” слоев

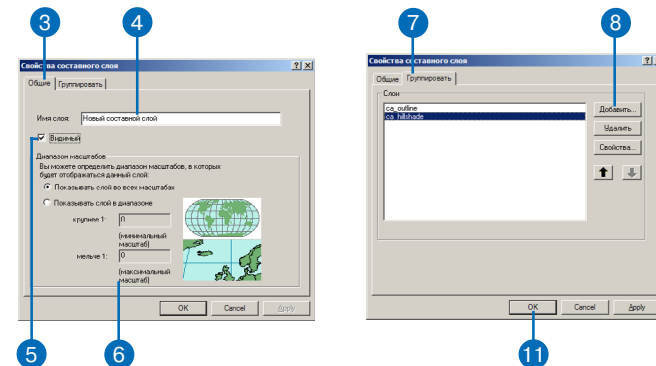
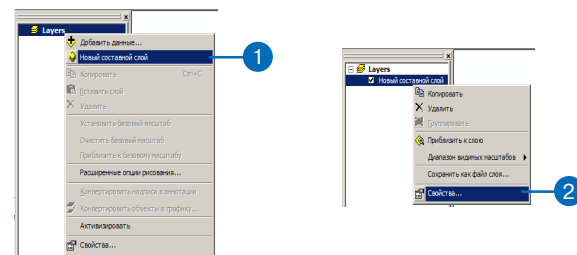
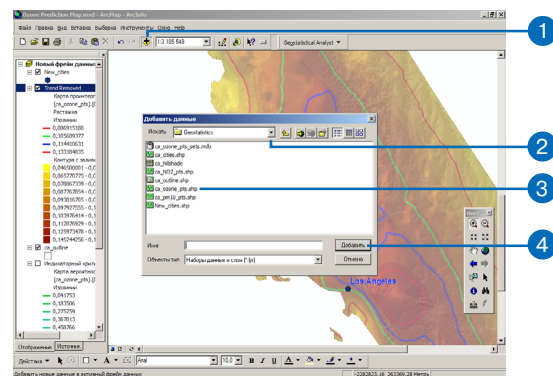
Слои, которые вы добавляете, могут быть “спрятаны”. Чтобы отобразить слой, который вы не видите, нажмите на нем правой клавишей мыши и выберите опцию Приблизить к слою.

## Добавление слоев

1. На стандартной панели инструментов ArcMap нажмите кнопку Добавить данные.
2. В меню Искать в... перейдите к папке, в которой хранится нужный слой.
3. Выберите слой.
4. Нажмите Добавить.

## Добавление группы слоев

1. В таблице содержания ArcMap нажмите правую клавишу мыши на строке Слои и выберите опцию Новый составной слой.
2. В таблице содержания выберите созданный новый составной слой, правой клавишей мыши откройте контекстное меню и выберите опцию Свойства.
3. Отобразите содержимое закладки Общие.
4. Необязательный шаг: назовите составной слой.
5. Необязательный шаг: отметьте галочкой опцию Видимый, чтобы сделать составной слой видимым.
6. Необязательный шаг: определите диапазон масштабов.
7. Откройте закладку Группировать.
8. Нажмите кнопку Добавить и перейдите к набору данных, который вы хотите добавить в составной слой.
9. Нажмите Добавить.
10. Продолжайте добавлять в группу требуемые наборы данных, повторяя шаги 8, 9.
11. Нажмите ОК.





## Работа со слоями на карте

В ArcMap существует множество инструментов, которые позволяют вам работать со слоями на карте. Некоторые из наиболее часто употребляемых и особенно полезных для пользователя, работающего с модулем Geostatistical Analyst, - включение и выключение отображения слоя, контроль за порядком отображения слоев, изменение масштаба и области отображения слоя. Эти инструменты могут быть использованы для анализа или для создания качественного картографического материала.

При использовании этих инструментов для анализа вы можете изучить исходный точечный слой, значения которого будут проинтерполированы, а также исследовать результирующий геостатистический слой.

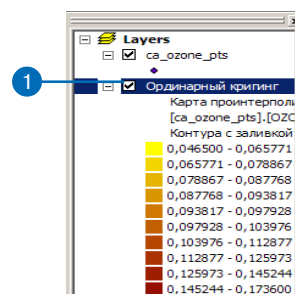
Если вы удовлетворены представлением результатов, вы можете сохранить их в виде карты.

### См. также

*См. также книгу Использование ArcMap для информации о дополнительных инструментах, которые регулируют отображение слоя на карте.*

## Включение и выключение слоя

1. Поставьте галочку в окошке, расположенном слева от названия слоя, чтобы отобразить его, и уберите галочку, чтобы выключить отображение слоя.

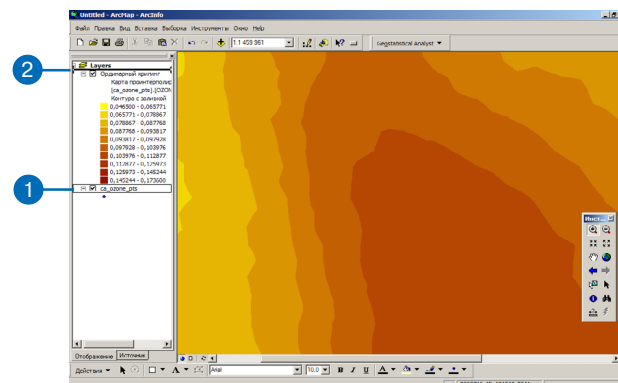


## Перемещение слоя с целью изменения порядка отображения данных

1. В таблице содержания выделите слой и, удерживая, перетяните его вверх или вниз.

Черная линия показывает, куда будет помещен слой.

2. Отпустите клавишу мыши в том месте, где вы хотите поместить слой.



## Масштабирование и изменение области отображения слоя

1. Откройте меню Вид на панели меню ArcMap, выберите опцию Панели инструментов и отметьте опцию Инструменты.
2. Для исследования карты могут быть использованы различные инструменты данной панели.

Увеличить  
Фиксированное увеличение  
Переместить  
Предыдущий экстенд  
Выбрать объекты  
Идентифицировать объекты  
Измерить



Уменьшить  
Фиксированное уменьшение  
Полный экстенд  
Следующий экстенд  
Выбрать элементы  
Найти  
Гиперссылка

## Управление слоями

ArcMap предоставляет набор инструментов для управления слоями данных. Хотя эти инструменты не помогают в исследовании данных или их анализе, они помогают упорядочить работу с картой. Наиболее полезны при работе с модулем Geostatistical Analyst следующие функции: сохранение, переименование, копирование и удаление геостатистических слоев.

Вы можете захотеть переименовать геостатистический слой, поскольку по умолчанию он создается с названием, совпадающим с методом, использованным при его создании и соответствующим порядковым номером (например, Ординарный кригинг\_2). Это может привести к путанице, если вы строите несколько поверхностей с применением одного метода, но различных параметров, или одного метода для различных наборов данных.

Копирование геостатистического слоя особенно полезно, когда вы хотите создать результирующую поверхность другого типа с теми же параметрами модели.

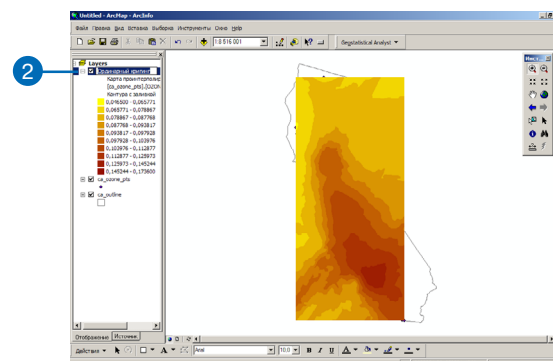
Полезно также удалять промежуточные геостатистические слои, которые вы использовали при исследовании параметров модели.

## Изменение названия слоя

1. Выберите слой в таблице содержания.
2. Повторно нажмите на названии слоя.

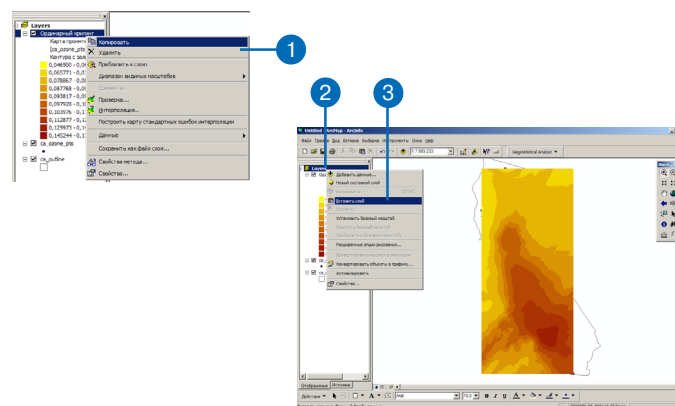
Название будет выделено, что позволит изменить его.

3. Напечатайте новое название.



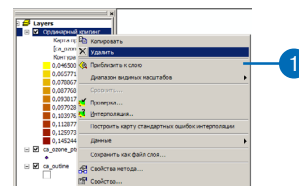
## Копирование слоя

1. Нажмите правую клавишу мыши, выбрав слой, который вы хотите скопировать, и выберите опцию Копировать.
2. Нажмите правую клавишу мыши на строке Слои.
3. Выберите Вставить слой(и).



## Удаление слоя

1. Нажмите правую клавишу мыши, выбрав слой, который вы хотите удалить и воспользуйтесь опцией Удалить.



## Просмотр геостатистических слоев в ArcCatalog

В ArcCatalog можно просмотреть существующие геостатистические слои и выполнить над ними определенные операции. ArcCatalog позволяет быстро пролистать данные и установить связи с базами данных и папками, размещенными как на вашем компьютере, так и на сетевых ресурсах.

В ArcCatalog вы можете выполнить предварительный просмотр отображения геостатистического слоя, либо увидеть связанные с ним метаданные.

### Подсказка

#### Метаданные

Стратегия исчерпывающих метаданных существенна для отслеживания географических или геостатистических данных.

### Подсказка

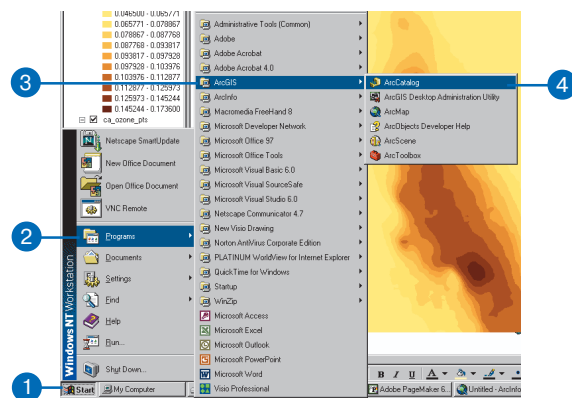
#### Доступ к ArcCatalog

Открыть ArcCatalog можно также, нажав на соответствующую иконку на стандартной панели инструментов ArcMap.

## Запуск ArcCatalog и модуля Geostatistical Analyst

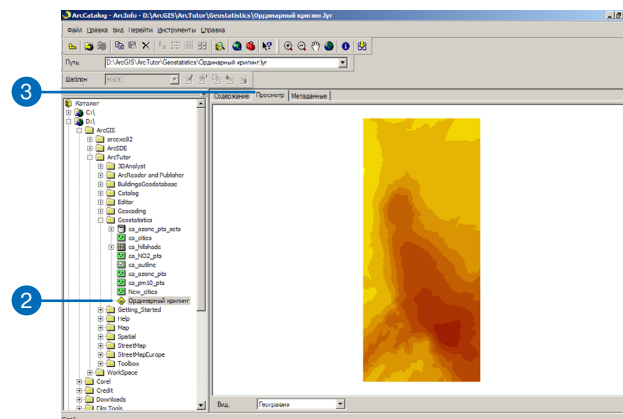
1. Нажмите кнопку Пуск на панели задач Windows.
2. Перейдите в меню Программы.
3. Выберите ArcGIS.
4. Выберите ArcCatalog.

Откроется окно ArcCatalog. В ArcCatalog, в меню Инструменты выберите опцию Дополнительные модули и отметьте галочкой модуль Geostatistical Analyst. Нажмите Закрывать.



## Предварительный просмотр данных

1. Запустите ArcCatalog.
2. В таблице содержания перейдите к нужному геостатистическому слою
3. Выберите закладку Просмотр.



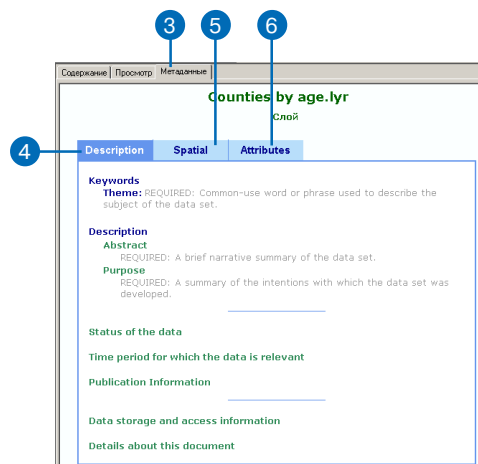


## См. также

Обратитесь к книге Using ArcCatalog (Использование ArcCatalog) для дополнительной информации о метаданных и способах их создания.

## Просмотр метаданных слоя

1. Откройте ArcCatalog.
2. В таблице содержания перейдите к нужному геостатистическому слою.
3. Выберите закладку Метаданные.
4. Выберите закладку Описание, чтобы просмотреть общее описание слоя.
5. Выберите закладку Пространственная привязка, чтобы изучить пространственные характеристики слоя, такие как координаты границ области отображения.
6. Выберите закладку Атрибуты, чтобы получить другую информацию о слое.
7. Введите любые метаданные или измените существующие.



## Отображение геостатистического слоя

Геостатистический слой может быть отображен четырьмя способами.

Контур с заливкой - это полигональное представление геостатистического слоя. При таком графическом отображении предполагается, что все точки, находящиеся внутри полигона, имеют одно и то же значение.

Грид - это растровое представление геостатистического слоя. Предполагается, что интерполяция выполняется для точки, являющейся центром каждой ячейки, или для каждой ячейки вычисляется среднее значение.

Для отображения геостатистического слоя могут также использоваться изолинии. Вы можете показывать линии с различным качеством (черновик или чистовое оформление).

При использовании растяжки (отмывки) создается рельефное изображение геостатистического слоя.

Одновременно можно применять несколько способов отображения.

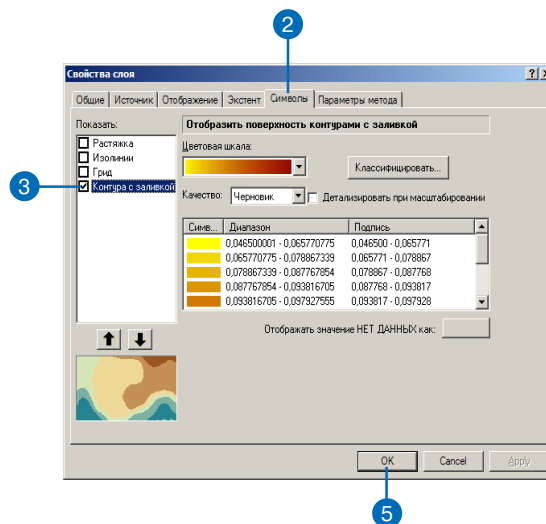
### Подсказка

#### Отображение нескольких наборов данных

Чтобы видеть одновременно два набора данных, отобразите один из них изолиниями и наложите их на поверхность, построенную для другого набора данных.

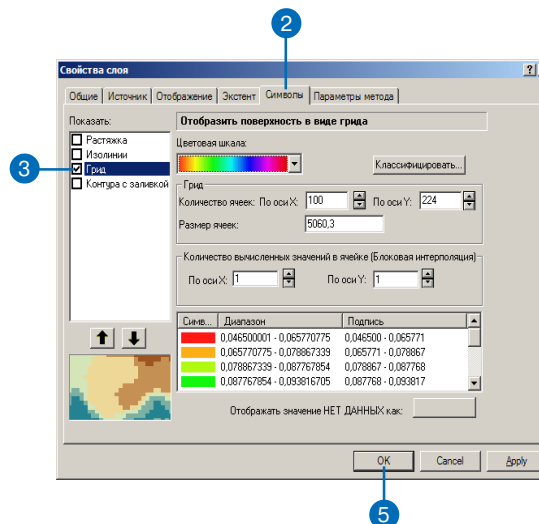
## Отображение геостатистического слоя методом контуров с заливкой

1. В таблице содержания ArcMap выберите геостатистический слой, нажмите правую клавишу мыши и выберите опцию Свойства.
2. Откройте закладку Символы.
3. В меню Показать: выберите способ отображения Контур с заливкой и поставьте галочку напротив соответствующей строки.
4. Определите требуемые параметры.
5. Нажмите OK.



## Отображение геостатистического слоя в виде грида

1. В таблице содержания ArcMap выберите геостатистический слой, нажмите правую клавишу мыши и выберите опцию Свойства.
2. Откройте закладку Символы.
3. В меню Показать: выберите способ отображения Грид и поставьте галочку напротив соответствующей строки.
4. Определите требуемые параметры.
5. Нажмите OK.



## Подсказка

**Фактор Z (высота), используемый при отмывке**  
Увеличение фактора Z увеличивает пропорцию единиц измерения  $z$  относительно единиц измерения  $x$  и  $y$ . Если значения  $z$  очень маленькие, а  $x$  и  $y$  - большие, вам необходимо будет задать большое значение фактора Z, чтобы увидеть какие-либо изменения.

## Подсказка

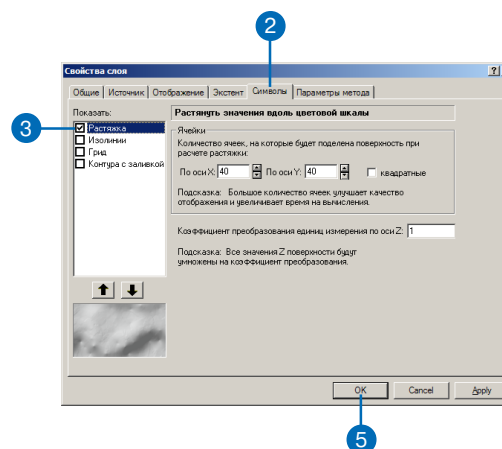
**Совместное отображение нескольких способов**  
Если использовать одновременно контура с заливкой и отмывку, можно получить очень эффектное отображение данных. Панель инструментов Эффекты и опции свойств слоя позволяют определить различные дополнительные параметры для отображения данных.

## Подсказка

**Быстрый доступ к свойствам слоя**  
Чтобы быстро открыть диалог Свойства слоя сделайте двойной щелчок мышью на названии слоя в таблице содержания ArcMap.

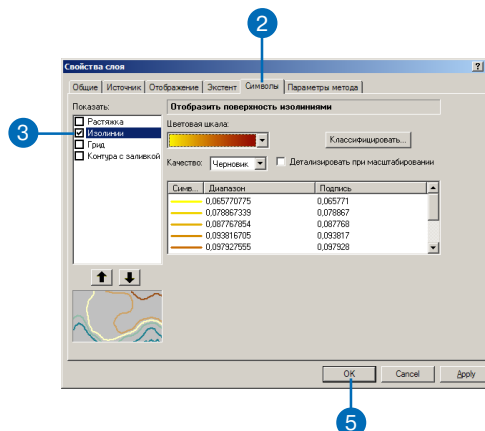
## Отображение геостатистического слоя способом отмывки

1. В таблице содержания ArcMap выберите геостатистический слой, нажмите правую клавишу мыши и выберите опцию Свойства.
2. Откройте закладку Символы.
3. В меню Показать: выберите способ отображения Растяжка и поставьте галочку напротив соответствующей строки.
4. Определите требуемые параметры.
5. Нажмите OK.



## Отображение геостатистического слоя способом изолиний

1. В таблице содержания ArcMap выберите геостатистический слой, нажмите правую клавишу мыши и выберите опцию Свойства.
2. Откройте закладку Символы.
3. В меню Показать: выберите способ отображения Изолинии и поставьте галочку напротив соответствующей строки.
4. Определите требуемые параметры.
5. Нажмите OK.



## Изменение СИМВОЛОГИИ геоestatистического СЛОЯ

ArcMap предоставляет множество инструментов, которые помогут вам при отображении слоев. Воспользовавшись этими инструментами, вы сможете создавать красиво оформленные карты, и, что более важно, изучать и анализировать данные слоя различными способами.

ArcMap отображает слой с использованием параметров, определяемых системой по умолчанию. Однако, вы можете подобрать другое оформление, заменив символы, предложенные по умолчанию. Цветовая схема отображения данных может быть выбрана из набора существующих цветовых схем, или создана в интерактивном режиме.

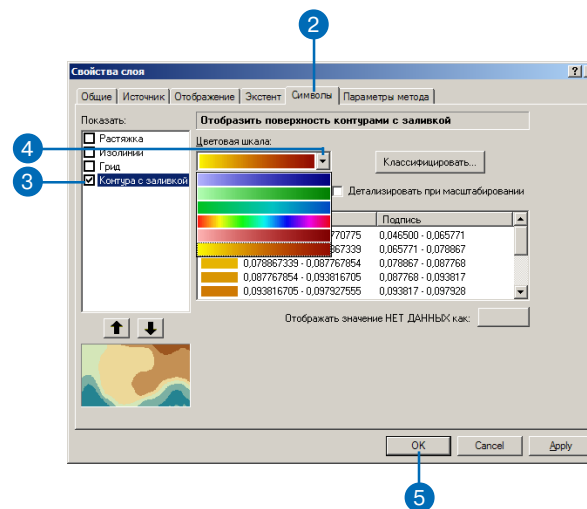
### Подсказка

#### Масштабирование

Если вы увеличиваете масштаб отображения карты, размер градуированных символов исходного слоя точечных пространственных объектов не будет меняться. Если вы хотите, чтобы они увеличивались с увеличением масштаба, вам необходимо определить базовый масштаб. В контекстном меню выберите опцию Диапазон видимых масштабов.

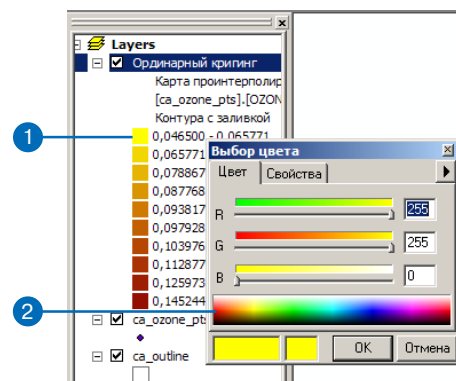
## Изменение цветовой схемы

1. Выберите геоestatистический слой и, нажав правую клавишу мыши, откройте диалог Свойства.
2. Откройте закладку Символы.
3. В списке Показать: выберите способ отображения Изолинии, Грид или Контур с заливкой.
4. В меню Цветовая шкала выберите цветовую схему.
5. Нажмите OK.



## Интерактивное изменение цвета

1. Сделайте щелчок правой клавишей мыши на изображении символа в легенде геоestatистического слоя.
  2. Выберите цвет для символа.
- Все значения на изображении, показанные этим символом, изменят цвет на выбранный.



# Классификация данных

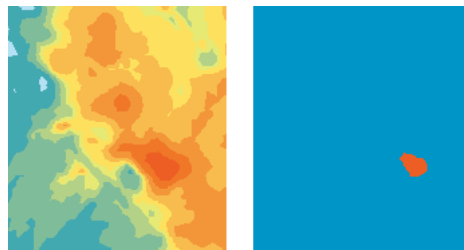
Когда вы осуществляете классификацию данных, вы группируете сходные пространственные объекты в классы путем присвоения одного и того же символа каждому объекту, попадающему в один и тот же класс. Объединение пространственных объектов в классы позволяет вам легче опознавать участки данных. Диапазоны значений определяют, какие пространственные объекты попадут в класс, что в свою очередь влияет на отображение данных на карте. Меняя границы класса, вы можете создавать карты, на которых одни и те же пространственные явления будут выглядеть по-разному. Вы можете установить границы классов вручную или воспользоваться стандартной схемой классификации.

## Почему нужно устанавливать границы класса вручную?

Устанавливайте границы классов вручную, если вы хотите отобразить на карте объекты, отвечающие какому-либо определенному критерию, или если вы сравниваете значения пространственных объектов с какой-либо специфической, значимой величиной. Для этого, вам необходимо вручную задать верхнюю и нижнюю границу для каждого класса.

Вы можете также вручную классифицировать данные в том случае, если хотите усилить какой-либо определенный диапазон значений, например, превышающих пороговое (критическое) значение или находящихся ниже этого порогового значения. Например, перед вами может стоять задача особым образом выделить участки, высота которых ниже определенного уровня, и вследствие этого они подвержены затоплению.

Интерактивное присвоение границ классов может быть также полезным при обособлении и выделении диапазонов данных. Например, если значения вашего набора данных колеблются от 0.0465 до 0.1736, и вы хотите отделить высокие значения, вы можете отнести все значения, меньше 0.15 к одному классу, а все значения, больше 0.15 ко второму классу.



На рисунке показано как выбранный диапазон данных может быть выделен цветом с использованием интерактивного определения границ классов.

## Использование стандартной схемы классификации

То, как установлен диапазон значений класса и его границы, определяет, сколько данных попадет в каждый класс и как будет выглядеть результирующая карта. В схеме классификации существует два главных компонента: количество классов, на которые будут разбиты значения и метод, по которому устанавливаются границы классов. Количество классов зависит от цели исследования. Правила, по которым значения относятся к тому или иному классу, однако требуют некоторого разъяснения. Для геостатистического слоя существуют три стандартных способа разбиения значений на классы:

- Метод равных интервалов
- Равновеликая классификация (метод квантилей)
- Метод регулируемых, “умных” квантилей

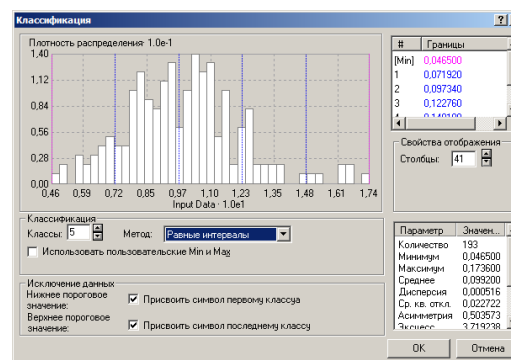
Описание этих методов приведено на следующих страницах.

## Метод равных интервалов

При этом методе все возможные значения разбиваются на интервалы с равным диапазоном значений. Поскольку обычно в наборе данных мало очень низких и очень высоких значений, в краевые классы попадает меньшее количество объектов. Эту опцию целесообразно использовать для того, чтобы подчеркнуть изменения в краевых классах. Возможно, его лучше применять к таким диапазонам значений, как процентное соотношение или температура. Эта опция наиболее используется для карт вероятности и стандартной ошибки интерполяции.



0.0465 - 0.0719  
0.0719 - 0.0973  
0.0973 - 0.1227  
0.1227 - 0.1481  
0.1481 - 0.1736

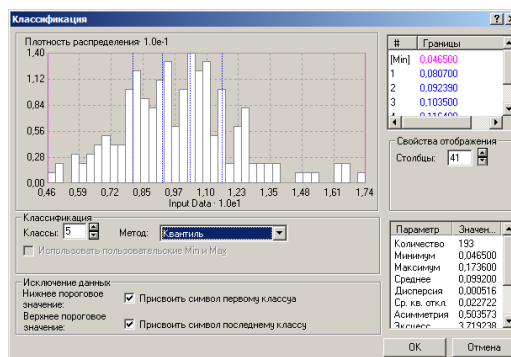


## Равновеликая классификация - квантиль

Диапазон возможных значений делится на интервалы таким образом, чтобы в каждый интервал попадало одинаковое количество значений. Классы, расположенные как по краям, так и в середине, имеют одинаковое количество значений. Поскольку, как правило, интервалы имеют более широкий диапазон значений для краевых классов, эта опция полезна для выделения изменений в средних значениях распределения.

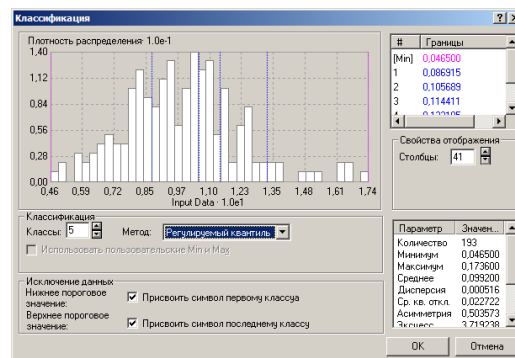
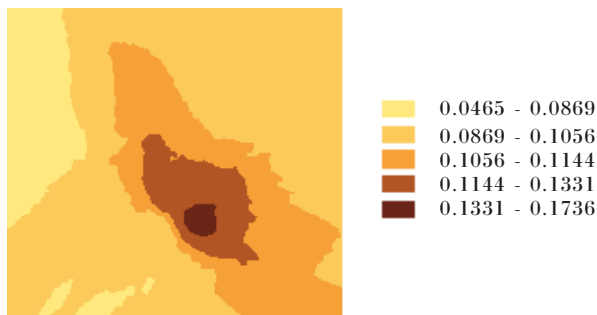


0.0465 - 0.0807  
0.0807 - 0.0923  
0.0923 - 0.1035  
0.1035 - 0.1164  
0.1164 - 0.1736



## Регулируемые квантили

Регулируемые квантили используются для описания классов, основанных на естественном объединении значений данных. Границы классов (точки прерывания) определяются при анализе групп и свойств, присущих данным. Пространственные объекты делятся на классы, границы которых устанавливаются в тех местах, где есть относительно большие скачки в значениях данных, и следовательно, группы со сходными значениями попадают в один класс. Этот метод является компромиссом между методами равных интервалов и квантилей. Интервалы, образуемые при использовании этого метода, не равновеликие и в отличие от квантилей, диапазон которых растянут для крайних классов, содержат меньшее количество значений. Эта опция пытается найти баланс между выделением изменений в средних значениях и в экстремальных значениях. Этот метод следует использовать для таких наборов данных, как, например, данные о количестве осадков, более половины значений которого может быть равно нулю.





# Классификация данных

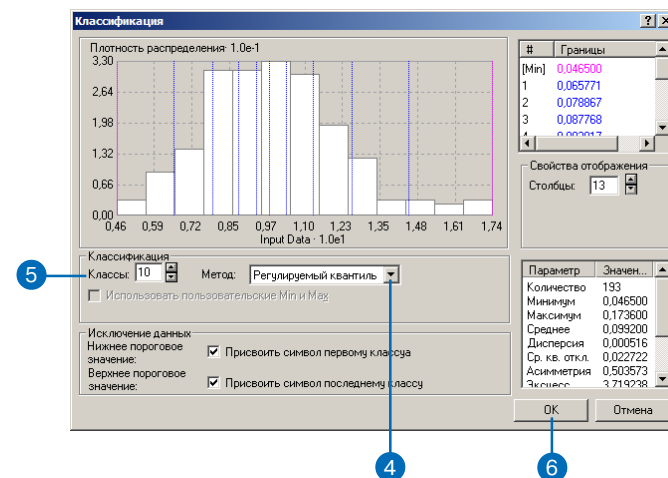
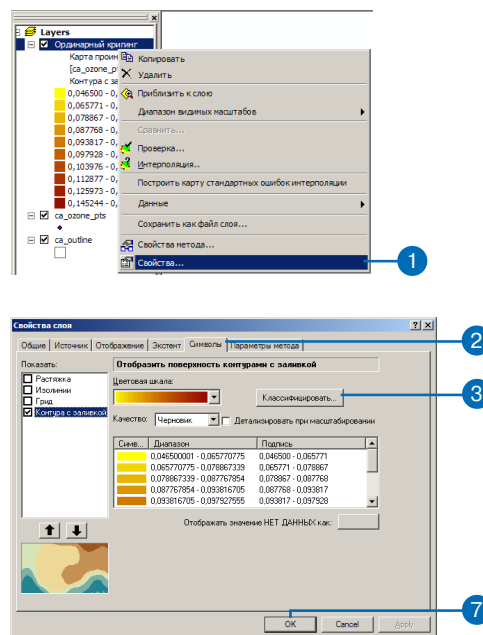
Когда вы классифицируете данные, вы можете воспользоваться одной из стандартных схем классификации, используемых для работы с геостатистическими слоями, или создать пользовательские классы, границы которых будут определены вами.

Если вы хотите воспользоваться одним из стандартных методов, просто выберите схему классификации и установите количество классов.

Если вы хотите определить свои собственные классы, вы можете вручную изменить границы классов или задать точные значения границ классов, подходящие для ваших данных.

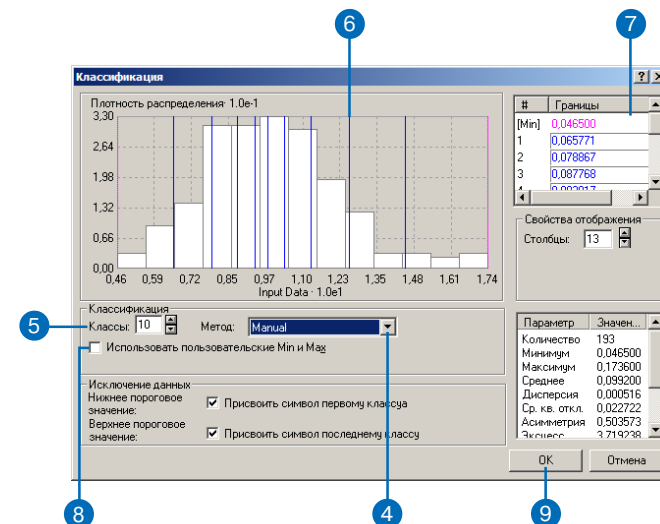
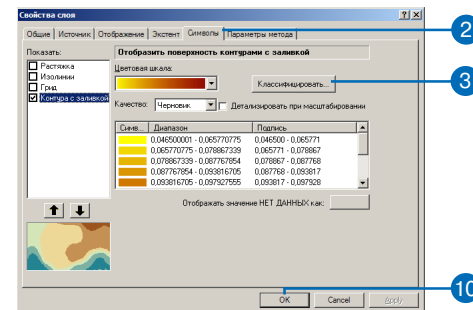
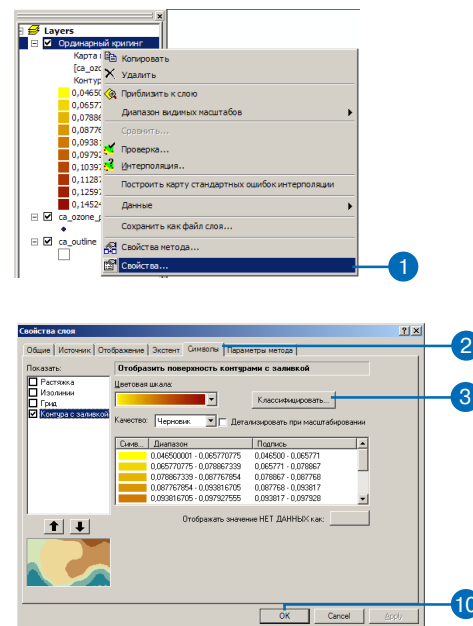
## Применение стандартного метода классификации

1. В таблице содержания выберите геостатистический слой и, нажав правую клавишу мыши, откройте диалог Свойства.
2. Откройте закладку Символы.
3. Нажмите кнопку Классифицировать.
4. Из меню Метод выберите метод классификации.
5. В окне Классы измените значение на необходимое количество классов.
6. Нажмите OK в диалоге Классификация.
7. Нажмите OK в диалоге Свойства слоя.



## Изменение границ класса вручную

1. В таблице содержания выберите геостатистический слой и, нажав правую клавишу мыши, откройте диалог Свойства.
2. Откройте закладку Символы.
3. Нажмите кнопку Классифицировать.
4. Из меню Метод выберите Вручную.
5. В окне Классы измените значение на необходимое количество классов.
6. Передвиньте границы класса в требуемое положение (синие отрезки).
7. Или наберите точные значения границ классов.
8. Или отметьте опцию Пользовательские Min и Max, затем наберите минимальное и максимальное значения, включаемые в классификацию.
9. Нажмите ОК в диалоге Классификация.
10. Нажмите ОК на закладке Символы.



## Определение масштабов, при которых будет отображаться геостатистический слой

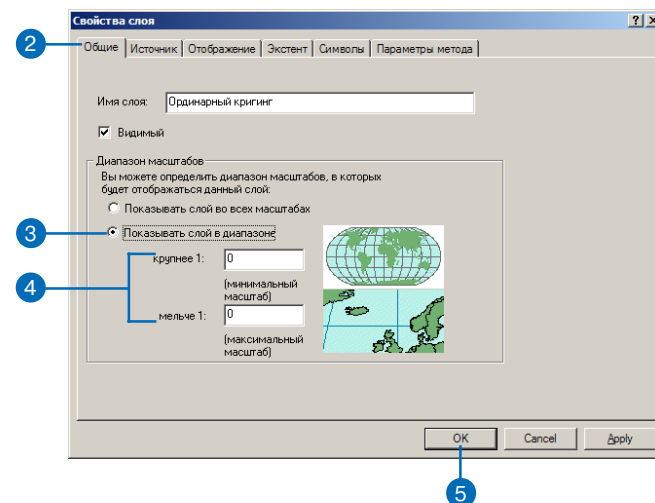
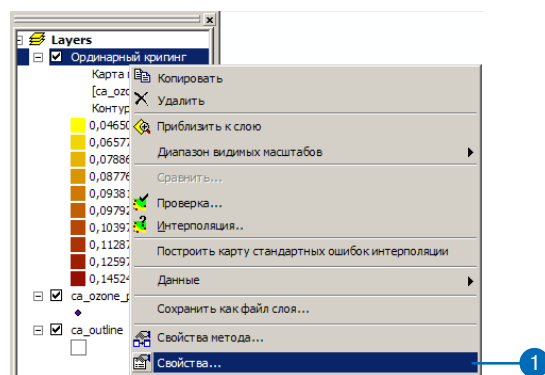
ArcMap не отображает ландшафт один к одному, с сохранением его природных размеров; это означает, что масштаб отображения ландшафта должен соответствовать размеру экрана. Километр на поверхности земли может соответствовать одному сантиметру на экране. Масштаб - уменьшение (или увеличение), необходимое для отображения интересующей вас территории.

Если вы уменьшаете или увеличиваете масштаб карты, масштаб отображения меняется. Возможно, что вам не нужно, чтобы все слои отображались для всех масштабов. Например, при отображении карты мира необязательно показывать границы округов.

Вы можете контролировать то, какие слои отображаются при определенном масштабе, установив диапазон масштаба для каждого слоя в диалоге Свойства.

## Определение диапазона масштаба

1. На выбранном геостатистическом слое нажмите правую клавишу мыши и выберите опцию Свойства.
2. Откройте закладку Общие.
3. Отметьте опцию Показывать слой в диапазоне масштабов:
4. Установите диапазон масштаба, введя значения в окошках “мельче” и “крупнее”.
5. Нажмите ОК.



## Экстраполяция значений в точках, расположенных за пределами изучаемой территории

По умолчанию, модуль Geostatistical Analyst интерполирует значения переменной во всех точках, попадающих внутрь минимального ограничивающего пространство прямоугольника. Минимальный ограничивающий пространство прямоугольник - это наименьший из возможных прямоугольник, ограничивающий все исходные опорные точки. Однако, этот прямоугольник может не покрывать целиком исследуемую территорию. Процесс создания карты проинтерполированных значений, выходящей за пределы этой ограничивающей рамки, носит название экстраполяции. Результирующий геостатистический слой, полученный в результате экстраполяции, будет покрывать ту область, которую определите вы.

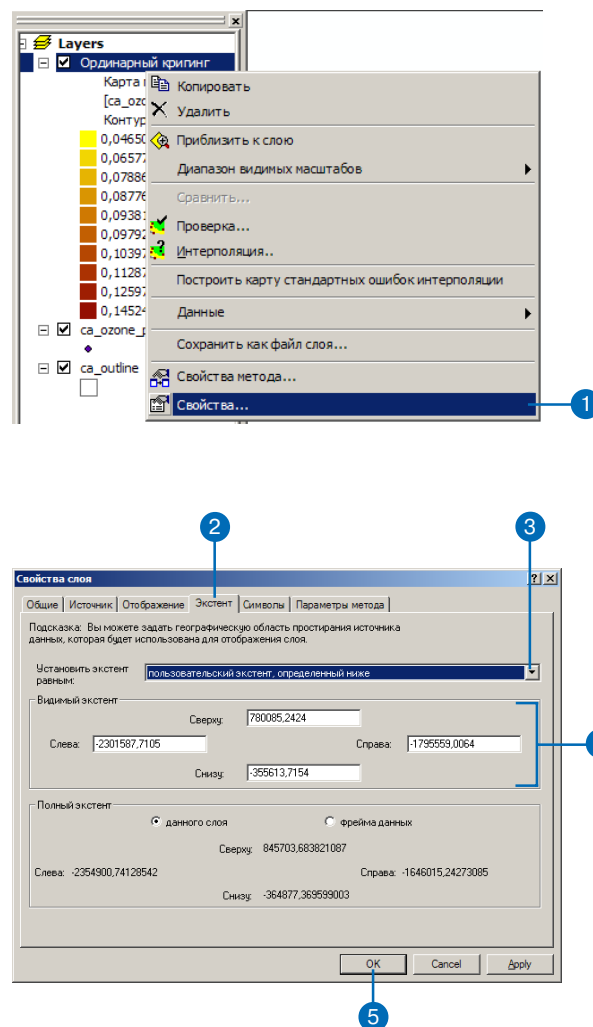
Обратите внимание, что не рекомендуется экстраполировать значения для точек, удаленных на значительное расстояние. Допустимо экстраполировать значения, расположенные близко к действительным опорным точкам (например, вокруг границы геостатистического слоя), но любая дальнейшая экстраполяция может привести к недостоверным результатам.

### Подсказка

*Чтобы вернуться к области отображения, предложенной по умолчанию, после определения новой области отображения, нажмите По умолчанию в закладке Экстент.*

## Экстраполяция значений

1. В таблице содержания ArcMap выберите геостатистический слой, значения которого вы хотите экстраполировать, нажмите правую клавишу мыши и выберите опцию Свойства.
2. Откройте закладку Экстент.
3. В меню Установить экстент равным: выберите опцию “пользовательский экстент, определенный ниже”.
4. В окне Видимый экстент введите новые значения.  
Или используйте область отображения любого из используемых слоев.
5. Нажмите ОК.



## Сохранение и экспорт геостатистических слоев

Геостатистический слой ArcMap включает ссылку на данные, которые хранятся на диске, символюгию, используемую для отображения слоя, и другую информацию, относящуюся к способу создания слоя.

Если сеанс ArcMap сохранен, слои в таблице содержания и их сопутствующие характеристики хранятся в файле с расширением .mxd. Однако, если вы хотите использовать геостатистический слой в другом сеансе ArcMap с сохранением символюгии, слой может быть сохранен в файл с расширением .lrg. Файл с расширением .lrg не копирует источник данных, а только ссылается на него.

Чтобы сделать геостатистический слой постоянным, копия может быть записана на диске (за исключением характеристик слоя) либо как грид ArcInfo, либо как шейп-файл.

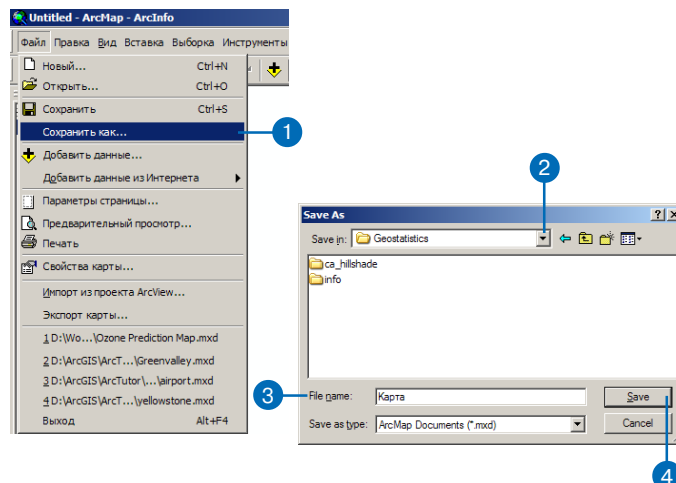
### Подсказка

#### Передача геостатистического слоя третьим лицам

Если вы планируете передавать свою карту третьим лицам, им понадобится не только документ карты, но и данные, на которые ссылаются ваши слои.

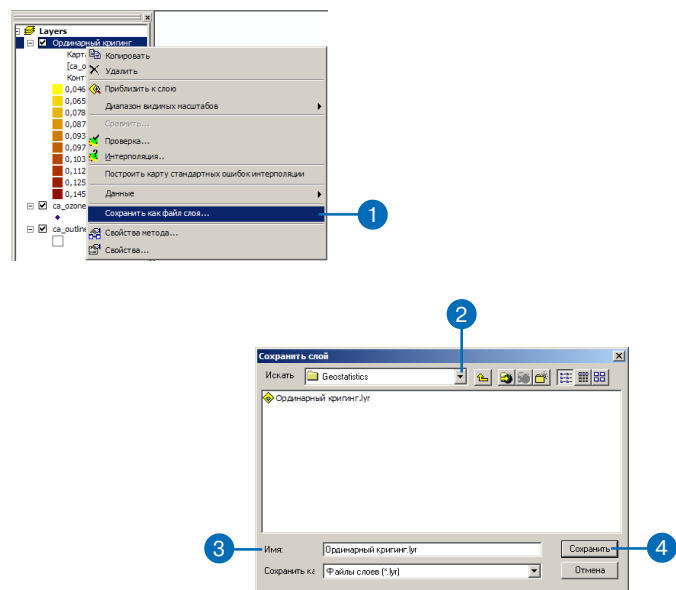
## Сохранение композиции карты

1. В меню Файл выберите опцию Сохранить как.
2. Перейдите в директорию, в которой вы хотите сохранить карту.
3. Измените название карты, если это необходимо.
4. Нажмите Сохранить.



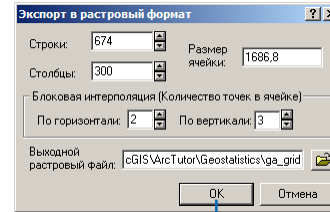
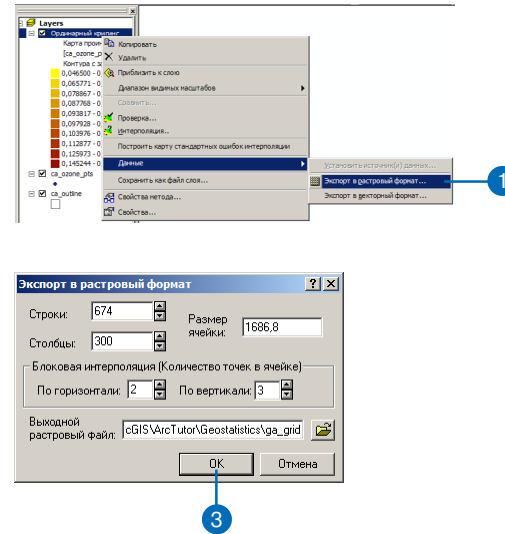
## Сохранение индивидуальных слоев

1. В таблице содержания ArcMap выберите геостатистический слой, который вы хотите сохранить, нажмите правую клавишу мыши и выберите опцию Сохранить как файл слоя....
2. Перейдите в директорию, в которой вы хотите сохранить слой.
3. Измените название слоя, если это необходимо.
4. Нажмите Сохранить.



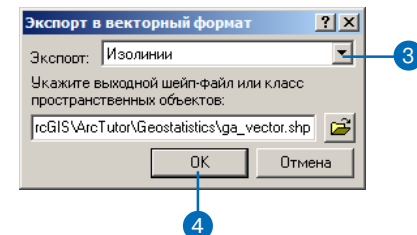
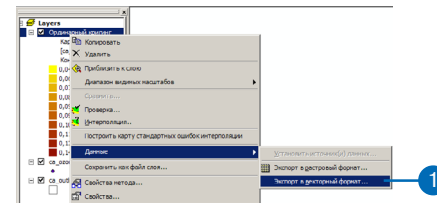
## Экспорт геостатистического слоя в растровый формат

1. В таблице содержания ArcMap нажмите правую клавишу мыши на выбранном геостатистическом слое, выберите опцию Данные, а затем Экспорт в растровый формат.
2. Установите необходимые свойства результирующего растрового изображения, такие как число строк и столбцов, размер ячейки, количество интерполируемых значений для каждой ячейки в меню Блоковая интерполяция при выполнении блоковой интерполяции, а также название и место растрового файла на диске.
3. Нажмите OK.



## Экспорт геостатистического слоя в векторный формат

1. В таблице содержания ArcMap нажмите правую клавишу мыши на выбранном геостатистическом слое, выберите опцию Данные, а затем Экспорт в векторный формат.
2. Задайте требуемый формат (например, шейп-файл, персональная база геоданных, база данных SDE).
3. В меню Экспорт выберите Изолинии или Контура с заливкой.
4. Нажмите OK.





# Дополнительные инструменты геостатистического анализа

## 9

### В ЭТОЙ ГЛАВЕ

- Изменение параметров геостатистического слоя: свойства метода
- Интерполирование значений для заданных точек
- Выполнение проверки геостатистического слоя, созданного из поднабора данных
- Наложение данных для выполнения более точной интерполяции

В ArcMap и в модуль Geostatistical Analyst включено множество инструментов, призванных помочь в выполнении геостатистического анализа. Вы можете менять параметры модели, найти значения интерполируемого показателя в заданной точке, выполнить проверку для поднаборов данных, пространственно разделить свои данные, выполнить интерполяцию для каждого из выделенных наборов, а затем скомбинировать результаты. В этой главе мы не приводим исчерпывающий список инструментов, которые могут помочь вам при выполнении геостатистического анализа. Вместо этого, в ней дано описание наиболее часто используемых при осуществлении анализа инструментов. Но при этом вы должны отдавать себе отчет в том, что поскольку модуль Geostatistical Analyst интегрирован в ArcMap, существует бесконечное множество функций, которые вы можете и будете использовать при анализе. По мере того, как вы будете узнавать ArcMap и модуль Geostatistical Analyst, вы будете открывать для себя все большее количество инструментов, отвечающих вашим потребностям в построении точных поверхностей.



## Изменение параметров геостатистического слоя: свойства метода

После построения поверхности в модуле Geostatistical Analyst, вы имеете возможность просмотреть его в ArcMap. Воспользовавшись многочисленными инструментами отображения геостатистического слоя, вы можете обнаружить, что некоторые участки вновь созданной поверхности не отвечают вашим представлениям об изучаемой территории. Вы можете принять решение усовершенствовать поверхность, изменив параметры, использованные для ее построения, а не создавать новую поверхность. Через свойства метода вы можете вернуться к диалогам, изменить любые параметры модели и проанализировать вновь полученные результаты.

### Подсказка

#### Изучение свойств метода

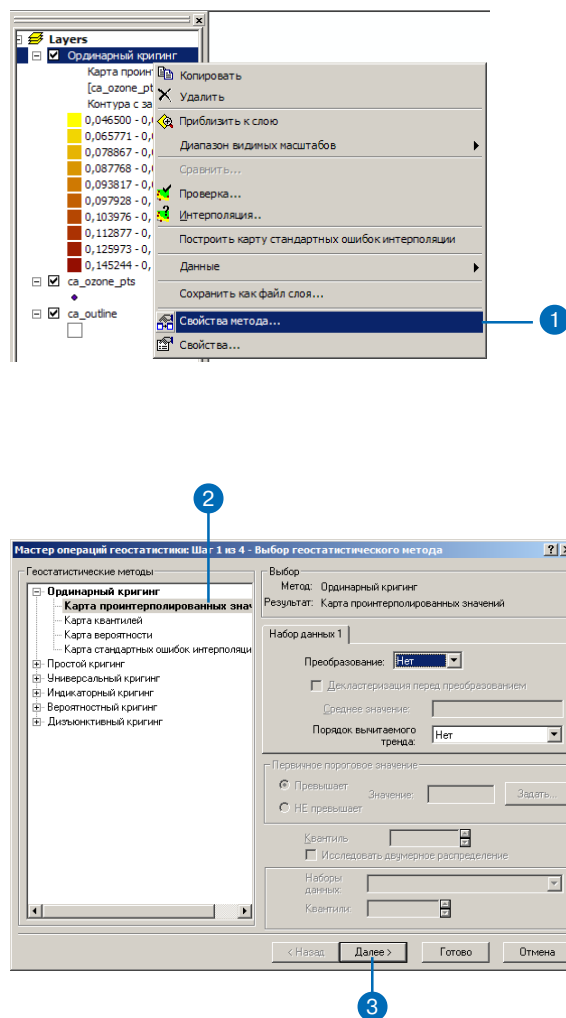
Диалог *Свойства метода* поможет вам лучше понять то, как различные опции влияют на результирующую поверхность.

## Использование свойств метода

1. В таблице содержания ArcMap щелкните правой клавишей мыши на выбранном слое и выберите опцию Свойства метода.

То, какие опции доступны, зависит от способа построения исходной поверхности. Эти шаги соответствуют построению поверхности вероятностей значений на основе метода кригинга.

2. Выберите необходимый метод кригинга.
3. В диалоге Выбор геостатистического метода нажмите Далее.
4. Необязательный шаг: изменить какие-либо параметры модели и в диалоге Моделирование вариограммы/ковариации нажмите Далее.
5. Необязательный шаг: задайте новую область поиска соседства и в диалоге Поиск соседства нажмите Далее.
6. Оцените результаты перекрестной проверки. Улучшился ли результат? Если нет, то повторите шаги со 2 по 5. Если да, нажмите Готово в диалоге Перекрестная проверка.
7. В диалоге Информация о результирующем слое нажмите ОК.



## Интерполирование значений для заданных точек

Иногда вы заинтересованы в интерполировании значений в заданной точке или в нескольких определенных точках, а не обязательно для всей поверхности. Например, вы обеспокоены уровнем радиации в районе определенного дома или хотите знать высоту в намечаемом месте строительства обзорной башни.

Если вы хотите знать предполагаемые значения в нескольких определенных точках, самый простой способ - воспользоваться всплывающими на карте подсказками. Однако, если вы хотите сохранить проинтерполированные значения в определенной точке в результирующем слое, чтобы их можно было использовать для последующего анализа, лучше воспользоваться диалогом Интерполяция.

### Подсказка

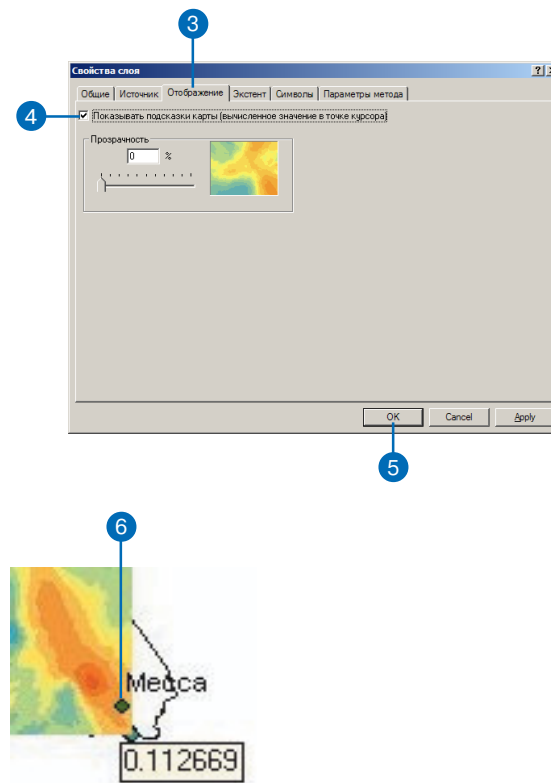
#### Выбор точек с использованием таблицы атрибутов

Помимо использования инструмента выбора, интересующие вас точки могут быть выбраны с помощью таблицы атрибутов (нажмите правой клавишей мыши на выбранном слое в таблице содержания ArcMap и выберите опцию Открыть таблицу атрибутов).

## Использование всплывающих подсказок

1. Создайте геостатистический слой с использованием любого из методов, описанных в главах 5 или 6.
2. Нажмите правую клавишу мыши на выбранном геостатистическом слое и выберите опцию Свойства.
3. Откройте закладку Отображение.
4. Отметьте опцию Показывать подсказки карты.
5. Нажмите ОК.
6. Поместите курсор в точке слоя, для которой вы хотите получить информацию.

Отобразится значение поверхности в данной точке.



## Подсказка

Другой способ просмотреть результаты

Результаты интерполяции можно просмотреть также в ArcCatalog™.

## Интерполяция значений для заданного объекта точечного слоя

1. В таблице содержания ArcMap выберите геостатистический слой, созданный с использованием данных наблюдений, нажмите правую клавишу мыши и выберите опцию Интерполяция.
2. В меню Входные данные выберите слой, содержащий точки, для которых вы хотите получить проинтерполированные значения.
3. Определите директорию, в которой будет храниться полученный набор данных, нажав кнопку с изображением папки, расположенную рядом с окном “Укажите результирующий шейп-файл или класс пространственных объектов”.
4. Пролитайте папки или напечатайте название директории, в которой будет храниться результирующий файл, и назовите его.
5. Нажмите Сохранить и ОК.
6. После подсказки добавьте файл с проинтерполированными значениями в таблицу содержания ArcMap.
7. Выберите слой с проинтерполированными значениями, нажмите правую клавишу мыши и выберите Открыть таблицу атрибутов, чтобы отобразить полученные результаты.

В таблице будут показаны проинтерполированные значения в заданных точках.

**Интерполяция**

Подсказка: Интерполирует значения в известной точке с использованием заданной модели.

Информация на входе

Входные данные: Test Locations

Поле X: Shape

Поле Y: Shape

Укажите результирующий шейп-файл или класс пространственных объектов:

OK Отмена

**Интерполяция**

Подсказка: Интерполирует значения в известной точке с использованием заданной модели.

Информация на входе

Входные данные: Test Locations

Поле X: ca\_ozone\_pts

Поле Y: ca\_outline

Укажите результирующий шейп-файл или класс пространственных объектов:

OK Отмена

**Выберите выходной набор данных**

Искать: Geostatistics

ca\_ozone.shp  
ca\_ozone\_pts.shp  
ca\_outline.shp  
ca\_ozone\_pts.shp  
ca\_pm10\_pts.shp  
New\_ozone.shp

Имя: Test Locations

Сохранять к: Shapefile

Сохранить Отмена

**Атрибуты Test Locations**

ID	PTS	Shape	LATITUDE	LONGITUDE	ELEVATION	OZONE	Проверено	Шейп-файл
1	Точка		39.1447	-122.2055	134	0.0465	0	
2	Точка		39.1451	-122.2051	132	0.0464	0	
3	Точка		39.1461	-122.1878	91	0.04	0	
4	Точка		39.1508	-122.2053	23	0.0379	0	
5	Точка		39.1508	-121.834	86	0.0386	0	
6	Точка		39.1508	-122.842	18	0.0588	0	
7	Точка		39.1514	-122.2675	7	0.0469	0	
8	Точка		39.1512	-122.1678	6	0.0474	0	
9	Точка		39.1436	-122.702	47	0.053	0	
10	Точка		41.7291	-122.6394	802	0.0439	0	
11	Точка		39.1521	-118.2037	27	0.0409	0	
12	Точка		39.1502	-120.108	1475	0.0459	0	
13	Точка		36.9772	-121.8117	73	0.0466	0	
14	Точка		36.9335	-121.088	20	0.0575	0	
15	Точка		37.8223	-122.2658	9	0.0474	0	
16	Точка		35.0282	-120.5014	87	0.0445	0	
17	Точка		36.033	-122.9219	488	0.07	0	
18	Точка		36.8991	-121.993	28	0.0709	0	
19	Точка		36.1241	-120.632	5	0.0719	0	
20	Точка		38.6336	-122.8038	81	0.07128	0	
21	Точка		39.1036	-120.0564	48	0.0715	0	
22	Точка		36.9337	-121.7614	67	0.0727	0	
23	Точка		36.8447	-119.197	890	0.074	0	
24	Точка		34.7888	-120.6861	48	0.0796	0	
25	Точка		38.3116	-122.2842	22	0.0761	0	

## Выполнение проверки для геостатистического слоя, созданного на основе поднабора данных

Наиболее точный способ оценить качество результирующей поверхности - сравнить интерполируемые значения для определенных точек с измеренными в полевых условиях. По разным причинам (время, деньги, и т.д.), часто бывает невозможным вернуться в изучаемый район и набрать независимый проверочный набор данных. Одно из решений - разделить исходный набор данных на две части. Одна часть может быть использована для моделирования, то есть для создания результирующей поверхности, а другая - для тестирования или проверки результирующей поверхности.

### Подсказка

#### Разделение набора данных

*Процентное соотношение, в котором должен быть разделен набор данных, должно быть основано на количестве возможных опорных точек. Вам необходимо достаточное количество точек для создания результирующей поверхности и выполнения значимой проверки полученной поверхности. Следовательно, если у вас небольшое исходное количество опорных точек, их может не хватить для корректного разделения набора данных.*

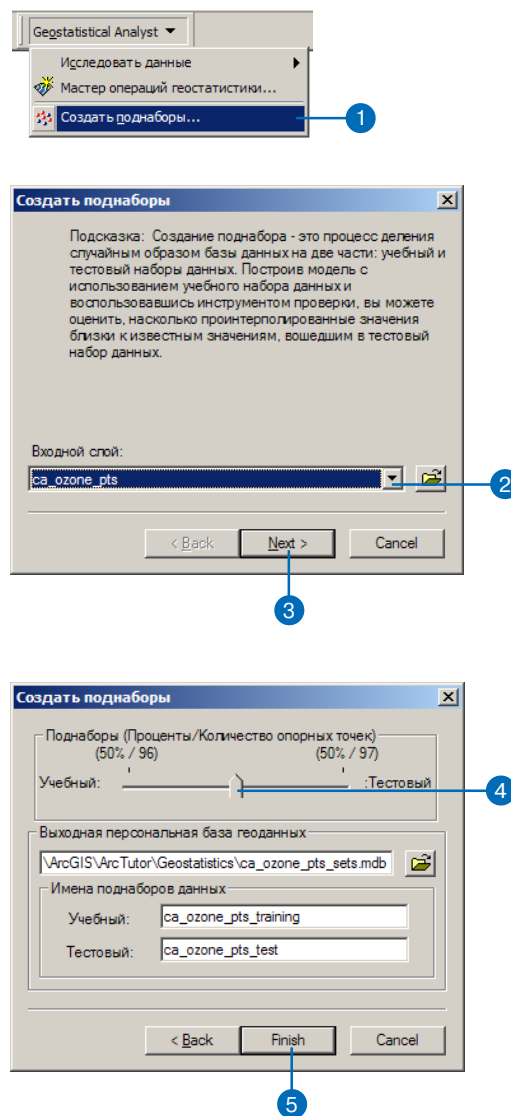
## Создание поднаборов данных

1. На панели инструментов Geostatistical Analyst выберите опцию Создать поднаборы.
2. Из меню Слой выберите слой, данные которого вы хотите разделить.
3. Нажмите Далее.
4. Передвиньте движок в положение, которое определит требуемое соотношение данных в учебном и тестовом наборе.

По умолчанию, результирующему набору данных присваивается название в соответствии с условием: “исходный файл набора данных” + “\_sets.mdb”, например, inputpoints\_sets.mdb, где inputpoints - название исходного набора данных, содержащего точки.

5. Нажмите Готово.

Учебные и тестовые наборы данных образуют две таблицы в персональной базе геоданных.



## Подсказка

### Выполнение проверки с помощью Мастера операций геостатистики

Учебный и тестовый наборы данных могут быть введены непосредственно в первый диалог Мастера операций геостатистики “Выберите входные данные и метод”. После того, как в диалогах Мастера были выбраны соответствующие параметры, в последующем диалоге будет показана статистика проверки, которую можно проанализировать перед построением окончательной поверхности.

## Подсказка

### Просмотр суммарной статистики

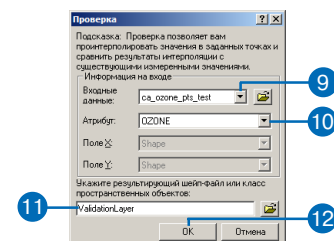
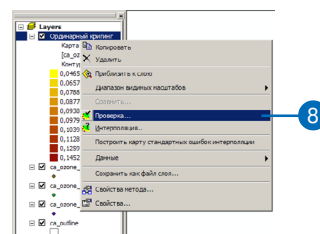
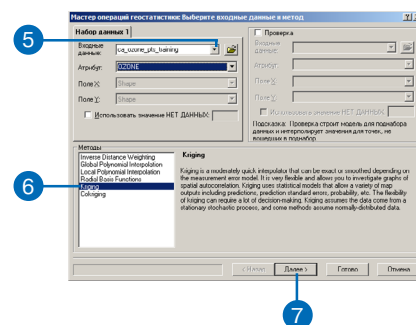
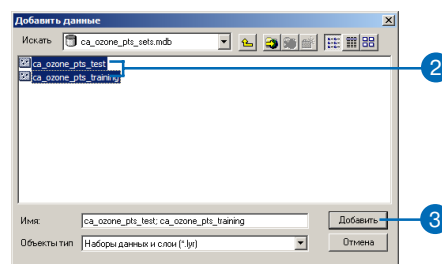
Суммарная статистика из проверки может быть просмотрена с использованием либо ArcMap, либо ArcCatalog.

## См. также

См. также Главу 7, ‘Использование аналитических инструментов при построении поверхностей’, для информации о том, что такое перекрестная проверка и проверка, и о содержании результирующей статистики.

## Выполнение проверки с использованием поднаборов данных

1. На стандартной панели инструментов ArcMap нажмите кнопку Добавить данные. Перейдите к папке, в которой был сохранен поднабор данных (если его нет в таблице содержания ArcMap).
2. Выберите учебный и тестовый слои (удерживая клавишу Shift).
3. Нажмите Добавить.
4. На панели инструментов Geostatistical Analyst выберите опцию Мастер операций геостатистики.
5. В окне Входные данные выберите учебный набор данных.
6. Выберите соответствующий метод.
7. Нажмите Далее и пройдите через все диалоги, чтобы построить поверхность.
8. Нажмите правую клавишу мыши, выделив вновь созданный геостатистический слой, и выберите опцию Проверка.
9. В окне Входные данные выберите тестовый набор данных.
10. В окне Атрибут выберите тот атрибут, который был использован при построении поверхности.
11. Наберите название результирующего (проверочного) набора данных и определите его место на диске.
12. Нажмите OK. ►



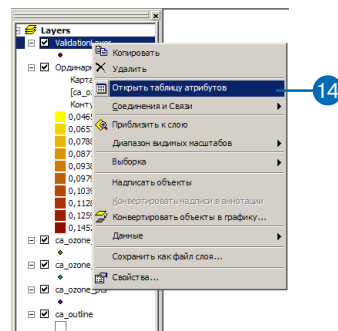
## Подсказка

### Результирующие файлы

В результате выполнения проверки будут созданы два файла: шейп-файл, который содержит тестовые опорные точки и связанные с ними проверочные данные, и таблица, в которой приведена суммарная статистика проверки, аналогичная представленной в диалоге перекрестной проверки (см. Главу 7, ‘Использование аналитических инструментов при построении поверхностей’).

13. После подсказки добавьте слой проверки в таблицу содержания ArcMap.
14. Выберите новый слой и, нажав правую клавишу мыши, выберите опцию Открыть таблицу атрибутов.

На экране отобразятся результаты проверки, которые вы сможете оценить.



№	Shape*	LATITUDE	LONGITUDE	ELEVATION	OZONE	Измерение
1	Точка	39.1447	-121.2065	134	0.0465	
2	Точка	37.7661	-122.3978	5	0.05	
3	Точка	37.9508	-122.3569	23	0.05789	
4	Точка	36.9395	-121.6354	36	0.0586	
5	Точка	37.8014	-122.2672	7	0.0589	
6	Точка	37.012	-122.1929	0	0.0606	
7	Точка	41.7293	-122.6254	802	0.06389	
8	Точка	33.9291	-118.2097	27	0.06409	
9	Точка	39.3302	-120.1808	1676	0.06539	
10	Точка	36.5722	-121.8117	73	0.0666	
11	Точка	36.9635	-121.9888	28	0.06789	
12	Точка	35.0202	-120.5614	60	0.06949	
13	Точка	36.9658	-121.993	28	0.0709	
14	Точка	35.3226	-120.6546	46	0.0719	
15	Точка	38.9444	-118.97	1905	0.0736	
16	Точка	38.3115	-122.2942	22	0.0754	
17	Точка	34.448	-118.342	231	0.0765	
18	Точка	38.2989	-122.4567	20	0.0772	
19	Точка	38.6838	-121.1627	98	0.07819	
20	Точка	34.9507	-120.4341	0	0.0789	
21	Точка	32.64	-117.083	56	0.0807	
22	Точка	37.9729	-122.5184	11	0.081	
23	Точка	36.2269	-121.1153	116	0.0819	
24	Точка	37.3724	-122.0767	24	0.0816	
25	Точка	36.5086	-116.9477	125	0.08229	





# Приложение

# A

## В ЭТОМ ПРИЛОЖЕНИИ

- Детерминистские методы
- Интерполяция по методу взвешенных расстояний
- Интерполяция по методу глобального полинома
- Интерполяция по методу локальных полиномов
- Интерполяция по методу радиальных базисных функций
- Геостатистические методы
- Диалог декластеризации
- Диалог моделирования распределения
- Диалог вариограммы/ковариации
- Диалог двумерного распределения
- Формулы кригинга
- Диалог перекрестной проверки

В основе модуля Geostatistical Analyst лежит множество формул и уникальных концепций выполнения задач. В данном приложении и приведены эти формулы и теоретические понятия. Предполагается, что до чтения этого приложения вы имели опыт в математике и геостатистике.

По тексту приложения приводятся ссылки на наиболее известные учебники и статьи, в которых вы сможете более детально познакомиться с математической составляющей методов, использованных в модуле Geostatistical Analyst. В некоторых случаях, когда подробности нелегко отыскать в учебниках, мы приводим более детальное описание метода.

## Детерминистские методы

### Интерполяция по методу взвешенных расстояний

Интерполяция по методу взвешенных расстояний - относительно простой метод, и его подробное описание приведено в Главе 5, ‘Детерминистские методы интерполяции пространственных данных’.

### Интерполяция по методу глобального полинома

Интерполяция по методу глобального полинома использует методы множественной регрессии для всех данных. Результат или поверхность тренда подбирается для координат  $x$  и  $y$ , являющихся ковариатами. Для тренда первого порядка, модель выглядит следующим образом:

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \varepsilon(x_i, y_i),$$

где  $Z(x_i, y_i)$  - фактическое значение в точке  $(x_i, y_i)$ ,  $\beta_j$  - параметры, а  $\varepsilon(x_i, y_i)$  - случайная ошибка. Для тренда второго порядка, модель выглядит следующим образом:

Для тренда третьего порядка, модель выглядит так:

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i + \varepsilon(x_i, y_i).$$

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i + \beta_6 x_i^3 + \beta_7 y_i^3 + \beta_8 x_i^2 y_i + \beta_9 x_i y_i^2 + \varepsilon(x_i, y_i),$$

и так далее, до 10-ой степени (максимальная из используемых модулем Geostatistical Analyst). Регрессионные модели подбираются путем оценки параметров  $\{\beta_j\}$  на основе использования метода наименьших квадратов, описание которого можно найти во многих учебниках по геостатистике, например в учебнике Snedecor и Cochran (1989).

### Интерполяция по методу локальных полиномов

Интерполяция по методу локальных полиномов аналогична интерполяции по методу глобального полинома за тем исключением, что используются данные только в пределах определенных “окон”, а не для всей территории, поэтому подбираются локальные тренды и используются значения весов. Окно можно передвигать, и значение поверхности в центре окна, назовем его  $\mu_0(x, y)$ , оценивается для каждой точки. Метод наименьших квадратов используется для минимизации суммы,

$$\sum_{i=1}^n w_i (Z(x_i, y_i) - \mu_0(x_i, y_i))^2,$$

где  $n$  - количество точек, попадающих в окно. В данном случае,  $w_i$  - весовой коэффициент,

$$w_i = \exp(-3d_{i0}/a),$$

где  $d_{i0}$  - расстояние между точкой и центром окна, а  $a$  - параметр, контролирующий скорость убывания весовых коэффициентов с расстоянием. И наконец,  $\mu_0(x_i, y_i)$  - значение полинома.

Для полинома первой степени:

$$\mu_0(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i,$$

Для полинома второй степени:

$$\mu_0(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i,$$

и так далее. Минимизация осуществляется для параметров  $\{\beta_j\}$ . Параметры оцениваются повторно в том случае, если центр окна и, следовательно, все окно смещается (Gandin, 1963).

## Интерполяция с использование радиальных базисных функций

Модуль Geostatistical Analyst использует набор из  $n$  базисных функций, по одной для каждой опорной точки. Интерполятор - это линейная комбинация базисных функций,

$$\hat{\mathcal{Z}}(\mathbf{s}_0) = \sum_{i=1}^n \omega_i \phi(\|\mathbf{s}_i - \mathbf{s}_0\|) + \omega_{n+1}$$

где  $\phi(r)$  - радиальная базисная функция,  $r = \|\mathbf{s}_i - \mathbf{s}_0\|$  - Евклидово расстояние между интерполируемой точкой  $\mathbf{s}_0$  и каждой опорной точкой  $\mathbf{s}_i$ , а  $\{\omega_i; i = 1, 2, \dots, n + 1\}$  - оцениваемые значения весов.

Пусть  $\mathbf{w} = (\omega_1, \omega_2, \dots, \omega_n)$ , которые вычисляются путем решения системы уравнений,

$$\begin{pmatrix} \Phi & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \omega_{n+1} \end{pmatrix} = \begin{pmatrix} \mathbf{z} \\ 0 \end{pmatrix}$$

где  $\Phi$  - матрица с  $i,j$ -ым элементом  $\phi(\|\mathbf{s}_i - \mathbf{s}_j\|)$  для пары опорных точек  $i,j$ ,  $\mathbf{1}$  - вектор столбца, состоящий из единиц, а  $\mathbf{z}$  - вектор столбца, содержащий данные. Если  $\phi$  - вектор, содержащий  $\phi(\|\mathbf{s}_i - \mathbf{s}_0\|)$ , интерполятор равен,

$$\hat{\mathcal{Z}}(\mathbf{s}_0) = \mathbf{w}'\phi + \omega_{n+1}.$$

Где  $\omega_{n+1}$  - параметр смещенности.

Следует использовать аналогичный интерполятор,

$$\hat{\mathcal{Z}}(\mathbf{s}_0) = \lambda'\mathbf{z},$$

где  $\lambda$  решает уравнение,

$$\begin{pmatrix} \Phi & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ m \end{pmatrix} = \begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix},$$

преимущество которого состоит в том, что он показывает весовые коэффициенты для всех данных. Веса отображаются в диалоге Поиск соседства.

В модуле Geostatistical Analyst используются следующие радиальные функции:

1. Полностью регуляризованный сплайн,

$$\phi(r) = -\sum_{n=1}^{\infty} \frac{(-1)^n (\sigma \cdot r)^{2n}}{n!n} = \ln(\sigma \cdot r/2)^2 + E_1(\sigma \cdot r/2)^2 + C_E,$$

где  $\ln$  - натуральный логарифм,  $E_1(x)$  - экспоненциальный интеграл (Abramowitz and Stegun, 1965, стр. 227), а  $C_E$  - константа Эйлера (Abramowitz and Stegun, 1965, стр. 255),

2. Функция сплайна с натяжением,

$$\phi(r) = \ln(\sigma \cdot r/2) + K_0(\sigma \cdot r) + C_E,$$

где  $K_0(x)$  - модифицированная функция Бесселя (Abramowitz and Stegun, 1965, стр. 374),

3. Мультиквadratic,

$$\phi(r) = (r^2 + \sigma^2)^{1/2},$$

4. Обратный мультиквadratic,

$$\phi(r) = (r^2 + \sigma^2)^{-1/2},$$

5. Плоский сплайн,

$$\phi(r) = (\sigma \cdot r)^2 \ln(\sigma \cdot r).$$

Оптимальный параметр сглаживания  $\sigma$  определяется путем минимизации среднеквадратичных ошибок вычислений с использованием перекрестной проверки.

Радиальные базисные функции описаны Бишопом (Bishop, 1995, стр. 164). Развернутое описание радиальных функций и их связей со сплайнами и методами кригинга можно найти в работах Cressie (1993, стр. 180) и Chiles и Delfiner (1999, стр. 272).

## Геометрическая анизотропия

Геометрическая анизотропия учитывается как преобразование координат:

$$\mathbf{s}^+ = \begin{pmatrix} \sqrt{r} & 0 \\ 0 & 1/\sqrt{r} \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \mathbf{s}$$

где  $\theta$  - угол поворота и  $r$  - соотношение размеров малой и большой оси результирующего эллипса. Расстояние затем рассчитывается как  $\|\mathbf{s}_i^+ - \mathbf{s}_0^+\|$ .

# Геостатистические методы

## Диалог Declustering (Декластеризация)

### Декластеризация по ячейкам

По разным причинам данные могли быть отобраны неравномерно, с различной плотностью опорных точек для разных участков территории. Для некоторых методов преобразований, таких, как преобразование по методу нормальных меток, важно, чтобы гистограмма выборки правильно отражала гистограмму всей совокупности. Решением для неравномерных выборок является присвоение данным весовых значений, при этом данным, отобранным на территориях с высокой плотностью точек, присваиваются меньшие веса. За более детальной информацией обратитесь к книгам Journel (1983), Isaaks и Srivastava (1989, стр. 421), Cressie (1993, стр. 352), и Goovaerts (1998, стр. 76).

### Метод по умолчанию—Индекс Морисита (Morisita's index)

По умолчанию в модуле Geostatistical Analyst используется метод грида с квадратными ячейками, размер которых определяется исходя из максимального значения индекса Морисита (Morisita's index) (Morisita, 1959; см. также Cressie, 1993, стр. 590), где индекс Морисита - функция размера ячейки.

### Полигональный метод

Возможно также использовать метод, который позволяет присваивать весовые значения в зависимости от размера полигонов, окружающих каждую точку. Полигоны образуются путем нахождения всех возможных точек, которые ближе к данной опорной точке, чем к любой другой опорной точке. Таким образом, каждая опорная точка имеет полигон влияния. В математике такие полигоны носят название диаграмм Вороного и полигонов Тиссена. За разъяснениями обратитесь к работам Isaaks и Srivastava (1989, стр. 238), и Goovaerts (1998, стр. 79). В модуле Geostatistical Analyst внешняя граница несколько больше, чем наименьший (без поворота) прямоугольник, который охватывает все точки. Прямоугольник образуется путем суммирования наибольшей х-координаты и у-координаты и  $1/2 * \sqrt{S/N}$ , где S - площадь прямоугольника, а N - число наборов данных. Самые маленькие значения координат х и у несколько уменьшаются аналогичным

образом. Внешняя граница оказывает значительное влияние на весовые коэффициенты крайних точек.

## Диалог моделирования распределения

### Преобразование по методу нормальных меток

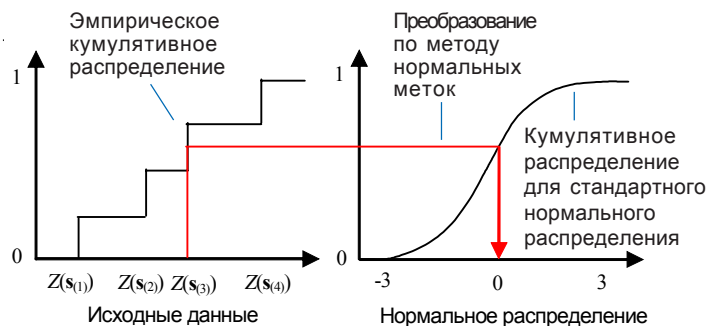
Для некоторых методов кригинга важно, чтобы данные подчинялись закону нормального распределения. Один из способов привести распределение данных к нормальному - использовать преобразование по методу нормальных меток. За разъяснениями обратитесь к книгам Journel and Huijbregts (1978, стр. 478), Isaaks и Srivastava (1989, стр. 469), Cressie (1993, стр. 281), Rivoirard (1994, стр. 46), Goovaerts (1998, стр. 266), и Chiles and Delfiner (1999, стр. 380). Однако многие иллюстрации в этих источниках вводят в заблуждение, поскольку на них показано, что функция кумулятивного распределения необработанных данных является непрерывной. В действительности это ступенчатая функция. Пусть статистический ряд данных выглядит следующим образом:  $Z(s_{(1)}), Z(s_{(2)}), \dots, Z(s_{(n)})$ , где  $Z(s_{(1)})$  - самое низкое значение, а  $Z(s_{(n)})$  - самое высокое значение. Предположим, что у нас есть только четыре значения ( $n = 4$ ); тогда эмпирическая функция кумулятивного распределения будет выглядеть примерно следующим образом:



Эта функция может быть сглажена различными способами. Также обратите внимание, что весовые значения по оси у - не должны быть приращениями ( $1/n$ ), если используется декластеризация по методу ячеек. Для выполнения преобразований в модуле Geostatistical Analyst предусмотрено несколько методов.

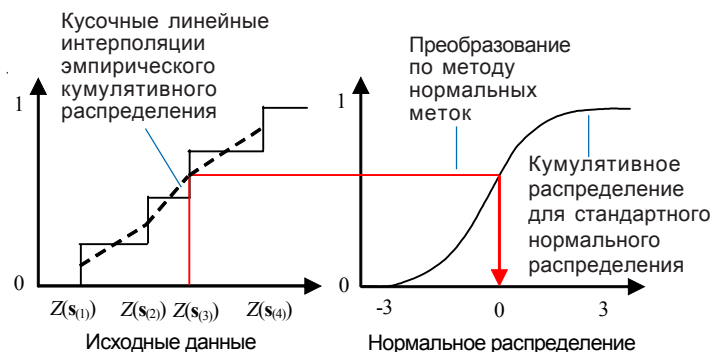
## Прямой способ

Он использует значения, которые соответствуют половине “шага” графика функции эмпирического распределения. На рисунке показано соответствие между исходными и трансформированными данными.



## Линейный метод

Линейный метод выполняет линейную интерполяцию по частям для исходной функции кумулятивного распределения. Суть этого метода легко понять, изучив рисунок:



## Смешанный Гауссов метод

Смесь из Гауссовых распределений может быть использована для сглаживания функции плотности вероятности. Модель функции плотности вероятности выглядит следующим образом:

$$p(z) = \sum_{i=1}^K \alpha_i p_i(z; \mu_i, \sigma_i^2)$$

где

$$\sum_{i=1}^K \alpha_i = 1$$

и

$$p_i(z; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(z - \mu_i)^2}{2\sigma_i^2}\right]$$

Параметры  $\alpha_i$ ,  $\mu_i$ , и  $\sigma_i$  оцениваются по степени максимального сходства, предполагая смесь из нормальных распределений и независимых данных. Кумулятивное распределение вычисляется через числовое интегрирование,

$$P(z) = \int_{-\infty}^z p(x) dx$$

и между  $P(z)$  и кумулятивным распределением для стандартного нормального распределения устанавливается соответствие, аналогично тому, как это делается при прямом и линейном методах.



## Диалог Вариограмма/Ковариация

### Определения вариограммы

Определение вариограммы приводится практически во всех работах по геостатистике (например, Journel и Huijbregts, 1978, стр. 31; Cressie, 1993, стр. 58; Goovaerts, 1997, стр. 68; Armstrong, 1998, стр. 19; Chiles и Delfiner, 1999, стр. 31; Stein, 1999, стр. 39). Обратите внимание, что некоторые авторы определяют ее как  $2g(\mathbf{h})$ , а другие как  $g(\mathbf{h})$ . Принято, что  $2g(\mathbf{h})$  - это вариограмма, а  $g(\mathbf{h})$  - полувариограмма.

### Определения ковариации

Определения функции ковариации применительно к пространственному анализу также приводятся практически во всех работах по геостатистике (например, Journel и Huijbregts, 1978, стр. 31; Isaaks и Srivastava, 1989, стр. 221; Cressie, 1993, стр. 53; Goovaerts, 1997, стр. 68; Armstrong, 1998, стр. 19; Chiles и Delfiner, 1999, стр. 30; Stein, 1999, стр. 15).

### Оценка вариограммы

Эмпирическая вариограмма может служить оценкой теоретической величины, определяемой вариограммой. Описание оценки при помощи эмпирической вариограммы также приведено практически во всех работах по геостатистике (например, Journel и Huijbregts, 1978, стр. 194; Isaaks и Srivastava, 1989, стр. 60; Cressie, 1993, стр. 69; Goovaerts, 1997, стр. 82; Armstrong, 1998, стр. 47; Chiles и Delfiner, 1999, стр. 36; Stein, 1999, стр. 39).

### Оценка ковариации

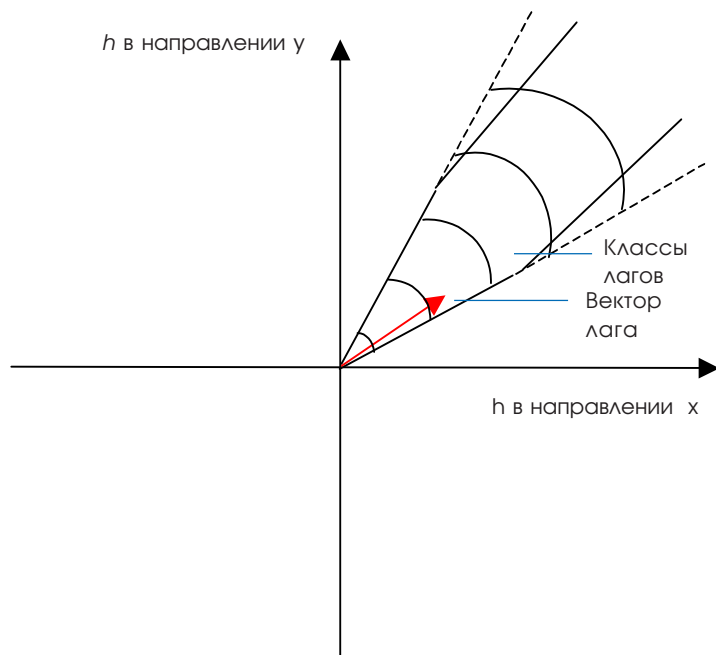
Эмпирическая ковариация служит для оценки теоретической величины, определяемой ковариационной функцией. Описание оценки с помощью эмпирической ковариации дано в многих работах по геостатистике (например, Journel и Huijbregts, 1978, стр. 192; Isaaks и Srivastava, 1989, стр. 59; Cressie, 1993, стр. 70; Goovaerts, 1997, стр. 86; Chiles и Delfiner, 1999, стр. 31; Stein, 1999, стр. 39).

### Бининг (объединение) оценок вариограммы и ковариации в классы лагов

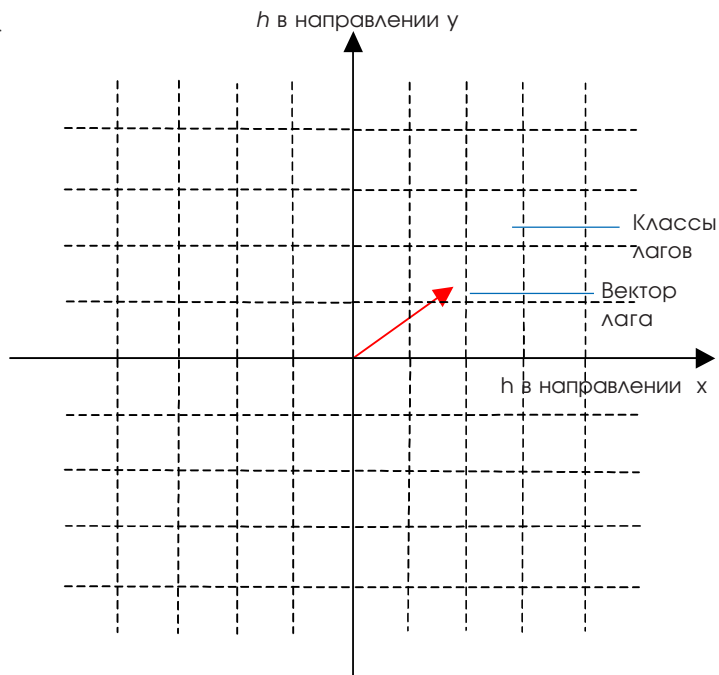
Оценки эмпирической вариограммы и ковариации обычно объединяются в классы лагов, основываясь на значении вектора  $\mathbf{h} = (h_x, h_y)^T$ , который разделяет пары точек, а затем значения вариограммы или ковариации усредняются для каждого бина. Чаще всего, бининг выполняется по радиальным секторам; мы назовем этот способ “методом сектора”.

Например, на рисунке показан метод, наиболее часто используемый в программных продуктах, включая программы GSLIB

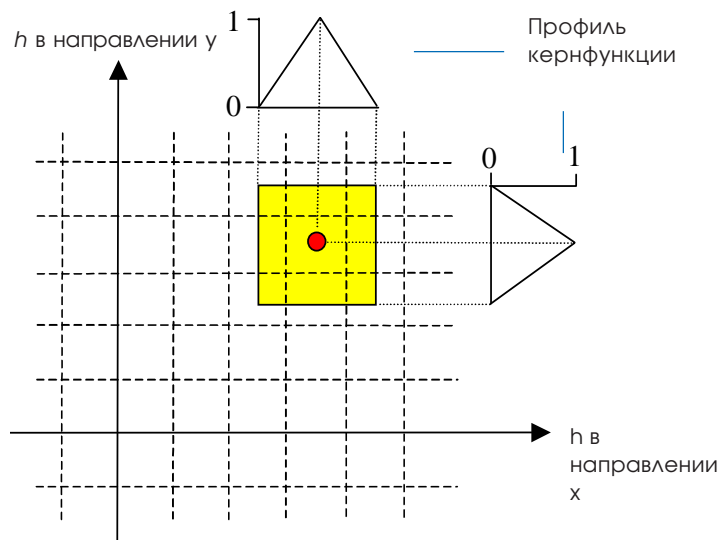




(Deutsch и Journel, 1992, стр. 45), Splus (*Splus Spatial Stats User Manual - Руководство пользователя*), и SAS (Technical Report - Технический отчет). В отличие от этого метода, в модуле Geostatistical Analyst лаги определяются по регулярной сетке; назовем этот способ “методом грида”.



Однако для данных, отобранных по регулярной сетке, определение границ бинов представляет определенную проблему. Для того, чтобы обойти эту проблему и сгладить эмпирическую вариограмму, модуль Geostatistical Analyst для определения взвешенных значений вариограммы для каждой ячейки использует метод кернфункции (значения определяются с учетом того, насколько близко точка расположена к центру ячейки). Значения весов для ячейки, содержащей точку, могут быть найдены как произведение двух маргинальных профилей, как показано на рисунке на стр. 255. Все ячейки вычисляются одинаково. Обратите внимание, что для любого лага будет вычислено четыре весовых коэффициента, и их сумма будет равна единице.



На рисунке видно, что любой вектор лага, который попадает в область, закрашенную желтым цветом, вносит вклад в класс (бин) лага с центром в красной точке. Чем ближе к центру бина, тем выше вес. То же самое выполняется для всех классов.

## Модели вариограммы и ковариации

### Связь между вариограммой и ковариацией

В дальнейшем будут даны формулы для вариограмм. Воспользовавшись связями между моделями вариограммы и ковариации, по вариограммам легко получить значения ковариации, и наоборот. Для внутренних стационарных процессов:

$$C(\mathbf{h}; \boldsymbol{\theta}) = \gamma(\infty; \boldsymbol{\theta}) - \gamma(\mathbf{h}; \boldsymbol{\theta}),$$

и

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = C(\mathbf{0}; \boldsymbol{\theta}) - C(\mathbf{h}; \boldsymbol{\theta}),$$

где  $\gamma(\infty; \boldsymbol{\theta})$  - порог вариограммы, а  $C(\mathbf{0}; \boldsymbol{\theta})$  - начало ковариационной функции. Эти связи поддерживаются только для вариограмм, имеющих порог; все модели вариограмм в модуле Geostatistical Analyst имеют порог.

### Геометрическая анизотропия

Геометрическая анизотропия может быть построена с использованием преобразования

$$\gamma_a(\mathbf{h}; \boldsymbol{\theta}, \Theta) = \gamma(\|\Theta \mathbf{h}\|; \boldsymbol{\theta}),$$

где  $\Theta$  - матрица  $2 \times 2$ , а  $\gamma(\mathbf{h}; \boldsymbol{\theta})$  - одна из моделей изотропной вариограммы, приведенной ниже. Вектор  $\mathbf{h} = (h_x, h_y)$  - вращается и масштабируется в новую систему координат, где область влияния вариограммы - эллипс. В модуле Geostatistical Analyst большой радиус влияния соответствует длинной оси эллипса, а малый радиус влияния - короткой оси эллипса. Параметры большого и малого радиусов влияния - это значения, при котором радиус влияния равен порогу для моделей, которые достигают пороговых значений, или 95 процентов значения порога для моделей, которые асимптотически приближаются к пороговому значению. Для более подробной информации обратитесь к следующей литературе: Journel и Huijbregts (1978, стр. 175), Isaaks и Srivastava (1989, стр. 377), Cressie (1993, стр. 64), Goovaerts (1997, стр. 90), или Armstrong (1998, стр. 28), Chiles и Delfiner (1999, стр. 93).

### Линейные комбинации моделей

Модели вариограммы могут представлять собой линейную комбинацию вариограмм:

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \gamma_1(\mathbf{h}; \boldsymbol{\theta}_1) + \gamma_2(\mathbf{h}; \boldsymbol{\theta}_2) + \dots$$

Модуль Geostatistical Analyst позволяет комбинировать до трех моделей, помимо модели, учитывающей эффект саморodka.

### Модель эффекта саморodka

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \begin{cases} 0 & \text{for } \mathbf{h} = 0 \\ \theta_s & \text{for } \mathbf{h} \neq 0 \end{cases}$$

где  $\theta_s \geq 0$ .

### Круговая модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \begin{cases} \frac{2\theta_s}{\pi} \left[ \frac{\|\mathbf{h}\|}{\theta_r} \sqrt{1 - \left(\frac{\|\mathbf{h}\|}{\theta_r}\right)^2} + \arcsin \frac{\|\mathbf{h}\|}{\theta_r} \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases} \quad \begin{matrix} \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \text{for } \theta_r < \|\mathbf{h}\| \end{matrix}$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния.

### Сферическая модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \begin{cases} \theta_s \left[ \frac{3}{2} \frac{\|\mathbf{h}\|}{\theta_r} - \frac{1}{2} \left(\frac{\|\mathbf{h}\|}{\theta_r}\right)^3 \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния.

### Тетрасферическая модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \begin{cases} \frac{2\theta_s}{\pi} \left[ \arcsin \left( \frac{\|\mathbf{h}\|}{\theta_r} \right) + \frac{\|\mathbf{h}\|}{\theta_r} \sqrt{1 - \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^2} + \frac{2}{3} \frac{\|\mathbf{h}\|}{\theta_r} \left( 1 - \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^2 \right)^{\frac{3}{2}} \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния.

### Пентасферическая модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \begin{cases} \theta_s \left[ \frac{15}{8} \frac{\|\mathbf{h}\|}{\theta_r} - \frac{5}{4} \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^3 + \frac{3}{8} \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^5 \right] & \text{for } 0 \leq \|\mathbf{h}\| \leq \theta_r \\ \theta_s & \text{for } \theta_r < \|\mathbf{h}\| \end{cases}$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния.

### Экспоненциальная модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{u}) = \theta_s \left[ 1 - \exp \left( - \frac{3 \|\mathbf{h}\|}{\theta_r} \right) \right] \text{ для всех } \mathbf{h},$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния.

### Гауссова модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{u}) = \theta_s \left[ 1 - \exp \left( - 3 \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^2 \right) \right] \text{ для всех } \mathbf{h},$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния. Поскольку эта модель нестабильна, если отсутствует самородок, по умолчанию модуль Geostatistical Analyst добавляет небольшое значение самородка, равное 1/1000 вычисленной дисперсии выборки.

### Рациональная квадратичная модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{u}) = \theta_s \frac{19 \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^2}{1 + 19 \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^2} \text{ для всех } \mathbf{h},$$

где  $\theta_s \geq 0$  - параметр частичного порога, а  $\theta_r \geq 0$  - параметр радиуса влияния.

### Модель эффекта дыры

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{u}) = \begin{cases} 0 & \text{for } \mathbf{h} = \mathbf{0} \\ \theta_s \frac{1 - \sin(2\pi \|\mathbf{h}\| / \theta_r)}{\sin(2\pi \|\mathbf{h}\| / \theta_r)} & \text{for } \mathbf{h} \neq \mathbf{0} \end{cases}$$

где  $\theta_s \geq 0$ .

### Модель К- Бесселя

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{u}) = \theta_s \left[ 1 - \frac{(\Omega_{\theta_k} \|\mathbf{h}\| / \theta_r)^{\theta_k}}{2^{\theta_k-1} \Gamma(\theta_k)} K_{\theta_k} (\Omega_{\theta_k} \|\mathbf{h}\| / \theta_r) \right] \text{ для всех } \mathbf{h},$$

где  $\theta_s \geq 0$ ,  $\theta_r \geq 0$ ,  $\theta_k \geq \Omega_{\theta_k}$  - значение, вычисленное таким образом, что  $\gamma(\theta_r) = 0.95 \theta_s$  для любого  $\theta_k$ ,  $\Gamma(\theta_k)$  - гамма-функция,

$$\Gamma(y) = \int_0^\infty x^{y-1} \exp(-x) dx$$

и  $K_{\theta_k}(\bullet)$  модифицированная функция Бесселя второго порядка  $\theta_k$  (Abramowitz и Stegun, 1965, стр. 374).

## Модель J-Бесселя

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \theta_s \left[ 1 - \frac{2^{\theta_d} \Gamma(\theta_d + 1)}{(\Omega_{\theta_d} \|\mathbf{h}\| / \theta_r)^{\theta_d}} J_{\theta_d} (\Omega_{\theta_d} \|\mathbf{h}\| / \theta_r) \right] \text{ для всех } \mathbf{h}$$

где  $\theta_s \geq 0$ ,  $\theta_r \geq 0$ ,  $\theta_d \geq 0$  должно удовлетворять условию,

$$\min_{B > 0, \gamma(B) = \theta_s, \gamma'(B) < 0} B = \theta_r$$

$\Gamma(\theta_k)$  - гамма-функция,

$$\Gamma(y) = \int_0^\infty x^{y-1} \exp(-x) dx$$

и  $J_{\theta_d}(\bullet)$  - функция J-Бесселя (Abramowitz и Stegun, 1965, стр. 358).

## Устойчивая модель

Модель вариограммы выглядит следующим образом:

$$\gamma(\mathbf{h}; \mathbf{n}) = \theta_s \left[ 1 - \exp \left( -3 \left( \frac{\|\mathbf{h}\|}{\theta_r} \right)^{\theta_e} \right) \right] \text{ для всех } \mathbf{h},$$

где  $\theta_s \geq 0$  и  $0 \leq \theta_e \leq 2$ . Поскольку эта модель нестабильна, если отсутствует самородок, по умолчанию модуль Geostatistical Analyst добавляет небольшое значение самородка, равное 1/1000 вычисленной дисперсии выборки.

## Модели взаимной ковариации

Модели взаимной ковариации в модуле Geostatistical Analyst используют модели корегинализации, что означает, что они принадлежат к тому же семейству, что и ковариационные формы моделей вариограммы, перечисленные ранее. Взаимные вариограммы (кросс-вариограммы) не используются в модуле Geostatistical Analyst. Традиционная взаимная вариограмма (Matheron, 1965) может быть применена только при определенных условиях (Journel и Huijbregts, 1978, стр. 236; Cressie, 1993, стр. 67; Ver Hoef и Cressie, 1993), и даже в этом случае не является оптимальной. Взаимная ковариация допускает модели, которые могут иметь некоторые пространственные сдвиги (Journel и Huijbregts, 1978, стр. 41; Ver Hoef и Cressie, 1993), а эмпирическая поверхность взаимной ковариации позволяет пользователю визуально определить и оценить эти смещения.

Однотипные модели для взаимной ковариации (иногда преподносимые как кроссвариограммы, но при этом их идеи и модели адаптированы под взаимную ковариацию) описаны в работах Journel и Huijbregts (1978, стр. 40), Isaaks и Srivastava (1989, стр. 390), Goovaerts, (1997, стр. 107), и Chiles и Delfiner (1999, стр. 339). Модуль Geostatistical Analyst применяет эти модели, допуская пространственное смещение между двумя переменными (Ver Hoef и Cressie, 1993). Это добавляет к модели два параметра, определяющих сдвиг по оси x и по оси y.

## Подбор моделей вариограммы и ковариации

Алгоритм подбора начинается с получения предварительной оценки диапазона данных, и этот процесс носит название первой стадии. Используйте  $Z_j^k(\mathbf{s}_i)$  для обозначения  $j$ -го измерения переменной типа  $k$  в  $i$ -той точке пространства  $\mathbf{s}_i$ .

### Стадия 1

Сначала модуль Geostatistical Analyst сопоставляет каждый набор данных,  $\tilde{Z}_j^k(\mathbf{s}_i) = Z_j^k(\mathbf{s}_i) / s_k$  где  $s_k$  - стандартное отклонение выборки. Стадия 1 начинается с предположения, что мы имеем дело с изотропной моделью; вычисляется эмпирическая вариограмма (или ковариация) для сопоставленных данных  $\tilde{Z}_j^k(\mathbf{s}_i)$  с использованием метода сектора (как было описано ранее в разделе 'Биннинг (объединение) оценок вариограммы и ковариации в классы лагов') для большого диапазона классов лагов, которые увеличиваются в геометрической прогрессии. Классы лагов образуются из интервалов  $[d^{k-1/2}, d^{k+1/2})$ , где  $d = 1.25$ , и  $k$  меняется от наименьшего аппаратного целого до наибольшего. Центр каждого класса лагов принимается равным  $d^k \cosh(\frac{1}{6} \log d^k)$ . Очевидно, что многие классы лагов останутся пустыми, и модуль Geostatistical Analyst использует только те, которые содержат данные. Назовем это эмпирической (взаимной) ковариацией  $\mathcal{E}_{ij}(\mathbf{h}_k)$ , где  $i$  обозначает  $i$ -ый тип переменной,  $j$  обозначает  $j$ -ый тип переменной, и  $k$  обозначает  $k$ -ый класс лагов. Первая итерация оценок параметров выполняется путем минимизации,

$$\sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^{n_{ij}} w_{ij}(\mathbf{h}_k) \left( \tilde{\mathcal{C}}_{ij}(\mathbf{h}_k; \mathbf{u}_{ij}^{(1)}) - \mathcal{E}_{ij}(\mathbf{h}_k) \right)^2 \quad (1)$$

для  $\boldsymbol{\theta}$ , где  $\boldsymbol{\theta}_{ij}$  вектор параметров для ковариационной функции  $i, j$ , и  $\boldsymbol{\theta}$  содержит все ковариационные параметры, где

$$w_{ij}(\mathbf{h}_k) = N_{ij}(\mathbf{h}_k) / \sum_{m=1}^{n_{ij}} N_{ij}(\mathbf{h}_m) \quad (2)$$

и  $N_{ij}(\mathbf{h}_k)$  - число пар в эмпирической (взаимной) ковариационной функции для переменных  $i$  и  $j$  в классе  $k$ . Назовем эту оценку  $\boldsymbol{\theta}^{(1)}$ . В следующей итерации модуль Geostatistical Analyst использует взвешенные наименьшие квадраты Кресси (1985) путем повторной минимизации (1), только в этот раз,

$$\varpi_{ij}(\mathbf{h}_k; \mathbf{u}_{ij}^{(1)}) = \frac{N_{ij}(\mathbf{h}_k)}{\tilde{\mathcal{C}}_{ii}(\mathbf{0}; \mathbf{u}_{ii}^{(1)}) \tilde{\mathcal{C}}_{jj}(\mathbf{0}; \mathbf{u}_{jj}^{(1)}) + \tilde{\mathcal{C}}_{ij}^2(\mathbf{h}_k; \mathbf{u}_{ij}^{(1)})}$$

а затем эти веса нормируются таким образом, что каждому значению (взаимной) ковариации присваиваются равные веса,

$$w_{ij}(\mathbf{h}_k) = \varpi_{ij}(\mathbf{h}_k; \mathbf{u}_{ij}^{(1)}) / \sum_{m=1}^{n_{ij}} \varpi_{ij}(\mathbf{h}_m; \mathbf{u}_{ij}^{(1)}) \quad (4)$$

Назовем эту оценку  $\boldsymbol{\theta}^{(2)}$ . Обратите внимание, что если вместо ковариации мы используем вариограмму,  $\mathbf{u}_{ii}^{(2)}$  равно

$$\arg \min_{\boldsymbol{\theta}_{ii}} \left[ \sum_{k=1}^{n_{ij}} w_{ii}(\mathbf{h}_k) \left( \tilde{\gamma}_{ii}(\mathbf{h}_k; \mathbf{u}_{ii}) - \mathcal{E}_{ii}(\mathbf{h}_k) \right)^2 \right] \quad (5)$$

где  $w_{ii}(\mathbf{h}_k)$  определяется по формуле (2), а затем  $\mathbf{u}_{ii}^{(2)}$  вычисляется по формуле (5) с весовыми значениями как в формуле (4), но теперь

$$\varpi_{ij}(\mathbf{h}_k; \mathbf{u}_{ii}^{(1)}) = \frac{N_{ij}(\mathbf{h}_k)}{\gamma_{ii}^2(\mathbf{h}; \mathbf{u}_{ii}^{(1)})}$$

Оценки  $\boldsymbol{\theta}^{(1)}$  и  $\boldsymbol{\theta}^{(2)}$  - это два шага в алгоритме, который использует несколько итераций для пересчета весовых значений по методу наименьших квадратов.

Оценка  $\boldsymbol{\theta}^{(2)}$  используется только для оценки диапазона для размера лага, предложенного по умолчанию для метода грида, при выполнении оценки эмпирической вариограммы или ковариации. Число лагов, предлагаемое по умолчанию, равно 12, следовательно, размер лага для метода грида принимается равным "2\*диапазон/12".

## Шаг 2

Шаг 2 по сути повторяет шаг 1, но работает с эмпирической вариограммой или (взаимной) ковариацией сопоставленных данных  $\tilde{Z}_j^k(\mathbf{s}_i)$  которые используют метод грида (описанный ранее в разделе ‘Объединение (бининг) оценок вариограммы и ковариации’), где размер лага, предложенный по умолчанию, определяется по оценке диапазона из итерации  $\boldsymbol{\theta}^{(2)}$  первого шага. Он также допускает использование анизотропии и линейных комбинаций из не более, чем трех моделей (взаимной) ковариации или вариограммы, с учетом эффекта саморodka для классов лагов для каждого набора данных,

$$\tilde{C}_{ij}(\mathbf{h}; \mathbf{u}) = \sum_{u=1}^S B_u(i, j) \rho_u(\mathbf{h}; \mathbf{u}_u).$$

В данной формуле,  $B_u(i, j)$  - параметр частичного порога, а также  $i, j$  -тый компонент  $\mathbf{B}_u$ , а  $T \times T$  положительно-определенная матрица, в которой  $T$  - число  $\delta \hat{i} \hat{i} \hat{a}$  переменных,  $S$  - число различных моделей (взаимной) ковариации, используемых в линейной комбинации, а функция  $\rho_u(\mathbf{h}; \boldsymbol{\Phi}_u)$  - нормализованная модель ковариации;  $\rho_u(\mathbf{0}; \boldsymbol{\Phi}_u) = 1$ , где  $\boldsymbol{\Phi}_u$  - параметры, которые обычно контролируют диапазон (и/или форму) модели ковариации. Как и ранее,  $\boldsymbol{\theta}$  включает все параметры. Третья итерация оценок параметров,  $\boldsymbol{\theta}^{(3)}$  осуществляется путем минимизации формулы (1) с весами (2) для эмпирической ковариации с использованием метода грида, а затем четвертая итерация  $\boldsymbol{\theta}^{(4)}$  выполняется путем минимизации (1) с весами, взятыми из формул (4) и (3) для эмпирической вариограммы, использующей метод грида. Эти формулы очевидно меняются, если мы используем вариограммы, как показано для шага 1. Теперь вернемся к исходному масштабу. Окончательные модели (взаимной) ковариации выглядят следующим образом:

$$C_{ij}(\mathbf{h}) = s_i s_j \tilde{C}_{ij}(\mathbf{h}; \mathbf{u}_{ij}^{(4)}),$$

и для вариограмм:

$$\gamma_{ii}(\mathbf{h}) = s_i^2 \tilde{\gamma}_{ii}(\mathbf{h}; \mathbf{u}_{ii}^{(4)}).$$

Если пользователь меняет какие-либо параметры (например, размер лага), оценки пересчитываются, начиная со второго шага.

## Диалог двумерного распределения

Дизъюнктивный кригинг требует, чтобы все пары данных подчинялись двумерному нормальному распределению. Это условие тяжело проверить на практике. Модуль Geostatistical Analyst предоставляет визуальный инструмент, который помогает оценить выполнение условия двумерной нормальности. Теоретическая кривая, как функция лага, может быть построена на основе различных пороговых значений для функции кумулятивного распределения. (см. Deutsch и Journel, 1992, стр. 139 и Goovaerts, 1998, стр. 265). Эту теоретическую кривую можно сравнить с эмпирической вариограммой на основе индикаторов. Более широко, если графики КК показывают нормальное распределение компоненты многомерной случайной величины, и похоже, что данные подчиняются двумерному нормальному распределению, разумно предположить и наличие многомерного нормального распределения данных. Таким образом, проверка на двумерное нормальное распределение может быть использована для простого кригинга, что позволяет пользователю удостовериться в том, что карты квантилей и вероятностей основаны на разумных предположениях.



## Формулы кригинга

Честь разработки методов кригинга принадлежит многим авторам. Кригинг с ковариацией эквивалентен методу лучшей линейной несмещенной интерполяции (best linear unbiased prediction - BLUP); методы простого, ординарного и универсального кригинга были разработаны такими учеными, как Wold (1938), Холмогоров (1941), Wiener (1949), Гандин (1959), Goldberger (1962), и Henderson (1963). Методы пространственной интерполяции с использованием вариограмм были сформулированы Гандиным (1959, 1963) и Matheron (1962, 1969). Обратитесь к работам Journel (1983) по индикаторному кригингу, Sullivan (1984) по вероятностному кригингу, и Matheron (1976) по дизъюнктивному кригингу. В работе Cressie (1990) дана подробная информация об основах кригинга.

Модуль Geostatistical Analyst включает интерполяторы, которые могут учитывать ошибку измерений. Эти модели рассматриваются в работах Гандина (1959, 1960, 1963) и Cressie (1986, 1988, 1993, стр. 127–135). К этим же моделям относятся и часто приводимые во многих учебниках по геостатистике так называемые “жесткие” интерполяторы. Под понятием “жесткий” подразумевается, что если интерполяция выполняется для точки, в которой отбирались данные, вычисленное значение будет в точности равно значению, полученному при выполнении измерений в этой точке, и для нее стандартная ошибка интерполяции равна нулю. Это может приводить к созданию странных карт, поскольку в опорных точках проинтерполированные значения будут как бы “подскакивать”. При наличии ошибок измерений, у пользователя может возникнуть желание “отфильтровать” измерения и составить более “сглаженную” карту проинтерполированных значений. Моделирование ошибок измерений возможно только для ординарного, простого и универсального кригинга. Описание всех моделей приведено далее. Начнем с замечания. Иногда, необходимо будет иметь дело с множественными опорными точками, с множественными измерения-

ми в одной точке (ошибкой измерений) для множественных типов переменных (при кокригинге). Для обозначения  $j$ -ого измерения переменной типа  $k$  в  $i$ -ой точке пространства  $s_i$  мы используем обозначение  $Z_j^k(s_i)$ .

## Оценка ошибки измерений

Если есть несколько наблюдений (измерений) для каждой опорной точки, модуль Geostatistical Analyst может оценить ошибку измерений. Формула для вычисления ошибки измерений следующая,

$$\sigma_{ME}^2 = \frac{\sum_{s_i \in D} \sum_{j=1}^{n_i} (Z_j(s_i) - \bar{Z}(s_i))^2}{N - n_D},$$

где  $D$  - набор всех опорных точек, в котором выполнено более одного измерения,  $Z_j(s_i)$  -  $j$ -ое измерение в точке  $s_i$ ,  $\bar{Z}(s_i)$  - среднее значение в точке  $s_i$ ,  $n_i$  - число наблюдений в точке  $s_i \in D$ ,  $N = \sum_i n_i$  для всех  $s_i$  в  $D$ , и  $n_D$  - число опорных точек в  $D$ . Движок в диалоге по умолчанию устанавливается в положении, которое соответствует этому значению; это означает, что он установлен на значении 100 ( $\sigma_{ME}^2$  / самородок) процентов. Пользователи при желании могут заменить  $\sigma_{ME}^2$ ; при выполнении оценки устанавливается, что эффект самородка не может быть меньше, чем  $\sigma_{ME}^2$ ; когда пользователь меняет значение самородка или значение, тогда сохраняется неравенство “самородок  $\geq \sigma_{ME}^2$ ”.

## Ординарный кригинг

Заинтересованный читатель может обратиться к работе Cressie (1993, стр. 127–135) за дополнительными разъяснениями по использованию кригинга с ошибкой измерений; далее приведена сжатая версия того, как ординарный кригинг реализован в модуле Geostatistical Analyst. Как и в Главе 6, предположим, что данные - это реализация пространственно автокоррелирующего процесса плюс независимые случайные ошибки:

$$Z_t(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon_t(\mathbf{s}),$$

но теперь разложим случайные ошибки,

$$\varepsilon_t(\mathbf{s}) = Y(\mathbf{s}) + \eta(\mathbf{s}) + \delta_t(\mathbf{s}),$$

где  $Z_t(\mathbf{s})$  обозначает  $t$ -ую реализацию в точке  $\mathbf{s}_t$ , и пусть  $n_t$  - число измерений в точке  $\mathbf{s}_t$ . Часто  $n_t = 1$ , и если  $n_t > 1$ , формируется модель ошибки измерений. Предполагается, что:

- $\mu(\mathbf{s}) = m$  - неизвестное детерминистское среднее значение.
- $Y(\mathbf{s})$  - сглаженный стационарный процесс второго порядка, чей радиус влияния для автокорреляции можно определить с помощью эмпирической вариограммы или ковариации.
- $E(Y(\mathbf{s})) = 0$ .
- $\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = C_y(\mathbf{h})$ , и нет дополнительного эффекта самородка в процессе  $Y(\mathbf{s})$ .
- $h(\mathbf{s})$  сглаженный стационарный процесс второго порядка, радиус влияния вариограммы которого настолько близок к 0, что он меньше, чем все действительные расстояния между опорными и интерполируемыми точками.
- $E(h(\mathbf{s})) = 0$ .
- $\text{Cov}(h(\mathbf{s}), h(\mathbf{s}+\mathbf{h})) = C_h(\mathbf{h})$  with  $C_h(\mathbf{1}') = 0$ .
- $d_t(\mathbf{s})$  продолжительный (белый) шум, состоящий из ошибок измерений.

- $E(d_t(\mathbf{s})) = 0$ , для всех  $\mathbf{s}$  и  $t$ .
- $\text{Cov}(d_t(\mathbf{s}), d_u(\mathbf{s}+\mathbf{h})) = s^2$  если  $\mathbf{h} = \mathbf{0}$  и  $t = u$ , в противном случае он равен 0.
- $Y(\cdot)$ ,  $h(\cdot)$ , и  $d(\cdot)$  независимы.

Предположим здесь, что эффект самородка, обозначаемый как  $v$ , состоит из двух частей: вариации на микроуровне и ошибки измерений. То есть,  $v = C_\eta(\mathbf{0}) + \sigma^2$ . Из этой модели вы можете вывести, что

$$\text{Cov}(Z_t(\mathbf{s}), Z_u(\mathbf{s}+\mathbf{h})) = \begin{cases} C_y(\mathbf{h}) + C_\eta(\mathbf{h}) & \text{if } \mathbf{h} \neq \mathbf{0} \\ C_y(\mathbf{0}) + C_\eta(\mathbf{0}) & \text{if } \mathbf{h} = \mathbf{0} \text{ and } t \neq u \\ C_y(\mathbf{0}) + C_\eta(\mathbf{0}) + \sigma^2 & \text{if } \mathbf{h} = \mathbf{0} \text{ and } t = u \end{cases}$$

Если существует ошибка измерений, вы захотите интерполировать фильтрованные (без шумов) величины  $S(\mathbf{s}_0) \equiv \mu + Y(\mathbf{s}_0) + \eta(\mathbf{s}_0)$  в точках  $\mathbf{s}_0$ ; то есть, вычесть ошибку измерений. Если ошибки измерений нет,  $S(\mathbf{s}_0) = Z(\mathbf{s}_0)$ . Ординарный кригинг с ошибкой измерений применяется для линейного интерполятора,

$$\mathcal{E}(\mathbf{s}_0) = \boldsymbol{\lambda}'\mathbf{z},$$

затем минимизируем,

$$E(S(\mathbf{s}_0) - \boldsymbol{\lambda}'\mathbf{z})^2,$$

где  $\mathbf{z}$  - вектор из наблюдаемых данных, а  $\boldsymbol{\lambda}$  - вектор из весов кригинга. Условие несмещенности,

$$E(S(\mathbf{s}_0) - \boldsymbol{\lambda}'\mathbf{z}) = 0,$$

подразумевает, что  $\boldsymbol{\lambda}'\mathbf{1} = 1$ , а это в свою очередь приводит к необходимости использования при минимизации множителя Лагранжа. Таким образом, получаем уравнения кригинга,

$$\begin{pmatrix} \mathbf{y}_z & \mathbf{1} \\ \mathbf{1}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ m \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ 1 \end{pmatrix}$$

где  $m$  - множитель Лагранжа,  $\Sigma_z$  - ковариационная матрица для данных, и  $\mathbf{c}$  равно  $\text{Cov}(\mathbf{z}, S(\mathbf{s}_0)) = \text{Cov}(\mathbf{z}, Y(\mathbf{s}_0) + \eta(\mathbf{s}_0))$ . Предположив, что радиус влияния  $\eta(\cdot)$  очень близок к 0, вы можете допустить, что  $\text{Cov}(\mathbf{z}, \eta(\mathbf{s}_0)) = \mathbf{0}$  для всех фактических расстояний, за исключением случая, когда  $\mathbf{s}_0 = \mathbf{s}_i$ , где  $\mathbf{s}_i$  - одна из точек наблюдения; тогда  $\text{Cov}(Z(\mathbf{s}_i), \eta(\mathbf{s}_i)) = C_\eta(\mathbf{0})$ , что требует оценки. Общий эффект самородка может быть оценен, но напомним, что он состоит из двух частей,  $v = \sigma^2 + C_\eta(\mathbf{0})$ . Если есть независимая оценка  $\sigma^2$ , тогда вы можете оценить  $C_\eta(\mathbf{0}) = v - \sigma^2$ . Это эквивалентно определению той части эффекта самородка, которая соответствует ошибке измерений, и той части, которая соответствует вариации на микроуровне;  $0 \leq \pi < 1$ , и могут быть установлены тождества  $\sigma^2 = \pi v$  и  $C_\eta(\mathbf{0}) = (1 - \pi)v$ . Если для одной точки есть несколько измерений, ошибка измерений может быть оценена так, как это было показано ранее.

После того, как определены  $\sigma^2$  и  $C_\eta(\mathbf{0})$ , переходите к решению уравнений кригинга. Если весь эффект самородка - это вариация на микроуровне  $\eta(\cdot)$  (т.е., нет ошибки измерений), то в результате решения уравнений кригинга, мы получим жесткий кригинг. Для  $\lambda$  получаем,

$$\lambda = \Sigma_z^{-1}(\mathbf{c} - \mathbf{1}m) \text{ где } m = (\mathbf{1}' \Sigma_z^{-1} \mathbf{c} - 1) / \mathbf{1}' \Sigma_z^{-1} \mathbf{1},$$

для интерполятора ординарного кригинга. Подставив в это уравнение  $\lambda$ , получим среднеквадратичную ошибку интерполяции,

$$\begin{aligned} E(S(\mathbf{s}_0) - \lambda' \mathbf{z})^2 \\ &= C_y(\mathbf{0}) + C_\eta(\mathbf{0}) - \lambda' \mathbf{c} - m, \\ &= C_y(\mathbf{0}) + (1 - \pi)v - \lambda' \mathbf{c} - m, \end{aligned}$$

следовательно, стандартные ошибки интерполяции равны:

$$\mathcal{E}_S(\mathbf{s}_0) = \sqrt{C_y(\mathbf{0}) + (1 - \pi)v - \lambda' \mathbf{c} - m}$$

## Вычисление нового значения для перекрестной проверки и проверки

При выполнении перекрестной проверки, вы не хотите интерполировать  $S(\mathbf{s}_0)$ , версию данных без шумов, но должны интерполировать  $Z_u(\mathbf{s}_0)$ , с ошибкой измерений, для того, чтобы стандартные ошибки вычислений отразили среднеквадратичную ошибку вычислений, полученную в результате выполнения перекрестной проверки. Вычисление “нового значения” выполняется для линейного интерполатора

$$\mathcal{Z}_u(\mathbf{s}_0) = \lambda' \mathbf{z},$$

затем минимизируем,

$$E(Z_u(\mathbf{s}_0) - \lambda' \mathbf{z})^2.$$

Предположим, что если  $\mathbf{s}_0 = \mathbf{s}_i \in D$ , то  $u > n_i$ . Аналогично предыдущему примеру, получаем уравнения кригинга,

$$\begin{pmatrix} \mathbf{Y}_z & \mathbf{1} \\ \mathbf{1}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \lambda \\ m \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ 1 \end{pmatrix},$$

где  $m$  - множитель Лагранжа,  $\Sigma_z$  - ковариационная матрица для данных и  $\mathbf{c}$  равно  $\text{Cov}(\mathbf{z}, Z_u(\mathbf{s}_0)) = \text{Cov}(\mathbf{z}, Y(\mathbf{s}_0) + \eta(\mathbf{s}_0) + \delta_u(\mathbf{s}_0))$ . Для  $\lambda$  получаем,

$$\lambda = \Sigma_z^{-1}(\mathbf{c} - \mathbf{1}m), \text{ где } m = (\mathbf{1}' \Sigma_z^{-1} \mathbf{c} - 1) / \mathbf{1}' \Sigma_z^{-1} \mathbf{1}.$$

Обратите внимание, что когда  $\mathbf{s}_0 = \mathbf{s}_i \in D$ , интерполированное значение  $\mathcal{Z}_u(\mathbf{s}_i)$  как правило  $i$  равно ни одному из наблюдаемых значений  $z_i(\mathbf{s}_i)$ ;  $t \leq n$ . Однако, подставив  $\lambda$ , чтобы найти среднеквадратичную ошибку интерполяции, получим

$$\begin{aligned} E(Z_u(\mathbf{s}_0) - \lambda' \mathbf{z})^2 \\ &= C_y(\mathbf{0}) + C_\eta(\mathbf{0}) + \sigma^2 - \lambda' \mathbf{c} - m, \\ &= C_y(\mathbf{0}) + v - \lambda' \mathbf{c} - m, \end{aligned}$$

следовательно, стандартные ошибки интерполяции равны

$$\mathcal{E}_Z(\mathbf{s}_0) = \sqrt{C_y(\mathbf{0}) + v - \lambda' \mathbf{c} - m}.$$

Эти ошибки должны быть сопоставлены со стандартными ошибками интерполяции  $\mathcal{E}_S(\mathbf{s}_0)$  для версии данных без шумов. Обратите внимание, что когда  $\mathbf{s}_0 = \mathbf{s}_i$  для одной из точек наблюдений  $\mathbf{s}_i \in D$ , ни одна из стандартных ошибок вычислений не будет равна 0. Для более подробной информации о кригинге с новым значением, обратитесь к работе Krivoruchko, Gribov, и Ver Hoef, 2000.

## Карты вероятностей и квантилей

Если случайные ошибки подчиняются закону нормального распределения и обладают стационарностью (либо второго порядка, либо внутренней), то ошибка интерполяции  $\mathcal{F}(\mathbf{s}_0) - S(\mathbf{s}_0)$  тоже подчиняется закону нормального распределения, с нулевым средним значением и дисперсией, равной  $\mathcal{E}_S^2(\mathbf{s}_0)$ . Нормальность распределения позволяет строить карты вероятностей или квантилей.

## Простой кригинг

Здесь приведено сжатое описание реализации простого кригинга в модуле Geostatistical Analyst. Чтобы понять, что представляет собой модель и какие делаются предположения, обратитесь к разделу, посвященному ординарному кригингу. Для простого кригинга одно предположение меняется на следующее:

$\mu(s)$  - известная, детерминистская средняя функция.

При ординарном же кригинге, необходимо интерполировать отфильтрованную (без шумов) величину  $S(s_0) \equiv \mu(s_0) + Y(s_0) + \eta(s_0)$  в точке  $s_0$ .

### Интерполяция с ошибкой измерений

Простой кригинг с ошибкой измерений получается для линейного интерполятора,

$$(s_0) = \lambda'z + k$$

затем минимизируем,

$$E(S(s_0) - \lambda'z - k)^2 = \text{Var}[Y(s_0) + \eta(s_0) - \lambda'z] + [\mu(s_0) - \lambda'\mu - k]^2,$$

где  $\mu$  - вектор из известных средних для всех полученных данных. Минимизация достигается при определении  $k = \mu(s_0) - \lambda'\mu$  и  $\lambda = \Sigma_z^{-1}c$ , где  $\Sigma_z$  ковариационная матрица данных, и  $c$  равно  $\text{Cov}(z, Y(s_0) + \eta(s_0))$ . Затем получаем интерполятор для простого кригинга,

$$(s_0) = \lambda'z + k = c'\Sigma_z^{-1}(z - \mu) + \mu(s_0).$$

Подставляем, чтобы получить среднеквадратичные ошибки интерполяции,

$$\begin{aligned} E(S(s_0) - \lambda'z - k)^2 &= C_y(\mathbf{0}) + C_\eta(\mathbf{0}) - c'\Sigma_z^{-1}c \\ &= C_y(\mathbf{0}) + (1 - \pi)v - c'\Sigma_z^{-1}c, \end{aligned}$$

и стандартные ошибки интерполяции равны:

$$\sigma_S(s_0) = \sqrt{C_y(\mathbf{0}) + (1 - \pi)v - \pi'c}$$

## Вычисление нового значения для перекрестной проверки

При выполнении перекрестной проверки нежелательно интерполировать значения версии данных “без шумов” — наоборот, следует интерполировать значения с ошибкой измерений с тем, чтобы стандартные ошибки интерполяции отразили среднеквадратичную ошибку интерполяции, полученную при выполнении перекрестной проверки. Вычисление “нового значения” выполняется для линейного интерполятора,

$$u(s_0) = \lambda'z + k,$$

затем минимизируем,

$$E(Z_u(s_0) - \lambda'z - k)^2.$$

Предположим, что если  $s_0 = s_i \in D$ , то  $u > n_i$ . Повторяя выполненные ранее операции, получаем интерполятор кригинга,

$$u(s_0) = \lambda'z + k = c'\Sigma_z^{-1}(z - \mu) + \mu(s_0),$$

со среднеквадратичными ошибками интерполяции,

$$\begin{aligned} E(Z_u(s_0) - \lambda'z - k)^2 &= C_y(\mathbf{0}) + C_\eta(\mathbf{0}) + \sigma^2 - c'\Sigma_z^{-1}c \\ &= C_y(\mathbf{0}) + v + c'\Sigma_z^{-1}c, \end{aligned}$$

в результате, стандартные ошибки интерполяции равны:

$$\sigma_Z(s_0) = \sqrt{C_y(\mathbf{0}) + v - \pi'c}.$$

Эти ошибки необходимо сравнить со стандартными ошибками интерполяции для версии данных без шумов. Обратите внимание, что когда  $s_0 = s_i$  для одной из точек наблюдений  $s_i \in D$ , ни одна из стандартных ошибок интерполяции не будет равна 0.

## Карты вероятности и квантилей

Если данные подчиняются закону связанного многомерного нормального распределения, где

$$\begin{bmatrix} S(s_0) \\ \mathbf{z} \end{bmatrix} \sim N \left( \begin{bmatrix} \mu(s_0) \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} C_y(\mathbf{0}) + C_\eta(\mathbf{0}) & \mathbf{c} \\ \mathbf{c}' & \mathbf{Y}_z \end{bmatrix} \right)$$

то  $\mathcal{E}(s_0)$  - условное ожидание,  $E(S(s_0) | \mathbf{z})$ , а известное свойство многомерного нормального распределения  $\mathcal{E}(s_0)$  состоит в том, что

$$S(s_0) | \mathbf{z} \sim N(\mathbf{c}'\Sigma_z^{-1}(\mathbf{z} - \boldsymbol{\mu}) + \mu(s_0), C_y(\mathbf{0}) + C_\eta(\mathbf{0}) - \mathbf{c}'\Sigma_z^{-1}\mathbf{c}).$$

Поскольку проинтерполированные значения также подчиняются закону нормального распределения, возможно вычислить значения вероятностей, или по аналогии построить карту квантилей. Также обратите внимание, что условное ожидание  $E(S(s_0) | \mathbf{z})$  - наилучший интерполятор из всех интерполяторов, линейных и нелинейных, так как имеет наименьшую среднеквадратичную ошибку интерполяции.

## Универсальный кригинг

Предположим, что у нас есть следующая модель:

$$Z_i(s) = [x(s)]' \beta + \varepsilon_i(s),$$

теперь разложим на составляющие случайные ошибки:

$$\varepsilon_i(s) = Y(s) + \eta(s) + \delta_i(s),$$

где  $X$  - предполагаемая матрица и  $\beta$  - вектор параметров, другими словами, она аналогична модели ординарного кригинга, с теми же допущениями, за исключением одного,

$\mu(s) = [x(s)]' \beta$ , где  $x(s)$  - вектор ковариат наблюдений и  $b$  - вектор из *неизвестных* параметров.

### Интерполяция с ошибкой измерений

и для ординарного кригинга, интерполируем фильтрованную (без шумов) величину  $S(s_0) \equiv [x(s_0)]' \beta + Y(s_0) + \eta(s_0)$  в точке  $s_0$ . Матрица  $X$  имеет столбец из единиц, другие столбцы содержат полиномиальные функции пространственных координат в точке  $s$ . Универсальный кригинг с ошибкой измерений получается для линейного интерполятора,

$$S(s_0) = \lambda' z,$$

затем минимизируем,

$$E(S(s_0) - \lambda' z)^2,$$

где  $z$  - вектор данных наблюдений, а  $\lambda$  - вектор весов кригинга. Условие несмещенности выглядит следующим образом:

$$E(S(s_0) - \lambda' z) = 0,$$

значит  $X' \lambda = x(s_0)$ . Повторив операции как для ординарного кригинга, получаем уравнения универсального кригинга,

$$\begin{pmatrix} Y_z & X \\ X' & 0 \end{pmatrix} \begin{pmatrix} l \\ m \end{pmatrix} = \begin{pmatrix} c \\ x(s_0) \end{pmatrix}$$

где  $m$  - вектор из множителей Лагранжа,  $\Sigma_z$  - ковариационная матрица данных и  $c$  равно  $\text{Cov}(z, S(s_0))$ . Вычислив  $\lambda$ , получаем интерполятор для универсального кригинга,

$$\lambda = \Sigma_z^{-1} (c - X m), \quad m = (X' \Sigma_z^{-1} X)^{-1} (X' \Sigma_z^{-1} c - x(s_0)).$$

Подставляем, чтобы получить среднеквадратичные ошибки интерполяции,

$$\begin{aligned} E(S(s_0) - \lambda' z)^2 \\ = C_y(0) + C_\eta(0) - \lambda' (c + X m), \\ = C_y(0) + (1 - \pi) v - \lambda' (c + X m), \end{aligned}$$

и стандартные ошибки интерполяции равны:

$$\sigma_S(s_0) = \sqrt{C_y(0) + (1 - \pi) v - \lambda' (c + X m)}.$$

### Вычисление нового значения для перекрестной проверки

При выполнении перекрестной проверки не интерполируйте  $S(s_0)$ , значения версии данных “без шумов”; наоборот, следует интерполировать значения с ошибкой измерений  $Z_u(s_0)$  с тем, чтобы стандартные ошибки интерполяции отразили среднеквадратичную ошибку интерполяции, полученную при выполнении перекрестной проверки. Вычисление “нового значения” выполняется для линейного интерполятора,

$$Z_u(s_0) = \lambda' z,$$

затем минимизируем,

$$E(Z_u(s_0) - \lambda' z)^2.$$

Предположим, что если  $s_0 = s_i \in D$ , то  $u > n_i$ . Повторяя выполненные ранее операции, получаем уравнения универсального кригинга,

$$\begin{pmatrix} Y_z & X \\ X' & 0 \end{pmatrix} \begin{pmatrix} l \\ m \end{pmatrix} = \begin{pmatrix} c \\ x(s_0) \end{pmatrix}$$

Вычислив  $\lambda$ , получаем интерполятор для универсального кригинга,

$$\lambda = \Sigma_z^{-1} (c - X m), \quad m = (X' \Sigma_z^{-1} X)^{-1} (X' \Sigma_z^{-1} c - x(s_0)).$$



Обратите внимание, что когда  $s_0 = s_i \in D$ , вычисленное значение  $\hat{Z}_u(s_0)$ , как правило, *и* равно ни одному из значений наблюдений  $z_i(s_i); i \leq n_i$ . Подставляя  $\lambda$ , получаем среднеквадратичную ошибку интерполяции,

$$\begin{aligned} E(Z_u(s_0) - \lambda'z)^2 &= \\ &= C_y(0) + C_\eta(0) + \sigma^2 - \lambda'(c + X\mathbf{m}), \\ &= C_y(0) + v - \lambda'(c + X\mathbf{m}), \end{aligned}$$

в результате, стандартные ошибки интерполяции равны:

$$\sqrt{C_y(0) + v - \lambda'(c + X\mathbf{m})}.$$

Эти ошибки необходимо сравнить со стандартными ошибками интерполяции для версии данных без шумов. Обратите внимание, что когда  $s_0 = s_i$  для одной из точек наблюдений  $s_i \in D$ , ни одна из стандартных ошибок интерполяции не будет равна 0.

## Карты вероятности и квантилей

Если случайные ошибки подчиняются закону нормального распределения и обладают стационарностью (либо второго порядка, либо внутренней), то ошибки вычислений  $\hat{Z}_u(s_0) - Z(s_0)$  тоже подчиняются закону нормального распределения, с нулевым средним значением и дисперсией, равной  $\sigma_s^2(s_0)$ . Нормальность распределения позволяет строить карты вероятностей или квантилей.

## Логарифмически нормальный линейный кригинг

Если вы выбираете логарифмическое преобразование, для ординарного, простого и универсального кригинга вы можете использовать логарифмически нормальный кригинг (или логнормальный кригинг), который реализован в соответствии с описанием, приведенном в работе Кресси (Cressie, 1993). Формулы интерполяции приведены на следующих страницах:

Ординарный кригинг—уравнение 3.2.40, стр. 135 (Cressie, 1993)

Простой кригинг—второе уравнение на стр. 136 (Cressie, 1993)

Универсальный кригинг—Cressie (1993) уравнение 3.2.40 сводится к,

$$\hat{\mu}_Y(\mathbf{Z}; \mathbf{s}_0) = \exp\{\hat{\mu}_Y(\mathbf{Z}; \mathbf{s}_0) + \hat{\sigma}_Y^2(\mathbf{s}_0)/2 - \mathbf{m}_Y'[\mathbf{x}(\mathbf{s}_0)]\}$$

где  $\mathbf{m}$  - вектор из множителей Лагранжа из уравнений универсального кригинга, и  $\mathbf{x}(\mathbf{s}_0)$  - вектор из ковариат в точке  $\mathbf{s}_0$ , для которой интерполируется значение. Дисперсия вычисления дана в работе Кресси (Cressie, 1993) в уравнении 3.2.41, где:

Для ординарного кригинга -  $\mu_Y$  меняется на:

$$\hat{\mu}_Y = \mathbf{1}' \mathbf{Y}_Y^{-1} \mathbf{Y} / (\mathbf{1}' \mathbf{Y}_Y^{-1} \mathbf{1}),$$

Для простого кригинга -  $\mu_Y$  известно,

Для универсального кригинга -  $\mu_Y$  меняется на:

$$\hat{\mu}_Y(\mathbf{s}_0) = [\mathbf{x}(\mathbf{s}_0)]' (\mathbf{X}' \mathbf{Y}_Y^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_Y^{-1} \mathbf{Y}$$

где вектор  $\mathbf{Y} = \log(\mathbf{Z})$ , и предполагается, что каждый элемент вектора  $\mathbf{Y}$  подчиняется нормальному распределению.

## Трансгауссов кригинг

Если выбрано преобразование по методу Вох—Сох или арксинуса, для ординарного, простого и универсального кригинга вы можете использовать трансгауссов кригинг, который реализован в соответствии с описанием, приведенном в работе Кресси (Cressie, 1993, стр. 137).

## Индикаторный кригинг

Индикаторный кригинг - это нелинейный метод, и для индикаторов может быть применена только жесткая форма (т.е., с неотфильтрованной ошибкой измерений) ординарного кригинга. Предположим, что данные относятся к пространственно коррелированному процессу,

$$Z(\mathbf{s}) = \mu + \varepsilon(\mathbf{s}),$$

и бинарная (0 или 1) вероятностная переменная образована с использованием порогового значения,

$$Z^1(\mathbf{s}) = I(Z(\mathbf{s}) > c_1),$$

где  $I(\text{условие})$  - индикаторная функция, которая равна 1, если *условие* верно, и 0, если *условие* неверно. Предположим, что бинарные данные также относятся к пространственно коррелированному процессу (с возможным эффектом саморodka),

$$Z^1(\mathbf{s}) = \mu_1 + \varepsilon^1(\mathbf{s}).$$

Индикаторный кригинг - это ординарный кригинг (с нулевой ошибкой измерений) бинарных переменных  $Z^1(\mathbf{s})$ , и, следовательно, для  $Z(\mathbf{s})$  не предпринимается попыток фильтрации ошибок измерений. Может быть использовано другое пороговое значение,

$$Z^2(\mathbf{s}) = I(Z(\mathbf{s}) > c_2),$$

для модели

$$Z^2(\mathbf{s}) = \mu_2 + \varepsilon^2(\mathbf{s}).$$

Теперь воспользуемся кокригингом для обеих бинарных переменных для вычисления значений  $Z^1(\mathbf{s}_0)$ . Теория и формулы приведены в работах Journel (1983), Isaaks и Srivastava (1989), Cressie (1993, стр. 281), Goovaerts (1997, стр. 293), и Chiles и Delfiner (1999, стр. 381).

## Вероятностный кригинг

Как и индикаторный кригинг, вероятностный кригинг (Sullivan, 1984; Cressie, 1993, стр. 283; Goovaerts, 1997, стр. 301; Chiles и Delfiner, 1999, стр. 385) - это нелинейный метод, и при использовании вероятностного кригинга не существует очевидных способов для фильтрации ошибки измерений. Предположим, что полученные данные - это реализация пространственно коррелированного процесса, плюс независимые случайные ошибки,

$$Z(s) = \mu + \varepsilon(s),$$

и на основании порогового значения создана бинарная (0 или 1) вероятностная переменная,

$$Z^1(s) = I(Z(s) > c),$$

где  $I(\text{условие})$  - индикаторная функция, которая равна 1, если *условие* верно, и 0, если *условие* неверно. Предположим, что бинарные данные также относятся к пространственно коррелированному процессу (с возможным эффектом саморodka),

$$Z^1(s) = \mu_1 + \varepsilon^1(s).$$

Затем воспользуемся кокригингом для вычисления  $Z^1(s_0)$  с использованием  $\{Z^1(s)\}$  в качестве первой переменной и исходных значений  $\{Z(s)\}$  в качестве второй переменной в уравнениях кокригинга. Обратитесь за дополнительной информацией к разделу, посвященному кокригингу.

## Дизъюнктивный кригинг

Для дизъюнктивного кригинга (Matheron, 1976), интерполятор выглядит следующим образом:

$$\hat{Z}(s_0) = \sum_{s \in D} g_s(Z(s)),$$

где  $g_s(Z(s))$  - некая функция переменной  $Z(s)$ . Модуль Geostatistical Analyst использует следующий интерполятор,

$$\hat{Z}(s_0) = f_0 + \sum_{k>0} f_k \hat{H}_k(Y(s_0)),$$

где

$$\hat{H}_k(Y(s_0)) = \sum_{i=1}^n \lambda_{ki} H_k(Y(s_i)),$$

$f_i$  и  $\lambda_{ki}$  - коэффициенты,  $H_k(Y(s_i))$  - многочлены Хермита (Hermite), а  $Y(s_i)$  и  $Y(s_j)$  подчиняются двумерному нормальному распределению. Переменная  $Y(s)$  может быть преобразована (т.е., дизъюнктивный кригинг может быть логнормальным и трансгауссовым) и позволяет пользователю изучить предположение о двумерной нормальности. Теория и практика дизъюнктивного кригинга довольно сложны; модуль Geostatistical Analyst следует методике, приведенной в работе Ривуарапа (Rivoirard, 1994).

## Кокригинг

Кокригинг может быть использован при наличии нескольких переменных. Точные формулы кокригинга приведены в работах: Journel и Huijbregts (1978, стр. 324), Isaaks и Srivastava (1989, стр. 400), Cressie (1991, стр. 138), Goovaerts (1997, стр. 224), и Chiles и Delfiner (1999, стр. 298). Ординарный кокригинг, простой кокригинг и универсальный кокригинг допускают использование моделей, учитывающих ошибку измерений, как и соответствующие им модели кригинга. Так же, как и индикаторный кокригинг, вероятностный кокригинг и дизъюнктивный кокригинг могут выполнять только жесткую интерполяцию (т.е., ошибка измерений не фильтруется).

### Ординарный, простой и универсальный кокригинг

Модель универсального кокригинга является наиболее общей, поэтому предположим, что

$$Z_j^k(\mathbf{s}) = [\mathbf{x}_k(\mathbf{s})]' \boldsymbol{\beta}_k + Y^k(\mathbf{s}) + \eta^k(\mathbf{s}) + \delta_j^k(\mathbf{s}),$$

где  $\mathbf{X}_k$  - предполагаемая матрица, и  $\boldsymbol{\beta}_k$  - вектор параметров для переменной  $k$ -ого типа, со следующими допущениями:

- $Y^k(\mathbf{s})$  - сглаженный второго порядка стационарный процесс, радиус влияния вариограммы которого больше, чем наименьшее расстояние между опорными точками.
- $E(Y^k(\mathbf{s})) = 0$ .
- $\text{Cov}(Y^k(\mathbf{s}), Y^m(\mathbf{s}+\mathbf{h})) = C_y^{km}(\mathbf{h})$ , при этом  $C_y^{km}(\infty) = 0$  (т.е., отсутствует дополнительный эффект самородка в процессе  $Y^k(\mathbf{s})$ ).
- $\eta^k(\mathbf{s})$  - сглаженный второго порядка стационарный процесс, радиус влияния вариограммы которого так близок к нулю 0, что он меньше всех фактических расстояний между опорными точками и точками, для которых интерполируются значения.
- $E(\eta^k(\mathbf{s})) = 0$ .
- $\text{Cov}(\eta^k(\mathbf{s}), \eta^m(\mathbf{s}+\mathbf{h})) = C_\eta^{km}(\mathbf{h})$ , когда  $k$  равно  $m$ , при этом  $C_\eta^{km}(\mathbf{h}) = 0$ .
- $\text{Cov}(\eta^k(\mathbf{s}), \eta^m(\mathbf{s}+\mathbf{h})) = 0$ , когда  $k$  не равно  $m$ .

- $\delta_j^k(\mathbf{s})$  - шумы, образуемые ошибками измерений.
- $E(\delta_j^k(\mathbf{s})) = 0$ , для всех  $k$  и  $j$ .
- $\text{Cov}(\delta_j^k(\mathbf{s}), \delta_j^k(\mathbf{s}+\mathbf{h})) = \sigma_k^2$  если  $\mathbf{h} = \mathbf{0}$ ; в противном случае, она равна 0.
- $\text{Cov}(\delta_j^k(\mathbf{s}), \delta_i^k(\mathbf{s}+\mathbf{h})) = 0$  для  $i$ , не равных  $j$ .

$Y^k(\cdot)$ ,  $\eta^l(\cdot)$ , and  $\delta^n(\cdot)$  независимы друг от друга для всех  $k, l$ , и  $m$ . Предположим, что в данном случае эффект самородка  $v_i$  состоит из двух частей: вариации на микроуровне и ошибки измерений; то есть,  $v_k = C_\eta^{kk}(\mathbf{0}) + \sigma_k^2$ . Также обратите внимание, что для  $\eta^k(\cdot)$  и  $\eta^m(\cdot)$  нет общей информации, поэтому их взаимную ковариацию можно установить, равной 0. На основании этой модели вы можете прийти к выводу, что

$$\text{Cov}(Z_j^k(\mathbf{s}), Z_t^m(\mathbf{s}+\mathbf{h})) = \begin{cases} C_y^{km}(\mathbf{h}) & \text{if } k \neq m \\ C_y^{kk}(\mathbf{h}) + C_\eta^{kk}(\mathbf{h}) & \text{if } k = m \text{ and } \mathbf{h} \neq \mathbf{0} \\ C_y^{kk}(\mathbf{0}) + C_\eta^{kk}(\mathbf{0}) & \text{if } k = m \text{ and } \mathbf{h} = \mathbf{0} \text{ and } j \neq t \\ C_y^{kk}(\mathbf{0}) + C_\eta^{kk}(\mathbf{0}) + \sigma_k^2 & \text{if } k = m \text{ and } \mathbf{h} = \mathbf{0} \text{ and } j = t \end{cases}$$

Для упрощения, рассмотрим только два типа переменных; выводы легко обобщить для большего количества типов переменных. Вычислим отфильтрованную (без шумов) величину  $S^1(\mathbf{s}_0) \equiv [\mathbf{x}_1(\mathbf{s}_0)]' \boldsymbol{\beta}_1 + Y^1(\mathbf{s}_0) + \eta^1(\mathbf{s}_0)$  в точке  $\mathbf{s}_0$ . Кокригинг в модуле Geostatistical Analyst получается для линейного интерполатора,

$$\hat{S}^1(\mathbf{s}_0) = \mathbf{L}_1' \mathbf{z}_1 + \mathbf{L}_2' \mathbf{z}_2$$

затем минимизируем,

$$E(Z(\mathbf{s}_0) - [\mathbf{L}_1' \mathbf{z}_1 + \mathbf{L}_2' \mathbf{z}_2])^2.$$

Выполнив те же операции, что и для ординарного кригинга, получаем следующие уравнения кокригинга:

$$\begin{pmatrix} \mathbf{Y}_z & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{l} \\ \mathbf{m} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{x}_1(\mathbf{s}_0) \end{pmatrix},$$

где

$$\Sigma_z = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}, X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}, c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix},$$

$m_1$  и  $m_2$  - множители Лагранжа,  $\Sigma_{km}$  - ковариационная матрица для данных  $z_k$  и  $z_m$ , а  $c_k$  равно  $\text{Cov}(z_k, S_1(s_0))$ . Решив уравнение для  $\lambda$ , получим,

$$\lambda = \Sigma_z^{-1} (c - X m), \text{ где } m = (X' \Sigma_z^{-1} X)^{-1} (X' \Sigma_z^{-1} c - x_1(s_0)).$$

Подставив  $\lambda$ , получим среднеквадратичную ошибку интерполяции,

$$\begin{aligned} E(S^1(s_0) - \lambda' z)^2 \\ = C_y^{11}(0) + C_\eta^{11}(0) - \lambda' (c + X m), \\ = C_y^{11}(0) + (1 - \pi_1) v_1 - \lambda' (c + X m), \end{aligned}$$

следовательно, стандартные ошибки интерполяции равны:

$$\sigma_{s1}(s_0) = \sqrt{C_y^{11}(0) + (1 - \pi_1) v_1 - \lambda' (c + X m)}.$$

Если

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

тогда ординарный кокригинг сводится к особому случаю универсального кокригинга. Для простого кригинга,

$$\hat{f}_1(s_0) = \lambda' z + k = c' \Sigma_z^{-1} (z - \mu) + \mu_1(s_0),$$

а стандартные ошибки интерполяции вычисляются по формуле,

$$\sigma_{s1}(s_0) = \sqrt{C_y^{11}(0) + (1 - \pi_1) v_1 - c' Y_z c}.$$

Интерполяция новых значений для перекрестной проверки может быть выполнена так же, как и для ординарного, простого и универсального кригинга.

## Индикаторный, вероятностный и дизъюнктивный кокригинг

Индикаторный, вероятностный и дизъюнктивный кокригинг относятся к нелинейным методам, и только жесткая форма (т.е., ошибка измерений не отфильтровывается) ординарного кокригинга может быть использована для этих методов. Индикаторный кокригинг - это просто кокригинг для индикаторов; обратитесь к работам: Cressie (1993, стр. 283), Goovaerts (1997, стр. 297), и Chiles и Delfiner (1999, стр. 386). Вероятностный кокригинг, помимо использования исходных данных, формирует индикаторы для двух типов переменных, а затем пользуется уравнениями ординарного кокригинга. Дизъюнктивный кокригинг - это обобщение метода дизъюнктивного кригинга для двумерных Гауссовых распределений (Muge и Cabecadas, 1989; Chiles и Delfiner, 1999, стр. 419).

## Диалог Перекрестная проверка

### Перекрестная проверка

Перекрестная проверка состоит в последовательном удалении из общей совокупности значения одной опорной точки, а затем интерполяции значения этой точки с использованием оставшихся данных. Затем, проинтерполированное значение может быть сопоставлено с фактическим (значением наблюдения) для оценки того, насколько хорошо работает модель интерполяции. Обратите внимание, что в геостатистике модели вариограммы, как правило, не оцениваются заново каждый раз, когда удаляется значение одной опорной точки. Более подробную информацию о выполнении перекрестной проверки вы можете получить в работах Isaaks и Srivastava (1989, стр. 351), Cressie (1993, стр. 101), Goovaerts (1997, стр. 105), Armstrong (1998, стр. 115), Chiles и Delfiner (1999, стр. 111), и Stein (1999, стр. 215).

### Суммарная статистика перекрестной проверки

Суммарная статистика и графики могут быть получены при сравнении проинтерполированного значения с фактическим значением данных в ходе выполнения перекрестной проверки. Пусть  $\hat{z}(s_i)$  - проинтерполированное значение, полученное в результате перекрестной проверки,  $z(s_i)$  - фактическое значение, полученное в результате наблюдений, и  $\sigma(s_i)$  - стандартная ошибка интерполяции для точки  $s_i$ . Тогда некоторые из значений суммарной статистики, представленной в модуле Geostatistical Analyst равны:

1. Средние ошибки интерполяции,

$$\frac{\sum_{i=1}^n (\hat{z}(s_i) - z(s_i))}{n}.$$

2. Среднеквадратичные ошибки интерполяции,

$$\sqrt{\frac{\sum_{i=1}^n (\hat{z}(s_i) - z(s_i))^2}{n}}.$$

3. Средняя стандартная ошибка кригинга,

$$\sqrt{\frac{\sum_{i=1}^n \sigma(s_i)^2}{n}}.$$

4. Средние нормированные ошибки интерполяции,

$$\frac{\sum_{i=1}^n (\hat{z}(s_i) - z(s_i)) / \sigma(s_i)}{n}.$$

5. Среднеквадратичные нормированные ошибки интерполяции,

$$\sqrt{\frac{\sum_{i=1}^n [(\hat{z}(s_i) - z(s_i)) / \sigma(s_i)]^2}{n}}.$$





# Приложение

## В ЭТОМ ПРИЛОЖЕНИИ

В этом приложении дан обзор методов, которые можно использовать в модуле Geostatistical Analyst. В таблице на следующей странице даны характеристики каждого метода, кратко сформулированы его достоинства и недостатки, и приведены типы создаваемых с его помощью поверхностей. Сравнив различия между этими методами, вы можете определить, какой из них вам лучше использовать для ваших приложений.

# В

# Сравнение методов, используемых в модуле Geostatistical Analyst

Метод	Детерминистский/ Стохастический	Типы результи- рующих поверхностей	Время вычислений/ Быстрый	Жесткий интерpolator время моделирования <sup>1</sup>	Преимущества	Недостатки	Ограничения <sup>2</sup>
Метод взвешенных расстояний	Детерминист- ский	Проинтерпо- лированных значений	Быстрый/ Быстрый	Да	Решения для меньшего числа параметров	Нет оценки ошибок интерполяции; приводит к образованию глаз буйвола ("bulls eyes") вокруг опорных точек	Нет
Глобального полинома	Детерминист- ский	Проинтерпо- лированных значений	Быстрый/ Быстрый	Нет	Решения для меньшего числа параметров	Нет оценки ошибок интерполяции; поверхность может быть слишком сглаженной; краевые точки имеют большое влияние	Нет
Локальных полиномов	Детерминист- ский	Проинтерпо- лированных значений	Умеренно быстрый/ средний	Нет	Решения для большого числа параметров	Нет оценки ошибок интерполяции; может быть слишком автоматическим	Нет
Радиальные базисные функции	Детерминист- ский	Проинтерпо- лированных значений	Умеренно быстрый/ средний	Да	Гибкий, возможен автоматический выбор некоторых параметров	Нет оценки ошибок интерполяции; может быть слишком автоматическим	Нет
Кригинг	Стохастич- еский	Проинтерпо- лированных значений; стандартных ошибок интерполяции; вероятности; квантилей	Умеренно быстрый/ относительно медленный	Да - без ошибки измерений; Нет - с ошибкой измерений	Очень гибкий; позволяет оценить пространственную автокорреляцию; можно вычислить стандартные ошибки интерполяции; много решений по параметрам	Требуется принятия многих решений по методам преобразований, трендам, моделям, параметрам и областям соседства	Данные должны относиться к стационарному стохастическому процессу; некоторые методы требуют, чтобы данные подчинялись закону нормального распределения
Кокригинг	Стохастич- еский	Проинтерпо- лированных значений; стандартных ошибок интерполяции; вероятности; квантилей	Средний/ самый медленный	Да - без ошибки измерений; Нет - с ошибкой измерений	Очень гибкий; может использовать информацию нескольких; наборов данных; по- зволяет оценить взаи- мную пространственную корреляцию; много решений по параметрам	Требуется принятия многих решений по методам преобразований, трендам, моделям, параметрам и областям соседства	Данные должны относиться к стационарному стохастическому процессу; некоторые методы требуют, чтобы данные подчинялись закону нормального распределения

1. Время вычислений - время, которое компьютер затрачивает на построение поверхности. Время моделирования включает время, необходимое пользователю для принятия решений по выбору параметров модели и определению области поиска соседства.

2. Мы предполагаем, что все методы интерполируют сглаженную поверхность с использованием данных, содержащих шумы.

## Описание

**IDW (Метод взвешенных расстояний):** IDW - быстрый детерминистский метод, который выполняет жесткую интерполяцию. Для этого метода необходимо определить небольшое количество параметров модели. Он может быть полезен для предварительного взгляда на интерполируемую поверхность. Однако, при использовании этого метода невозможно оценить ошибки интерполяции. Кроме того, метод взвешенных расстояний может привести к образованию “глаз буйвола” (“bulls eyes”) вокруг опорных точек. Для данных нет никаких ограничений.

**Метод глобального полинома:** Метод глобального полинома - быстрый детерминистский метод, который выполняет сглаженную (не жесткую) интерполяцию. Для этого метода необходимо определить небольшое количество параметров модели. Лучше всего он подходит для моделирования поверхностей, меняющихся медленно и постепенно. Однако, при использовании этого метода невозможно оценить ошибки интерполяции. Поверхность может получиться слишком сглаженной. Опорные точки, расположенные по краям области интереса, могут оказывать большое влияние на результирующую поверхность. Для данных нет никаких ограничений.

**Метод локальных полиномов:** Метод локальных полиномов - относительно быстрый детерминистский метод, который выполняет сглаженную (не жесткую) интерполяцию. Он более гибкий, чем метод глобального полинома, но требует принятия решений для большего количества параметров. При использовании этого метода невозможно оценить ошибки интерполяции. Метод позволяет строить поверхности проинтерполированных значений, сопоставимые с поверхностями, получаемыми при использовании кригинга с учетом ошибок измерений. Методы локальных полиномов не позволяют вам изучить автокорреляцию данных, что делает их менее гибкими и более автоматическими, чем методы кригинга. Для данных нет никаких ограничений.

**Радиальные базисные функции:** Использование радиальных базисных функций - это относительно быстрый детерминистский метод, который выполняет жесткую интерполяцию. Метод радиальных базисных функций более гибкий, чем метод взвешенных расстояний, но требует принятия решений для большего числа параметров. При использовании этого метода невозможно оценить ошибки интерполяции. Метод позволяет строить поверхности проинтерполированных значений, сопоставимые с поверхностями, получаемыми при использовании жесткой формы кригинга. Радиальные базисные функции не позволяют вам изучить автокорреляцию данных, что делает их менее гибкими и более автоматическими, чем методы кригинга. Радиальные базисные функции не требуют введения ограничений на используемые данные.

**Кригинг:** Кригинг - это относительно быстрый метод интерполяции, который может быть жестким, если данные не содержат ошибки измерений, или сглаженным, если данные содержат ошибку измерений. Он очень гибкий и допускает изучение пространственной автокорреляции данных. Поскольку кригинг использует статистические модели, он позволяет строить различные карты, включая карты проинтерполированных значений, карты стандартных ошибок интерполяции, карты вероятностей и карты квантилей. Гибкость кригинга может потребовать принятия большего, относительно других методов интерполяции, количества решений, однако, вы можете воспользоваться и параметрами, предлагаемыми по умолчанию. Кригинг предполагает, что данные относятся к стационарному стохастическому процессу, и некоторые методы требуют, чтобы данные подчинялись закону нормального распределения.

**Кокригинг:** Кокригинг - это относительно быстрый метод интерполяции, который может быть жестким, если данные не содержат ошибки измерений, или сглаженным, если данные содержат ошибку измерений. Он может использовать информацию из нескольких наборов данных. Кокригинг - это очень гиб-

кий метод, который позволяет вам изучить пространственную автокорреляцию и взаимную корреляцию данных. Поскольку кокригинг использует статистические модели, он позволяет строить различные карты, включая карты проинтерполированных значений, карты стандартных ошибок интерполяции, карты вероятностей и карты квантилей. Гибкость кокригинга требует принятия наибольшего из всех других методов количества решений по параметрам, однако, вы можете воспользоваться параметрами, предлагаемыми по умолчанию. Кокригинг предполагает, что данные относятся к стационарному стохастическому процессу, и некоторые методы требуют, чтобы данные подчинялись закону нормального распределения.