

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

ГЛАВНОЕ УПРАВЛЕНИЕ ОБРАЗОВАНИЯ,
НАУКИ И КАДРОВОЙ ПОЛИТИКИ

Учреждение образования
«БЕЛОРУССКАЯ ГОСУДАРСТВЕННАЯ
ОРДЕНОВ ОКТЯБРЬСКОЙ РЕВОЛЮЦИИ
И ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
СЕЛЬСКОХОЗЯЙСТВЕННАЯ АКАДЕМИЯ»

Т. Н. Мыслыва, О. А. Куцаева

МЕТОДОЛОГИЯ И СОВРЕМЕННЫЕ ПРОБЛЕМЫ ГЕОМАТИКИ

*Допущено Министерством образования Республики Беларусь
в качестве учебного пособия для студентов учреждений
высшего образования по специальности магистратуры
«Землеустройство, кадастры, геодезия и геомастика»*

Горки
БГСХА
2022

УДК 91:519.8
ББК 26.8в6
М95

*Рекомендовано методической комиссией
землеустроительного факультета 30.03.2021 (протокол № 7)
и Научно-методическим советом БГСХА 28.04.2021 (протокол № 8)*

Авторы:

доктор сельскохозяйственных наук, доцент *Т. Н. Мыслыва*;
старший преподаватель *О. А. Куцаева*

Рецензенты:

кандидат географических наук, доцент *Д. М. Курлович*;
кандидат технических наук, доцент *О. В. Кравченко*

Мыслыва, Т. Н.
М95 Методология и современные проблемы геоматики : учебное
пособие / Т. Н. Мыслыва, О. А. Куцаева. – Горки : БГСХА,
2022. – 269 с.
ISBN 978-985-882-200-2.

Учебное пособие представляет собой последовательное изложение теоретического материала в рамках изучения учебной дисциплины «Методология и современные проблемы геоматики» с примерами его практической реализации. Теоретический материал сопровождается широким набором иллюстраций и тестовыми заданиями для контроля знаний.

Для студентов учреждений высшего образования по специальности магистратуры «Землеустройство, кадастры, геодезия и геоматика».

**УДК 91:519.8
ББК 26.8в6**

ISBN 978-985-882-200-2

© УО «Белорусская государственная
сельскохозяйственная академия», 2022

ВВЕДЕНИЕ

Представляемое учебное пособие содержит систематизированные научно-технические сведения для изучения дисциплины «Методология и современные проблемы геоматики». Цель учебного пособия – получение специальных знаний в области геоматики, современных геоинформационных технологий, методов создания и использования географических информационных систем (ГИС), а также выработка методических навыков выполнения научных исследований на основе полученных знаний. В семи главах данного издания дано компактное изложение фундаментальных понятий геоматики, рассмотрены вопросы методики выполнения геопространственного анализа и работы с геопространственными данными, приведена характеристика геостатистики как метода анализа геопространственных данных и примеры прикладного использования ее функциональных возможностей. Отдельные главы посвящены вопросам геопространственного, в том числе трехмерного моделирования, и анимации геопространственных данных. Каждая из семи глав включает краткую аннотацию ее содержания, перечень основной и дополнительной литературы, перечень тем для самостоятельного изучения и список литературных и информационных источников к ним, а также тестовые задания с ответами для проверки знаний по рассматриваемой теме. Приведенный в учебном пособии материал сопровождается широким набором практических примеров и иллюстраций. Особое внимание в данном издании уделено вопросам моделей пространственных данных, созданию баз геоданных, освоению технологий изучения пространственных взаимосвязей объектов и явлений на основе векторных и растровых моделей данных, кластерному геопространственному анализу.

В результате изучения материала, представленного в учебном пособии, студент должен закрепить и развить универсальную профессиональную компетенцию УК-2: владеть методами и средствами интеграции современных информационных технологий сбора, обработки, использования и анализа пространственных данных. Настоящее издание разработано для студентов учреждений высшего образования, обучающихся по специальности 1-56 80 01 Землеустройство, кадастры, геодезия и геоматика, а также ориентировано на студентов различной специализации (физическая и социально-экономическая география, экология и природопользование; картография и геоинформатика).

1. ВВЕДЕНИЕ В ГЕОМАТИКУ

1.1. Геомастика как научная дисциплина, технология и сфера производственной деятельности. Цель и задачи изучения дисциплины.

1.2. Место геомастики в системе наук. Понятие и функции геоинформационных систем и технологий.

1.3. Возможности применения геоинформационных систем и технологий.

Количество часов, отведенных на изучение главы, – 1.

Аннотация темы.

Тематика лекции посвящена характеристике геомастики как научной дисциплины и сферы производственной деятельности. В пункте 1.1 дано определение понятия «геомастика», охарактеризованы области ее применения, а также приведены цели и задачи изучения учебной дисциплины «Методология и современные проблемы геомастики». В пункте 1.2 очерчено место геомастики в системе наук, охарактеризованы ее области исследования и основные компоненты, а также раскрыто определение понятия «геоинформационная система» и дана краткая характеристика современных проприетарных и оупенсорсных ГИС. В пункте 1.3. приведено описание и примеры применения функциональных возможностей ГИС в сфере землеустройства и кадастра.

Литературные и информационные источники

1. Ананьев, Ю. С. Геоинформационные системы / Ю. С. Ананьев. – Томск: Изд-во ТПУ, 2003. – 70 с.
2. Геостатистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.
3. Курлович, Д. М. Геоинформационные технологии. Лабораторный практикум / Д. М. Курлович, Н. В. Жуковская, О. М. Ковалевская. – Минск: БГУ, 2015. – 160 с.
4. Митчелл, Э. Руководство по ГИС-анализу. / Э. Митчелл. – ESRI: 2000. – Ч. 1: Пространственные модели и взаимосвязи. – 170 с.
5. ТКП 610-2017 (33520) «Земельно-информационная система Республики Беларусь. Порядок создания и ведения (эксплуатации, обновления)» – Введ. с 01.09.2017. – Минск, Госкомимущество Республики Беларусь, 2017. – 50 с.

Internet-ресурсы

6. Домашняя страница проекта Quantum GIS [Электронный ресурс]. – Режим доступа: <https://www.qgis.org/en/site/>.

7. Сайт Государственного комитета по имуществу Республики Беларусь [Электронный ресурс]. – Режим доступа: <http://www.gki.gov.by/ru/>

8. Сайт ГП «Белгеодезия» [Электронный ресурс]. – Режим доступа: <http://www.geo.by/ru>.

9. Сайт ГП «БелПСХАГИ» [Электронный ресурс]. – Режим доступа: <https://www.belaerogis.by>.

10. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.

11. Сайт РУП «Белкартография» [Электронный ресурс]. – Режим доступа: <http://belkarta.by>.

12. Учебник Quantum GIS [Электронный ресурс]. – Режим доступа: https://wiki.gis-lab.info/w/%D0%A3%D1%87%D0%B5%D0%B1%D0%BD%D0%B8%D0%BA_Quantum_GIS.

1.1. Геоматика как научная дисциплина, технология и сфера производственной деятельности. Цель и задачи изучения дисциплины

Технический прогресс существенно увеличил и расширил возможности сбора, обработки и анализа данных, в частности, пространственных и моделирования геотехнических и геоэкологических систем, в том числе и в масштабе всей планеты. Данный факт привел к экспоненциальному росту объемов накопления данных о Земле и к острой необходимости развития новых методов анализа, синтеза и обработки таких данных. Как следствие признания научным сообществом растущей потребности в пространственной информации для науки и практики, а также для реализации положений системного подхода к изучению процессов и явлений в объектах окружающего мира возникла **геоматика** – синтетическая дисциплина, связанная со сбором, распространением, хранением, анализом, обработкой, представлением географической (геопространственной) информации и состоящая из продуктов, услуг и инструментов, участвующих в сборе, интеграции и управлении географическими данными¹.

Термин «геоматика» впервые был использован в научной статье в 1975 г. французским геодезистом и фотограмметристом Бернаром Дюбуиссоном², чтобы отразить недавние изменения в работе геодезистов и фотограмметристов, связанные с развитием компьютеризации и геоинформационных систем. Этот термин впервые был введен в науч-

¹Географические данные – данные о пространственных объектах и/или их наборах.

²Dubuisson, Bernard. (1975). *Pratique de la Photogrammetrie et des Moyens Cartographiques derives des Ordinateurs*. (K. J. Dennison, Trans.). Paris: EditionsEyrolles.

ное использование на французском языке («géomatique») и в 1977 г. зарегистрирован Международным комитетом французского языка как неологизм.

На английском языке данный термин был популяризирован франко-канадским геодезистом Мишелем Паради в 1981 г. Университет Лавала (провинция Квебек, Канада) ввел этот термин в академическое использование в 1986 г. с введением программы бакалавриата по геомапике. Из Квебека он распространился на Университет Нью-Брансуика, а затем и на всю Канаду. Именно двуязычный характер Канады, скорее всего, стал важным фактором в принятии и распространении геомапики именно в этой стране.

Термин «геомапика» до сих пор не стандартизирован, поэтому в одних странах считается синонимом термина «геоинформатика», в других – тождествен понятию «наука о пространстве». Ряд стран оперируют понятием «инженерная геомапика³», являющимся тождественным понятию «геодезия», однако в более широком смысле подразумеваемая объединение геодезии и информатика.

Начиная с объединения геодезии и информатика, включая другие науки о Земле, где доминирующим являлось направление геодезических работ, в дальнейшем геомапика интегрировала науки о Земле на основе САПР⁴ и информатика. Сегодня преобладающим в ней является выявление пространственных отношений, получение геоанных и построение пространственных моделей для использования в разных сферах человеческой деятельности.

Области применения геомапики включают:

- аэромагнитные исследования;
- аэрогеофизику;
- аэронавигационное обслуживание;
- археологические исследования;
- исследования для приложений ГИС;
- управление прибрежными территориями;
- геоинформационное картографирование;

³Инженерная геомапика – работы в области кадастра, землепользования, инженерных изысканий, строительства, муниципального управления.

⁴Система автоматизированного проектирования (САПР) – автоматизированная система, реализующая информационную технологию выполнения функций проектирования, представляющая собой организационно-техническую систему, предназначенную для автоматизации процесса проектирования, состоящую из персонала и комплекса технических, программных и других средств автоматизации его деятельности.

- мониторинг и прогнозирование последствий природных бедствий и катастроф;
- управление инфраструктурой;
- землепользование и земельный кадастр;
- мониторинг и развитие природных ресурсов;
- городское и региональное планирование;
- океанографию;
- метеорологию и гидрометеорологию;
- мониторинг окружающей среды.

Современный **инженер-геоматик** – это специалист по сбору, обработке, анализу, визуализации и управлению пространственно связанной информацией для решения широкого круга технических и социальных задач.

Задачи изучения учебной дисциплины составляют:

- сбор, обработку и моделирование геоданных, а также их использование в географических информационных системах;
- связывание топографических геоданных с тематическими;
- обработку данных дистанционного зондирования и интеграцию их в ГИС;
- реализацию геоданных в мультимедийных продуктах;
- развитие многомерных геоинформационных продуктов;
- проектирование и создание геоинформационных продуктов.

1.2. Место геоматики в системе наук.

Понятие и функции геоинформационных систем и технологий

Одной из **особенностей развития геоматики** является то, что терминологически она тесно связана с науками о Земле, в то время как в аспекте приложения значительно шире используется в других областях.

Геоматика включает **инструменты и методы**, используемые в топографической съемке, дистанционном зондировании, картографии, географических информационных системах (ГИС), глобальных навигационных спутниковых системах (GPS, ГЛОНАСС, Galileo, Beidou), фотограмметрии, геофизике, географии и связанных формах картографирования земли. **Направления исследования геоматики** относят к следующим областям (рис. 1.1).

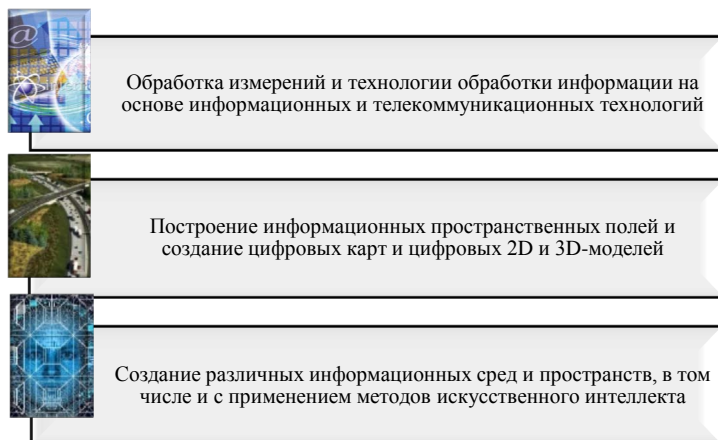


Рис. 1.1. Области исследования геоматики как науки

Геоматика имеет феноменологический аспект научного развития и является синтетической дисциплиной, а ее исследования строятся на основе интеграции знаний целого ряда других дисциплин и разработке теоретических основ на основе релевантных⁵ фундаментальных наук. Она синтезирует знания из следующих дисциплин: геодезия, геодинамика, гидрография, картография и цифровое картографирование, географические информационные системы (ГИС), дистанционное зондирование, фотограмметрия, землеустройство, земельный кадастр и кадастр недвижимости, геоинформатика (рис. 1.2).

Первые две компоненты определяют теорию и методологию геоматики, остальные три – ее прикладную составляющую.

Первую компоненту геоматики составляют основы информатики, математики и кибернетики, которые формируют теорию и методы структуризации, кодирования и передачи информации. Данная компонента поставлена первой потому, что в условиях усиливающейся информатизации общества и возрастания объемов информации, необходимой при анализе и управлении, геоматика способствует уменьшению информационной нагрузки на лицо, принимающее решение.

⁵Релевантность – степень соответствия найденного документа или набора документов информационным нуждам пользователя.

Вторая компонента геоматики включает теорию предметных областей лингвистики и географии, которые раскрывают принципы моделирования на основе абстракции, методов познания и интерпретации. Эту компоненту можно рассматривать как методологическую основу геоматики.



Рис. 1.2. Структура геоматики как синтетической дисциплины

Третьей компонентой геоматики считают теорию, технологии и методы отображения пространства, заимствованные из геодезии, теории измерений, фотограмметрии, дистанционного зондирования, картографии. Эта компонента отвечает в первую очередь за сбор, во вторую – за обработку информации.

К *четвертой компоненте* геоматики относят знания из таких областей применения, как городское и региональное управление, геомаркетинг, многочисленные инженерные приложения, например, маркшейдерское дело, которые формируют специфические требования для описания и анализа пространственной информации, обусловленные решением специальных задач. Эту компоненту можно рассматривать как прикладной компонент инженерного значения.

Пятая компонента включает общепринятые нормы применения и обработки информации из области экономики и права, которые опре-

деляют условия применения информационных технологий для решения социальных задач. Эта компонента геоматики связана с коммерческой деятельностью и управлением.

Все возможности геоматики реализуются посредством геоинформационных систем и технологий, являющихся ее основным инструментом.

Геоинформационная система (ГИС) – это информационная система, оперирующая пространственными данными и по пространственному охвату подразделяющаяся на глобальные, субконтинентальные, национальные, межнациональные, региональные, субрегиональные и локальные ГИС. Основные функции ГИС предусматривают следующее (рис. 1.3).

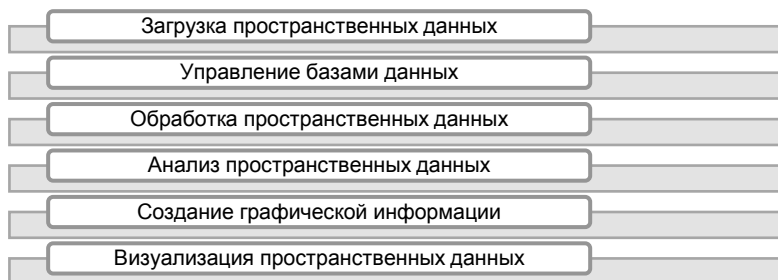


Рис. 1.3. Основные функции геоинформационных систем

Геоинформационные системы функционируют на базе **геоинформационных технологий** – совокупности приемов, способов и методов применения программно-технических средств обработки и передачи информации, позволяющей реализовать функциональные возможности геоинформационных систем. ГИС подразделяются на проприетарные и оупенсорсные. **Проприетарные ГИС** – программное обеспечение, являющееся частной собственностью авторов или правообладателей и распространяющееся по лицензии. Правообладатель проприетарной ГИС сохраняет за собой монополию на ее использование, копирование и модификацию, полностью или в существенных моментах. *К преимуществам* проприетарных ГИС можно отнести такие: постоянная техническая поддержка продукта; более стабильная работа по сравнению с оупенсорсными ГИС; гарантированное отсутствие вредоносных объектов (вирусов); автоматическое обновление; качественное использование всех возможностей оборудования. *Недостатки* проприетарных

ГИС: плата за лицензию; проприетарные протоколы работы устройства; зависимость от разработчика; невозможность изменить исходный код; ограничение на распространение и модификацию. Наиболее широко используемой в геомапике проприетарной ГИС является ArcGIS (разработчик: ESRI, США), представляющая собой полную систему, которая позволяет собирать, организовывать, управлять, анализировать, обмениваться и распределять географическую информацию (рис. 1.4).



Рис. 1.4. Структура и функциональные возможности проприетарной ГИС ArcGIS

Оупенсорные ГИС (открытые, свободные) – программное обеспечение с открытым исходным кодом, пользователи которого имеют права (свободы) на его неограниченную установку, запуск, свободное использование, изучение, распространение и изменение, а также распространение копий и результатов изменения. История свободных ГИС началась с 1978 г., когда службой охраны рыбных ресурсов и диких животных США была создана ГИС MOSS, которая сочетала в себе возможность работы как с растровыми, так и с векторными данными. Начало XXI в. ознаменовалось разработкой таких свободных ГИС, таких как SAGA – System for Automated Geoscientific Analyses (2001 г.), gvSIG (2003 г.), QGIS (2002 г.), ILWIS (2007 г.). Наиболее широко используемой оупенсорной ГИС является Quantum GIS (рис. 1.5).



Рис. 1.5. Структура и функциональные возможности проприетарной ГИС Quantum GIS

Данная ГИС изначально разрабатывалась как простой графический интерфейс для GRASS. Она работает на платформах Windows, Mac OS X и Linux и поддерживает векторные и растровые данные, а также базы геоданных и данные, предоставляемые различными картографическими веб-серверами. Функциональность постоянно увеличивается за счет добавления модулей, создаваемых на языках программирования C++ и Python. К *преимуществам* оупенсорсных ГИС можно отнести отсутствие платы за лицензию; быстрый темп разработки за счет привлечения разработчиков со всего мира; открытость и доступность данных. *Недостатками* оупенсорсных ГИС являются: отсутствие постоянной технической поддержки продукта; наличие ошибок и менее стабильная работа по сравнению с проприетарными ГИС; отсутствие полноценных справочных и обучающих материалов.

1.3. Возможности применения геоинформационных систем и технологий

ГИС являются мощным инструментом, позволяющим решать проблемы и принимать лучшие решения; выполнять успешное планирование и рационально использовать ресурсы; прогнозировать события; более эффективно обмениваться информацией; получать образование и мотивировать других, а также дают возможность *выполнять следующее*: создавать, обмениваться и использовать интеллектуальные карты; компилировать географическую информацию; создавать и управ-

лять базами географических данных; решать различные задачи с помощью пространственного анализа; создавать web-приложения на базе карт; производить связь и обмен информацией посредством визуализации геоданных.

С помощью ГИС-приложений, используемых для создания и редактирования наборов векторных и растровых данных, возможно создавать и редактировать различные цифровые карты, оцифровывать отсканированные с бумажных носителей карты и редактировать их (рис. 1.6).

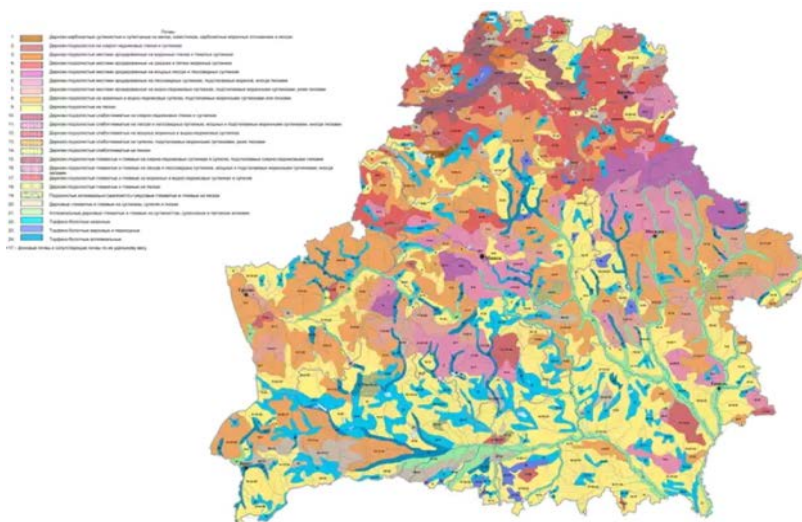


Рис. 1.6. Карта почв Республики Беларусь

С помощью GIS данные карт на бумажных носителях возможно трансформировать в цифровой вид; корректировать по данным ДЗЗ; объединять в единые покрытия; визуализировать в картографическом проекте (рис. 1.7).

Возможности современных геоинформационных систем позволяют осуществлять комплексную оценку, моделирование и прогнозирование состояния территорий и могут с успехом применяться для принятия управляющих решений по охране земельных ресурсов и рациональному их использованию.

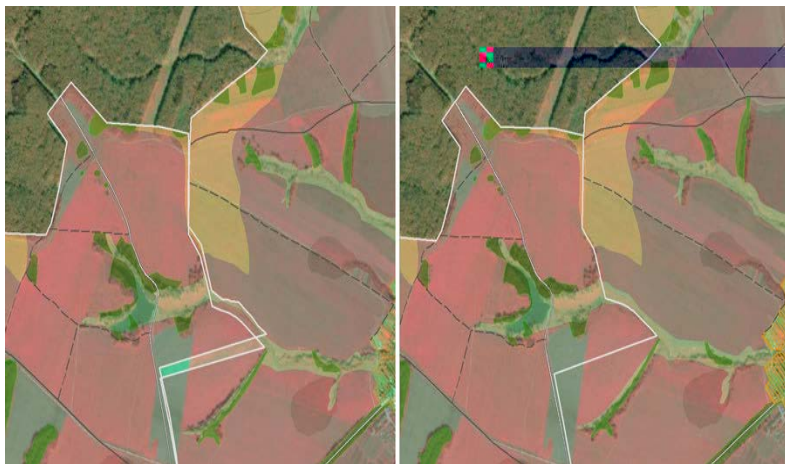


Рис. 1.7. Фрагмент цифровой почвенной карты до (слева) и после корректировки пространственной привязки

Применение ГИС-технологий в оптимизации земельных площадей фермерских хозяйств повышает эффективность получения оперативных и точных данных о земельной собственности и площади. Базовые возможности ГИС позволяют создавать цифровые модели рельефа, выполнять комплексный морфометрический анализ, моделирование и анализ эрозионной сети (рис. 1.8).

ГИС позволяют комбинировать векторные и растровые данные, а также визуализировать результаты определения содержания химических элементов в почве с одновременным созданием динамических баз геопространственных данных о почвенных свойствах и параметрах (рис. 1.9).

Среди основных направлений **использования ГИС в землеустройстве и земельном кадастре** можно выделить следующие:

1. Мониторинг состава, структуры и состояния земельных ресурсов и почвенного покрова земель.

2. Визуализация данных мониторинга, позволяющая обеспечить: одновременное отображение большого числа разнотипных данных; демонстрировать в результатах запроса кластеры, относительные размеры групп, схожесть и различие групп, выпадающие значения; взаимодействие с некоторым выбранным объектом с возможностью просмотра его положения и связей с контекстом; масштабируемость –

способность легко и быстро перемещаться между микро- и макро-представлением данных.

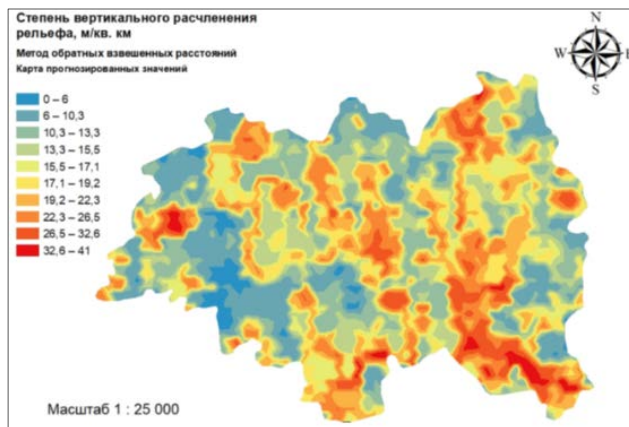


Рис. 1.8. Степень вертикального расчленения рельефа территории Горьковского района

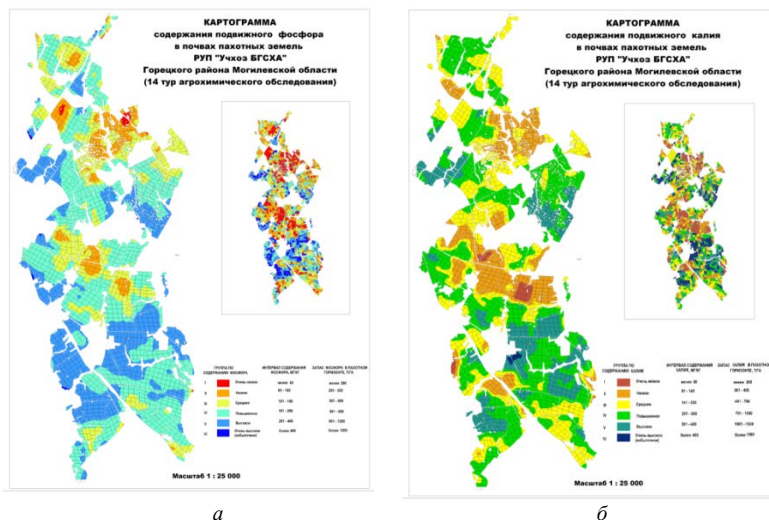


Рис. 1.9. Картограмма содержания подвижного фосфора (а) и калия (б) в почве

3. Прогнозирование и планирование развития территорий на основе оценки ресурсного потенциала земель, организация эффективного земледелия с элементами точного земледелия.

4. Моделирование рационального использования и охраны земельных ресурсов путем автоматизации расчетов количественных показателей земельных ресурсов и их последующей визуализации.

5. Качественная оценка земель, изучение их природно-экологического и экономического потенциала, а также динамики состояния под воздействием хозяйственной деятельности человека.

6. Территориальное планирование, направленное на определение путей наиболее эффективного использования территорий в целях обеспечения устойчивого развития.

7. Информационное обеспечение и ведение земельного кадастра, выявление соответствия кадастровой и рыночной стоимости земель и земельных участков, предоставление информационных услуг населению и коммерческим организациям.

Геоинформационные системы являются действенным инструментом для полной и эффективной реализации функций внутрихозяйственного землеустройства, а их применение для целей оптимизации землепользования, трансформации земельных угодий повышает эффективность получения оперативных и точных данных как о количественном и качественном составе земель, так и об их правовом статусе.

Основным назначением ГИС в землеустройстве является создание цифровых карт и планов местности, баз геопространственных данных о качественных и количественных характеристиках земельных ресурсов, а также мониторинг земель. Среди *преимуществ использования ГИС в землеустройстве*: автоматизация получения данных об объектах, точность географической информации, возможность быстрой корректировки данных и обновление содержания, доступность использования через Internet, возможность пространственного анализа в ГИС, наглядность, создание картограмм, быстрый поиск в базе данных.

Примером эффективно функционирующей ГИС являются Геопортал земельно-информационной системы Республики Беларусь (Геопортал ЗИС) и Публичная кадастровая карта, доступ к которым осуществляется в онлайн режиме. Геопортал ЗИС – это полнофункциональная геоинформационная система, предназначенная для автоматизации хранения, обработки и предоставления пространственной информации для поддержки принятия решений по организации эффективной работы в области землеустройства, геодезии, картографии, зе-

мельного, лесного кадастра и кадастра недвижимости, градостроительства и архитектуры, телекоммуникаций, обслуживания трубопроводов, добычи и транспортировки нефти и газа, электрических сетей, экологии и природопользования, геологии и геофизики, железнодорожного и автомобильного транспорта, банковского дела, образования, государственного управления.

Использование функциональных возможностей ГИС для целей земельного кадастра включает: составление цифровых картографических моделей земельных участков на основе оцифровки исходящих аналоговых материалов и использования результатов натурных геодезических измерений; формирование атрибутов данных: разработку структуры и заполнение атрибутов базы геоданных и создание топологии площадных линейных и точечных объектов; организацию и построение тематических слоев, обеспечивающих возможности пространственного анализа и реализацию типовых запросов пользователей системы.

Вопросы для самостоятельного изучения

1. Применение функциональных возможностей геоинформационных систем и технологий в сельском хозяйстве.
2. Применение функциональных возможностей геоинформационных систем и технологий в сфере мониторинга и охраны окружающей среды.

Тестовые задания для проверки знаний

1. *Вставьте пропущенное словосочетание.* Геоматика – дисциплина, связанная со сбором, распространением, хранением, анализом, обработкой, представлением географической информации и состоящая из продуктов, услуг и инструментов, участвующих в сборе, интеграции и управлении _____.

- 1) информационными потоками;
- 2) статистическими данными;
- 3) данными геодезических изысканий;
- 4) географическими данными;
- 5) облачными данными;
- 6) сетевыми потоками данных.

2. Выберите нужные даты и укажите их правильную последовательность. Термин «геоматика» впервые был использован в научной статье в _____ г., зарегистрирован как неологизм в _____ г. и был популяризирован на английском языке в _____ г.

- 1) 1983;
- 2) 1981;
- 3) 1977;
- 4) 1979;
- 5) 1975;
- 6) 1978;
- 7) 1980.

3. Правильным или нет является утверждение? Одной из особенностей развития геоматики является то, что ее теоретической основой являются науки о Земле, в то время как в прикладном аспекте она довольно широко используется в других областях.

- 1) да;
- 2) нет.

4. Правильным или нет является утверждение? Функциональные возможности геоматики реализуются посредством геоинформационных систем и технологий, являющихся ее основным инструментом.

- 1) да;
- 2) нет.

5. Правильным или нет является утверждение? Геоматика имеет феноменологический аспект научного развития и не является синтетической дисциплиной.

- 1) да;
- 2) нет.

6. Дополните определение. Система, оперирующая пространственными данными и по пространственному охвату подразделяющаяся на глобальные, субконтинентальные, национальные, межнациональные, региональные, субрегиональные и локальные ГИС, называется:

- 1) информационная система;
- 2) геоинформационная система;
- 3) автоматизированная система;
- 4) географическая система;
- 5) интеллектуальная система.

7. Дополните определение. Совокупность приемов, способов и методов применения программно-технических средств обработки и передачи информации, позволяющая реализовать функциональные возможности геоинформационных систем, называется:

- 1) информационная технология;
- 2) автоматизированная технология;
- 3) геоинформационная технология;
- 4) инновационная технология;
- 5) интеллектуальная технология.

8. *Правильным или нет является утверждение?* ГИС позволяют комбинировать векторные и растровые данные, а также визуализировать результаты определения содержания химических элементов в почве с одновременным созданием статических баз геопространственных данных о почвенных свойствах и параметрах.

- 1) да;
- 2) нет.

9. *Правильным или нет является утверждение?* ГИС, являющиеся платными и требующие наличия лицензии на использование, имеющие широчайший спектр функций и возможностей, а также постоянно совершенствующиеся разработчиками, называют оупенсорсными.

- 1) да;
- 2) нет.

10. *Дополните определение.* Полнофункциональная геоинформационная система, предназначенная для автоматизации хранения, обработки и предоставления пространственной информации для поддержки принятия решений по организации эффективной работы в области землеустройства, геодезии, картографии, земельного, лесного кадастра и кадастра недвижимости, градостроительства и архитектуры, телекоммуникаций, обслуживания трубопроводов, добычи и транспортировки нефти и газа, электрических сетей, экологии и природопользования, геологии и геофизики, железнодорожного и автомобильного транспорта, банковского дела, образования, государственного управления – это:

- 1) публичная кадастровая карта;
- 2) база геоданных Национального кадастрового агентства;
- 3) база геоданных Государственного комитета по имуществу Республики Беларусь;
- 3) регистр стоимости земель и земельных участков;
- 4) геопортал земельно-информационной системы.

2. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ ГЕОМАТИКИ

2.1. Основные понятия и термины геоматики. Пространственно-распределенные данные.

2.2. Метрика в пространстве. Пространственное разрешение.

2.3. Сеть мониторинга и кластерность. Декластеризация данных.

2.4. Пространственная непрерывность данных. Стационарность данных.

2.5. Геостатистическое оценивание данных. Проверка качества модели методом кросс-валидации.

Количество часов, отведенных на изучение главы, – 3.

Аннотация темы.

Тематика лекции посвящена базовым понятиям и терминам геоматики, а также смежных областей, в частности геостатистики. В пункте 2.1 даны определения понятия пространственно распределенных данных, а также характеристика моделей данных, используемых в геоматике. В пункте 2.2 сделан экскурс в смежную область, связанную с пространственными данными – метрику пространства, а также определено понятие «пространственное разрешение». В пункте 2.3. приведено описание сети мониторинга и рассмотрены вопросы кластеризации и декластеризации геопространственных данных. Пункт 2.4 посвящен одному из важнейших понятий геоматики – пространственной непрерывности данных, а также различным видам стационарности и связанным с ними предположениями. Содержание пункта 2.5 посвящено вопросам геостатистического оценивания данных и выполнению кросс-валидации для проверки качества моделей.

Литературные и информационные источники

1. Введение в методы анализа данных по окружающей среде / М. Ф. Каневский [и др.] // Элементарное введение в геостатистику. Сер. Проблемы окружающей среды и природных ресурсов. – 1999. – № 11. – С. 1–12.

2. Геостатистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.

3. Геостатистический анализ данных в экологии и природопользовании (с применением пакета R): учеб. пособие / А. А. Савельев [и др.]. – Казань: Казанский университет, 2012. – 120 с.

4. Митчелл, Э. Руководство по ГИС-анализу / Э. Митчелл. – ESRI: 2000. – Ч. 1: Пространственные модели и взаимосвязи. – 170 с.

5. Морисита, М. Ид-индекс, мера рассеивания людей / М. Морисита // Исследования по экологии населения. – 1962. – Вып. 4(1). – С. 1–7.

6. Самсонов, Т. Е. Визуализация и анализ географических данных на языке R [Электронный ресурс] / Т. Е. Самсонов. – М.: Географический факультет МГУ, 2017. – Режим доступа: <https://tsamsonov.github.io/r-geo-course>.

2.1. Основные понятия и термины геоматики. Пространственно-распределенные данные

Различные сферы деятельности оперируют значительными объемами различной пространственно распределенной информации, собранной в базы и банки данных, в частности big data⁶. Задача ее интерпретации, анализа и дальнейшего использования является чрезвычайно важной и требует комплексного системного подхода. Статистическое моделирование пространственных явлений, являющееся одним из важных составляющих геоматики, позволяет обобщить имеющиеся измерения и получить модель их распределения в пространстве. Любые **данные** являют собой информацию, представленную в виде, пригодном для обработки автоматическими средствами при возможном участии человека. Данные о пространственных объектах и (или) их наборах называют **пространственно-распределенными данными** (геоинформационными, геопространственными данными, геоданными). Такие данные, кроме информации о значении переменной, содержат также сведения о ее локализации, как правило координаты места положения, а также информацию о взаимном расположении переменных (топологию). Если кроме информации о местоположении присутствует информация о времени получения данных, то данные называют **пространственно-временными**.

Примеры пространственно-распределенных данных:

– распаханность земель сельскохозяйственного значения в пределах административного района;

⁶Big data (большие данные) – обозначение структурированных и неструктурированных данных огромных объемов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами. В качестве определяющих характеристик для этих данных выделяют «три V»: объем (англ. *volume*, в смысле величины физического объема), скорость (англ. *velocity*, в смысле как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (англ. *variety*, в смысле возможности одновременной обработки различных типов структурированных и полуструктурированных данных).

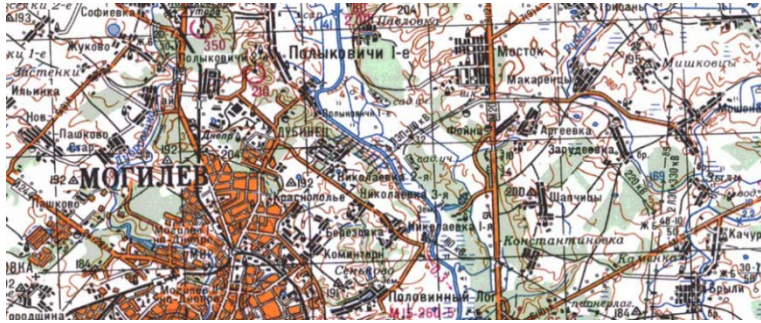
- плотность загрязнения почвы радионуклидами;
- средняя стоимость земельного участка в пределах административно-территориальной единицы.

Примеры пространственно-временных данных:

- плотность загрязнения почвы по турам агрохимических обследований;
- кадастровая стоимость земель сельскохозяйственного назначения по турам кадастровой оценки;
- динамика количества земель крестьянских (фермерских) хозяйств за определенный период.

Пространственно-распределенные данные могут быть **непрерывными** (плотность почвы, электропроводимость почвы, концентрация в почве загрязняющих веществ, содержание в почве элементов питания) или **категориальными** (тип почвы, вид земель, тип оценочной зоны). Для анализа пространственно-распределенных данных разного типа используются различные подходы. Для описания однотипных пространственно-распределенных данных (объектов) и топологических отношений между ними используют **модель пространственных данных**. В геоматике оперируют преимущественно двумя моделями пространственных данных – **растровой** и **векторной**.

Растровая модель пространственных данных – цифровое представление пространственных объектов в виде совокупности ячеек растра (пикселей) с присвоенными им значениями класса объекта. Она предполагает позиционирование объектов с указанием их положения в соответствующей растру прямоугольной матрице единообразно для всех типов пространственных объектов (точек, мультиточек, линий, полилиний, полигонов и поверхностей, аннотаций и объектов-размеров). Иными словами – растровая модель пространственных данных оперирует элементарными местоположениями. Каждой ячейке такой растровой модели соответствует одинаковый по размерам, но разный по характеристикам участок поверхности геопространственного объекта. Данная модель разбивает всю изучаемую территорию (или поверхность) на отдельные элементы регулярной сетки или ячейки растра, каждая из которых содержит только одно атрибутивное значение. Основное назначение растровых моделей геопространственных данных – непрерывное отображение поверхности (рис. 2.1).



a



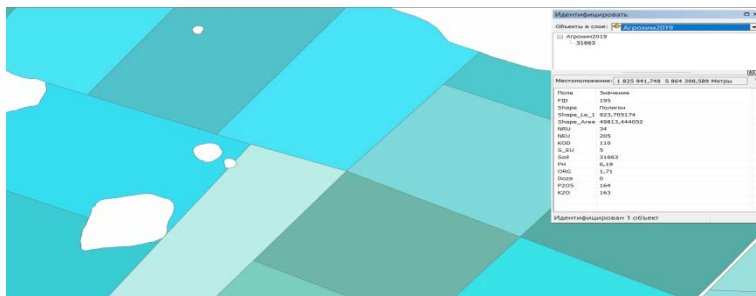
б



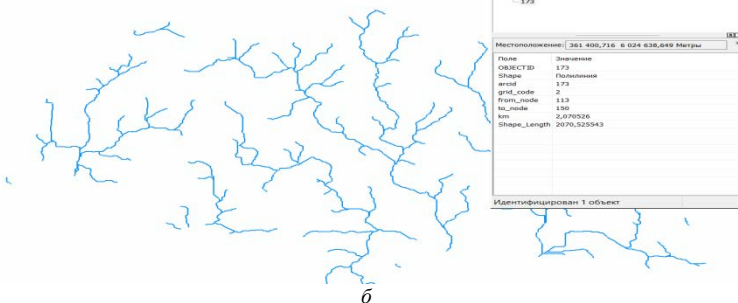
в

Рис. 2.1. Примеры растровых моделей данных: топографическая карта (*a*); спутниковый снимок (*б*); ортофотомозаика (*в*)

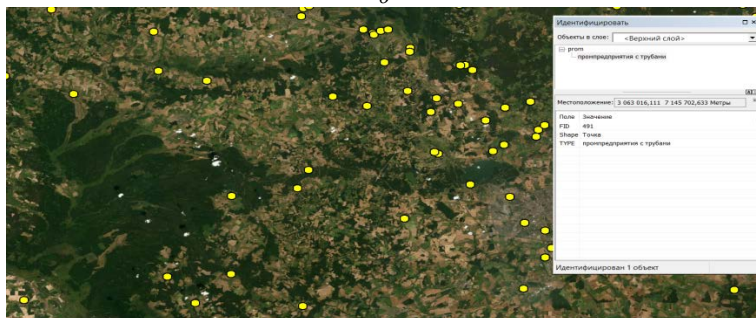
Векторная модель пространственных данных – цифровое представление пространственных объектов в виде множества дискретных объектов с заданными характеристиками. Векторные модели бывают топологическими и нетопологическими (рис. 2.2).



a



б



в

Рис. 2.2. Примеры векторных моделей данных: полигональный слой (а); линейный слой (б); точечный слой (в)

Векторная топологическая модель состоит из описания отдельных объектов, а также из описаний топологии – отношений отдельных объектов между собой, тогда как **векторная нетопологическая модель** содержит только информацию о пространственном местоположении объекта (переменной). В *векторных топологических моделях* все объекты имеют определенное заданное размещение относительно друг друга и не могут произвольно размещаться в пространстве. Векторная топологическая модель данных содержит три основных типа объектов: узлы, дуги и полигоны. Каждый из этих объектов имеет уникальный идентификатор ID, с помощью которого устанавливаются взаимные связи между объектами. Узлы являются обычными точечными объектами, характеризуемыми координатами на плоскости (x, y) , а дуги – линейными объектами – ломаными линиями, соединяющими пару узлов покрытия и проходящими через последовательность промежуточных точек. Полигоны (области) являются площадными объектами. Они характеризуются набором контуров, каждый из которых, в свою очередь, описывается последовательностью дуг покрытия.

В *векторных нетопологических моделях* все объекты являются полностью независимыми друг от друга и могут произвольно размещаться в пространстве, а пространственная информация представляется наборами элементарных графических примитивов – точек, линий, полигонов.

Наиболее широко распространенной векторной нетопологической моделью пространственно-распределенных данных является **шейп-модель**. В шейп-модели допустимы четыре вида данных: точки, линии, полигоны и мультиточки. При этом в пределах одного слоя, представленного в виде шейп-модели, допускаются объекты только одного вида. Соответствующие слои при этом называются точечными, линейными, полигональными (площадными) и мультиточечными. Под слоем в данном случае подразумевают совокупность однотипных пространственных объектов, определенных в одной модели данных на общей территории и в общей системе координат. Данные в шейп-модели могут быть определены на плоскости – тогда они характеризуются двумя координатами (x, y) , а также в 3-мерном (координаты (x, y, z)) или (x, y, m)) и 4-мерном (координаты (x, y, z, m)) пространстве. Координата z , как правило, содержит сведения о высотных характеристиках, а координата m обычно используется в качестве так называемой меры для линий, когда значение меры описывает расстояние от некоторого начала (пикетажное расстояние).

2.2. Метрика в пространстве. Пространственное разрешение

Геоматика оперирует регионализированными данными – измерениями, обладающими координатной привязкой, которая может быть:

– *пространственной*, определяющей географическое положение измерения (географические координаты) в пространстве или его относительное положение по отношению к другим объектам (специальная координатная система для определенной местности);

– *временной*, определяющей время проведения измерения (абсолютное или относительное);

– *пространственно-временной*, т. е. и пространственной, и временной одновременно.

Основное требование к координатной системе – ее метричность, т. е. координаты должны сопровождаться метрикой, возможностью вычислять расстояния между точками. Для двумерного евклидова пространства R^2 метрика такова, что расстояние между точками пространства $X_1 = (x_1; y_1)$ и $X_2 = (x_2; y_2)$ определяется евклидовой нормой:

$$\|(x_1; y_1)(x_2; y_2)\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (2.1)$$

Евклидово расстояние между точками $X_1 = (x_{11}, \dots, x_{1n})$ и $X_2 = (x_{21}, \dots, x_{2n})$ в n -мерном пространстве вычисляется аналогично двумерному случаю:

$$\|(\bar{X}_1)(\bar{X}_2)\| = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}. \quad (2.2)$$

Важной метрической характеристикой является масштаб, различие величин которого связано с измерениями: например, если рассматривать распространение загрязнения в почве, то горизонтальные пространственные области распространяются на десятки километров, в то время как рассматриваемая глубина при анализе почвы не превышает 0,3 м, а геологические слои могут иметь толщину от нескольких сантиметров до сотен метров. При работе с различными масштабами обычно производится нормировка координат – переход к другой системе, где размерности соизмеримы, например, линейное преобразование на отрезок.

Пространственные переменные всегда определяются в конкретной области пространства – в *геометрическом поле*. Пространственную переменную V можно рассматривать как функцию точки пространства x : $Z = Z(x)$. Однако чаще всего интерес представляют не точечные, а средние значения какой-либо величины $Z(x)$ в пределах малой обла-

сти пространства – *геометрической базы*. Например, для такого признака, как содержание органического вещества либо питательных веществ в почве, геометрической базой является объем пробы. В геоматике база должна быть определена весьма точно: необходимо знать ее объем, форму и ориентацию в пространстве, поскольку если изменяется геометрическая база, то возникает новая пространственная переменная, близкая к предыдущей, но не совпадающая с ней:

$$Z^*(x_0) = \frac{1}{S} \int_{S(x_0)} Z(x) dx. \quad (2.3)$$

Теория пространственных переменных, которая называется геостатистикой, позволяет предсказывать характеристики переменной Z^* , связанной с геометрической базой в поле S , по известным характеристикам другой точечной переменной V , заданной в поле X , отличном от поля S .

Одним из ключевых свойств пространственно распределенных данных является их *пространственное разрешение*. При исследовании того либо иного пространственного явления очень важно, чтобы имеющиеся данные могли адекватно отразить его. Обычно под **пространственным разрешением** понимается наименьший размер особенности, которую могут отражать данные и пространственные оценки. Разрешение интерполяционной пространственной оценки на регулярной сетке характеризуется размером ячейки, а если сетка оценивания нерегулярная, то ее разрешение можно охарактеризовать распределением расстояний между узлами сетки.

Пространственное разрешение растров определяет размер ячеек набора растровых данных и соотношение количества пикселей на экране и в изображении при текущем масштабе карты. Например, один пиксел на экране может быть результатом пересчета девяти пикселей в изображении в один (разрешение раstra 1:9). В этом случае каждый пиксел на экране должен отображать девять ячеек раstra, поэтому изображение не будет четким и детализированным.

Степень детализации (пространственных объектов (явлений)) раstra часто зависит от размера ячейки или пространственного разрешения, раstra. Ячейка раstra должна быть достаточно мала для отображения мелких деталей, но при этом достаточно велика, чтобы хранение и компьютерный анализ были эффективны. Большое число пространственных объектов, более мелкие пространственные объекты или объекты с большей детальностью в экстендах пространственных объектов могут быть представлены растром с меньшим размером ячеек.

Однако больше не всегда означает лучше, поскольку чем меньше размер ячейки, тем больше размер набора растровых данных, отображающего данную поверхность; поэтому для его хранения требуется больший объем дискового пространства, что увеличивает и время обработки. Более высокое пространственное разрешение подразумевает большее количество ячеек на единицу площади; на рис. 2.3 слева показано изображение с более высоким пространственным разрешением, чем справа.



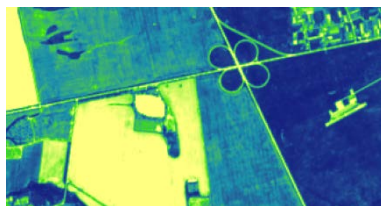
Рис. 2.3. Изображение с высоким (слева) и низким (справа) пространственным разрешением

Пространственное разрешение относится к размеру ячейки, представляющей покрываемый ею участок поверхности земли. Поэтому, если участок отображается ячейкой размером 5×5 м, то разрешение составляет 5 м. Чем выше разрешение растра, тем меньше размер ячейки, и следовательно, выше детализация отображаемой территории. Обратная ситуация происходит с масштабом: чем меньше масштаб, тем хуже детализация. Например, ортофотоснимок, отображаемый в масштабе 1:2000, показывает больше деталей (проявляющихся при увеличении масштаба отображения), чем тот же снимок, отображаемый в масштабе 1:24 000 (при уменьшении масштаба отображения). Однако если размер ячейки этого снимка составляет 5 м, его разрешение останется неизменным вне зависимости от масштаба отображения,

поскольку физический размер ячейки (площадь поверхности земли, отображаемая в одной ячейке) не меняется (рис. 2.4).



Размер ячейки 10 м, масштаб 1:50 000



Размер ячейки 10 м, масштаб 1:10 000

Рис. 2.4. Влияние масштаба на отображение растровых данных с одинаковым пространственным разрешением

Детализация отображения растра зависит от его пространственного разрешения, а не от масштаба изображения, т. е. при одинаковом масштабе более детальным будет растровое изображение с большим разрешением.

2.3. Сеть мониторинга и кластерность. Декластеризация данных

Мониторинг как общенаучное понятие представляет собой постоянное отслеживание какого-либо процесса или явления для установления его соответствия первоначальным предположениям или желаемому результату. Цель мониторинга – постоянный или периодический контроль за объектом наблюдения, предусматривающий возможность предупреждения негативных или нежелательных изменений состояния объекта. В процессе мониторинга создается *мониторинговая сеть*, состоящая из пунктов, в которых с установленной периодичностью выполняются те или иные наблюдения. Например, в Республике Беларусь мониторинг фоновое загрязнения почв проводится на сети, включающей 90 пунктов наблюдения, распределенных по всей территории страны и расположенных в луговых биогеоценозах с ненарушенным почвенным покровом. Мониторинг загрязнения почв сельскохозяйственных угодий остаточными количествами хлорорганических пестицидов проводится на 116 пунктах наблюдения на пахотных землях сельскохозяйственного назначения в пределах 29 административных районов всех областей республики. Периодичность мониторинго-

вых наблюдений составляет 1 раз в 3–5 лет в зависимости от принадлежности и назначения земель и их функционального использования⁷.

Мониторинговые сети бывают *регулярными* и *нерегулярными*. Адекватная структура мониторинговой сети – важнейшая часть любого мониторингового исследования, связанного с моделированием и оценкой на основании данных, получаемых из естественных ресурсов, или других явлений, происходящих на местности. Выбор типа сети мониторинга обуславливается видом мониторинга, точностью, предъявляемой к его данным, периодичностью мониторинговых наблюдений, свойствами объекта мониторинга.

Для визуализации сети наблюдений используют ряд способов, простейшим и наиболее общепринятым из которых является нанесение точек (пунктов наблюдений) на плоскость пространственных координат, при этом цвет нанесенной точки может соответствовать измеренной в ней величине (рис. 2.5, а).

Если сеть мониторинга имеет зоны с заметно более высокой плотностью измерений, чем остальная область, то сеть мониторинга **кластерная**⁸. Для визуализации кластерной структуры сети мониторинга используется *триангуляция Делоне*⁹ – система треугольников с вершинами в точках измерений, непересекающимися ребрами и минимальным количеством тупоугольных треугольников. Триангуляция используется для создания поверхности из точек, расположенных на нерегулярной сети (случайно, нерегулярно). Результатом триангуляции Делоне является создание TIN-модели (TriangulatedIrregular Network – триангуляционная нерегулярная модель) – векторного способа отображения поверхности (рис. 2.5, б). Такой способ визуализации позволяет качественно обособить области с повышенной плотностью изме-

⁷ Источник: <http://rad.org.by/monitoring/soil.html>.

⁸Кластер (cluster) – область повышенной плотности точек измерений пространственной функции.

⁹Делоне Борис Николаевич (1890–1980) – советский математик, профессор МГУ, член-корреспондент АН СССР. Разрабатывал теории правильного разбиения пространства, приведения квадратичных форм, решетчатых покрытий пространства сферами. Доказал фундаментальную теорему теории стереоэдров, по которой число различных комбинаторно-геометрических типов разбиений n -мерного евклидова пространства на выпуклые нормальные стереоэдры является конечным. Занимался также математическими вопросами кристаллографии и, в частности, установил наличие 24 типов трехмерных решеток в зависимости от комбинаторного строения области Вороного – Дирихле и расположения элементов симметрии относительно нее. Разрабатывал вопросы приведения квадратичных форм и теорию решетчатых покрытий пространства сферами.

рений – с кластерами. Кроме того, триангуляция Делоне строит систему соседства: точки, которые соединены друг с другом ребрами треугольников, являются ближайшими соседями по отношению друг к другу. Еще одним видом визуализации сети мониторинговых наблюдений являются *полигоны Вороного*, или, как их еще называют, *разбиение Тиссена*, *ячейки Дирихле* или области влияния (рис. 2.5, в).

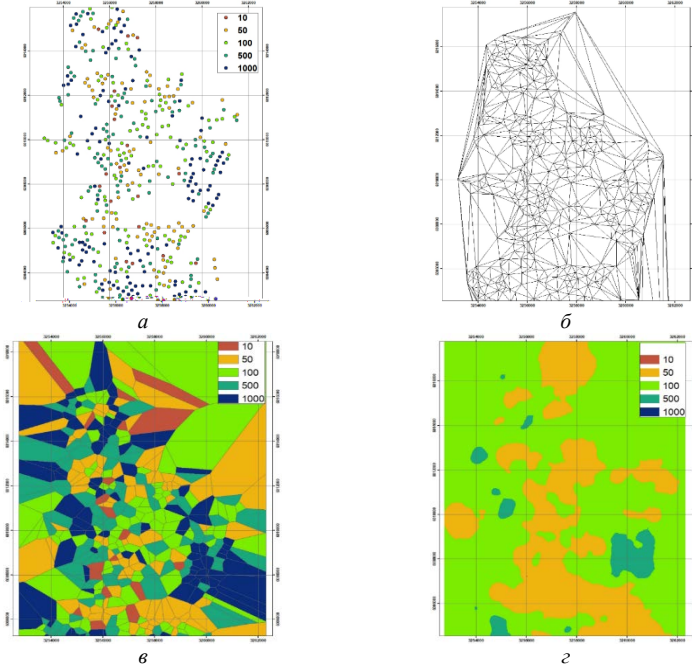


Рис. 2.5. Диаграмма расположения точек измерений (а), триангуляция сети мониторинга (б), полигоны Вороного (в) и контуры данных измерений по триангуляции (г)

Полигон Вороного P_i , построенный для точки измерений x_i , характеризуется тем, что содержит те и только те точки, расстояние от которых до точки x_i меньше или равно расстоянию до любой другой точки измерений x_j . При построении полигонов Вороного используется система соседства, полученная в процессе триангуляции Делоне. Границы полигона Вороного P_i состоят из отрезков серединных перпендику-

ляров, проведенных к сторонам треугольников Делоне. Полигоны Вороного можно использовать как разрывную интерполяционную оценку (оценка по ближайшему соседу). Для этого каждой точке, попавшей в полигон, присваивается значение, соответствующее его материнской точке. Эти полигоны также используются в задачах пространственной классификации – классификация (по ближайшему соседству).

Триангуляция также является основой для построения простейшей линейной интерполяции, поскольку три точки в пространстве (вершины треугольников) однозначно определяют плоскость, в пределах которой значения функции вычисляются согласно геометрическим принципам (рис. 2.5, з).

С целью выявления наличия кластерных структур (сгущений) или разреженностей в сети мониторинга (наборе точек измерений) проводят *анализ сети мониторинга*. Простейшими методами такого анализа являются описание топологии сети с помощью гистограммы расстояний между точками (рис. 2.6, а) и гистограммы площадей полигонов Вороного (рис. 2.6, б). Гистограмма в данном случае – это график числа каких-либо событий (числа пар или числа полигонов), попавших в какой-либо интервал значений.

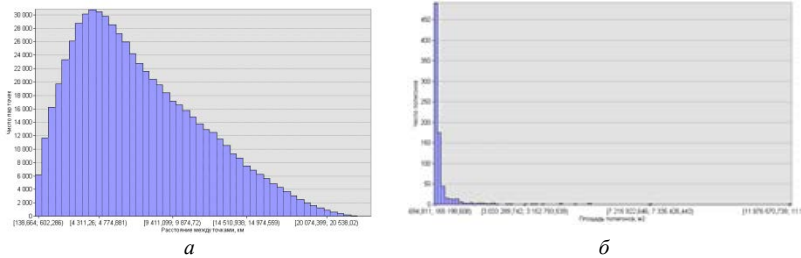


Рис. 2.6. Гистограмма расстояний между точками (а) и гистограмма площадей полигонов Вороного (б)

При равномерном распределении точек в пространстве число пар должно быть одинаково для всех расстояний или уменьшаться при увеличении расстояния за счет граничного эффекта. Рост числа пар с ростом расстояния между точками свидетельствует о наличии кластеров. Гистограмма площадей полигонов для регулярной сети должна представлять собой дельта-функцию (один пик), поскольку все полигоны одного размера. Любые искажения (широкий пик, длинный

хвост, несколько пиков) означают присутствие каких-либо особенностей в сети мониторинга.

Еще одним методом, позволяющим выполнить анализ сети мониторинга, является статистический подход, рассматривающий точки измерений как случайный точечный процесс. Охарактеризовать распределение точек (пунктов мониторинга) возможно с использованием статистических индексов. Примером такого подхода является диаграмма Морисита. Индекс Морисита¹⁰ вычисляется для области, разбитой на прямоугольные ячейки равного размера, по формуле

$$I_{\delta} = Q \frac{\sum_{i=1}^Q n_i(n_i-1)}{N(N-1)}, \quad (2.4)$$

где N – полное число точек сети мониторинга;

Q – число ячеек разбиения;

$n_i (i = 1, 2, \dots, Q)$ – число точек сети мониторинга, попавших в i -ю ячейку.

Этот индекс характеризует вероятность того, что при выборе двух случайных точек наблюдений они окажутся в одной ячейке. Диаграмма Морисита представляет собой зависимость индекса Морисита от размера ячейки разбиения. Существуют три типа диаграммы Морисита, которые позволяют судить о характеристиках сети мониторинга:

– если на диаграмме величина индекса Морисита с ростом размера ячейки также возрастает и стремится к 1, тогда распределение точек наблюдений сети мониторинга является равномерным;

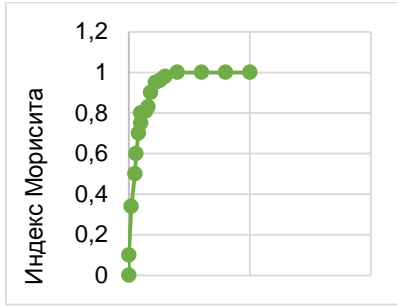
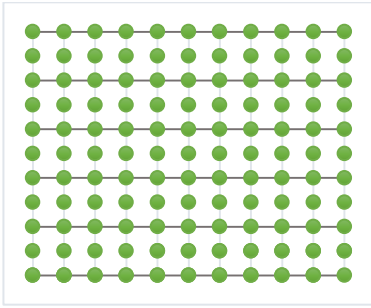
– если на диаграмме величина индекса Морисита не зависит от размера ячейки и примерно равна ≈ 1 (колеблется около 1), то распределение точек наблюдений сети мониторинга случайно и не имеет кластерных структур;

– если на диаграмме величина индекса Морисита с ростом размера ячейки уменьшается или растет выше 1, то распределение точек сети мониторинга кластерное и требует декластеризации.

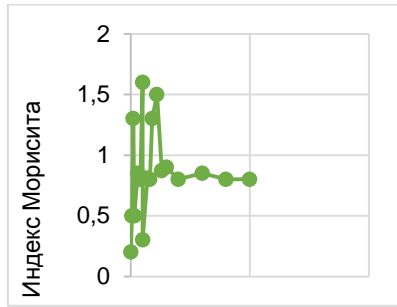
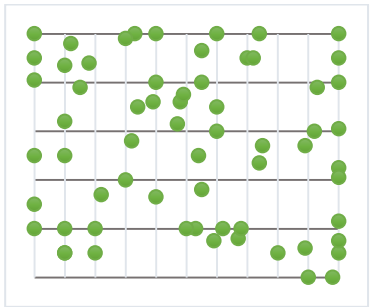
Иными словами, индекс Морисита равняется $I_{\delta} = 1$ при случайном типе распределения точек (объектов) в пространстве, $I_{\delta} < 1$ – при равномерном, и $I_{\delta} > 1$ – при групповом.

На рис. 2.7 приведены примеры диаграмм Морисита для различных типов сетей мониторинга.

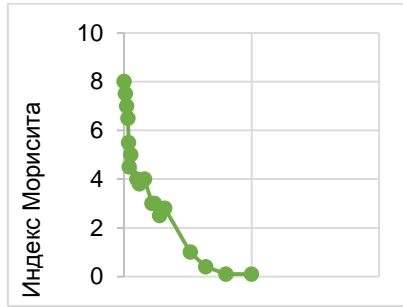
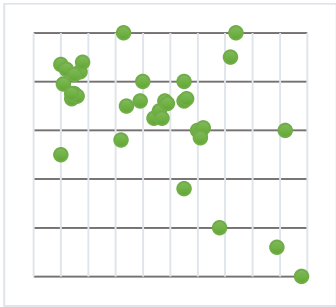
¹⁰Масаки Морисита (1913–1997) – японский эколог, известный ученый в области популяционной экологии, почетный профессор Киотского университета. Основоположник популяционной экологии Японии.



a



б



в

Рис. 2.7. Примеры диаграммы Морисита для различных сетей мониторинга: регулярная равномерная сеть (*a*), произвольная сеть со слабой кластерной структурой (*б*), произвольная слабо связанная кластеризованная сеть (*в*)

В частности, в случае мониторинговых наблюдений, выполняемых на регулярной равномерной сетке, диаграмма имеет вид гладкой кри-

вой логарифмического типа, стремящейся к единице (рис. 2.7, а). При наличии многочисленных кластеров в плотной неравномерной сети мониторинга кривая Морисита изобилует точками перегиба, которые характеризуют размеры различных кластеров (рис. 2.7, б). В случае нерегулярной сети мониторинга с несколькими четко выраженными кластерами кривая Морисита имеет более гладкий вид и уменьшается, стремясь к единице (рис. 2.7, в). Размер кластеров в этом случае характеризуют точки изменения кривизны.

В случае если нерегулярная сеть мониторинга сильно кластеризована, необходимо выполнить декластеризацию – присвоение весов значениям пространственной функции в точках в зависимости от характера сети мониторинга. Такая операция позволит сделать сеть точек наблюдений нерегулярной сети более равномерной, что является непременным условием для выполнения дальнейшей интерполяции и построения интерполяционных поверхностей. Различают два основных метода декластеризации – **полигональную** и **ячеистую**. Оба метода обеспечивают определение весов, приписываемых каждому элементу выборки. В дальнейшем одномерные параметры генеральной совокупности (математическое ожидание, дисперсия) оцениваются с учетом этих весов: чем больше вес, тем больше вклад данного элемента выборки в оценку. Метод **полигональной декластеризации** связывает с каждым элементом выборки вес, пропорциональный площади его многоугольника влияния (каждая точка внутри такого многоугольника ближе к данному элементу, чем к любому другому из той же выборки). Сумма площадей многоугольников масштабируется единицей, и кластеризованные точки (полигоны с наименьшей площадью) получают меньшие веса, чем некластеризованные (рис. 2.8).

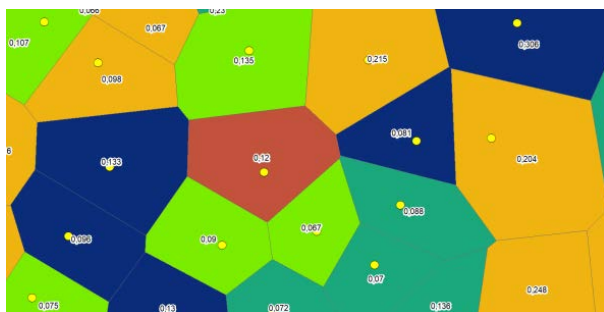


Рис. 2.8. Полигональная декластеризация (числа обозначают декластеризующие веса)

При использовании данного метода необходимо решить задачу построения границ многоугольников влияния, поскольку если данные не окружены границей (например точки, находящиеся на ребрах полигона либо очень близко прилегающие к ним), то они часто могут получать высокие весовые коэффициенты.

Метод **ячейстой декластеризации** предусматривает разделение всей области, покрытой мониторинговой сетью, на прямоугольные ячейки, при этом каждая точка выборки получает вес, обратно пропорциональный числу точек, попавших в одну и ту же ячейку (рис. 2.9).

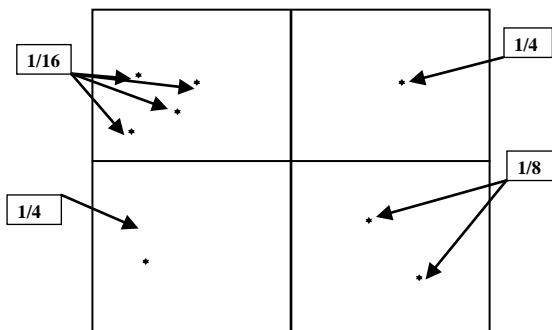


Рис. 2.9. Ячейстая декластеризация (показанные числа обозначают декластеризующие веса)

Вес, присваиваемый i -й точке, вычисляется по формуле

$$w_i = 1 / (N \cdot n_i), \quad (2.5)$$

где N – общее число ячеек;

n_i – число точек выборки, попавших в i -ю ячейку.

Поскольку сумма всех весов равна единице, а элементы выборки, попавшие в одну ячейку, получают одинаковые веса, далее оценивание выполняют в два шага:

1) в каждой ячейке оценивается локальное математическое ожидание как среднее попавших в нее n_i точек выборки;

2) оценивается глобальное математическое ожидание как среднее из оценок локальных математических ожиданий по N ячейкам.

Если кластеризация наблюдается в области с высокими значениями, размер ячеек выбирают как дающий наименьшее глобальное среднее. Если кластеризация наблюдается в области низких значений, раз-

мер ячеек выбирается как дающий наибольшее значение глобального среднего. Если кластеризация наблюдается в обеих областях, размер ячеек для декластеризации выбирают так, чтобы в областях, где нет кластеризации, в ячейку попадала одна точка данных. Метод ячеистой декластеризации реализуется по умолчанию в модуле Geostatistical Analyst ГИС ArcGIS версии 10.x.

Для декластеризации параметров двумерных распределений (полувариограмма, коррелограмма, ковариация) взвешиваются не кластеризованные точки выборки, а кластеризованные пары точек. Для того чтобы «разбить» кластеры, для расчетов используют небольшое число точек из каждого кластера, т. е. создают выборку, поэтому данный метод декластеризации носит название **выборочная декластеризация**.

2.4. Пространственная непрерывность данных. Стационарность данных

Пространственная непрерывность, которая означает, что близко расположенные в пространстве измерения, скорее всего, будут иметь близкие значения, является важным свойством пространственно распределенных данных. Она существует в большинстве геофизических явлений и выражает степень изменения переменной в пространстве. Пространственная непрерывность имеет статистический характер – непрерывность обычно наблюдается в среднем: когда точка X стремится к X_0 , лишь среднее значение $|f(X) - f(X_0)|^2$ стремится к нулю. Иными словами, объекты, явления и процессы, которые расположены ближе в пространстве, являются более подобными между собой по сравнению с теми, которые более удалены друг от друга.

Обычно пространственная непрерывность описывается с помощью *корреляционных* и *ковариационных* функций (статистических моментов), выражающих меру этой непрерывности. Одной из наиболее популярных функций, применяемых для количественного описания пространственной непрерывности и моделирования пространственной корреляции, является **вариограмма** – статистический двухточечный момент второго порядка, характеризующий степень различия данных в зависимости от расстояния между ними (зависимость квадрата разности значений функции в точках от вектора расстояния между точками). Соответственно, чем ближе значения данных (меньше разница между ними), тем больше значение вариограммы.

Вариация разницы значений переменной в двух точках как функция расстояния и направления между ними описывается следующей зависимостью:

$$\gamma(x; x+h) = 0,5 \text{Var}[Z(x) - Z(x+h)] = 0,5E[Z(x) - Z(x+h)]^2, \quad (2.6)$$

где $Z(x)$ и $Z(x+h)$ – значения в точках, разделенные вектором h .

Для $N(h)$ экспериментальных точек, разделенных вектором h , зависимость принимает следующий вид:

$$N(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i+h)]^2. \quad (2.7)$$

Пространственная непрерывность непосредственно тесно связана со **стационарностью данных**. Стационарность, или постоянство, – свойство процесса либо явления не изменять свои характеристики со временем. Пространственная стационарность в строгом смысле означает, что распределения случайной величины в двух различных зонах области распределения являются идентичными. **Пространственная нестационарность** заключается в меняющемся характере функции распределения данных в зависимости от местоположения точек измерения.

Случайная функция стационарна в строгом смысле, если ее закон распределения инвариантен относительно сдвига, т. е. $\{Z(x_1), Z(x_2), \dots, Z(x_k)\}$ и $\{Z(x_1+h), Z(x_2+h), \dots, Z(x_k+h)\}$ имеют одинаковый k -мерный закон распределения при любом h . Однако на практике это предположение редко выполняется, поэтому обычно оно заменяется гипотезой о стационарности второго порядка или гипотезой о стационарности приращений.

Случайная функция $Z(x)$ обладает стационарностью второго порядка, если:

- математическое ожидание существует и не зависит от местоположения x :

$$E\{Z(x)\} = m, \forall x; \quad (2.8)$$

- для каждой пары случайных величин $\{Z(x), Z(x+h)\}$ ковариация существует и зависит только от разности координат h :

$$C(h) = E\{Z(x+h) \cdot Z(x)\} - m^2, \forall x. \quad (2.9)$$

Многие природные явления имеют бесконечный разброс значений, для которых не существует дисперсия или ковариация. Однако существует дисперсия для приращений значений, а значит, существует и может быть определена вариограмма для самого явления. Случайная функция $Z(x)$ обладает стационарностью приращений, если:

– математическое ожидание существует и не зависит от местоположения x :

$$E\{Z(x)\} = m, \forall x; \quad (2.10)$$

– для любого h приращение $Z(x + h) - Z(x)$ имеет конечную дисперсию, не зависящую от x :

$$\text{Var}\{Z(x + h) - Z(x)\} = E\{[Z(x + h) - Z(x)]^2\} = 2 \gamma(h), \forall x. \quad (2.11)$$

В условиях стационарности второго порядка корреляция между измерениями в двух точках, как уже указывалось, предполагается зависящей только от разности местоположений этих точек. С точки зрения пространственных корреляций это означает, что различные регионы статистически подобны, что, кстати, позволяет интерпретировать различные регионы как различные реализации стохастической региональной функции и делать статистические выводы. Таким образом, значения измерений, проведенных в некотором конечном множестве точек, могут быть исследованы с точки зрения поведения разности между ними. Всевозможные пары точек могут быть рассортированы по классам в соответствии с разностью их координат $h = x_i - x_j$, называемой **лагом** (или лэгом – *lag*). Для близких точек разность значений функции в них обычно меньше и растет с увеличением расстояния между точками.

2.5. Геоestatистическое оценивание данных.

Проверка качества модели методом кросс-валидации

Геопространственная статистика – раздел математической статистики, который связан с численным описанием переменных, распределенных в географическом пространстве и времени. Наиболее часто инструменты геоestatистики используются для решения задачи интерполяции – восстановления сплошного поля распределения случайной величины по ограниченному множеству данных в точках наблюдений. **Геоestatистическое оценивание данных** подразумевает оценку статистических параметров, которые определяют пространственное и (или)

временное распределение и зависимость соответствующих переменных. Данный процесс предусматривает: определение наличия ошибок и выбросов в данных; оценку базовых статистических закономерностей; проведение корреляционного анализа при наличии нескольких переменных. С этой целью определяются:

1) *минимальное и максимальное расстояния окрестности поиска ближайшего соседства*, дающие возможность подобрать оптимальную величину окрестности поиска при пространственном моделировании (инструмент геообработки «Пошаговая пространственная автокорреляция»);

2) *глобальный индекс Морана (I)*, позволяющий определить, имеет ли место явление кластеризации по отношению к атрибутивным данным, каковыми являются сведения о физико-химических и агрохимических свойствах почвы (инструмент геообработки «Пространственная автокорреляция»);

3) *общий индекс Getis-OrdG* для оценки общей структуры и тренда геоданных, а также степени кластеризации высоких и (или) низких значений выборки (инструмент геообработки «Высокая (Низкая) кластеризация»);

4) *индекс Getis-OrdG**, позволяющий установить наличие кластеризации данных с высокими и низкими значениями и выполнить анализ кластеров и выбросов для установления отличий атрибутивных значений каждого рабочего участка от его окрестности, а окрестности рабочего участка – от остальной территории (инструмент геообработки «Анализ горячих точек») (см. подробно главу 4).

Дальнейший пространственный анализ предполагает исследование и моделирование **пространственной корреляции**¹¹ между данными по одной или нескольким переменным. Основной геостатистической моделью, которая в том или ином виде используется во всех методах геостатистики, является **кригинг**¹² (kriging) – линейный интерполятор, использующий для получения оценки значения функции в некоторой

¹¹Пространственная корреляция – зависимость между значениями пространственно распределенной функции от взаимного расположения точек.

¹²Кригинг (kriging) – геостатистический метод пространственной интерполяции, основанный на линейной регрессии, обладает несмещенностью и минимальной дисперсией оценки. Метод назван в честь южноафриканского горного инженера Дэниела Крига, занимавшегося созданием геологических карт по ограниченному набору данных в некоторой области.

точке пространства x_0 экспериментально измеренные значения этой функции в других точках¹³:

$$Z^*(x_0) = \sum_{s=1}^{N(x)} w_i(x_0) Z(x_i), \quad (2.12)$$

где $Z^*(x_0)$ – базовый линейный регрессионный оценщик;

$Z(x_i)$ – известное значение i -й переменной;

$w_i(x_0)$ – вес (весовой коэффициент) переменной.

Для определения весов $w_i(x_0)$ могут использоваться различные детерминистические методы, например, веса могут выбираться обратно пропорциональными расстоянию от измеренной точки до оцениваемой или в соответствии с каким-либо другим предположением о природе связей в данных. Веса кригинга не зависят от известных значений функции, а определяются моделью пространственной корреляции.

Валидация – проверка качества работы модели с помощью данных, неиспользованных для ее настройки. При использовании той или иной модели интерполяции крайне важно правильно подобрать значения модельно-зависимых параметров. Для кригинга такими параметрами являются параметры модели вариограммы. При работе с реальными данными не всегда удастся сразу выбрать теоретическую модель экспериментальной вариограммы. Для проверки качества выбранной модели используют различные количественные методы: кросс-валидацию (cross-validation), метод складного ножа¹⁴ (jack-knife), бутстреп¹⁵ (bootstrap).

Кросс-валидация (cross-validation) – наиболее простой и часто используемый метод подбора оптимальных параметров модели интерполяции с помощью оценки значения в точке измерения без учета самого измерения в этой точке.

Выполняется кросс-валидация следующим образом:

– из базы данных временно изымается одна точка, и для нее проводится оценка значения;

¹³Более подробная информация приведена в главе 6.

¹⁴Метод складного ножа (jack-knife) является общим случаем кросс-валидации, когда оценивание проводится не в одной, а в нескольких точках измерений, данные о которых предварительно изымаются из рассмотрения.

¹⁵Бутстреп (bootstrap) состоит в оценке на основе случайных выборок из набора данных. Выборки делаются из исходного набора случайным образом. Выбранная точка не изымается, она может попасть в выборку несколько раз. Оценка проводится по оставшимся невыбранными точкам.

– полученное значение сравнивается с известным, и вычисляется невязка – разница между измеренными и оцененными значениями:

$$\Delta Z(x) = Z(x) - Z^*(x); \quad (2.13)$$

– первые два шага проводятся для всех точек базы данных.

Полученные невязки $\Delta Z(x)$ могут быть графически представлены в виде карты (карты невязок), по которой можно посмотреть, в каких зонах метод интерполяции срабатывает лучше, а в каких хуже.

Точность методов интерполяции определяют по величине средней ошибки (ME), среднеквадратичной ошибки (RMSE), средней нормированной ошибки (MSE), средней стандартизированной ошибки (ASE) и среднеквадратичной нормированной ошибки (RMSS):

$$ME = \frac{\sum_{i=1}^N (Q_i - S_i)}{N}; \quad (2.14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_i - S_i)^2}{N}}; \quad (2.15)$$

$$MSE = \sqrt{\frac{\sum_{i=1}^N \frac{(Q_i - S_i)}{\sigma^2(S_i)}}{N}}; \quad (2.16)$$

$$ASE = \sqrt{\frac{\sum_{i=1}^N \sigma^2(S_i)}{N}}; \quad (2.17)$$

$$RMSS = \frac{RMSE}{\Delta}, \quad (2.18)$$

где Q_i – наблюдаемое значение;

S_i – предсказанное значение;

N – объем выборки;

σ – дисперсия выборки;

Δ – диапазон, равный разнице между максимальным и минимальным наблюдаемыми значениями.

Вопросы для самостоятельного изучения

1. Понятие пиксела и воксела. Характеристики пиксела.
2. Алгоритмы сжатия растровой информации. Понятие о форматах файлов для хранения растровых изображений (*tiff, bmp, jpeg, gif, pcx*).
3. Понятие топологических отношений. Матрица девяти пересечений.

Тестовые задания для контроля знаний

1. *Дополните определение.* Данные, в которых кроме информации о местоположении присутствует информация о времени их получения, называют:

- 1) пространственно-распределенными;
- 2) пространственно-временными;
- 3) геопространственными;
- 4) геоинформационными;
- 5) структурированными.

2. *Дополните определение.* Цифровое представление пространственных объектов в виде совокупности ячеек раstra (пикселов) с присвоенными им значениями класса объекта называют:

- 1) модель пространственных данных;
- 2) цифровая модель пространственных данных;
- 3) модель геопространственных данных;
- 4) растровая модель пространственных данных;
- 5) векторная модель пространственных данных.

3. *Выберите правильный вариант ответа.* Какие данные в качестве определяющих характеристик используют систему «volume – velocity – variety»?

- 1) большие данные;
- 2) структурированные данные;
- 3) геопространственные данные;
- 4) пространственно-временные данные;
- 5) пространственно-распределенные данные.

4. *Дополните определение.* Цифровое представление пространственных объектов в виде множества дискретных объектов с заданными характеристиками называют:

- 1) модель пространственных данных;
- 2) цифровая модель пространственных данных;
- 3) модель геопространственных данных;
- 4) растровая модель пространственных данных;
- 5) векторная модель пространственных данных.

5. *Выберите правильные характеристики для векторной топологической модели:*

1) все объекты модели имеют определенное заданное размещение относительно друг друга;

2) все объекты модели могут произвольно размещаться в пространстве;

3) модель данных содержит три основных типа объектов: узлы, дуги и полигоны;

4) все объекты являются полностью независимыми друг от друга;

5) все объекты модели не могут произвольно размещаться в пространстве;

6) пространственная информация представляется наборами элементарных графических примитивов.

6. Выберите правильные характеристики для векторной нетопологической модели:

1) все объекты модели имеют определенное заданное размещение относительно друг друга;

2) все объекты модели могут произвольно размещаться в пространстве;

3) модель данных содержит три основных типа объектов: узлы, дуги и полигоны;

4) все объекты являются полностью независимыми друг от друга;

5) все объекты модели не могут произвольно размещаться в пространстве;

6) пространственная информация представляется наборами элементарных графических примитивов.

7. Правильным или нет является утверждение? Шейп-модель является наиболее широко распространенной векторной нетопологической моделью пространственно-распределенных данных.

1) да;

2) нет.

8. Выберите правильный вариант(ты) ответа. Координатная привязка может быть:

1) стандартной;

2) пространственной;

3) геопозиционированной;

4) временной;

5) пространственно-временной.

9. Правильным или нет является утверждение? Геометрическая база является составной частью геометрического поля.

1) да;

2) нет.

10. Дополните определение. Наименьший размер особенности, которую могут отражать данные и пространственные оценки, называют:

1) пространственное геопозиционирование;

2) пространственное разрешение;

- 3) пространственное распределение;
- 4) пространственная модель;
- 5) пространственное поле.

11. *Дополните определение.* Сеть мониторинга, имеющую зоны с заметно более высокой плотностью измерений, чем остальная область, называют:

- 1) регулярной;
- 2) нерегулярной;
- 3) геопозиционированной;
- 4) кластерной;
- 5) пространственно-временной.

12. *Выберите правильный вариант(ты) ответа.* С целью выявления наличия кластерных структур или разреженностей в сети мониторинга проводят ее анализ, предусматривающий:

- 1) описание топологии сети с помощью гистограммы расстояний между точками;
- 2) построение вариограммы пунктов мониторинга;
- 3) построение гистограммы площадей полигонов Вороного;
- 4) расчет статистического индекса Морисита;
- 5) декластеризацию пунктов мониторинга.

13. *Дополните определение.* Операцию, позволяющую сделать сеть точек наблюдений нерегулярной сети более равномерной посредством присвоения весов значениям пространственной функции в точках в зависимости от характера сети мониторинга, называют:

- 1) триангуляция;
- 2) делинеация;
- 3) декластеризация;
- 4) поляризация;
- 5) классификация.

14. *Дополните определение.* Различают следующие виды декластеризации:

- 1) полигональная;
- 2) статистическая;
- 3) ячеистая;
- 4) выборочная;
- 5) триангуляционная.

15. *Дополните определение.* Статистический двухточечный момент второго порядка, характеризующий степень различия данных в зависимости от расстояния между ними, называют:

- 1) корелограмма;
- 2) вариограмма;
- 3) диаграмма;
- 4) корреляция;
- 5) регрессия.

16. *Правильным или нет является утверждение?* Идентичное распределение случайной величины в двух различных зонах области наблюдения называют пространственной нестационарностью данных.

- 1) да;
- 2) нет.

17. *Правильным или нет является утверждение?* Изменение характера функции распределения данных в зависимости от местоположения точек измерения называют пространственной нестационарностью данных.

- 1) да;
- 2) нет.

18. *Дополните определение.* Раздел математической статистики, который связан с численным описанием переменных, распределенных в географическом пространстве и времени, называют:

- 1) геоинформатика;
- 2) геоматика;
- 3) теория вероятностей;
- 4) геопространственная статистика;
- 5) прикладная статистика.

19. *Выберите правильный вариант(ты) ответа.* Геостатистическое оценивание данных предусматривает:

- 1) определение наличия ошибок и выбросов в данных;
- 2) моделирование геопространственных данных;
- 3) оценку базовых статистических закономерностей;
- 4) проведение корреляционного анализа при наличии нескольких переменных;
- 5) выполнение анализа мониторинговой сети.

20. *Дополните определение.* Наиболее простой и часто используемый метод подбора оптимальных параметров модели интерполяции с помощью оценки значения в точке измерения без учета самого измерения в этой точке называют:

- 1) корреляция;
- 2) регрессия;
- 3) кросс-валидация;
- 4) интерполяция;
- 5) координация.

3. ВВЕДЕНИЕ В ГЕОПРОСТРАНСТВЕННЫЙ АНАЛИЗ

3.1. Пространственные объекты и пространственные атрибуты. Анализ местоположения объектов.

3.2. Анализ распределения числовых показателей. Карты плотности. Поиск объектов внутри области. Анализ окружения и пространственных изменений.

3.3. Основные виды векторного ГИС-анализа.

3.4. Основные виды растрового ГИС-анализа.

3.5. Основные операции моделирования в ГИС.

Количество часов, отведенных на изучение главы, – 8.

Аннотация темы.

Тематика лекции посвящена базовым основам выполнения геопространственного анализа. В пункте 3.1 дано общее определение понятия «пространственный объект», охарактеризованы основные виды пространственных объектов, приведены сведения об основных типах атрибутивных данных. В пункте 3.2 рассмотрены особенности анализа количественных показателей геоданных и их классификации различными методами. Кроме того, детально рассмотрены условия использования, преимущества и недостатки методов создания карт плотности распределения объектов, поиска объектов внутри области, оценки окружения объектов и анализа пространственных изменений. Пункт 3.3. посвящен характеристике таких видов векторного ГИС-анализа, как элементарный пространственный анализ, пространственная статистика, расширенный пространственный анализ, сетевой анализ. В пункте 3.4 рассмотрены особенности выполнения растрового ГИС-анализа: интерполяции растра, анализа гипсометрических поверхностей, анализа плотности и расстояний на основе растров, использования функций растровой алгебры. Основные операции моделирования в ГИС рассматриваются в пункте 3.5, где, в частности, освещены вопросы имитационного моделирования.

Литературные и информационные источники

1. Григорьев, И. AnyLogic за три дня: практ. пособие по имитационному моделированию / И. Григорьев. – AnyLogic, 2017. – 273 с.

2. Курлович, Д. М. Геоинформационные методы анализа и прогнозирования погоды: учеб.-метод. пособие / Д. М. Курлович. – Минск: БГУ, 2013. – 191 с.

3. Митчелл, Э. Руководство по ГИС-анализу / Э. Митчелл. – ESRI: 2000. – Ч. 1: Пространственные модели и взаимосвязи. – 170 с.

4. Borshchev, A. The Big Book of Simulation Modeling. Multimethod modeling with AnyLogic 6 / A. Borshchev. – AnyLogic North America, 2013. – 614 с.

5. Jensen J. R. Introductory Geographic Information Systems / J. R. Jensen, R. R. Jensen. – Boston: Pearson, 2013. – 195 p.

6. Silverman, B. W. Density estimation for statistics and data analysis / B. W. Silverman. – London – New York: Chapman & Hall, 1986. – 175 p.

Internet-ресурсы

7. Официальный сайт компании AnyLogic [Электронный ресурс]. – Режим доступа: <http://www.anylogic.com/anylogic/help/>.

8. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.

3.1. Пространственные объекты и пространственные атрибуты. Анализ местоположения объектов

Геопространственный анализ – процесс поиска пространственных закономерностей в распределении географических данных и взаимосвязей между объектами. Основные задачи *геопространственного анализа*:

– анализ местоположения объектов – поиск, где размещаются объекты;

– анализ распределения числовых показателей – выявление, где больше, где меньше;

– построение карт плотности – картографирование плотности;

– поиск объектов внутри области – поиск того, что внутри;

– анализ окружения – поиск того, что рядом;

– анализ пространственных изменений – картографирование изменений. Все объекты окружающего мира в геопространственном анализе представлены в виде набора **пространственных**¹⁶ и **атрибутивных**¹⁷ **данных**, носящих общее название «геопространственные данные». Геопространственные данные, в свою очередь, содержат четыре *интегрированных компонента*:

¹⁶Пространственные данные – цифровые данные о пространственных объектах, включающие сведения об их местоположении и свойствах, пространственных и непространственных атрибутах.

¹⁷Атрибутивные данные – информация в цифровом или текстовом виде о количественных и качественных характеристиках объектов или явлений.

– *географическое положение* (размещение) пространственных объектов (представляется 2-, 3- или 4-мерными координатами в географически соотнесенной системе координат (широта и долгота);

– *атрибуты* – свойство, качественный или количественный признак, характеризующий пространственный объект (но не связанный с его местоположением);

– *пространственные отношения* (определяют внутренние взаимоотношения между пространственными объектами);

– *временные характеристики* (представляются в виде сроков получения данных, определяют их жизненный цикл, изменение местоположения или свойств пространственных объектов во времени).

Элементом пространственных данных, моделирующим конкретный объект реального мира и содержащим пространственные (координаты, длину или периметр, площадь) и атрибутивные (характеристики) данные, является **пространственный объект**.

Пространственные объекты имеют качественные и количественные характеристики и подразделяются на *дискретные* (имеют определенное положение в пространстве и форму) (рис. 3.1) и *непрерывные* (поверхности, или географические поля) (рис. 3.2).

Описание однотипных пространственных объектов, топологических отношений между ними и дополнительных знаний об объектах содержится в модели пространственных данных. Наиболее распространенными способами цифрового описания пространственных объектов являются *растровые* и *векторные модели данных*¹⁸.

Различают следующие пространственные объекты:

– **точки** – нульмерные (точечные) объекты, характеризующиеся координатами на плоскости или в пространстве;

– **мультиточки** – нульмерные (точечные) объекты, состоящие из нескольких (не менее одной) точек;

– **линии** (полилинии, полиполилинии) – одномерные (линейные) объекты, состоящие из последовательности (не менее двух) точек, соединенных между собой отрезками (сегментами, дугами); не все последовательные точки могут соединяться между собой отрезками, а потому объекты данного типа могут иметь разрывы, т. е. быть топологически несвязанными (топологически связанные линии обычно называют *полилиниями*, а несвязанные – *полиполилиниями*);

¹⁸Более подробная информация содержится в главе 2.

– **полигоны** – двумерные (площадные) объекты, состоящие из нескольких (не менее одного) контуров, заданных в виде последовательности замкнутых линий, и частей плоскости внутри контуров;

– **сложные фигуры** (фигуры оформления, объекты САПР из систем автоматизированного проектирования) – разнообразные 0-, 1-, 2- и 3-мерные фигуры, используемые в ГИС для оформления. На практике используются прямоугольники, эллипсы, дуги эллипса, сплайны, внедренные изображения (в виде растров и метафайлов), OLE-объекты (объекты для вставки на карту произвольных графических изображений по технологии OLE операционной системы Windows), различные текстовые надписи – аннотации, указатели, размерные линии, а также специальные объекты для оформления карт в ГИС (масштабные линейки, стрелки направления на север, легенды карты, фрагменты других карт).

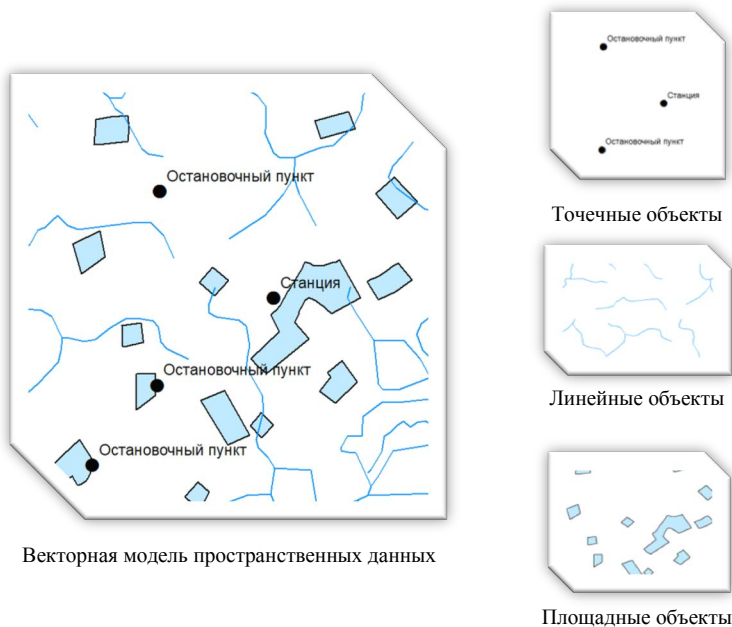
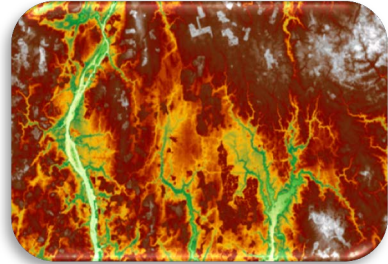


Рис. 3.1. Дискретные пространственные объекты



Данные дистанционного зондирования



Интерполяция по точечным данным

Рис. 3.2. Непрерывные пространственные объекты

Аннотации – это подписи, содержащие параметры отображения текста. Помимо текстовой строки каждой аннотации там хранятся и другие свойства, например, точки фигуры для размещения текста, его шрифт и точечный размер, а также другие свойства отображения. Аннотация может также быть связанной с надписываемыми объектами и может содержать подклассы. На рис. 3.3 показана карта, на которой линии горизонталей и объекты аннотаций перекрывают друг друга. Однако при использовании маски для аннотации части контурных линий становятся скрытыми, тогда как гипсометрическая цветовая заливка рельефа остается видимой.

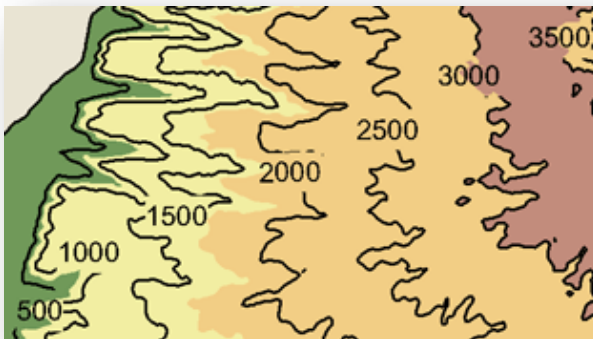


Рис. 3.3. Пример использования аннотаций

Объекты-размеры – специальный тип аннотации, показывающий специфические длины или расстояния, например, для указания длины стороны здания, участка земли или расстояния между двумя объектами. *Выровненные объекты-размеры* идут параллельно основной линии и показывают точное расстояние от начальной до конечной точки измерения. *Линейные объекты-размеры* не представляют истинное расстояние между начальной и конечной точками измерений. Линейные размеры могут быть вертикальными, горизонтальными или расположенными под углом. Вертикальная линия размера отображает расстояние по вертикали между начальной и конечной точками измерений. Горизонтальная – аналогичное расстояние по горизонтали. Линия размера, проходящая под углом к базовой линии, отображает собственную длину, а не размер базовой линии (рис. 3.4).

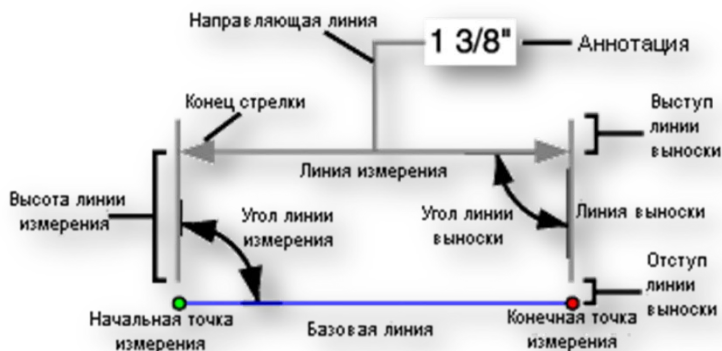


Рис. 3.4. Пример использования объектов-размеров

Мультиточки – пространственные объекты, состоящие из более чем одной точки, используемые для управления массивами очень больших совокупностей точек, таких как кластеры точек LiDAR¹⁹, которые могут содержать миллиарды пунктов (рис. 3.5).

¹⁹Лидар (LIDAR – Light Identification, Detection and Ranging) – технология получения и обработки информации дистанционного зондирования с помощью активных оптических систем (лазеров), использующих в том числе явления отражения света от поверхности Земли с проведением высокоточных измерений x , y , z координат.

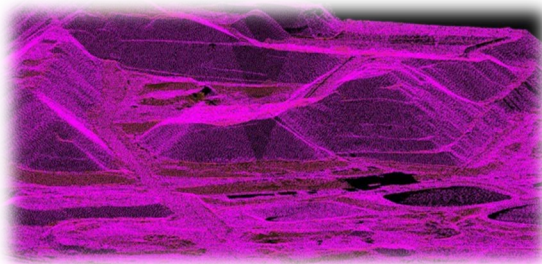


Рис. 3.5. Лазерное сканирование карьера для создания точной модели рельефа

Мультипатчи – это 3D-геометрия, используемая для представления внешней поверхности или оболочки объектов, которые занимают дискретную область или объем в трехмерном пространстве. Они охватывают плоские 3D-окружности и треугольники, используемые в комбинации для моделирования трехмерной оболочки, и могут применяться для представления как простых объектов, например, сфер и кубов, так и сложных объектов, таких как изоповерхности зданий (рис. 3.6).

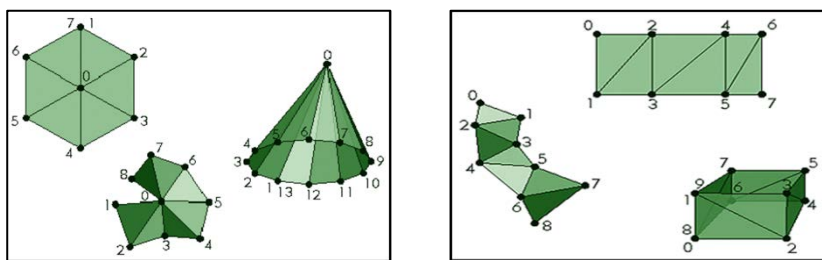


Рис. 3.6. Пример использования мультипатчей

Классы пространственных объектов – однородные совокупности однотипных объектов, каждый из которых имеет одинаковое пространственное представление в виде точек, линий или полигонов и общий набор атрибутивных полей, хранящихся в таблице базы данных. Они могут быть собраны в набор классов, но могут существовать

и по отдельности внутри базы геоданных²⁰. Все классы объектов в наборе имеют общую систему координат. Таблицы могут содержать дополнительные атрибуты для класса пространственных объектов или пространственную информацию, такую как адреса или координаты x , y , z .

Атрибуты – это числовые или символьные характеристики, не связанные с местоположением, содержащиеся в базе данных, которые могут относиться как к самим геометрическим примитивам (точки, линии, полигоны), так и к объектам, составленным из этих примитивов. Атрибутивные данные являются важнейшими элементами аналитических возможностей ГИС. Данные, хранящиеся в атрибутах, принадлежат, как правило, к целым, вещественным и символьным типам. *Значения атрибутов* подразделяются на следующие типы: *категории*; *ранги*; *численность*; *абсолютные величины*; *относительные значения*.

Категории представляют собой группы схожих объектов. Все объекты с одинаковым значением для категории ведут себя одинаково и отличаются от объектов других категорий. Например, можно разделить на категории дороги по типам: автомагистрали, скоростные автомобильные дороги, обычные автомобильные дороги, автомобильные дороги низших категорий. Значения категорий можно представить, используя числовые коды или текстовый идентификатор.

Ранги используются для сортировки объектов по порядку от большего к меньшему и устанавливают относительный порядок для объектов. Ранги часто используют в тех случаях, когда затруднительно произвести непосредственные измерения или если количественная мера характеризуется комбинацией нескольких факторов. Например, трудно количественно охарактеризовать качество почвы по комплексу показателей с различными единицами измерения. В этом случае можно просто установить, что участки, имеющие максимальные значения показателей, имеют более высокий ранг, нежели участки с их минимальными значениями.

Численность и абсолютные величины дают представление о некоторых суммарных количественных характеристиках объектов. *Численность* характеризует реальное число объектов на местности. *Абсолютная величина* представляет собой некоторую суммарную величину, связанную с каждым объектом (процентное соотношение, кратность). Использование этих количественных характеристик позволяет количе-

²⁰Более подробная информация содержится в главе 4.

ственно сравнивать друг с другом объекты с различными характеристиками.

Относительные значения показывают взаимные отношения между двумя количественными величинами и вычисляются делением одной количественной величины на другую для каждого объекта. Наиболее часто применяются такие относительные значения, как *средние значения, пропорции и плотности*. Средние значения используются для сравнения областей, содержащих небольшое число объектов, с областями, в которых таких объектов много. Соответственно, деление площади пахотных земель, находящихся в пределах административного района, на площадь сельскохозяйственных угодий даст значение пропорции количества пахотных земель в каждом из районов области. Плотность, в свою очередь, показывает распределение объектов или величин, отнесенных к единице площади. В качестве *характеристик атрибутов* геопространственных объектов могут выступать буквы, числа, индексы, абсолютные, относительные, условные (баллы, индексы) показатели. Атрибутивные данные геопространственных объектов представляются в форме специальных *атрибутивных таблиц*, состоящих из строк и столбцов (рис. 3.7).

The image shows a screenshot of a geospatial attribute table. The table has columns for OBJECTID, Shape, OBJECTID, LandType, LandCode, Name, Area, and SHAPE_Length. A red box highlights a specific row (row 7) and a specific column (LandCode). Callouts point to these elements: 'Название атрибута' points to the 'LandCode' header, 'Значение атрибута' points to the value '3' in the highlighted cell, 'Строка данных' points to the entire highlighted row, and 'Столбец данных' points to the highlighted column.

OBJECTID	Shape	OBJECTID	LandType	LandCode	Name	Area	SHAPE_Length
1	ПолYGON	73739	103	121	4	12812	2503.205302
2	ПолYGON	74529	103	123	6.8	4829	1294.175956
3	ПолYGON	74900	101	3	43	10867	5138.790672
4	ПолYGON	76967	101	3	2	160893	351.950875
5	ПолYGON	81546	103	121	4	018616	1198.676021
6	ПолYGON	81781	103	122	3	373324	1876.670167
7	ПолYGON	81802	101	3	4	323506	882.104411
8	ПолYGON	85905	103	121	4	234228	1236.354718
9	ПолYGON	86911	103	121	41	12517	3236.336005
10	ПолYGON	88293	103	121	107	284978	8887.266748
11	ПолYGON	88530	103	123	212	104136	11574.121042
12	ПолYGON	88532	101	3	121	109635	15360.803274
13	ПолYGON	88533	103	121	5	417847	3852.428276
14	ПолYGON	88536	103	121	4	874162	2520.444678
15	ПолYGON	88542	103	121	2	649002	1925.193077
16	ПолYGON	88546	101	3	6	20105	1306.771918
17	ПолYGON	88547	101	3	5	244512	1840.560383
18	ПолYGON	88553	103	121	2	926074	1856.072365
19	ПолYGON	88558	101	3	65	565314	4301.164861
20	ПолYGON	88559	101	3	125	096145	4715.849436

Рис. 3.7. Структура атрибутивной таблицы геоданных

Для оперативной и корректной обработки геопространственных данных принято, что каждая запись в таблице атрибутов объектов со-

держит описание одного объекта. В зависимости от содержания атрибутивных данных в атрибутивной таблице создается одно из возможных типов полей: символьное – любая комбинация алфавитно-цифровых символов; числовое – любые символы, которые составляют допустимое целое или вещественное число; дата. Каждая запись атрибутивной таблицы содержит также уникальный идентификатор объекта (ID), который имеет уникальные значения для каждой точки, дуги или полигона.

Атрибутивные таблицы позволяют картографировать и визуализировать геопространственные данные. С их помощью можно классифицировать или категорировать атрибуты для обозначения слоя. Например, можно использовать различные цвета для представления каждого типа землепользования в слое земельных участков. Кроме того, можно использовать атрибутивные значения для формирования текста и подписывания каждого объекта участка. На рис. 3.8 поля севооборота раскрашены в зависимости от сельскохозяйственной культуры, выращиваемой на них, и надписаны названиями культур.

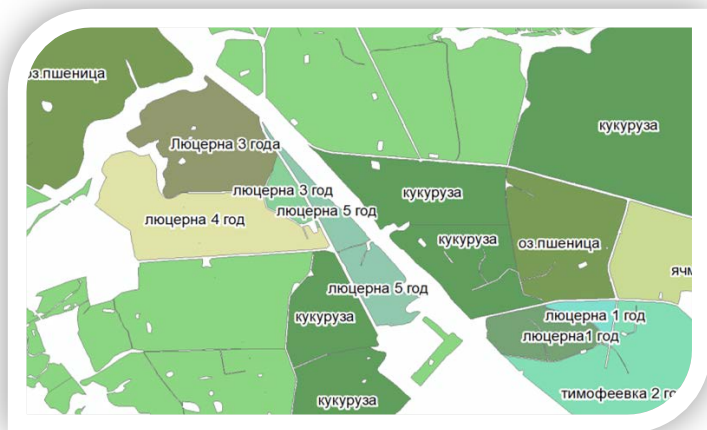


Рис. 3.8. Визуализация данных атрибутивной таблицы

Атрибутивные данные позволяют выполнять пространственный анализ и пространственные запросы. Например, можно изучить распределение объектов с определенными атрибутами, создав запрос для

выбора объектов, содержащих атрибуты, которые необходимо изучить. На рис. 3.9 земельные участки с содержанием меди, большим или равным 3 мг/кг, выбраны с использованием атрибутивного запроса.

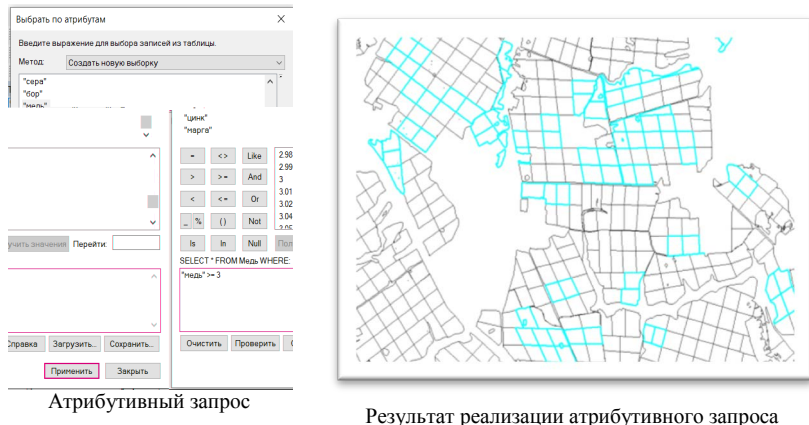


Рис. 3.9. Визуализация атрибутивного запроса

Анализ местоположения объектов помогает определить области, соответствующие определенным заданным критериям и установить причины и следствия пространственных взаимосвязей. Например, используя ГИС, торговая сеть может ежемесячно оценивать распределение количества посещений магазинов и их повторяемость в пределах города и соответственно координировать свою работу.

Чтобы исследовать закономерности в распределении геопространственных данных, необходимо отобразить объекты слоя, используя различные виды символов, а также определить способ отображения, основываясь на том, какого рода информация необходима и как будут использоваться результаты ее визуализации. Целью выполнения анализа местоположения объектов является:

- получение информации о том, где в пределах целевой территории имеются объекты определенного рода, а где их нет (например, наличие банкоматов определенного банка в пределах микрорайона города);
- учет типа объекта при анализе и выявление частоты повторений встречи объектов одного типа в том же месте (например, частота посещения аптек жителями определенного микрорайона города старше 60 лет).

Визуализация результатов анализа местоположения объектов осуществляется посредством создания цифровых картографических изображений²¹.

Карта должна соответствовать аудитории, которой она адресована, и отображать объекты и явления с необходимой детализацией. В случае отображения геопространственных объектов по типам каждому из них присваивают код, идентифицирующий его категорию, например, земли сельскохозяйственного назначения: пахотные, луговые, под многолетними насаждениями. Многие категории являются иерархическими, т. е. более общие из них разделены на частные. Например, земли населенных пунктов могут иметь код общей категории и код частной категории, указывающий на конкретный тип функциональной зоны: жилая зона, общественно-деловая зона, производственная зона, зона инженерной и транспортной инфраструктур, ландшафтно-рекреационная зона. Количество отображаемых категорий должно соотноситься с основными принципами генерализации. Например, не целесообразно отображать в пределах одного картографического изображения более чем 6–7 слоев, поскольку отображение большего числа категорий значительно затруднит восприятие.

3.2. Анализ распределения числовых показателей. Карты плотности. Поиск объектов внутри области. Анализ окружения и пространственных изменений

Геопространственные объекты, как правило, имеют числовые значения, с помощью которых возможно определить объекты, отвечающие определенным критериям, выявить места, требующие вмешательства, или же установить закономерности распределения количественных показателей объектов.

Поиск геоданных с заданным числовым интервалом означает, что необходимо указать объекты, количественные показатели которых находятся в его пределах. Например, компания, продающая косметику по каталогу, заинтересована получить почтовые адреса молодых семей с относительно высоким доходом. Или служащие органов управления системы образования хотели бы знать количество учителей на 1 тыс. человек в каждом регионе, чтобы видеть, в какие области следует распределить молодых специалистов после окончания учебы.

²¹Цифровое картографическое изображение – логико-математическое представление в цифровой форме объектов картографирования и отношений между ними.

Отображение объектов, основанное на *анализе их количественных показателей*, добавляет новую информацию к сведениям об их местоположении. Например, информация о местах расположения больших предприятий может быть полезна при планировании маршрутов общественного транспорта. Однако отображение еще и числа работников на каждом предприятии более точно отразит картину данного распределения и поможет эффективнее спланировать объемы пассажирских перевозок.

Визуализация геопространственных объектов с однотипными значениями позволяет сопоставить количественные характеристики этих объектов, а знание типа объектов, требующих отображения, поможет выбрать оптимальный способ отображения.

Возможно отображать величины, связанные с дискретными объектами, непрерывными явлениями или данными, обобщенными по области. Дискретными объектами могут быть точки на местности, линии или области. Точечные и линейные объекты при визуализации обычно разделяются с помощью шкалы символов, в то время как области обозначают заливкой или штриховкой, соответствующей их количественным показателям.

Непрерывные явления могут представлять собой области или поверхности непрерывного распределения каких-либо значений (рельеф, почвенный покров). Области условной однородности непрерывных явлений обычно отображаются различными цветами, а изменения параметров поверхности могут быть обозначены оттенками цвета, изолиниями или трехмерной перспективой.

Числовые показатели при визуализации могут быть представлены *количеством (величиной), относительными величинами (отношениями), категориями или рангами*. Знание типа числового показателя позволяет выбрать лучший способ для отображения данных.

Количество и величина являются обобщенными характеристиками объектов. *Количество* обозначает фактическое число объектов в пределах области исследования. *Величина* – обобщенное значение, характеризующее каждый объект. Использование количества или величины позволяет численно сопоставлять однотипные объекты. Возможно отображать количество и величину как дискретных объектов, (например, число учащихся в каждой школе в пределах микрорайона), так и непрерывных явлений (например, среднегодовая температура воздуха в каком-либо регионе).

Отношения отражают связь между двумя величинами и образуются путем деления одной величины, характеризующей элемент, на другую. Использование отношений позволяет сгладить резкие различия между размерами областей или количеством объектов в них так, что отображение распределения этих величин становится понятнее. Наиболее широко используемые относительные величины – *среднее, отношения и плотность*. Среднее применяется для сравнения величин, обобщенно характеризующих какое-либо место. Относительные величины показывают, какую часть целого составляет данное количество и часто выражаются в процентах. Чтобы вычислить отношение, необходимо разделить одномерные величины. Например, деление числа жителей старше 65 лет в каждом административном районе на количество всего населения района даст относительное количество людей старше 65 лет в каждом районе. Показатель плотности отражает степень концентрации объектов в различных локализациях. Чтобы вычислить плотность, необходимо разделить значение параметра, обобщенного по площади, на величину этой площади, в результате чего будет рассчитано количество на единицу площади. Например, разделив население области на ее площадь в квадратных километрах, получают количество людей на квадратный километр. Плотность хорошо отражает распределения, когда размер областей, которыми оперируют при расчетах, имеет значительные изменения. Например, маленькие и большие по площади районы могут иметь приблизительно одинаковое количество людей, но разную их плотность на единицу площади.

Ранги позволяют упорядочить объекты в интервале от высокого до низкого и отражают относительные, а не измеренные величины. Ранги используют, когда прямые измерения затруднительны или количественная характеристика зависит от целого ряда факторов. Чтобы указывать ранги, можно использовать описательные признаки (например, очень высокий, умеренный, низкий, очень низкий) или числа (например, от 10 до 100).

Поскольку ранги относительны, они отображают только место объекта по порядку, а не определяют, насколько выше или ниже других его значение. Например, можно представить, что тальвег с рангом «3» выше, чем объекты с рангами «1» и «2», но ниже, чем тальвеги с рангом «4» и «5», но при этом неизвестно, насколько выше или ниже они друг от друга (рис. 3.10).

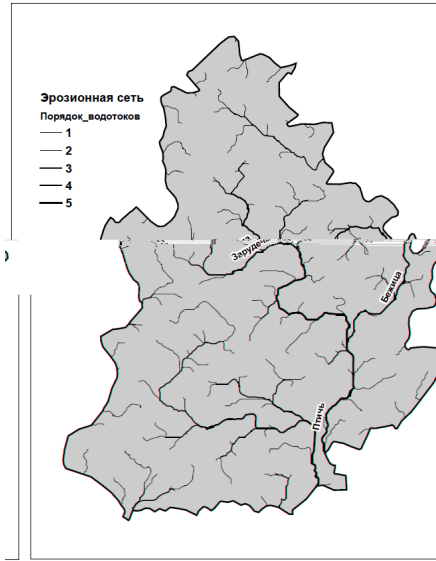


Рис. 3.10. Пример использования рангов

Ранги часто присваиваются на основании значения какого-либо атрибута объекта (типа, класса) или комбинации атрибутов. Например, на рис. 3.11 идентифицированным интервалам уровня кислотности почвы присвоены соответствующие ранги в зависимости от величины показателя, входящего в интервал.

После определения типа анализируемых геопространственных объектов необходимо выбрать *способ их визуализации*, присвоив каждому объекту *собственное значение (символ)* или *сгруппировав* значения объектов в *классы*. В процессе визуализации количественных показателей с целью выявления закономерностей распределения всегда существует проблема выбора между представлением точных значений данных или обобщением указанных значений по площади. Обычно общее число, количество и отношения группируются в классы, поскольку каждый геопространственный объект потенциально имеет различное значение. Если каждое значение при визуализации представлено уникальным символом, то оценить ситуацию можно лишь при небольшом их количестве. Если диапазон значений большой, то целесообразным является использование классов, позволяющее быстро сопоставить данные на больших территориях.

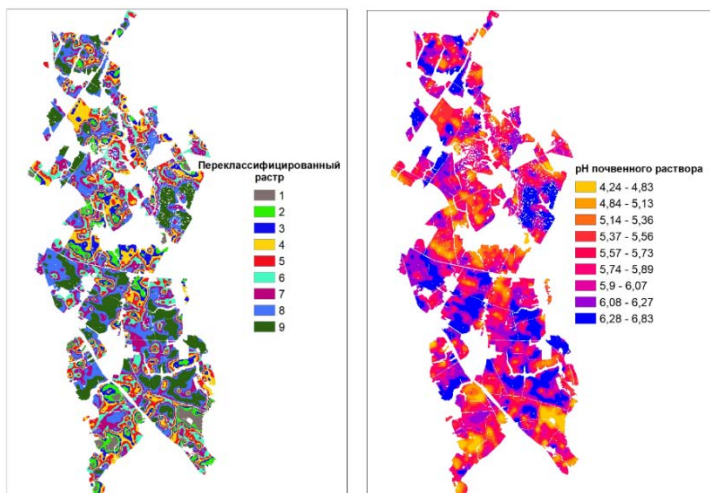


Рис. 3.11. Соответствие градаций pH почвенного раствора присвоенным рангам

Собственные значения позволяют осуществлять поиск закономерностей по непроверенным данным, когда работа с исследуемой областью осуществляется впервые или выполняется разведочный анализ данных. Их отображение необходимо также при принятии решения о группировке данных в классы. Если визуализируются ранги и их количество превышает восемь или девять, необходимо перегруппировать ранги в классы, поскольку обилие различных символов при визуализации затрудняет распознавание. Если отображается не больше 11–12 уникальных значений или количество отображаемых геопространственных объектов не превышает 20, целесообразно использовать собственные значения для отображения отношений или количественных показателей.

Классы объединяют геопространственные объекты со схожими значениями, приписывая им одинаковый символ. Их можно создавать вручную или использовать стандартную схему классификации. Необходимость выполнения *ручной классификации* возникает, когда требуется сгруппировать объекты, отвечающие специфическим критериям, или сравнить значения их атрибутов с конкретной характерной величиной. В этом случае определяется верхняя и нижняя граница интервала для каждого класса и каждому классу присваивается свой символ.

Если целью классификации данных является выявление закономерности пространственного распределения, используются *стандартные схемы разбивки данных* на интервалы. Среди наиболее широко применяемых стандартных схем различают: *естественную разбивку, квантиль, равные интервалы* и *стандартное отклонение*.

При **естественной разбивке** геопространственных данных на классы группировка значений выполняется в соответствии со скачками в значениях (рис. 3.12).

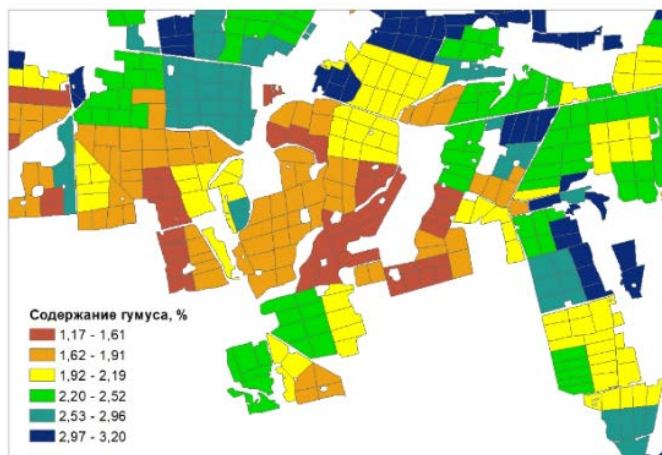


Рис. 3.12. Классификация геопространственных данных о содержании гумуса методом естественной разбивки

При разбивке по **квантилям**²² каждый класс содержит равное число геопространственных объектов, зависящее от установленного числа классов разбивки (рис. 3.13).

При использовании **равных интервалов** классификация осуществляется посредством разбивки ряда значений на интервалы, равные разности между максимальными и минимальными значениями (рис. 3.14).

²²Квантиль в геостатистике – значение, которое заданная случайная величина не превышает с фиксированной вероятностью. Если вероятность задана в процентах, то квантиль называется процентилем или перцентилем.

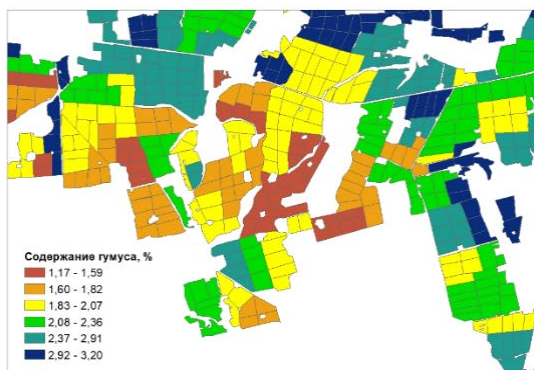


Рис. 3.13. Классификация геопространственных данных о содержании гумуса по квантилям

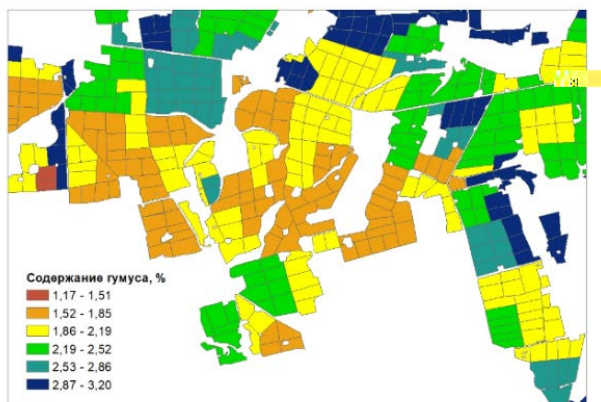


Рис. 3.14. Классификация геопространственных данных о содержании гумуса методом равных интервалов

Схема разбивки ряда геопространственных данных с использованием **стандартного отклонения** предусматривает распределение объектов на классы по величине отклонения их значений от среднего и иллюстрирует, насколько отличаются значения объектов от среднего (рис. 3.15).

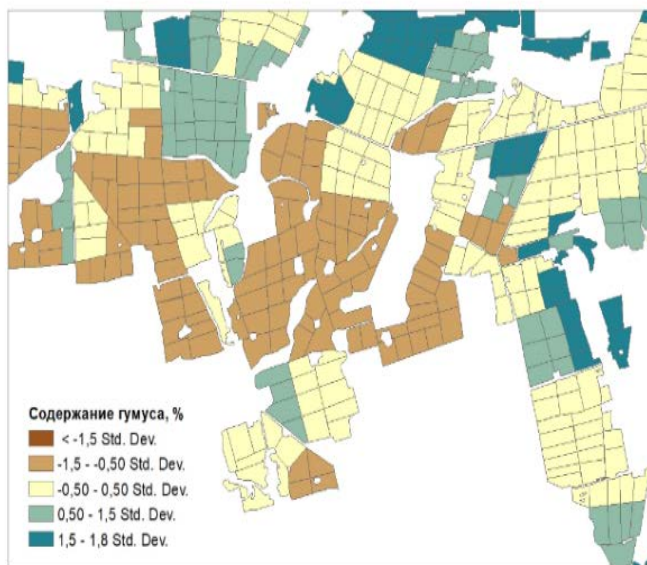


Рис. 3.15. Классификация геопространственных данных о содержании гумуса по величине стандартного отклонения

Каждый метод классификации геопространственных данных имеет как преимущества, так и недостатки и должен применяться после предварительного изучения данных с помощью методов геопространственной статистики, в частности, построения гистограммы частоты распределения данных и оценки распределения на нормальность (табл. 3.1).

Для определения оптимальной схемы классификации необходимо четко выяснить характер распределения геопространственных данных. Если данные распределены неравномерно (многие объекты имеют одинаковые или близкие значения и заметные перепады между значениями групп), следует использовать естественную разбивку. Если данные распределены равномерно и необходимо подчеркнуть разницу между объектами, целесообразно использовать равные интервалы или среднеквадратичное отклонение. Если геопространственные данные распределены равномерно, но необходимо показать относительные различия между объектами, следует использовать квантили.

Таблица 3.1. **Преимущества и недостатки различных методов классификации геопространственных данных**

Название метода	Сущность метода	Преимущества метода	Недостатки метода
1	2	3	4
Естественная разбивка	Находит группировки и закономерности, свойственные геопространственным данным, так что внутри класса попадают близкие между собой значения, а между классами устанавливается видимая разница	Эффективен при визуализации данных, имеющих неравномерное распределение, поскольку разбросанные в пределах области исследования данные с близкими значениями помещаются в один класс	Поскольку диапазоны, рассчитанные для каждого класса, характерны только для конкретной выборки, трудно сравнивать между собой классы из различных (по территориальному признаку) областей значений. Кроме того, если данные распределены равномерно, то выбор оптимального числа классов будет затруднен
Разбивка по квантилям	Геопространственные объекты упорядочиваются по принципу изменения их атрибута в интервале от максимального до минимального значения, их количество суммируется и делится на число классов, чтобы получить количество объектов в каждом классе	Наглядное сопоставление областей, размеры которых приблизительно равны; возможность эффективного отображения геопространственных объектов, значения которых равномерно распределены; возможность оценить относительное положение объекта среди его окружения	Геопространственные объекты с близкими значениями могут попасть в разные классы, особенно если значения расположены плотно, что может привести к необоснованному их разделению, и наоборот, несколько далеко расположенных смежных значений могут оказаться в одном классе, скрывая различия между объектами; если области имеют большую разницу в размерах, то их классификация квантилями может исказить реальные закономерности распределения

1	2	3	4
Равные интервалы	Каждый класс имеет равный диапазон значений, т. е. разность между максимальным и минимальным значениями одинакова для каждого класса	Равные интервалы более просты для понимания, так как диапазон для каждого класса одинаков; возможность отображения непрерывных данных, типа суммы осадков или температуры	Если сгруппированные данные распределены неравномерно, возможно скопление большого количества объектов в одном или двух классах и полное их отсутствие в других
Стандартное отклонение	Находится среднее по выборке; путем вычитания среднего из каждого значения вычисляется среднеквадратичное отклонение, которое возводится в квадрат. Полученные значения суммируются и делятся на число объектов, а из полученного выражения извлекается корень	Дает возможность судить о направлении отклонения параметра геопространственного объекта от среднего значения в большую или меньшую сторону; хорошо отображает данные, имеющие в общей массе небольшое отклонение от среднего (логнормальное или нормальное распределение)	Не показывает реальные характеристик объектов – только отклонения от среднего значения; очень высокие или низкие отдельные значения могут исказить значение среднего так, что большинство объектов окажется в одном классе

В процессе визуализации данных может оказаться, что в выборке присутствует несколько очень высоких или очень низких значений – *пространственных выбросов*. Эти аномальные значения могут исказить диапазоны классов, а следовательно, и те закономерности, которые были выявлены. Это особенно заметно при использовании схемы разбивки на равные интервалы или по среднеквадратичному отклонению, когда все значения, за исключением аномального, могут попасть в один класс. Использование естественной разбивки может изолировать аномальные значения в самом высоком или самом низком классе, но это все равно отразится на оставшихся классах.

Пространственные выбросы могут быть случайной величиной, обусловленной как несовершенством методики ее определения, так и ошибкой в базе данных. С другой стороны, наличие выбросов может

быть обусловлено и объективными факторами, например, применением различных доз минеральных удобрений в пределах отдельных рабочих участков, что неминуемо повлечет значительные различия в содержании растворимого фосфора в почве в пределах одного и того же земельного массива.

Существует несколько вариантов обращения с аномальными значениями в зависимости от того, как они влияют на остальные данные и на закономерности их пространственного распределения: 1) размещение каждого аномального значения в собственном классе (возможно, если аномальные значения существенно разбросаны); 2) группировка выбросов вместе в один класс (возможна, если группа аномальных значений локализована в одном месте); 3) группировка выбросов со следующим самым близким верхним или нижним классом, если они расположены не слишком далеко от других значений в этом классе.

Выпадающие значения важно идентифицировать, поскольку они могут быть как реальными аномалиями в явлении, так и причиной неправильного измерения значения. Если выпадающее значение является фактической аномалией в явлении, то оно может быть самой показательной точкой в исследовании и осмыслении явления. В случае если выпадающие значения вызваны ошибками во время ввода очевидно неправильных данных, они должны быть исправлены или удалены перед выполнением моделирования их пространственного распределения (рис. 3.16).

После определения оптимальной схемы классификации геопространственных данных определяют необходимое количество классов. Поскольку большинство пользователей при визуализации геопространственных данных может отличать до семи цветов, использование более семи классов делает затруднительным поиск объектов с близкими значениями. Обычно четырех или пяти классов достаточно, чтобы отразить основные закономерности распределения геоданных. Однако использование менее трех-четырех классов может привести к утере многих взаимосвязей между объектами, а следовательно, и ряда закономерностей.

Если изучаются новые данные с целью определения типа группировки объектов или выявления закономерностей их распределения, может потребоваться увеличить число классов. В этом случае диапазон каждого класса будет уже, а значения попавших в него объектов – ближе к своей реальной величине (рис. 3.17).

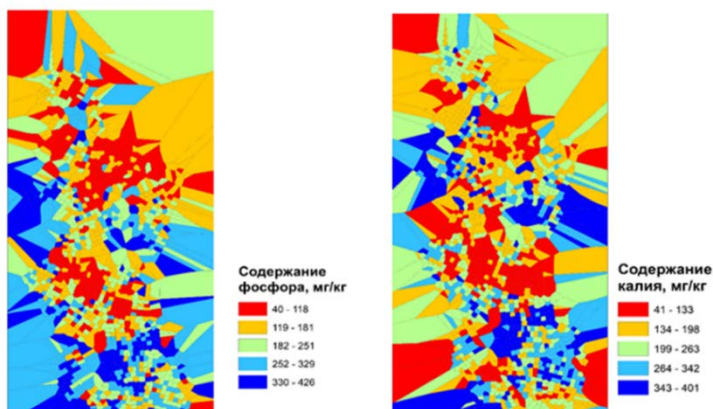


Рис. 3.16. Карты Вороного, иллюстрирующие наличие выбросов в содержании подвижных фосфора и калия в почве

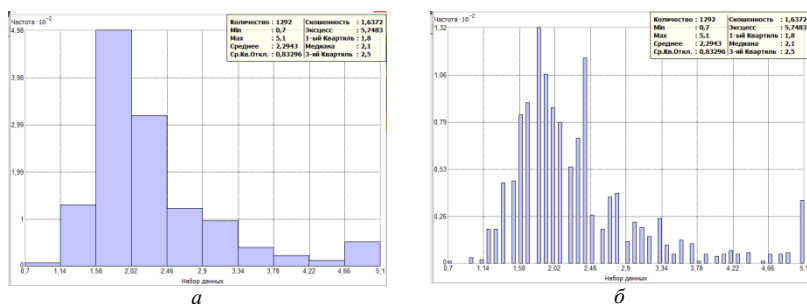


Рис. 3.17. Гистограмма содержания меди в почве при отображении 10 классов (а) и 100 классов (б) значений

Карты плотности позволяют визуально оценить концентрацию точечных либо линейных объектов в пределах какого-либо участка или всей территории. Отображение плотности дает возможность оценить изменение концентрации объектов по площади. Карты плотности довольно эффективны как при оценке характера распределения индивидуальных объектов, так и для оценки распределения объектов в пределах областей различных размеров, если, например, требуется сравнить между собой два региона с различной площадью, но с одинаковым количеством объектов, находящихся в их пределах. Создание кар-

ты плотности позволяет измерять концентрацию объектов на единицу площади и более точно сопоставлять области или определять, соответствуют ли они заданным критериям. Карту плотности можно создать как посредством построения плотности дискретных объектов, так и посредством создания поверхности плотности (табл. 3.2).

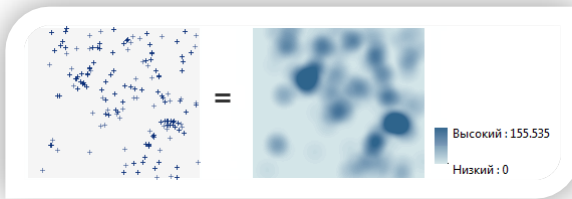
Таблица 3.2. Методы создания карт плотности распределения объектов

Название метода	Условие использования метода	Вид представления результата	Особенности применения
Карта плотности дискретных объектов	Имеются данные, агрегированные по площади, либо линии и точки, которые можно суммировать по области	Карта плотности, выполненная в виде залитых цветом областей или полей точек	Простой метод, однако не показывает точное местоположение конкретных значений плотности, особенно для больших областей; может требовать предварительной обработки данных
Поверхность плотности	Имеются точки, обозначающие места расположения отдельных объектов, точки сети наблюдений или линии	Оттененная поверхность плотности или изолинии	Дает более точное представление о значении плотности в любом месте области исследования, но требует больших затрат времени на обработку данных

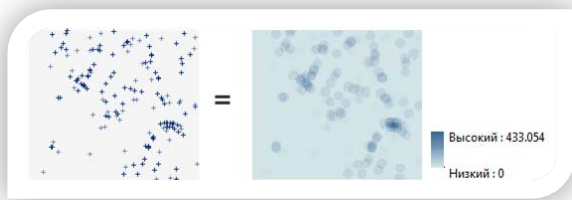
В ГИС возможно вычисление величины и построение карт плотности ядер, точек и линий (рис. 3.18).

При определении величины и построении *карты плотности ядер* с использованием функции ядра для подгонки плавно сужающейся поверхности к каждой точке или полилинии вычисляется количество точечных или полилинейных объектов на единицу площади. Примером использования может служить определение плотности инженерных коммуникаций на территории города.

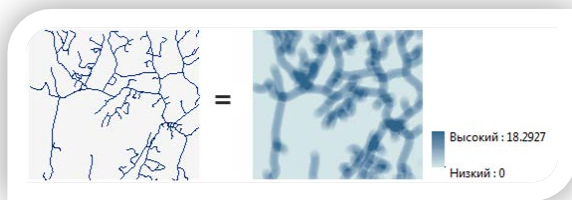
При определении *плотности точек* вычисляется количество точечных объектов на единицу площади, попадающих в окрестность вокруг каждой ячейки растра. Концептуально вокруг центра каждой ячейки растра определяется окрестность, после чего количество точек, попадающих в окрестность, подсчитывается и делится на площадь окрестности. Примером использования может служить определение плотности застройки на территории города.



a



б



в

Рис. 3.18. Вычисление плотности с использованием возможностей ГИС:
a – плотность ядер; *б* – плотность точек; *в* – плотность линий

При определении *плотности линий* вычисляется плотность линейных объектов в окрестности каждой ячейки выходного растра. Плотность вычисляется в единицах длины на единицу площади. Примером использования может служить определение плотности эрозионной сети на определенной территории.

Размер ячейки при построении поверхности плотности определяет, насколько грубо или детально будут проявляться закономерности в пределах исследуемой территории. Поскольку ячейки представляют собой квадраты, их размер определяется длиной любой из сторон. Размер ячейки подбирается таким образом, чтобы получить от 10 до 100 ячеек на единицу площади. Например, если вычисляется количество жителей на квадратный километр, а размер ячейки дается в метрах, тогда он должен находиться в пределах от 100 до 300 м.

Определяя размер ячейки, сначала приводят единицы измерения площади (квадратные километры) к единицам измерения ячейки (метры), затем, чтобы получить площадь ячейки, делят площадь области на планируемое на единицу площади количество ячеек. Поскольку размер ячейки определен длиной одной стороны квадрата, чтобы получить его величину, следует извлечь квадратный корень из площади ячейки. На рис. 3.19 представлены карты плотности промышленных объектов и капитальных сооружений на территории Республики Беларусь, построенные с использованием плотности точек (*а*) и плотности ядер (*б*).

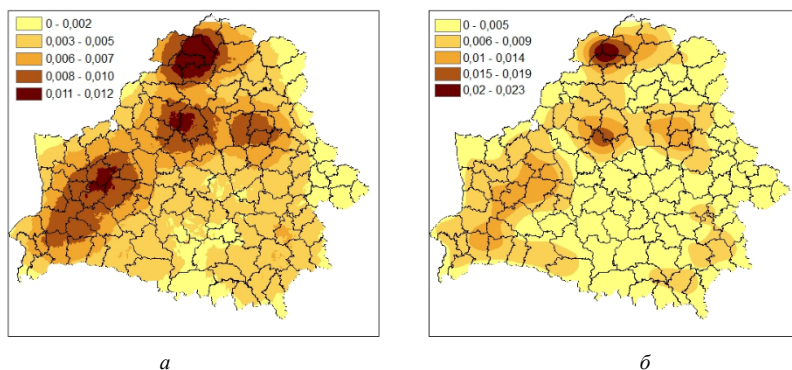


Рис. 3.19. Карты плотности промышленных объектов и капитальных сооружений на территории Республики Беларусь

Поиск объектов внутри области позволяет исследовать, что происходит в ее пределах, или обобщить информацию по каждой из исследуемых областей с целью их сопоставления между собой либо с какой-либо референтной областью. Контролируя пространство в пределах области, возможно оценить адекватность принимаемых мер. Например, определив и визуализировав область возможного распро-

странения радиоактивного загрязнения в случае аварии на атомной электростанции, можно оптимально организовать эвакуацию населения из зоны заражения.

Обобщение объектов по признаку принадлежности к какой-либо области позволяет сравнивать территории по количественным критериям. Например, визуализация распределения древостоев различной спелости в разрезе отдельных выделов позволит более эффективно планировать деятельность лесохозяйственного предприятия.

Существует три *метода поиска объектов внутри области* (табл. 3.3):

- 1) визуальное сопоставление границ области и объектов;
- 2) построение пространственного запроса;
- 3) поиск методом наложения слоев.

Таблица 3.3. Методы поиска объектов внутри области

Название метода	Условие использования метода	Типы объектов	Особенности применения
Визуальное сопоставление границ области и объектов	Визуальная оценка объектов, попадающих в определенную область	Точки, линии, полигоны, поверхности	Быстрый и простой метод, однако не позволяет оценивать атрибуты объектов, расположенных внутри области
Построение пространственного запроса	Получение перечня или суммы объектов внутри области	Точки, линии, полигоны	Не позволяет проанализировать несколько различных областей
Метод наложения слоев	Количественная оценка полного или частичного попадания объектов в пределы различных областей	Точки, линии, полигоны, поверхности	Эффективен для анализа как одной, так и нескольких областей, однако требует больших затрат времени для реализации

При *выборе метода поиска* следует руководствоваться следующим:

– если необходимо определить, какие объекты попали в пределы одной области, используют визуальное сопоставление границ области и объектов;

– при необходимости получить перечень или сумму объектов, полностью или частично попадающих в границы области, используют возможности пространственного запроса данных;

– наложение областей и объектов применяют в случаях, если: 1) имеется совокупность областей и необходимо проанализировать каждую; 2) имеется одна область, но в задачу входит получение перечня или обобщенного значения объектов, попадающих только в определенную ее часть; 3) имеется одна область, но речь идет об обработке непрерывных, а не дискретных величин.

Анализ окружения позволяет в пределах заданного расстояния оценить территорию, прилегающую к объекту, либо оценить пространственные изменения самого объекта. Он дает возможность контролировать события, происходящие в заданном окружении, определять территорию воздействия или оценивать его последствия.

При определении того, что находится на определенном от объекта расстоянии, идентифицируется область, на которую оказало влияние произошедшее событие или проводимая деятельность, а также содержащиеся в ней объекты. Например, при открытии нового супермаркета, возможно, потребуется оповестить об этом всех живущих в радиусе 500 м, поскольку именно они являются потенциальными клиентами торгового объекта. Оценка окружающей территории на заданном расстоянии также позволяет контролировать деятельность в заданном районе. Например, работник землеустроительной организации может контролировать деятельность строительной фирмы, не допуская постройку зданий и сооружений в водоохранной зоне.

Диапазон пространственных изменений объекта измеряется с помощью таких понятий, как расстояние, время или затраты. Определение того, что попадает в диапазон перемещений или изменения объекта, помогает определить зону воздействия. Например, органы государственного управления, осуществляющие регулирование и управление в сфере обеспечения пожарной безопасности, могут получить информацию о том, какие улицы в пределах населенного пункта находятся в трех минутах езды от той либо иной пожарной станции; аналитик рынка оптовых продаж может определить, сколько людей живет в пятнадцати минутах езды от предлагаемого места нового рынка оптовой торговли сельскохозяйственной продукцией.

Знание того, что находится в зоне перемещения или изменения объекта, дает возможность отобразить области, по различным причинам удобные или пригодные для специфического использования. Например, биолог, изучающий дикую природу, отображает на карте область, расположенную в 500 м от реки, а затем соотносит ее с типом вегетации, уклонами, уровнем залегания грунтовых вод, чтобы обозначить предполагаемый ареал водно-болотного угодья.

Различают три основных *метода поиска и оценки окружения* (табл. 3.4):

- 1) измерение кратчайшего расстояния;
- 2) измерение расстояния или затрат по сети;
- 3) расчет затрат по поверхности.

Таблица 3.4. Методы поиска и оценки окружения

Название метода	Условие использования метода	Объекты окружения	Результаты определения	Преимущества метода	Недостатки метода
Измерение кратчайшего расстояния	Определение области влияния вокруг объекта, создание границ области влияния либо выбор объектов в ее пределах	Точки, линии, полигоны	Дистанция	Относительно быстрый и простой	Дает только приблизительное представление о фактической длине пути
Измерение расстояния или затрат по сети	Измерение длины пути с учетом фиксированной инфраструктуры	Точки, линии	Длина пути или затраты на передвижение	Дает более точное представление о фактической длине пути	Требует наличия слоя сети точных маршрутов передвижения
Расчет затрат по поверхности	Оценка затрат на передвижение по пересеченной местности и диапазон пространственных перемещений объектов	Сплошной растр поверхности	Затраты на передвижение	Позволяет комбинировать несколько слоев для оценки затрат на передвижение	Требует определенной подготовки данных для построения поверхности затрат

При выборе метода оценки следует руководствоваться следующими принципами:

– использовать измерение кратчайшего расстояния, если необходимо определить область влияния или быстро оценить диапазон перемещения объекта;

– использовать оценку удаленности или затрат по сети, если требуется измерить длину пути по фиксированной инфраструктуре к объекту или от него;

– использовать подсчет затрат по поверхности, если оценивается передвижение по пересеченной местности.

Возможно также осуществить поиск расстояния между объектами окружения и различными источниками в следующих случаях:

– если необходимо видеть, какие объекты расположены поблизости от нескольких источников, а какие – только от одного. Например, менеджеры сети ресторанов быстрого обслуживания могут определить, в каких районах города достаточно много клиентов находятся рядом с несколькими ресторанами;

– если, зная ближайший источник, необходимо определить, какой источник является вторым или третьим по удаленности от каждого объекта. Например, можно узнать, где находятся ближайшие или вторые по удаленности от определенного места медицинские учреждения;

– при необходимости сопоставить фактор удаленности с другими факторами. Например, можно сравнить расстояние от места жительства каждого покупателя до магазина с числом его визитов в магазин, а затем обобщить эту информацию и определить зависимость количества визитов в магазин от расстояния до его места расположения.

Анализ пространственных изменений позволяет визуально оценить происходящие изменения, глубже понять причины, их вызывающие, и оценить последствия их воздействия. Изменения в геопрограммных данных визуализируются для того, чтобы спрогнозировать их масштаб и направление, принять оптимальные решения о мерах воздействия или оценить результаты практической деятельности. Визуализируя положение движущихся объектов в определенные моменты времени, можно более глубоко вникнуть в причины их поведения. Например, метеоролог может изучать маршруты ураганов и смерчей для прогнозирования их возникновения в будущем. Биолог, изучающий диких животных, может исследовать передвижения зубров за 24-часовой период для выяснения особенностей их питания и определения размеров территории, минимально необходимой для поддержания устойчивой популяции. Еще одна возможность, реализуемая с помощью пространственных изменений, – прогнозирование будущих потребностей в ресурсах. Например, начальник транспортной службы может изучать, каким образом из месяца в месяц изменяется количество жителей в различных районах города, чтобы планировать маршруты общественного транспорта. Отражая на карте условия до и после воздействия или происшествия, можно оценить эффективность принятых мер. Аналитик рынка розничной торговли может отслеживать из-

менения в объеме продаж до и после проведения рекламной кампании нового продукта, оценивая эффективность рекламы.

Геопространственные объекты могут изменять местоположение, геометрию или свойства. Отслеживать эти изменения возможно посредством идентификации местоположения и состояния объектов в каждый момент времени. Также можно рассчитать и отразить различия, произошедшие с каждым объектом за установленный период.

Различают три основных *метода анализа пространственных изменений* (табл. 3.5):

- 1) использование временных рядов;
- 2) построение карт слежения;
- 3) количественная оценка изменений.

Таблица 3.5. **Методы анализа пространственных изменений**

Название метода	Тип изменения	Модель временного ряда	Преимущества метода	Недостатки метода
Временные ряды	Движение или изменение свойств	Тенденция, Цикл «До» и «после»	Сильный визуальный эффект при условии значительных изменений; отражение состояния объекта в каждый момент	Для оценки изменений доступно только визуальное восприятие
Карты слежения	Движение	Тенденция, Цикл «До» и «после»	Хорошо визуализирует движение и слабые изменения	Труден для восприятия при большом количестве объектов
Количественная оценка изменений	Изменение свойств	Тенденция, Цикл «До» и «после»	Показывает действительную разницу в количестве или значениях	Не отражает действительных условий в каждый момент времени, а проводит сопоставление только между двумя датами

При *выборе метода анализа* следует руководствоваться следующим:

– если необходимо показать движение или изменение свойств объекта в течение двух или более временных отрезков области, используют временные ряды;

– если необходимо показать движения объекта между двумя или более моментами времени или отразить периодичность изменения, используют карты слежения;

– если необходимо показать численные различия в параметрах объекта между двумя моментами времени, используют количественную оценку изменений.

3.3. Основные виды векторного ГИС-анализа

Векторный ГИС-анализ используется при обработке цифровых векторных слоев с учетом атрибутов геопространственных объектов. Его применение наиболее целесообразно при работе с дискретными геообъектами, имеющими четкие границы в геопространстве. В векторном анализе данных используются такие *геометрические объекты*, как точки, линии и полигоны. Топология также может участвовать в качестве объекта анализа некоторых векторных данных (операции буферизации и наложения). Точность результатов векторного ГИС-анализа зависит от точности его объектов с точки зрения расположения и формы. Основными *видами векторного ГИС-анализа* являются: 1) элементарный пространственный анализ; 2) пространственная статистика; 3) расширенный пространственный анализ; 4) сетевой анализ.

Элементарный пространственный анализ предусматривает просмотр векторных геообъектов в среде ГИС, анализ атрибутивных данных пространственных объектов, картометрические измерения и визуализацию атрибутов посредством создания картографических изображений. Преимуществами *просмотра векторных геообъектов* в ГИС являются удобная навигация; возможность осуществления выборки и идентификации геообъектов по атрибуту; одновременный просмотр нескольких слоев различных геоданных. Основные операции *анализа атрибутивных данных* пространственных объектов включают выполнение поиска объекта и определение его местоположения по определенному атрибуту (рис. 3.20); осуществление выбора объектов по заданному атрибуту (рис. 3.21); сортировку объектов по их атрибутам и переключение выборки.

Осуществление выбора объектов по заданному атрибуту позволяет задать SQL-выражение запроса, которое используется для выбора объектов, удовлетворяющих критерию выборки. На рис. 3.21 в качестве атрибута использована площадь полигона, в качестве критерия выборки – полигоны с площадью большей или равной 50 га.

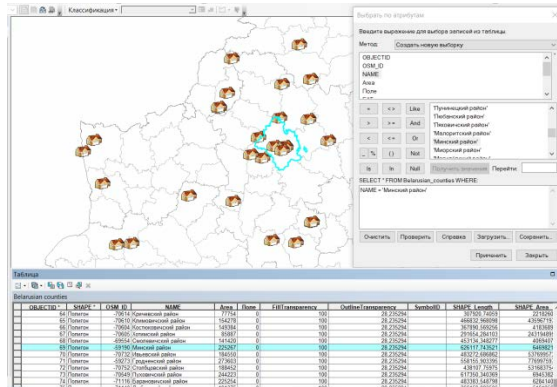


Рис. 3.20. Поиск объекта и определение его местоположения по определенному атрибуту (в качестве запроса задан поиск производителей органической продукции в Минском районе (атрибут))

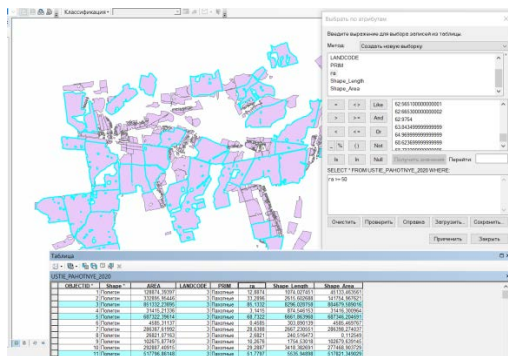


Рис. 3.21. Выбор объектов по заданному атрибуту

Сортировка объектов по их атрибутам и переключение выборки позволяет в первом случае отсортировать объекты в порядке убывания либо возрастания их атрибутивных значений, во втором – отобразить в атрибутивной таблице только выбранные по заданному атрибуту либо невыбранные записи. К *картометрическим измерениям* относят: определение расстояний между векторными геопространственными объектами; определение координат точечных объектов; определение длин линейных объектов; определение площадей и периметров полигональных объектов. На основе текстовых и количественных атрибутов геопространственных данных в ГИС возможно составление разно-

образных тематических слоев с помощью *картографических способов изображений*²³.

Среди **операций пространственной статистики** в ГИС выделяют: описательную статистику выборки геоданных, операции с базами атрибутивной информации, а также расширенные операции геостатистики²⁴. *Статистическая обработка атрибутов* позволяет выполнять расчет значений описательной статистики выборки (максимум, минимум, среднее, среднеквадратическое отклонение) массива числовых атрибутивных данных векторных геопространственных объектов (рис. 3.22).

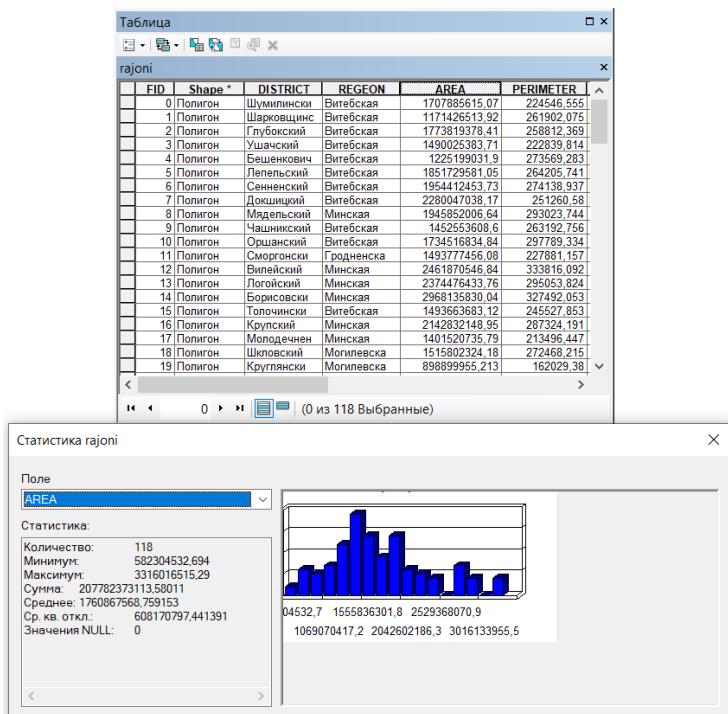


Рис. 3.22. Результат расчета значений описательной статистики выборки данных

²³Картографические способы изображения – системы условных обозначений, применяемые при создании карт, для показа пространственного размещения объектов, явлений, процессов, их сочетаний, связей и развития.

²⁴Более подробная информация приведена в главе 5.

При работе с базами геоданных, представленных различными слоями с отдельными атрибутивными таблицами, для связывания атрибутивных данных из разных слоев используют такие операции, как соединение по атрибуту и пространственное связывание таблиц. При соединении двух таблиц атрибуты из одной таблицы присоединяются к атрибутам другой на основании поля, общего для обеих таблиц, называемого ключевым. Пространственное связывание таблиц также описывает отношение между двумя таблицами на основании общего поля, но при этом добавление атрибутов одной таблицы в другую не производится. Соединение таблиц по атрибуту осуществляется посредством следующих типов отношений: 1) отношение «один-к-одному», когда одной записи в основной таблице соответствует одна запись в присоединяемой атрибутивной таблице; 2) отношение «много-к-одному», когда несколько записей основной таблицы атрибутов слоя соединяются с одной записью присоединяемой таблицы атрибутов слоя; 3) отношение «один-к-многим», когда одна запись основной таблицы атрибутов слоя соединяется с несколькими записями присоединяемой атрибутивной таблицы; 4) отношение «многие-к-многим», когда нескольким записям основной таблицы соответствуют несколько записей присоединяемой.

Пространственное связывание используется, если слои не имеют общего атрибутивного поля, и выполняется посредством связывания атрибутов двух слоев на основе местоположения объектов слоев. С его помощью возможно находить ближайшие объекты по отношению к другим объектам; идентифицировать, что расположено внутри объекта или что пересекает объект; вычислить, сколько точек находится внутри каждого полигона. Пространственное связывание отличается от соединения по атрибуту и классу отношений тем, что оно не динамично и требует сохранения результатов в новый выходной слой. При его осуществлении используют один из трех *типов ассоциаций*:

1) сопоставление каждого объекта с ближайшим объектом или объектами: в этой ассоциации можно либо добавить атрибуты ближайшего объекта, либо добавить множество численных атрибутов ближайших объектов;

2) сопоставление каждого объекта с объектом внутри: в этом случае добавляются атрибуты объекта, находящегося внутри текущего объекта;

3) сопоставление каждого объекта с пересекающим объектом или объектами: в данной ассоциации можно либо добавить атрибуты от-

дельного пересекающего объекта, либо множество численных атрибутов пересекающих объектов.

Расширенные операции геопространственной статистики предусматривают измерение пространственного распределения объектов, кластерный и регрессионный анализ²⁵.

Расширенный пространственный анализ предусматривает выполнение оверлейных операций, анализ близости, переклассификацию геопространственных данных²⁶, их генерализацию²⁷ и геообработку.

Генерализация в ГИС имеет два аспекта: генерализация информации в базе данных и геометрическая генерализация. Генерализация информации в базе геоданных может быть представлена как построение масштабного ряда для всех объектов карты. Геометрическая генерализация позволяет упрощать и улучшать векторные пространственные объекты с целью их отображения в более мелких масштабах (табл. 3.6).

Таблица 3.6. Характеристика операций геометрической генерализации

Название операции генерализации	Описание операции генерализации
1	2
Агрегирование точек	Создает полигональные объекты вокруг групп соседних точечных объектов
Агрегирование полигонов	Объединяет полигоны, находящиеся в пределах указанного расстояния друг от друга, в новые полигоны
Совмещение двойных линий в центральную линию	Вычисляет центральные линии на основе объектов, содержащих двойные линии, например, контуров проезжей части дорог, используя указанный допуск ширины
Сворачивание детализации дорог	Позволяет свернуть подробные конфигурации сегментов дороги, разрывающие общую линию дорожной сети (например, кольцевые развязки), заменяя их упрощенным обозначением

²⁵Более подробная информация содержится в главе 5.

²⁶Более подробная информация содержится в главе 4.

²⁷Термин «генерализация» происходит от французского слова *generalisation* – обобщение, которое, в свою очередь, является производным от латинского слова *generalis*, что значит «общий», «главный».

1	2
Отделение застроенных территорий	Создает отображение застроенных территорий путем отделения плотных скоплений зданий на мелкомасштабных картах
Создание картографических частей	Разбивает общий класс входных пространственных объектов на ячейки в виде полигональных объектов, каждый из которых включает в себя ограниченное количество входных объектов в зависимости от плотности и рассредоточения входных объектов
Слияние фрагментов дорог	Создает один линейный объект дороги вместо совпадающей пары разделенных дорожных полос
Упрощение здания	Упрощает границы полигонов зданий, сохраняя при этом их форму и размеры
Упрощение линии	Упрощает линию посредством удаления ее лишних изгибов с сохранением основной формы
Упрощение полигонов	Упрощает полигоны посредством удаления лишних изгибов с сохранением основной формы
Сглаживание линии	Улучшает визуальное и картографическое качество линий посредством сглаживания их углов
Сглаживание полигонов	Улучшает визуальное и картографическое качество линий границ полигонов посредством сглаживания их углов
Прореживание дорожной сети	Создает упрощенную дорожную сеть, которая сохраняет связность и общие особенности при отображении в меньших масштабах

Геообработка – это среда и набор инструментов для обработки геопространственной информации, позволяющие создавать новую информацию посредством выполнения операций над уже существующими данными.

Любое изменение или извлечение информации, которое выполняется при работе с геоданными, подразумевает решение задач геообработки. Геообработка включает в себя большое количество инструментов, дающих возможность выполнять пространственный анализ или управлять ГИС-данными в автоматическом режиме. Одни задачи, которые необходимо автоматизировать, могут быть обычными, например, преобразование группы данных из одного формата в другой; перепроецирование нескольких шейп-файлов. Решение других задач требует автоматизации последовательности выполнения операций моделирования и анализа комплексных пространственных отношений,

например, расчет оптимальных маршрутов для пассажирского либо грузового автотранспорта; прогнозирование распространения болезней и вредителей лесных насаждений; определение территорий, подверженных оползням; прогноз величины зоны затопления вследствие наводнения. Инструменты геообработки упрощают процесс обработки пространственных данных для моделирования различных явлений реального мира. Существует четыре типа инструментов: встроенные; модель; скрипт и специальный инструмент. Один из простейших способов авторизации и автоматизации технологического процесса и отслеживания задач геообработки – создание модели. Модель состоит как минимум из одного процесса или в большинстве случаев из нескольких последовательно связанных процессов. Представленная на рис. 3.23 геоинформационная модель структурно состоит из трех исходных параметров, 21 процесса и 23 результирующих растровых, векторных слоев и слоев в виде таблиц. Использование данной модели позволяет ускорить процесс анализа более чем в 30 раз по сравнению с процессом анализа, выполняемого посредством последовательного применения отдельных инструментов.

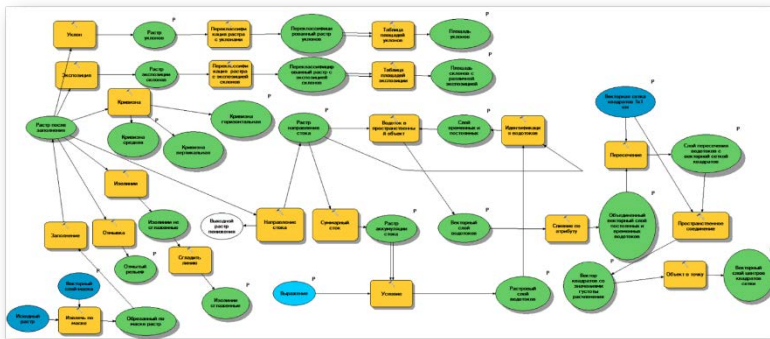


Рис. 3.23. Геоинформационная модель для автоматизации морфометрического анализа, извлечения системы водотоков и определения величины горизонтального расчленения рельефа

Различные инструменты являются основными строительными блоками рабочего процесса в модели. С их помощью выполняются операции с географическими или табличными данными. Когда инструменты добавляются в модель, они становятся элементами модели. Диалоговое окно инструмента – это интерфейс, в котором задаются параметры

инструмента. *Параметры инструмента* – это входные и выходные значения инструмента, а также спецификации, которые влияют на обработку инструмента. В каждом параметре есть элемент управления интерфейса пользователя, в котором вводится значение или путь. Некоторые из параметров являются обязательными, другие – опциональными. *Параметры модели* – это параметры, отображающиеся в диалоговом окне инструмента модели. Любую переменную модели можно преобразовать в параметр модели.

Сетевой анализ является одним из наиболее современных и востребованных направлений ГИС-анализа, основанным на теории графов²⁸ и топологии. Существующие или вновь проектируемые жилые дома, капитальные сооружения, кварталы, объекты инфраструктуры рассматриваются как *вершины*, а соединяющие их дороги, инженерные и коммуникационные сети, линии электропередачи – как *ребра*. Любая сеть, или граф²⁹, состоит из набора связанных вершин и ребер, а теория графов описывает, измеряет и сравнивает графы, или сети. Различают *неориентированные* и *ориентированные* графы. *Неориентированный граф* – абстрактный математический объект, представляющий собой множество вершин и набор ребер, т. е. соединений между парами вершин. Например, за множество вершин можно взять остановки общественного транспорта на маршруте, обслуживаемом некоторой транспортной компанией, а за множество ребер – регулярные рейсы этой транспортной компании в черте города. *Ориентированный граф* – граф, ребрам которого присвоено направление (направленные ребра часто именуются дугами).

Применение различных вычислений, производимых на графе, позволяет решить целый ряд задач, например:

- найти кратчайший объездной путь в случае возникновения препятствия на маршруте;
- спланировать оптимальный маршрут доставки товаров с учетом загруженности транспортной сети;
- оценить рациональность распределения сети торговых объектов;
- определить зону обслуживания (доступности) для торгового объекта либо объекта транспортной инфраструктуры, медицинского учреждения, отделения связи (рис. 3.24).

²⁸ Теория графов – раздел дискретной математики, изучающий свойства графов (граф представляется как множество вершин (узлов), соединенных ребрами).

²⁹ Термин «граф» впервые ввел английский математик Джеймс Джозеф Сильвестр в 1878 г.

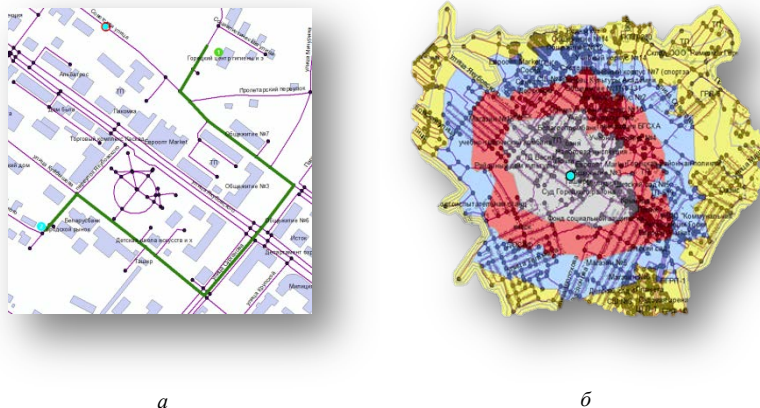


Рис. 3.24. Результат выполнения сетевого анализа:
 а – построение кратчайшего пути при возникновении препятствия на маршруте;
 б – определение зон обслуживания торгового объекта с различной удаленности

3.4. Основные виды растрового ГИС-анализа

Растровый анализ – это произведение вычислительных операций с растровыми данными с целью извлечения из них дополнительной информации. Наиболее часто растровый анализ применяется при работе с непрерывными геопространственными процессами или явлениями³⁰. Данный вид анализа осуществляется в следующих основных направлениях:

- интерполяция растра;
- анализ гипсометрических поверхностей;
- анализ плотности и расстояний;
- использование функций картографической растровой алгебры.

Интерполяция³¹ – это процесс прогнозирования значений для точек, где нет измеренных значений, по ограниченному набору опорных

³⁰ Более подробная информация содержится в главе 2.

³¹ Интерполяция, интерполирование (от лат. *inter-polis* – разглаженный, подновленный, обновленный; преобразованный) – нахождение неизвестных промежуточных значений некоторой функции по имеющемуся дискретному набору ее известных значений определенным способом. Термин «интерполяция» впервые употребил английский математик Джон Валлис (1616–1703) в своем трактате «Арифметика бесконечных» (1656).

точек с известными значениями. Ее фундаментальной теоретической основой является первый закон географии Уолдо Тоблера³², согласно которому все связано со всем остальным, но объекты, находящиеся рядом, более связаны между собой, чем отдаленные друг от друга. Иными словами, значения точек, расположенных ближе к точке, в которой произведены эмпирические наблюдения, будут более схожими, нежели значения более отдаленных от нее точек.

Интерполяция используется для создания поверхности по ограниченному числу значений какого-либо параметра. Каждый объект слоя точек – это место (пункт), где проводилось измерение процесса или явления. С помощью интерполяции рассчитываются значения между точками, в которых не производилось измерений (рис. 3.25).

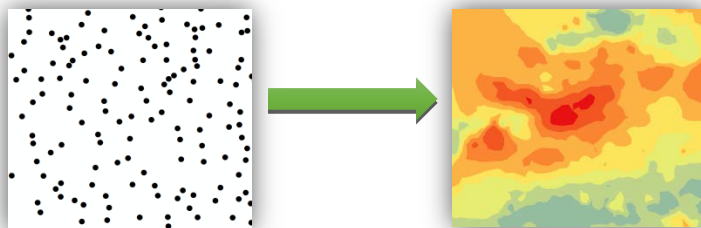


Рис. 3.25. Результат выполнения интерполяции

Ее можно использовать для вычисления неизвестных значений любых геопространственных точечных данных: высоты над уровнем моря, уровня осадков, температуры воздуха, концентрации химических веществ, содержания элементов питания в почве; плотности населения в некоторой географической области, мощности геологической формации. Инструменты интерполяции делятся на *детерминированные* и *геостатистические* (рис. 3.26).

³²Уолдо Рудольф Тоблер (1930–2018) – американский географ и картограф швейцарского происхождения.

ДЕТЕРМИНИРОВАННЫЕ создают поверхности из измеренных значений

Метод обратных взвешенных расстояний
(Inverse Distance Weighted)

Интерполяция по методу глобального
полинома (Global Polynomial Interpolation)

Интерполяция по методу локального
полинома (Local Polynomial Interpolation)

Радиальные базисные функции (Radial Basis
Functions)

Интерполяция с барьерами (использующая
в процессе интерполяции проницаемые или
полупроницаемые барьеры)

- - ядро диффузии (Diffusion kernel)
- - сглаживание ядра (Kernel smoothing)

ГЕОСТАТИСТИЧЕСКИЕ используют для построения поверхности статистические свойства измеренных значений

Ординарный кригинг (Ordinary Kriging)

Простой кригинг (Simple Kriging)

Универсальный кригинг (Universal Kriging)

Индикаторный кригинг (Indicator Kriging);

Вероятностный кригинг (Probability Kriging);

Дизъюнктивный кригинг (Disjunctive Kriging);

Площадная интерполяция (Areal
interpolation)

Эмпирический байесовский кригинг
(Empirical Bayesian Kriging)

Рис. 3.26. Методы интерполяции, используемые для моделирования пространственного распределения данных

Детерминированные методы интерполяции предполагают наличие заданной аналитической зависимости между значениями функции в пространстве. При использовании детерминированных методов предполагается, что анализируемые данные описываются некоторой детерминированной функцией $Z(x, \lambda)$, определенной на исследуемой области S , где $x \in S$ – координаты точки; λ – набор внутренних параметров геопространственной модели. При выполнении интерполяции необходимо, базируясь на известных данных ($Z_i = Z(x_i)$ – значения, измеренные в точках $x_i \in S$) и на другой контекстной информации об исследуемом явлении либо процессе, подобрать набор параметров λ и построить функцию $Z(x, \lambda)$ для всей исследуемой области S . После этого значение в любой точке, в которой отсутствуют данные эмпирических наблюдений, просто вычисляется по формуле.

В зависимости от количества используемых опорных точек (точек с известными значениями) выделяют:

– *глобальные детерминированные методы* – вычисляют искомые значения с использованием всего набора опорных точек (интерполяция по методу глобального полинома);

– *локальные детерминированные методы* – вычисляют неизвестные значения по опорным точкам, расположенным в окрестностях ис-

комой точки (интерполяция по методу обратных взвешенных расстояний (ОВР) и по методу локального полинома (сплайн)).

В зависимости от того, проходит или не проходит поверхность, построенная с помощью детерминированных методов интерполяции, через опорные точки, выделяют:

– *жесткие интерполяторы* – дают в опорной точке значение, равное измеренному (методы ОВР и сплайн);

– *нежесткие интерполяторы* – дают в опорной точке значение, отличное от измеренного, аппроксимируя значение в опорной точке (интерполяция по методу глобального полинома).

Достоинством детерминированных методов интерполяции является простота вычисления оценки по заданной параметрической формуле, а недостатком – отсутствие возможности характеризовать качество оценки, поскольку настройка параметров модели не предполагается или производится скрыто, исключая участие исследователя в данном процессе.

Геостатистические методы интерполяции³³ основываются на статистических моделях, предусматривающих анализ пространственной автокорреляции (статистических отношений между измеренными в точках значениями). Результатом такого подхода является то, что геостатистические методы не только позволяют создавать поверхности прогнозируемых значений, но и предоставляют оценку достоверности (точности) прогнозирования. Базовой геостатистической интерполяционной моделью является **кригинг**. Термин «кригинг» служит для обозначения семейства алгоритмов линейной пространственной регрессии. Данный интерполяционный метод назван в честь южноафриканского горного инженера Дэниела Крига (1919–2013), занимавшегося созданием геологических карт по ограниченному набору данных в некоторой области для анализа золотых месторождений Южной Африки.

Среди методов анализа гипсометрических поверхностей в ГИС наиболее широко используются *построение изолиний* и *расчет морфометрических характеристик рельефа*. *Изолинии* – это линии, которые соединяют местоположения с равным значением в наборе растровых данных, представляющем непрерывное явление, например, рельеф. Распределение линий изолинии показывает, как значения меняются по поверхности. По созданной методами интерполяции поверхности (например, цифровой модели рельефа) можно сформировать векторный слой изолиний (рис. 3.27).

³³ Более подробная информация содержится в главе 6.

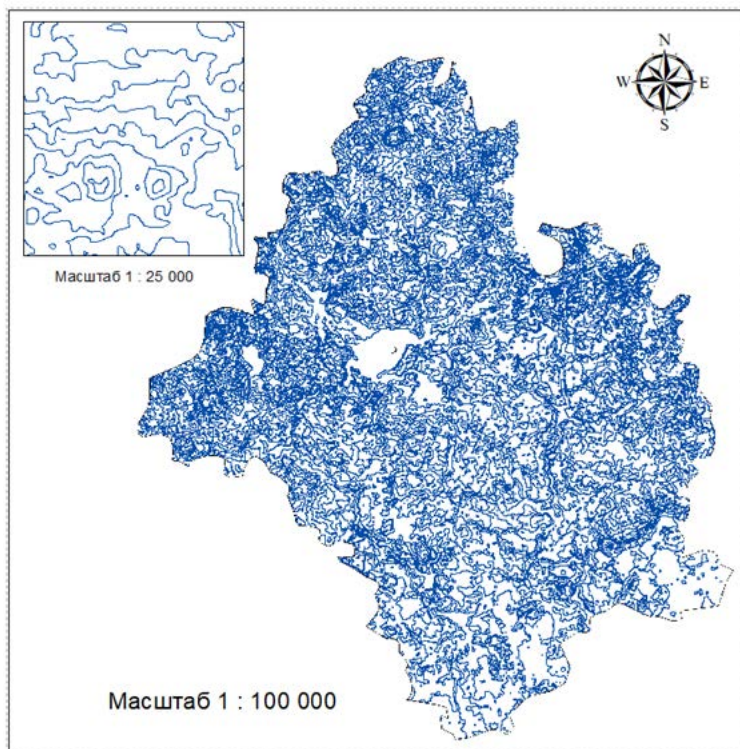


Рис. 3.27. Результат извлечения изолиний из цифровой модели рельефа Минского района

Морфометрический анализ цифровой модели рельефа выполняется с целью получения информации о крутизне и экспозиции склонов в пределах исследуемой территории, кривизне поверхности, густоте эрозионной. Он выполняется с учетом трех аспектов: линейного, площадного и рельефного (табл. 3.7).

Результаты выполнения комплексного морфометрического анализа могут использоваться при проведении оперативной и эффективной оценки критических свойств рельефа для целей прогнозирования использования земель и функционального зонирования территории, а также для межхозяйственного землеустройства.

Таблица 3.7. Основные аспекты морфометрических параметров цифровой модели рельефа территории

Название параметра	Символ	Формула расчета параметра
Линейные аспекты		
Площадь территории, км ²	A	Общая площадь территории
Длина бассейна, км	L_b	$L_b = 1,312 * A^{0,568}$, где A – площадь территории, км ²
Периметр бассейна, км	P	Внешняя граница территории
Порядок тальвегов	U	Иерархический ранг
Длина тальвегов, км	L_u	Длина тальвегов каждого порядка
Количество тальвегов, шт.	N_u	Количество тальвегов каждого порядка
Средняя длина тальвегов каждого порядка, км	L_{sm}	$L_{sm} = L_u / N_u$
Соотношение длин тальвегов	R_1	$R_1 = L_u / L_{u-1}$, где L_{u-1} – общая длина тальвегов следующего низшего порядка
Коэффициент бифуркации	R_b	$R_b = N_u / N_{u-1}$, где N_{u-1} – общее количество тальвегов следующего высшего порядка
Средний коэффициент бифуркации	R_{bm}	Средний уровень бифуркации тальвегов всех порядков
Площадные аспекты		
Плотность водосборной сети, км/км ²	D_d	$D_d = L_u / A$, где L_u – общая длина тальвегов
Текстура водосборной сети, шт/км	R_t	$R_t = N_u / P$, где N_u – общее количество тальвегов
Частота водосборной сети, шт/км ²	F_s	$F_s = N_u / A$
Коэффициент удлинения	R_e	$R_e = 2 * \sqrt{(A / \pi)} / L_b$
Коэффициент округлости	R_c	$R_c = (4 * \pi * A) / P^2$
Коэффициент формы водосбора	R_f	$R_f = A / L_b^2$
Коэффициент вытянутости водосбора	B_s	$B_s = L_b^2 / A$
Коэффициент компактности	C_c	$C_c = 0,2821 * P / A^{0,5}$
Длина водотоков на поверхности	L_g	$L_g = 1/2 D_d$
Коэффициент инфильтрации	I_f	$I_f = D_d * F_s$
Аспекты рельефа		
Амплитуда колебаний высот, м	B_h	Вертикальное расстояние между самой высокой и самой низкой точкой
Коэффициент рельефа	R_h	$R_h = B_h / L_b$
Коэффициент устойчивости рельефа	R_n	$R_n = B_h / D_d$

Анализ расстояний средствами растрового ГИС-анализа позволяет определить пространственную взаимосвязь каждой ячейки растра по отношению к источнику или набору источников и предусматривает:

- расчет расстояния до ближайшего объекта-источника для каждой ячейки растра;

- присвоение каждой ячейке идентификатора ближайшего источника;

- расчет направления до ближайшего источника для каждой ячейки растра;

- расчет расстояния до ближайшего источника для каждой ячейки растра со сведением до минимума стоимости, заданной в стоимостной поверхности;

- присвоение идентификатора ближайшего источника каждой ячейке растра со сведением до минимума заданной стоимости стоимостной поверхности;

- расчет наименьшей стоимости пути от источника к пункту назначения со сведением до минимума стоимости, определенной по стоимостной поверхности;

- определение для каждой ячейки растра соседней ячейки, которая располагается на пути наименьшей стоимости от источника к пункту назначения, со сведением до минимума стоимости, заданной в стоимостной поверхности;

- расчет расстояния до ближайшего источника для каждой ячейки растра со сведением до минимума горизонтальной стоимости, заданной в стоимостной поверхности, а также стоимости расстояния по поверхности с учетом рельефа местности и трудности вертикального перемещения, определяемой растром рельефа и вертикальными стоимостными параметрами;

- присвоение каждой ячейке растра идентификатора ближайшего источника со сведением до минимума горизонтальной стоимости в стоимостной поверхности, а также стоимостей с учетом рельефа расстояния поверхности и трудности вертикального перемещения, определяемых растром рельефа и вертикальными стоимостными параметрами;

- определение для каждой ячейки растра соседней ячейки, которая располагается на пути наименьшей стоимости от источника до назначения, со сведением до минимума горизонтальной стоимости, заданной в стоимостной поверхности, а также стоимости с учетом рельефа расстояния поверхности и трудности вертикального перемещения,

определяемых растром рельефа и вертикальными стоимостными параметрами;

– расчет суммы общей стоимости для двух входных растров стоимостного расстояния;

– вычисление длины линейных объектов по растровой поверхности с учетом рельефа.

Пример присвоения каждой ячейке растра идентификатора ближайшего источника (в данном примере – населенного пункта) представлен на рис. 3.28, где определены водосборные бассейны для агрогородков, расположенных на территории Горецкого района Могилевской области.

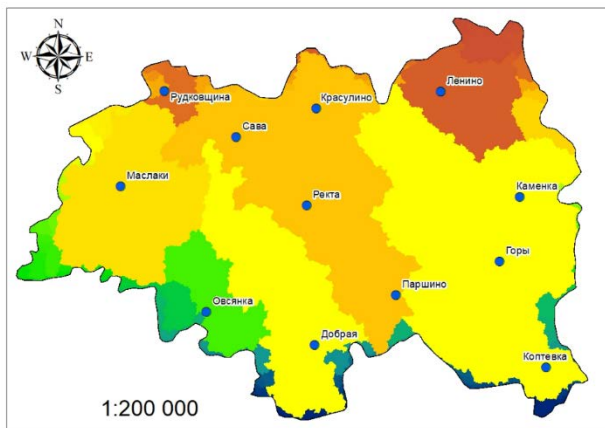


Рис. 3.28. Результат моделирования водосборных бассейнов для агрогородков, расположенных на территории Горецкого района

С помощью грид-моделей плотности в ГИС анализируется пространственное распределение точечных либо линейных векторных объектов, по результатам которого посредством интерполяции создается растровая поверхность. Примером такого вида растрового ГИС-анализа может служить расчет общего показателя расчленения рельефа территории (рис. 3.29). Пособством геоинформационного анализа определялась сумма длин всех горизонталей (предварительно извлеченных из растровой цифровой модели рельефа) в пределах единицы площади, каковой в данном примере являлась ячейка сетки размером 1×1 км. Итог расчета записывался в таблицу атрибутов результирующего слоя и

конвертировался в точечный векторный слой, по которому посредством выполнения интерполяции по методу обратных взвешенных расстояний производилось создание интерполяционной грид-модели.

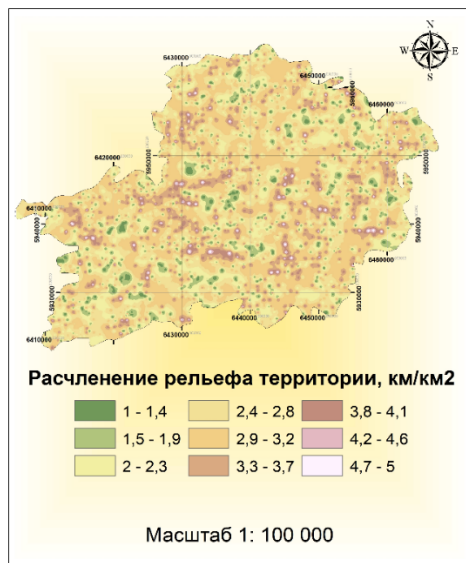


Рис. 3.29. Показатель общего расчленения рельефа территории Климовичского района

Растровая модель данных обладает рядом ограничений по сравнению с векторной моделью, поскольку не позволяет оперировать отдельными объектами, их границами, площадями. Однако она обладает внутренней топологией: ее ячейки соприкасаются друг с другом, что позволяет моделировать непрерывные в пространстве и динамические явления. Поскольку все ячейки растра имеют одинаковый размер, к ним возможно применять однотипные операции, которые будут давать предсказуемый результат вне зависимости от конкретной локации в пределах растра. Такие операции называют **операциями растровой алгебры**, или **алгебры карт**, под которой подразумевают набор примитивных операций в географической информационной системе, позволяющий двум или более растровым слоям схожих размеров создавать новый растровый слой с использованием алгебраических опера-

ций, таких как сложение, вычитание. Основными задачами картографической растровой алгебры являются переклассификация растровых данных, а также выполнение различных математических и статистических операций с гридами.

Классификация операций растрового анализа, предполагающая, что обработке подвергается каждая ячейка, и разделяющая все операции по охвату относительно текущей ячейки растра, была введена в 2012 г. американским профессором Ч. Д. Томлином³⁴. Согласно положениям этой классификации выделяют следующие операции растровой алгебры:

- локальные – анализируется одна ячейка растра или совпадающие в пространстве ячейки нескольких растров;
- фокальные (методы анализа соседства) – анализируются все ячейки в окрестности, которая может быть как фиксированной, так и расширенной, когда ее размер управляется внешними факторами, например, множеством объектов, до которых требуется вычислить расстояние;
- зональные – анализируются все ячейки в пределах зон, определяемых извне (например, еще одним растровым слоем);
- глобальные – анализируются все ячейки растра.

3.5. Основные операции моделирования в ГИС

Модели являются основой при обработке информации в геоинформационных технологиях и имеют важное прикладное значение, поскольку моделирование позволяет осуществить замену эксперимента математическими или информационными манипуляциями и выполнить довольно точный перенос результатов моделирования на объект исследования.

ГИС-моделирование является имитацией процессов, происходящих в реальном мире в определенный момент времени или в течение длительного периода с помощью функциональных возможностей геоинформационных систем и с использованием данных с географической привязкой.

В ГИС выделяют четыре основные группы моделирования: семантическое, инвариантное, эвристическое и информационное.

³⁴Источник: <https://tsamsonov.github.io/r-geo-course/raster-analysis.html>.

Семантическое моделирование взаимосвязано с задачами кодирования и лингвистического обеспечения и выполняется в процессе сбора геопространственной информации, ее структурирования и поиска ошибок.

Инвариантное моделирование основано на работе с полностью или частично унифицированными информационными элементами или структурами и предусматривает использование общих свойств моделируемых объектов (свойств типов или классов) независимо от специфических характеристик отдельных объектов. Данный вид ГИС-моделирования применяется при работе с базами геопространственных данных.

Эвристическое моделирование применяется при необходимости экспертных решений, учете дуальных свойств объектов на видеоизображениях и при решении специальных нетиповых задач и реализуется посредством использования возможностей машинного обучения.

Информационное моделирование связано с созданием и преобразованием различных форм информации (графической, текстовой) в вид, задаваемый пользователем, и реализуется посредством создания подсистемы документационного обеспечения.

Различают следующие типы ГИС-моделей (рис. 3.30).



Рис. 3.30. Типы ГИС-моделей

Процесс ГИС-моделирования представляет собой последовательное прохождение четырех этапов:

- первый этап – определение цели модели;
- второй этап – разбиение модели на элементы и определение свойств каждого элемента и взаимодействия между элементами;

- третий этап – внедрение и калибровка модели;
- четвертый этап – проверка модели на точность.

ГИС-моделирование осуществляется на основе декомпозиции исходных информационных данных с последующим синтезом общего модельного решения и базируется на *интеграции*, которая предусматривает как переход от автоматизации отдельных частных задач к комплексному решению задач, так и интеграцию задач, относящихся к различным этапам жизненного цикла моделируемого объекта. В процессах моделирования реализуется принцип единства информационной модели объекта как системно-организованной сущности.

Технологии моделирования в ГИС основаны на следующих принципах (рис. 3.31).

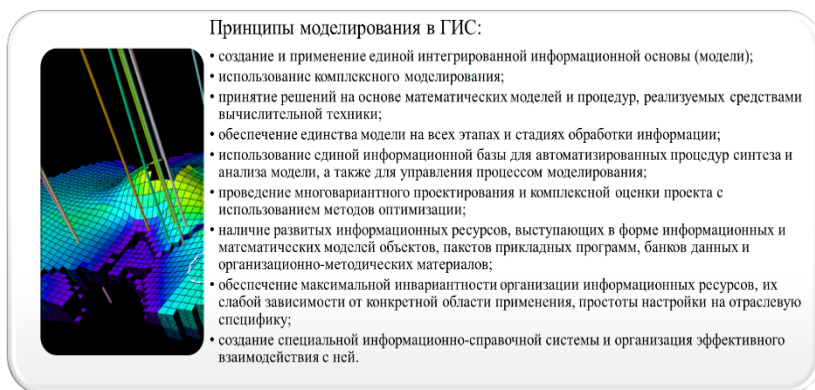


Рис. 3.31. Основные принципы моделирования в ГИС

Основные технологические подходы к осуществлению ГИС-моделирования классифицируются по уровню их интеграции и представлены следующими элементами:

- *процедура* – элементарная операция, выполняемая в процессе обработки информации (поиск по атрибуту, создание выборки данных);
- *задача* – совокупность процедур для получения одного вида геопространственных данных (агрегирование полигонов, фильтрация, отсечение шумов в пределах растра);
- *процесс* – совокупность задач, обеспечивающих реализацию типового цикла преобразования геоданных (постклассификационная обработка растровых данных);

– *функция* – группы специализированных задач, выполняющих взаимосвязанные действия, в ходе которых осуществляется получение комплекта специализированной ГИС-продукции (информации) (реализация моделей пригодности);

– *комплекс работ* – совокупность работ, заканчивающихся выпуском комплекта проектных документов (цифровой проект организации территории сельскохозяйственного предприятия);

– *интегрированные работы* – поддержка и автоматическое обновление базы данных; внесение данных в экспертную систему; выдача наряду с типовым комплектом документации прогнозов, рекомендаций, экспертных оценок; информационный обмен с сетями баз данных и ГИС (геопортал ЗИС, публичная кадастровая карта).

Имитационное моделирование – один из путей выбора оптимальных решений при анализе динамических систем, позволяющий смоделировать поведение объектов в системе посредством создания набора правил, согласно которым система переходит из одного состояния в другое.

Основными *преимуществами имитационного моделирования* являются:

– возможность анализировать системы и находить решения в тех случаях, когда такие методы, как аналитические вычисления и линейное программирование, не справляются с поставленной задачей;

– более простой, чем у аналитической модели, способ разработки, поскольку процесс создания модели будет инкрементальным³⁵ и модульным.

– подобие – структура имитационной модели, естественным образом отображает структуру моделируемой системы;

– возможность отслеживать все объекты системы, учтенные в выбранном уровне абстракции, добавлять метрики и проводить статистический анализ;

– возможность проигрывать модель во времени и анимировать ее поведение, что позволяет качественно улучшить и упростить процесс верификации модели и нахождения ошибок.

Имитационное моделирование нашло применение в различных отраслях и видах деятельности (рис. 3.32).

³⁵Инкрементальный – поэтапный, пошаговый, следующий временным графикам процесс моделирования, в котором различные части системы разрабатываются в разное время и с разными темпами.



Рис. 3.32. Области применения имитационного моделирования

В имитационном моделировании используются три методологии: дискретно-событийное моделирование, агентное моделирование и системную динамику, каждая из которых предусматривает наличие того либо иного уровня абстракции³⁶ (рис. 3.33).

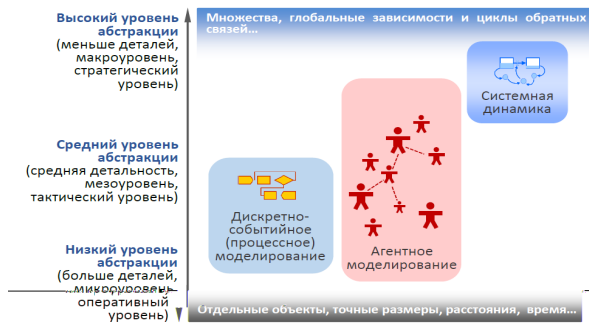


Рис. 3.33. Взаимосвязь методов имитационного моделирования и уровней абстракции

Дискретно-событийное моделирование (discrete-eventsimulation) – вид имитационного моделирования, в котором функционирование системы представляется как хронологическая последовательность событий, которые происходят в определенный момент времени и знамену-

³⁶ Абстракция – мысленное отвлечение, обособление от тех или иных сторон, свойств или связей предметов и явлений для выделения существенных их признаков.

ют собой изменение состояния всей системы. Данный метод поддерживает средний и низкий уровни абстракции.

Агентное моделирование (agent-based model) – метод имитационного моделирования, исследующий поведение децентрализованных агентов и то, как оно определяет поведение всей системы в целом. При этом под агентами понимаются различные по своей природе элементы, например, социальные агенты (люди, социальные группы), экономические агенты (государство, регионы, отрасли), технические агенты (автотранспорт, авиатранспорт), экосистемные агенты (пожары, наводнения). Поведение агентов определяется на индивидуальном уровне, а их глобальное поведение возникает как результат деятельности множества агентов (моделирование «снизу вверх»). Агентные модели могут быть как очень детализированными, когда агенты представляют физические объекты, так и предельно абстрактными, когда с помощью агентов моделируются конкурирующие компании или правительства государств.

Системная динамика – направление в изучении сложных систем, исследующее их поведение во времени и в зависимости от структуры элементов системы и взаимодействия между ними. Данный метод предполагает очень высокий уровень абстракции и используется для стратегического моделирования.

Для хранения набора типовых моделей и их элементов, дополнительной справочной информации используют специализированные базы геоданных, которые могут образовывать как распределенную, так и централизованную систему (банк данных). Решение задачи обмена информацией между базами геоданных реализуется посредством использования интегрированной информационной основы.

Выдача информации, полученной в ходе реализации процесса моделирования, осуществляется в форме представления графических моделей. При этом методы моделирования графики должны быть инвариантны к структуре графической базы данных и техническим средствам. При моделировании в ГИС выделяют следующие программно-технологические блоки: 1) операции преобразования форматов и представлений данных; 2) проекционные преобразования; 3) геометрический анализ; 4) оверлейные операции; 5) функционально-моделирующие операции³⁷.

³⁷Более подробная информация содержится в главе 4.

Вопросы для самостоятельного изучения

1. Технология получения и обработки информации дистанционного зондирования с помощью активных оптических систем.
2. Выполнение сетевого анализа в ArcGIS с помощью модуля Network Analyst (создание графа, построение оптимального маршрута, операции размещения-распределения объектов).
3. Особенности выполнения агентного моделирования на примере анализа рынка жилой недвижимости.
4. Геоинформационное моделирование в ArcGIS с помощью модуля Model Builder. Создание моделей пригодности.

Источники информации

1. Григорьев, И. Anylogic за три дня: практическое пособие по имитационному моделированию / И. Григорьев. – Москва: Anylogic, 2017. – 273 с.
2. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.

Тестовые задания для проверки знаний

1. *Дополните определение.* Процесс поиска пространственных закономерностей в распределении географических данных и взаимосвязей между объектами называют:

- 1) анализ распределения;
- 2) анализ местоположения объектов;
- 3) кластерный анализ;
- 4) анализ пространственных изменений;
- 5) геопространственный анализ.

2. *Правильным или нет является утверждение?* Атрибутивные данные – цифровые данные о пространственных объектах, включающие сведения об их местоположении и свойствах, пространственных и непространственных атрибутах.

- 1) да;
- 2) нет.

3. *Правильным или нет является утверждение?* Пространственные данные – информация в цифровом или текстовом виде о количественных и качественных характеристиках объектов или явлений.

- 1) да;
- 2) нет.

4. Выберите правильный вариант(ты) ответа. Геопространственные данные содержат следующие интегрированные компоненты:

- 1) атрибуты;
- 2) пространственные отношения;
- 3) географическое положение;
- 4) числовые характеристики;
- 5) временные характеристики.

5. Дополните определение. Наиболее распространенными способами цифрового описания пространственных объектов являются растровые и _____ модели данных.

- 1) геопространственные;
- 2) векторные;
- 3) непрерывные;
- 4) пространственные;
- 5) статистические.

6. Дополните определение. Нульмерные объекты, характеризующиеся координатами на плоскости или в пространстве, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

7. Дополните определение. Нульмерные объекты, состоящие из нескольких точек, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

8. Дополните определение. Одномерные объекты, состоящие из последовательности точек, соединенных между собой отрезками, называют:

- 1) мультиточки;

- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

9. *Дополните определение.* Топологически связанные одномерные объекты, состоящие из последовательности точек, соединенных между собой отрезками, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

10. *Дополните определение.* Двумерные объекты, состоящие из нескольких контуров, заданных в виде последовательности замкнутых линий, и частей плоскости внутри контуров, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

11. *Дополните определение.* Подписи, содержащие параметры отображения текста, например, точки фигуры для размещения текста, его шрифт и точечный размер, а также другие свойства отображения, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;

- 7) аннотации;
- 8) объекты-размеры.

12. Дополните определение. Специальный тип аннотации, показывающий специфические длины или расстояния, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

13. Дополните определение. 3D-геометрию, используемую для представления внешней поверхности или оболочки объектов, которые занимают дискретную область или объем в трехмерном пространстве, называют:

- 1) мультиточки;
- 2) линии;
- 3) полигоны;
- 4) мультипатчи;
- 5) точки;
- 6) полилинии;
- 7) аннотации;
- 8) объекты-размеры.

14. Дополните определение. Однородную совокупность однотипных объектов, каждый из которых имеет одинаковое пространственное представление в виде точек, линий или полигонов и общий набор атрибутивных полей, хранящихся в таблице базы данных, называют:

- 1) класс пространственных объектов;
- 2) атрибут пространственных объектов;
- 3) категория пространственных объектов;
- 4) ранг пространственных объектов;
- 5) численность пространственных объектов;
- 6) база данных пространственных объектов;
- 7) набор классов пространственных объектов;
- 8) тип пространственных объектов.

15. Дополните определение. Числовые, или символные, характеристики, не связанные с местоположением, содержащиеся в базе данных,

которые могут относиться как к самим геометрическим примитивам, так и к объектам, составленным из этих примитивов, называют:

- 1) категории;
- 2) ранги;
- 3) значения;
- 4) атрибуты;
- 5) численность.

16. Выберите правильное соответствие определения значению:

1) представляет собой группу схожих объектов	Д) относительная величина
2) используется для сортировки объектов по порядку от большего к меньшему	В) численность
3) характеризует реальное число объектов на местности	Г) абсолютная величина
4) представляет некоторую суммарную величину, связанную с каждым объектом (процентное соотношение, кратность)	А) категория
5) показывает взаимные отношения между двумя количественными величинами и вычисляется делением одной количественной величины на другую	Б) ранг

17. Дополните определение. Процесс определения областей, соответствующих определенным заданным критериям, и установления причины и следствия возникновения пространственных взаимосвязей называют:

- 1) анализ распределения;
- 2) анализ местоположения объектов;
- 3) кластерный анализ;
- 4) анализ пространственных изменений;
- 5) геопространственный анализ.

18. Дополните определение. Процесс выполнения группировки геопространственных данных в соответствии со скачками в их значениях называют:

- 1) разбивка по квантилям;
- 2) разбивка на равные интервалы;
- 3) естественная разбивка;
- 4) разбивка по величине стандартного отклонения.

19. Дополните определение. Процесс выполнения группировки геопространственных данных, при котором каждый класс содержит равное число геопространственных объектов, зависящее от установленного числа классов разбивки, называют:

- 1) разбивка по квантилям;

- 2) разбивка на равные интервалы;
- 3) естественная разбивка;
- 4) разбивка по величине стандартного отклонения.

20. *Дополните определение.* Процесс выполнения группировки геопространственных данных посредством разбивки ряда значений на интервалы, равные разности между максимальными и минимальными значениями, называют:

- 1) разбивка по квантилям;
- 2) разбивка на равные интервалы;
- 3) естественная разбивка;
- 4) разбивка по величине стандартного отклонения.

21. *Дополните определение.* Процесс выполнения группировки геопространственных данных посредством распределения объектов на классы по величине отклонения их значений от среднего называют:

- 1) разбивка по квантилям;
- 2) разбивка на равные интервалы;
- 3) естественная разбивка;
- 4) разбивка по величине стандартного отклонения.

22. *Правильным или нет является утверждение?* Карты плотности позволяют визуально оценить концентрацию точечных либо линейных объектов в пределах какого-либо участка или всей территории, а также измерить концентрацию объектов на единицу площади и более точно сопоставить области или определить, соответствуют ли они заданным критериям.

- 1) да;
- 2) нет.

23. *Выберите правильный вариант(ты) ответа.* Поиск объектов внутри области позволяет исследовать, что происходит в ее пределах, или обобщить информацию по каждой из исследуемых областей с целью их сопоставления между собой либо с какой-либо референтной областью и выполняется посредством:

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

24. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если необходимо определить, какие объекты попали в пределы одной области?

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

25. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если необходимо получить перечень или сумму объектов, полностью или частично попадающих в границы области?

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

26. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если имеется одна область, но необходимо получить перечень или обобщенное значение объектов, попадающих только в определенную ее часть?

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

27. *Выберите правильный вариант(ты) ответа.* Анализ окружения позволяет в пределах заданного расстояния оценить территорию, прилегающую к объекту, либо оценить пространственные изменения самого объекта, а также дает возможность контролировать события, происходящие в заданном окружении, определять территорию воздействия или оценивать его последствия и выполняется посредством:

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

28. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если необходимо определить область влияния или быстро оценить диапазон перемещения объекта?

- 1) расчета затрат по поверхности;

- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

29. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если требуется измерить длину пути по фиксированной инфраструктуре к объекту или от него?

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

30. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если требуется оценить передвижение по пересеченной местности?

- 1) расчета затрат по поверхности;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) применения наложения слоев;
- 5) построения пространственного запроса;
- 6) измерения расстояния или затрат по сети.

31. *Выберите правильный вариант(ты) ответа.* Анализ пространственных изменений, позволяющий визуально оценить происходящие изменения, глубже понять причины, их вызывающие, и оценить последствия их воздействия, выполняется посредством:

- 1) количественной оценки изменений;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) использования временных рядов;
- 5) построения карт слежения;
- 6) измерения расстояния или затрат по сети.

32. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если необходимо показать движение или изменение свойств объекта в течение двух или более временных отрезков области?

- 1) количественной оценки изменений;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) использования временных рядов;

- 5) построения карт слежения;
- 6) измерения расстояния или затрат по сети.

33. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если необходимо показать движения объекта между двумя или более моментами времени или отразить периодичность изменения?

- 1) количественной оценки изменений;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) использования временных рядов;
- 5) построения карт слежения;
- 6) измерения расстояния или затрат по сети.

34. *Выберите правильный вариант(ты) ответа.* Какой метод следует применить, если необходимо показать численные различия в параметрах объекта между двумя моментами времени?

- 1) количественной оценки изменений;
- 2) визуального сопоставления границ области и объектов;
- 3) измерения кратчайшего расстояния;
- 4) использования временных рядов;
- 5) построения карт слежения;
- 6) измерения расстояния или затрат по сети.

35. *Правильным или нет является утверждение?* Векторный ГИС-анализ используется при обработке цифровых векторных слоев с учетом атрибутов геопространственных объектов, а его применение наиболее целесообразно при работе с дискретными геообъектами, имеющими четкие границы в геопространстве.

- 1) да;
- 2) нет.

36. *Правильным или нет является утверждение?* Топология не может участвовать в качестве объекта векторного ГИС-анализа.

- 1) да;
- 2) нет.

37. *Дополните определение.* Анализ, предусматривающий просмотр векторных геообъектов в среде ГИС, анализ атрибутивных данных пространственных объектов, картометрические измерения и визуализацию атрибутов посредством создания картографических изображений, называют:

- 1) элементарный пространственный анализ;
- 2) пространственная статистика;
- 3) расширенный пространственный анализ;
- 4) сетевой анализ;

5) атрибутивный анализ.

38. *Дополните определение.* Анализ, позволяющий выполнять расчет значений описательной статистики выборки массива числовых атрибутивных данных векторных геопространственных объектов, называют:

- 1) элементарный пространственный анализ;
- 2) пространственная статистика;
- 3) расширенный пространственный анализ;
- 4) сетевой анализ;
- 5) атрибутивный анализ.

39. *Дополните определение.* Анализ, предусматривающий выполнение оверлейных операций, анализ близости, переклассификацию геопространственных данных, их генерализацию и геообработку, называют:

- 1) элементарный пространственный анализ;
- 2) пространственная статистика;
- 3) расширенный пространственный анализ;
- 4) сетевой анализ;
- 5) атрибутивный анализ.

40. *Дополните определение.* Анализ, основанный на теории графов и топологии, позволяющий найти кратчайший путь, спланировать оптимальный маршрут, оценить рациональность распределения сети объектов, определить зону доступности, называют:

- 1) элементарный пространственный анализ;
- 2) пространственная статистика;
- 3) расширенный пространственный анализ;
- 4) сетевой анализ;
- 5) атрибутивный анализ.

41. *Правильным или нет является утверждение?* Растровый анализ – это процесс прогнозирования значений для точек, где нет измеренных значений, по ограниченному набору опорных точек с известными значениями.

- 1) да;
- 2) нет.

42. *Правильным или нет является утверждение?* Геостатистические методы интерполяции предполагают наличие заданной аналитической зависимости между значениями функции в пространстве и используют абсолютные значения в точках для прогнозирования неизвестных промежуточных значений некоторой функции.

- 1) да;
- 2) нет.

43. *Правильным или нет является утверждение?* Детерминированные методы интерполяции предполагают наличие заданной аналитической зависимости между значениями функции в пространстве и используют анализ автокорреляции для прогнозирования неизвестных промежуточных значений некоторой функции.

- 1) да;
- 2) нет.

44. *Правильным или нет является утверждение?* Анализ расстояний средствами растрового ГИС-анализа позволяет определить пространственную взаимосвязь каждой ячейки растра по отношению к источнику или набору источников.

- 1) да;
- 2) нет.

45. *Правильным или нет является утверждение?* Растровая модель данных обладает рядом ограничений по сравнению с векторной моделью, поскольку не позволяет оперировать отдельными объектами, их границами, площадями.

- 1) да;
- 2) нет.

46. *Дополните определение.* Операцию растровой алгебры, с помощью которой анализируется одна ячейка растра или совпадающие в пространстве ячейки нескольких растров, называют:

- 1) локальная;
- 2) зональная;
- 3) расширенная;
- 4) фокальная;
- 5) глобальная.

47. *Дополните определение.* Операцию растровой алгебры, с помощью которой анализируются все ячейки в окрестности, которая может быть как фиксированной, так и расширенной, называют:

- 1) локальная;
- 2) зональная;
- 3) расширенная;
- 4) фокальная;
- 5) глобальная.

48. *Дополните определение.* Операцию растровой алгебры, с помощью которой анализируются все ячейки в пределах зон, определяемых извне, называют:

- 1) локальная;
- 2) зональная;
- 3) расширенная;

- 4) фокальная;
- 5) глобальная.

49. Дополните определение. Операцию растровой алгебры, с помощью которой анализируются все ячейки растра, называют:

- 1) локальная;
- 2) зональная;
- 3) расширенная;
- 4) фокальная;
- 5) глобальная.

50. Вставьте пропущенное словосочетание. ГИС-моделирование является имитацией процессов, происходящих в реальном мире, в определенный момент времени или в течение длительного периода, с помощью функциональных возможностей _____ и с использованием данных с географической привязкой.

- 1) геоинформационных систем;
- 2) имитационного моделирования;
- 3) компьютерных систем;
- 4) дискретно-событийного моделирования;
- 5) агентного моделирования.

51. Выберите правильный порядок иерархии элементов ГИС-моделирования.

- 1) процедура;
- 2) интегрированные работы;
- 3) функция;
- 4) процесс;
- 5) комплекс работ;
- 6) задача.

52. Выберите правильное соответствие определения значению:

1) взаимосвязано с задачами кодирования и лингвистического обеспечения и выполняется в процессе сбора геопространственной информации, ее структурирования и поиска ошибок	В) эвристическое моделирование
2) основано на работе с полностью или частично унифицированными информационными элементами или структурами и применяется при работе с базами геопространственных данных	А) семантическое моделирование
3) применяется при необходимости экспертных решений, учете дуальных свойств объектов на видеоизображениях и реализуется посредством использования возможностей машинного обучения	Г) информационное моделирование
4) связано с созданием и преобразованием различных форм информации (графической, текстовой) в вид, задаваемый пользователем	Б) инвариантное моделирование

53. *Дополните определение.* Вид имитационного моделирования, в котором функционирование системы представляется как хронологическая последовательность событий, происходящих в определенный момент времени и знаменующих собой изменение состояния всей системы, называют:

- 1) эвристическое моделирование;
- 2) имитационное моделирование;
- 3) семантическое моделирование;
- 4) дискретно-событийное моделирование;
- 5) агентное моделирование;
- 6) инвариантное моделирование;
- 7) информационное моделирование.

54. *Дополните определение.* Вид имитационного моделирования, исследующий поведение децентрализованных агентов и то, как оно определяет поведение всей системы в целом, называют:

- 1) эвристическое моделирование;
- 2) имитационное моделирование;
- 3) семантическое моделирование;
- 4) дискретно-событийное моделирование;
- 5) агентное моделирование;
- 6) инвариантное моделирование;
- 7) информационное моделирование.

4. РАБОТА С ГЕОПРОСТРАНСТВЕННЫМИ ДАННЫМИ

4.1. Операции переструктуризации данных. Агрегирование данных.

4.2. Геокодирование данных.

4.3. Классификация и переклассификация изображений.

4.4. Локальные, зональные и фокальные операции.

4.5. Операции вычислительной геометрии. Оверлейные операции. Построение буферных зон.

4.6. Базы геоданных и их свойства. Проектирование баз геоданных.

4.7. Подтипы, домены и их характеристика. Классы отношений геометрической сети и определение отношений между объектами. Создание правил отношений.

Количество часов, отведенных на изучение главы, – 6.

Аннотация темы.

Тематика лекции посвящена основам работы с геопространственными данными. В пункте 4.1 приведены сведения о сущности процесса переструктуризации растровых и векторных данных, а также их агрегирования с использованием функциональных возможностей ГИС. В пункте 4.2 рассмотрены особенности осуществления геокодирования, приведена характеристика основных стилей локаторов адресов и примеры практического применения возможностей геокодирования данных. В пункте 4.3 освещены вопросы, относящиеся к выполнению контролируемой и неконтролируемой классификации растровых данных и их переклассификации. Пункт 4.4 посвящен характеристике и особенностям использования локальных, зональных и фокальных операций и содержит характеристику инструментов, применяемых для их выполнения. Содержание пункта 4.5 посвящено вопросам проведения оверлейных операций и построения буферных зон, характеристике инструментов для их выполнения и особенностям реализации топологических наложений. В пункте 4.6 рассмотрена структура и свойства баз геоданных и основные этапы их проектирования. В пункте 4.7 рассмотрены понятия «подтип» и «атрибутивный домен», типы атрибутивных доменов, а также дана подробная характеристика классов простых и сложных отношений между геопространственными объектами и рассмотрено понятие кардиальности.

Литературные и информационные источники

1. Ананьев, Ю. С. Геоинформационные системы / Ю. С. Ананьев. – Томск: Изд-во ТПУ, 2003. – 70 с.
2. Поморцева, Е. Е. Проектирование баз геоданных. Лабораторный практикум: учеб. пособие / Е. Е. Поморцева. – Харьков: ХНУГХ им. А. Н. Бекетова, 2016. – 157 с.
3. Шнитко, С. Г. ГИС в геодезии: конспект лекций / С. Г. Шнитко. – Новополюцк: ПГУ, 2014. – 68 с.

Internet-ресурсы

4. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.
5. Учебник Quantum GIS [Электронный ресурс]. – Режим доступа: https://wiki.gislab.info/w/%D0%A3%D1%87%D0%B5%D0%B1%D0%BD%D0%B8%D0%BA_Quantum_GIS.

4.1. Операции переструктуризации данных. Агрегирование данных

Переструктуризация (конвертация) данных является собой процесс, предусматривающий превращение геопространственных данных из одного типа данных в другой. Наиболее широко используемыми операциями переструктуризации являются:

- преобразование векторных данных в растровые и наоборот;
- конвертация данных из таблиц Excel в атрибутивные таблицы геопространственных данных и обратно.

Конвертация файлов Microsoft Excel в атрибутивную таблицу позволяет получить выходные данные с плавающей точкой, а также данные в формате текста или даты, организованные в таблицу с вертикальной ориентацией (рис. 4.1).

Если входной столбец таблицы Excel содержит более одного типа данных или формата, выходное поле атрибутивной таблицы будет иметь текстовый формат. Переструктуризация данных позволяет также преобразовывать атрибутивные таблицы в таблицы в формате Excel. Любой класс пространственных объектов (из базы геоданных, шейп-файла или покрытия), содержащий точечные, линейные или полигональные объекты, может быть конвертирован в набор растровых данных. Для **переструктуризации растровых данных в векторные данные** используют три различных метода:

- конвертация растровых поверхностей в векторные пространственные объекты;
- конвертация поверхности TIN³⁸ в векторные пространственные объекты;

³⁸TIN-модель пространственных данных (TriangulatedIrregular Network) – это способ моделирования непрерывной поверхности точками и значениями в этих точках, выбранными с переменной плотностью.

Конвертация растровых поверхностей в векторные пространственные объекты осуществляется посредством преобразования растра в слой полигональных, линейных либо точечных объектов (рис. 4.2).

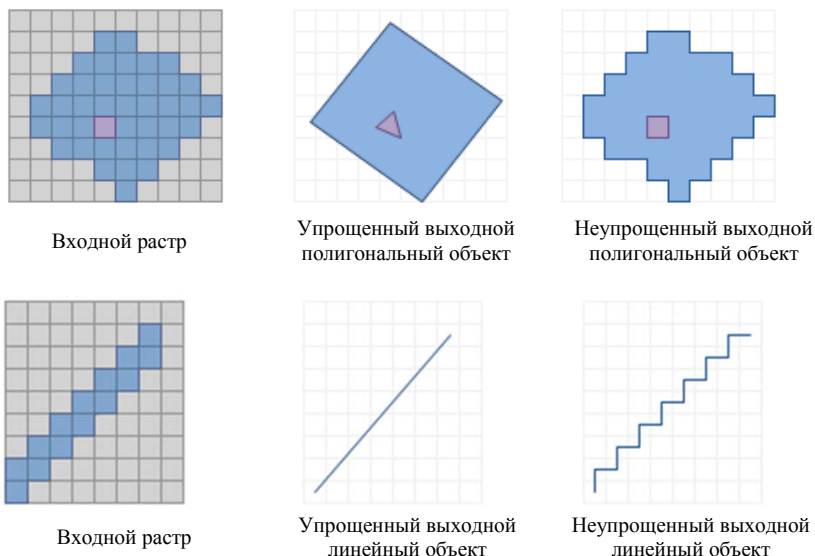


Рис. 4.2. Переструктуризация растровой поверхности в векторные пространственные объекты

При конвертации набора растровых данных, содержащего полигональные объекты, каждая группа смежных ячеек с одинаковым значением конвертируется в полигон. Границы полигонов при этом создаются из границ ячеек в растре, а ячейки с отсутствующими значениями во входном растре не конвертируются в выходные полигоны.

При конвертации набора растровых данных, содержащих линейные объекты, полилинии создаются из каждой ячейки во входном наборе растровых данных. Они размещаются так, чтобы проходить через центр каждой ячейки растра. Ячейки с отсутствующими значениями во входном наборе растровых данных в выходные полилинейные объекты не конвертируются.

При преобразовании растровой поверхности в векторный слой точечных объектов для каждой ячейки входного набора растровых данных создается точка в выходном классе объектов, расположенная в

центре ячейки, которую она представляет. Если данные в ячейке растра отсутствуют, то для такой ячейки точка создаваться не будет.

При конвертации поверхности TIN в векторные пространственные объекты треугольные грани из набора данных TIN экспортируются в полигональные объекты и формируется атрибутивная таблица, в которой для каждого треугольника указываются значения уклона, экспозиции, площади и периметра (рис. 4.3).

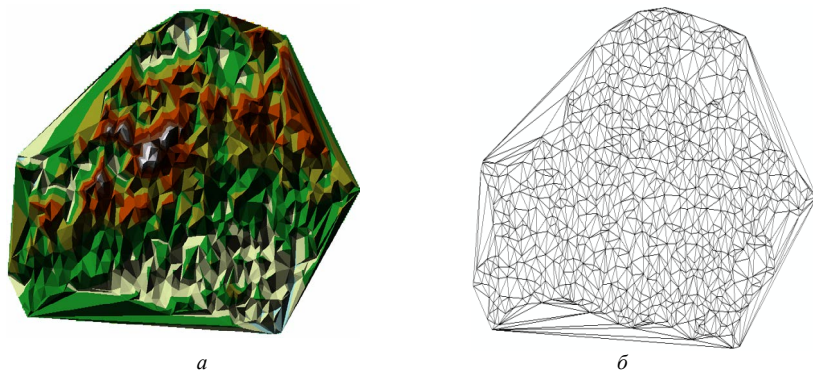


Рис. 4.3. Результат переструктуризации TIN-модели (а) в векторный слой полигональных пространственных объектов (б)

Конвертация набора данных Terrain производится только в точечный либо мультиточечный векторный класс пространственных объектов (рис. 4.4).

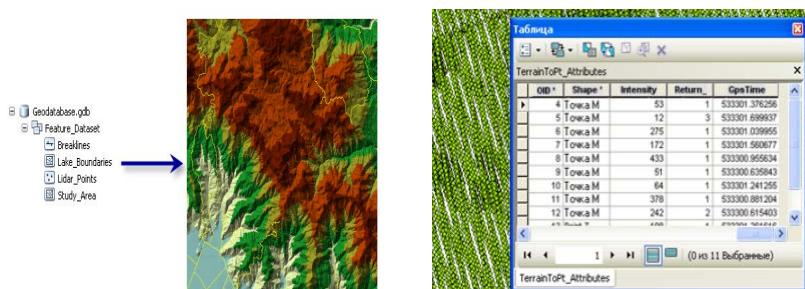


Рис. 4.4. Результат переструктуризации набора данных Terrain в новый класс пространственных объектов точек или мультиточек

Растровые данных могут также конвертироваться в другие типы файлов – файлы с ASCII и файлы со значениями с плавающей точкой. Файлы ASCII⁴⁰ состоят из заголовка, содержащего информацию о наборе ключевых слов, и значений ячеек в построчном порядке. Информация в файле ASCII сохраняется в текстовом файле и содержит следующие данные:

- количество столбцов и строк в растре, заданное файлом ASCII;
- координаты левого нижнего угла левой нижней ячейки растра;
- размер ячейки растра;
- значение, служащее для представления ячеек со значениями без данных;
- поток данных значений ячеек, где строка 1 данных – верхняя строка матрицы ячеек растра, строка 2 – следующая и т. д.

Файлы со значениями с плавающей точкой похожи на файлы ASCII, но содержат значения с плавающей точкой. Значения с плавающей точкой – это значения с десятичной точностью, которые могут быть как положительными, так и отрицательными. Файлы со значениями с плавающей точкой хранятся в виде файлов с 32-битными двоичными значениями со знаком.

Существует также ряд инструментов для конвертации геопространственных данных различных форматов в растровые выходные данные (табл. 4.1).

Таблица 4.1. Инструменты для конвертации данных различных форматов в растровые выходные данные

Название инструмента	Описание инструмента
1	2
ASCII в растр	Конвертирует ASCII файл, содержащий растровые данные, в набор растровых данных
ЦМР ⁴¹ в растр	Конвертирует цифровую модель рельефа в формате United States Geological Survey (USGS) в набор растровых данных
Объекты в растр	Конвертирует пространственные объекты в растры

⁴⁰ASCII (англ. American Standard Code for Information Interchange) – название таблицы (кодировки, набора), в которой некоторым распространенным печатным и непечатным символам сопоставлены числовые коды. Таблица была разработана и стандартизована в США в 1963 г.

⁴¹Цифровая модель рельефа – ЦМР (Digital Elevation Model – DEM) – цифровое представление топографической поверхности в виде растра или регулярной сети ячеек заданного размера.

1	2
Файл с плавающей точкой в растр	Конвертирует файл, содержащий бинарные значения с плавающей точкой, представляющие растровые данные, в набор растровых данных
Набор данных LAS ⁴² в растр	Создает растр, используя значения высоты, интенсивности или цвета RGB, на основе лидарных точек, на которые ссылается набор данных LAS
Мультипатч ⁴³ в растр	Конвертирует объекты мультипатч в набор растровых данных
Точки в растр	Конвертирует точечные пространственные объекты в набор растровых данных
Полигон в растр	Конвертирует полигональные пространственные объекты в набор растровых данных
Полилинии в растр	Конвертирует линейные пространственные объекты в набор растровых данных
Растр в другой формат	Конвертирует один или несколько наборов растровых данных в другой формат

Агрегирование данных – процесс пересчета входного растра с высоким разрешением в растр с более низким разрешением. Каждая выходная ячейка создаваемого при этом растра содержит сумму, минимальное, максимальное, среднее и медианное значения входных ячеек, которые попадают в пределы экстенда ячейки созданного растра. Процесс агрегирования выполняется в несколько этапов:

1) разрешение входного растра умножается на коэффициент, устанавливаемый параметром коэффициента размера ячейки, а результирующее значение соответствует разрешению ячейки выходного растра; коэффициент ячеек влияет на размер ячейки агрегированного растра, например, при коэффициенте 4 размер ячейки увеличится в 4 раза;

2) пространственный экстенд выходных ячеек переносится на входной растр (рис. 4.5, а);

3) определяются ячейки, для которых будут выполнены вычисления агрегирования; в вычисления по определению выходного значения ячейки включаются значения ячеек входного растра, которые попадают в экстенд выходной ячейки;

⁴²LAS – это открытый публичный стандартный формат файлов, использующийся для обмена лидарными данными; бинарный формат, поддерживающий специфическую информацию лидарных данных.

⁴³Мультипатчи – это 3D-геометрия, используемая для представления внешней поверхности, или оболочки, объектов, которые занимают дискретную область или объем в трехмерном пространстве. Они охватывают плоские 3D-окружности и треугольники, используемые в комбинации для моделирования трехмерной оболочки.

4) вычисляется выходное значение для каждой ячейки агрегированного растра посредством определения суммы, медианы, минимального, максимального или среднего значения ячеек входного растра, которые попадают внутрь пространственного экстенда выходной ячейки; агрегированное значение записывается в каждую ячейку выходного растра (рис. 4.5, б).

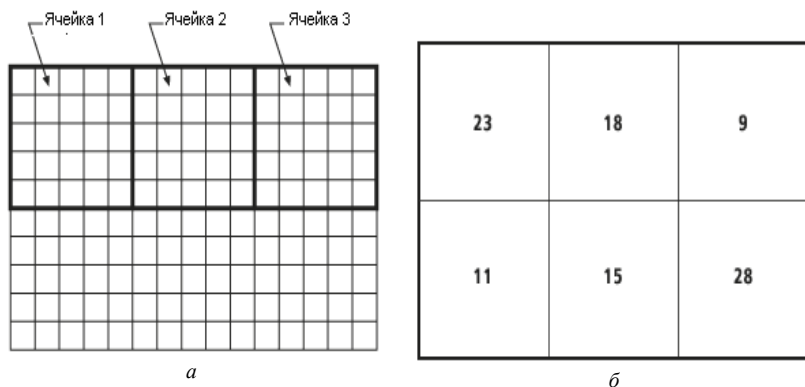


Рис. 4.5. Процесс агрегирования растра

Результатом агрегирования растровых данных является уменьшение пространственного разрешения растра и, как следствие, – уменьшение его размера (рис. 4.6).

В случае если значение, заданное для коэффициента размера ячейки, таково, что результат деления количества ячеек в строках или в столбцах на коэффициент размера ячейки дает остаток, пространственный экстенд входных ячеек не будет совпадать с экстендом выходного растра с учетом последней ячейки в его строках или в столбцах. Когда это происходит, то значение для последней выходной ячейки может быть вычислено с использованием доступных ячеек входного растра, которые попадают в экстенд, либо растр может быть обрезан посредством удаления последней ячейки выходного растра.

Выполнить агрегирование растра возможно и посредством статистики по блокам, когда расчет указанных статистических характеристик выполняется во входных ячейках внутри окрестностей, которые не пересекаются.

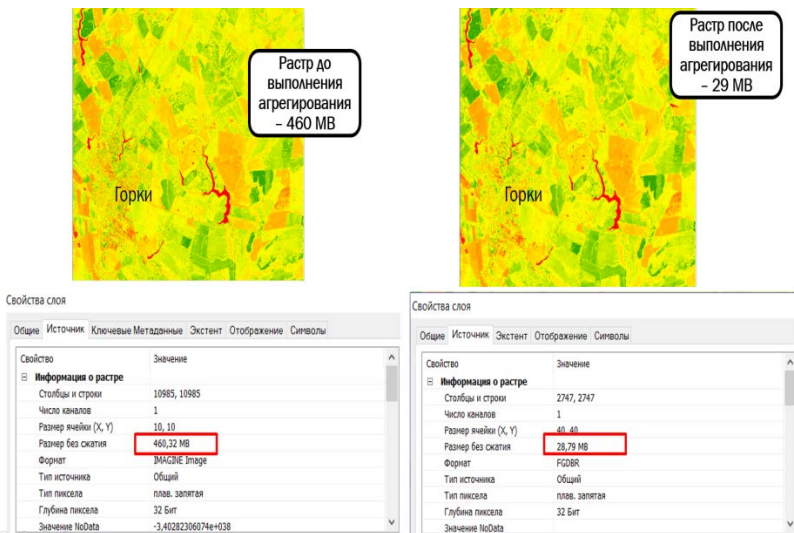


Рис. 4.6. Результат агрегирования растровых данных (размер ячейки раstra увеличился в 4 раза, объем растровых данных уменьшился в 16 раз)

4.2. Геокодирование данных

Геокодирование – способ получения геоданных из адресной информации, процесс, преобразующий описание местоположения (например, координаты, адрес или название места) в местоположение на местности. Оно позволяет получить координаты объектов, адреса которых известны. В связи со значительным объемом адресных баз данных эту операцию обычно производят сервисы геокодирования, наиболее известными из которых являются Google Geocoding⁴⁴, Яндекс. Карты, Геокодирование и OSM Nominatim⁴⁵.

В результате выполнения геокодирования образуются географические объекты с атрибутами, которые можно использовать для составления различных карт или выполнения пространственного анализа.

⁴⁴Библиотека, позволяющая найти объект на карте по его названию и получить координаты либо при наличии координат получить информацию о том, что находится в закоординированной локации.

⁴⁵Инструмент, используемый для поиска данных OSM по имени и адресу, а также для создания полных адресов точек OSM (обратное геокодирование).

Оно является одним из ключевых моментов в управлении пользовательскими данными и позволяет получить информацию клиентов и создать карту их местоположений. С помощью различных приложений эту информацию можно использовать многими способами:

- выполнять построение маркетинговых стратегий;
- выделять отдельные целевые классы клиентов;
- создавать маршрутные карты и инструкции по проезду.

Организации, работающие в сфере недвижимости, пользуются преимуществами распространения информации о доступных помещениях через Интернет. Комбинирование информации из базы данных о домах и различных веб-сервисов посредством геокодирования позволяет распространять пространственную и непространственную информацию среди широкого круга пользователей.

С помощью геокодирования можно быстро находить различные виды местоположений:

- достопримечательности или названия из географического справочника, такие как горы, мосты и магазины;
- координаты на основе долготы и широты или других систем привязки к местности;
- адреса в различных стилях и форматах, включая пересечения дорог, номера домов с названиями улиц и почтовыми кодами.

Геокодирование выполняет кодирование таблицы с адресами. Иными словами, геокодирование сопоставляет адреса с локатором адресов и сохраняет результат для каждой входной записи в новом классе точечных пространственных объектов (рис. 4.7).

Для выполнения этого процесса требуется таблица, в которой хранятся адреса, которые необходимо геокодировать, а также локатор адресов или составной локатор адресов.

Локатор адресов создается на основе определенного стиля локаторов. Он содержит свойства и параметры геокодирования, адресные атрибуты базовых данных и запросы для выполнения поиска. Локатор адресов также содержит набор правил разбора и сопоставления адресов, которые указывают подсистеме геокодирования, как именно следует проводить стандартизацию и сопоставление адресов. Существует несколько стилей локаторов адресов, большинство из которых использует стили, разработанные и применяемые в США (табл. 4.2).

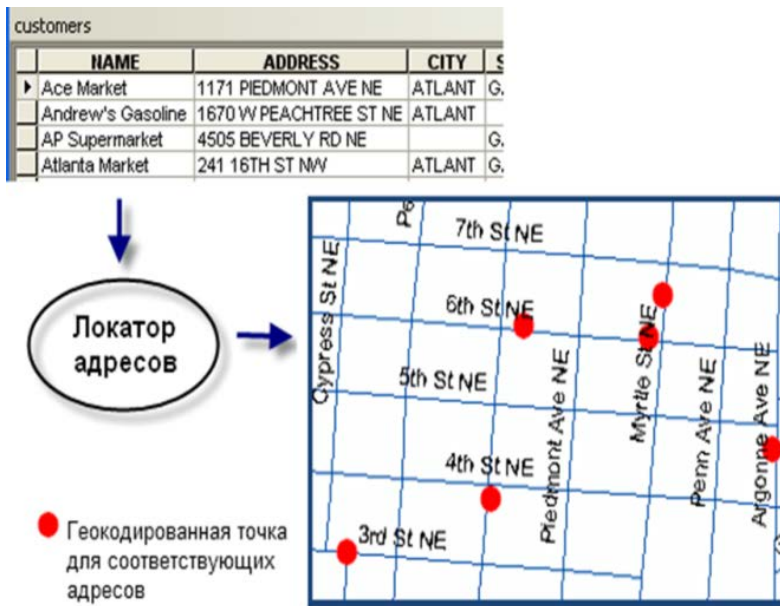


Рис. 4.7. Процесс геокодирования

Таблица 4.2. Характеристика основных стилей локоаторов адресов

Стили	Типичная геометрия набора базовых данных	Типичное представление набора базовых данных	Параметры поиска адреса	Примеры	Применение функций путевых расстояний
1	2	3	4	5	6
US Address – один диапазон	Линии	Один диапазон для каждой стороны сегмента улицы	Все элементы адреса хранятся в одном поле	Рытова 15 (15 Рытова ул.), улица Вишневая 35 (35 Вишневая ул.)	Поиск дома на улице, где сторона не имеет значения, либо стороны улицы сохранены как атрибуты в каждом сегменте улицы
Стиль локоатора адрес в США – дом (US Address – Single House)	Точки или полигоны	Каждый объект представляет один адрес	Все элементы адреса хранятся в одном поле	Переулок Строительный 21 (21 Строительный пер.)	Поиск участков, зданий или точек с адресом
US Address – название улицы	Линии	Каждый объект с названием улицы и дополнительным названием зоны	Элементы адреса без номера дома в одном поле	Переулок Кооперативный, Дрибин, Могилевская область (Кооперативный переулок, Дрибин, Могилевская область)	Поиск объектов по названию улицы
US Address – пятизначный почтовый индекс (ZIP-код)	Точки или полигоны	Регион ZIP-кода или центроид	Пятизначный почтовый индекс (ZIP-код)	213406	Поиск определенного местоположения по почтовому индексу (коду)

1	2	3	4	5	6
Общий – Город, Штат, Страна (General – City, State, Country)	Точки или полигоны	Город в границах страны	Название города и штата или его аббревиатура, название страны	Горки, Могилевская область, Республика Беларусь (Горки, Могилевская, Беларусь)	Поиск конкретного города в стране
Общий – Газетир (General – Gazetteer)	Точки или полигоны	Каждый объект представляет собой особое место, ориентир или пункт	Все элементы названия места в одном поле	Несвижский замок, Беларусь Брест, Республика Беларусь	Поиск географических названий мест или ориентиров в любом районе мира
Общий – Однополе (General – Single Field)	Точки или полигоны	Каждый объект идентифицируется по тексту, названию или индексу (коду)	Единая, определенная пользователем переменная	Кафе KFC (Cafe KFC) № 1, № 115	Поиск объектов, идентифицируемых по названию или по индексу (коду)

4.3. Классификация и переклассификация изображений

Классификация изображений – это процесс извлечения классов информации из многоканального растрового изображения. Растр, полученный в результате классификации изображения, можно использовать для создания тематических карт.

В **классификации изображений с обучением** используются спектральные сигнатуры⁴⁶, полученные из обучающих выборок. Процесс **классификации без обучения** находит спектральные классы (или кластеры) в многоканальном изображении без вмешательства аналитика. При выполнении классификации изображений с обучением создается **обучающая выборка** – области изображения, представляющие различные классы для классификации с обучением. Обучающие выборки являются примерами классов в изображении, показывающими инструментам классификации, как классифицировать остальные пиксели. На рис. 4.8 представлена схема процесса классификации изображений.

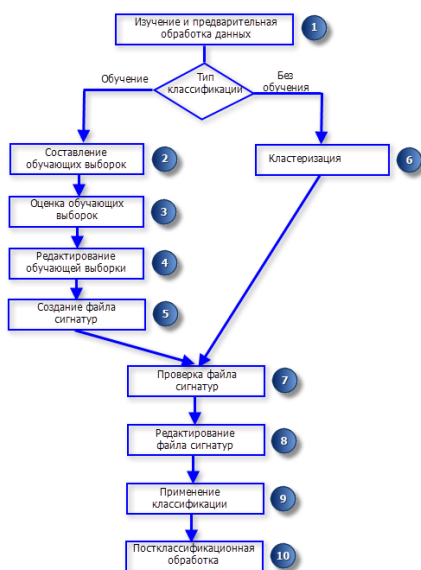


Рис. 4.8. Схема процесса классификации изображений

⁴⁶Файл, в который записываются сигнатуры спектра различных классов по ряду каналов. Сигнатура для каждого класса содержит средние и матрицу ковариаций, вычисленные на основе соответствующей обучающей выборки.

Переклассификация – процесс изменения значений ячеек раstra на альтернативные значения с использованием различных методов. Она выполняется в случае, если: нужно заменить значения входного раstra на новые значения (например, землепользования в области, измененное с течением времени); необходимо упростить информацию в растре (например, можно группировать различные типы леса в один класс леса); требуется изменение масштаба значений набора растров к общему масштабу; необходимо присвоение значений приоритета, чувствительности, последовательности или некоторых сходных критериев в растре. Возможно выполнить переклассификацию одного значения или группы значений, используя альтернативные поля на основе таких критериев, как заданные интервалы (например, группа значений в 10 интервалов) или по площади (например, группа значений в 10 групп, содержащих одинаковое количество ячеек).

Очень часто переклассификацию раstra используют, когда нужно выполнить поиск определенного участка по комплексу заданных параметров, например, участки с наиболее оптимальным набором агрохимических свойств почв. Причем количество классов в переклассифицированном растре может совпадать (рис. 4.9) либо не совпадать (рис. 4.10) с количеством таковых в исходном растре.

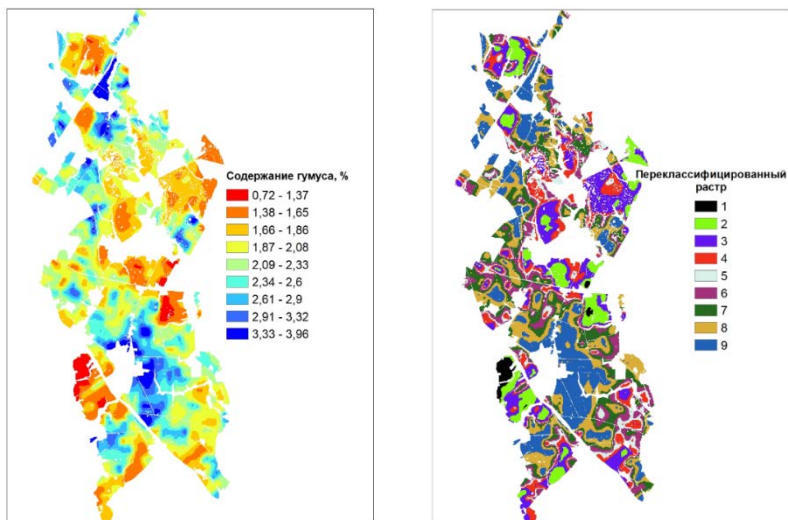


Рис. 4.9. Результат переклассификации раstra с совпадением количества классов (число градаций содержания гумуса – 9 соответствует количеству классов результирующего переклассифицированного раstra)

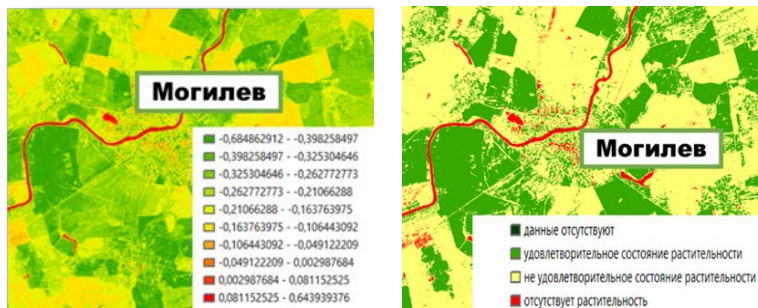


Рис. 4.10. Результат переклассификации растра без совпадения количества классов (число градаций величины вегетационного индекса не соответствует выделенным классам с различным типом развития растительности)

4.4. Локальные, зональные и фокальные операции

Локальные операции позволяют осуществлять геопространственный анализ в пределах одной ячейки растра. С помощью локальных инструментов возможно объединить входные растры, вычислить статистику или оценить критерий для каждой ячейки в выходном растре, основываясь на значениях каждой ячейки на нескольких входных растрах.

Наиболее часто используемой локальной операцией является комбинирование, позволяющее объединять в одно изображение до 20 отдельных растров (рис. 4.11).

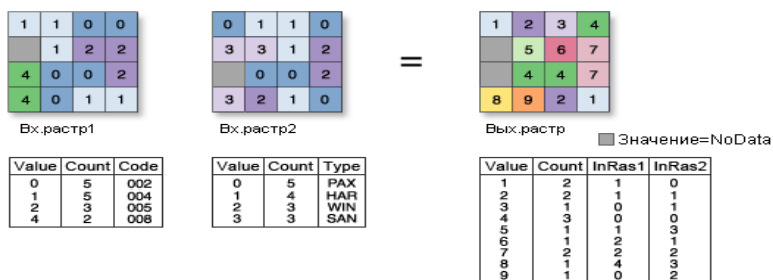


Рис. 4.11. Комбинирование растров

Характеристика локальных инструментов представлена в табл. 4.3.

Таблица 4.3. Характеристика локальных инструментов

Название инструмента	Функции инструмента
Статистика по ячейкам	Вычисляет статистику по ячейкам на основании значений из нескольких растров (большинство*, максимум, среднее, медиана, минимум, меньшинство**, диапазон, среднеквадратическое отклонение, сумма, разнообразие***).
Комбинировать	Комбинирует несколько растров, так что уникальное выходное значение присваивается каждой уникальной комбинации входных значений
Равно частоте	Оценивает по принципу «ячейка-за-ячейкой», во сколько раз значения в наборе растров равны другому растру
Частота больше, чем	Оценивает по принципу «ячейка-за-ячейкой», во сколько раз набор растров больше, чем другой растр
Максимальное положение	Определяет по принципу «ячейка-за-ячейкой» положение растра с максимальным значением в наборе растров
Частота меньше, чем	Оценивает по принципу «ячейка-за-ячейкой», во сколько раз набор растров меньше, чем другой растр
Минимальное положение	Определяет по принципу «ячейка-за-ячейкой» положение растра с минимальным значением в наборе растров
Распространенность	Определяет значение в списке аргументов, который находится на определенном уровне распространенности по принципу «ячейка-за-ячейкой». Конкретный уровень распространенности (количество повторов каждого значения) задается первым аргументом
Ранг	Значения из набора входных данных ранжируются по принципу «ячейка-за-ячейкой», и то, какие из них возвращаются, определяется значением входных растров ранга

*Большинство – значения, чаще всего встречающиеся среди входных растров.

**Меньшинство – значения, реже всего встречающиеся среди входных растров.

***Разнообразие – количество уникальных значений среди входных растров.

Зональные операции позволяют выполнять анализ, выходные данные которого являются результатом вычислений, выполненных на всех ячейках, принадлежащих каждой входной зоне. Зона может быть определена как одна область конкретного значения, но она также может состоять из нескольких отсоединенных элементов, или регионов, все из которых имеют одно значение. Зоны могут задаваться растром или наборами классов объектов. Примером зон могут служить контура почвенных разностей, менеджмент-зоны, выделяемые в пределах зем-

лепользования при внедрении системы точного земледелия, контура с тем либо иным почвенным показателем либо параметром в пределах поля.

Одни зональные операции оценивают количество определенных свойств геометрии или формы входных данных зоны и им не требуются другие входные данные. Другие зональные операции используют входные данные зоны для определения местоположений, для которых будут вычисляться другие параметры, например, статистика, площади или частота значений. Существуют также зональные операции, которые используются для заполнения указанных зон минимальным значением, находящимся на границе зоны (табл. 4.4).

Таблица 4.4. Характеристика зональных инструментов

Название инструмента	Функции инструмента
Таблица площадей	Вычисляет перекрестную таблицу площадей между двумя наборами данных и выдает результирующую таблицу
Заполнение зон	Выполняет заполнение зон, используя минимальное значение ячейки растра весов вдоль границы зоны
Зональная геометрия	Вычисляет для каждой зоны в наборе данных определенную геометрическую характеристику (площадь, периметр, толщину либо характеристики эллипса)
Зональная геометрия в таблице	Вычисляет для каждой зоны в наборе данных измерения геометрии (площадь, периметр, толщину и характеристики эллипса) и выдает результаты в виде таблицы
Зональная гистограмма	Создает таблицу и график гистограммы, показывающие плотность распределения значений ячеек на входных данных
Зональная статистика	Вычисляет статистику по значениям растра в пределах зон, определяемых другим набором данных
Зональная статистика в таблице	Суммирует значения растра в пределах зон, определяемых другим набором данных, и записывает результаты в таблицу

Наиболее часто используемой зональной операцией является построение таблицы площадей, позволяющей вычислить площадь каждого контура в пределах каждого земельного участка. В результате выполненных действий будет создана таблица, в которой поля FID 0, FID 1 ... FID N – порядковые номера участков, а значения в каждом поле – площадь контура земель в пределах каждого конкретного участка, соответствующая тому либо иному классу качества (рис. 4.12).

OBJECTID *	VALUE	FID 0	FID 1	FID 2	FID 3	FID 4	FID 5
1	1	0	0	0	0	0	0
2	2	0	0	0	0	0	0
3	3	27951,459209	53197,938494	82952,717652	86559,35755	55902,918418	58607,898341
4	4	23443,159336	20738,179413	0	0	0	0
5	5	0	0	0	0	0	0

Рис. 4.12. Фрагмент таблицы площадей

Фокальные операции позволяют выполнять анализ окрестности, создавая выходные значения для каждого положения ячейки на основании значения в этой точке и значений, определяемых в заданной окрестности. Окрестность может быть двух типов: движущаяся или радиус поиска. Движущаяся окрестность может быть либо перекрывающейся, либо неперекрывающейся (рис. 4.13).

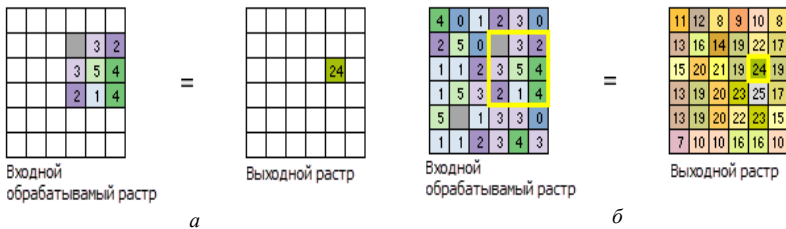


Рис. 4.13. Пример фокальной статистики при вычислении на одной ячейке входного растра (*a*) и на всех ячейках входного растра (*б*)

Окрестность может быть в форме кольца, окружности, прямоугольника и клина. В ее пределах можно вычислить следующие статистические параметры (рис. 4.14):

- среднее и медианное значения;
- большинство и разнообразие значений;
- максимальное и минимальное значения;
- диапазон значений;
- стандартное отклонение;
- сумма всех значений.

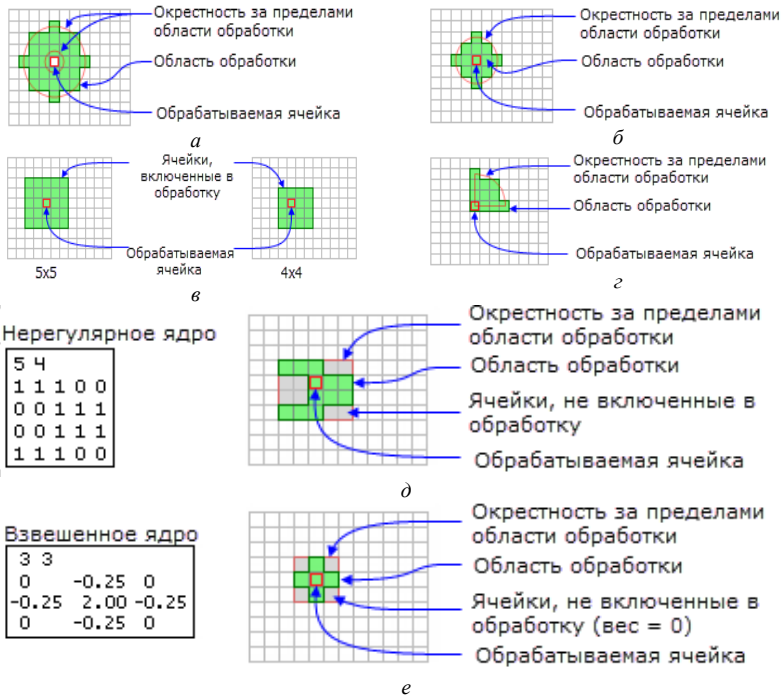


Рис. 4.14. Примеры окрестностей фокальной статистики: *а* – с кольцевой окрестностью (внутренний радиус – 1 ячейка, внешний радиус – 3 ячейки); *б* – с круговой окрестностью (радиус – 2 ячейки); *в* – с прямоугольной окрестностью; *г* – с клиновидной окрестностью (радиус – 3 ячейки, начальный угол = 0, конечный угол = 90); *д* – с окрестностью неправильной формы; *е* – с взвешенной окрестностью

4.5. Операции вычислительной геометрии. Оверлейные операции. Построение буферных зон

Вычислительная геометрия – раздел геоинформатики, в котором рассматриваются алгоритмы для решения геометрических задач. Наиболее широко используемыми операциями вычислительной геометрии являются оверлейные операции и построение буферных зон.

Оверлейные операции (overlay) – это действия, в результате которых выполняется объединение пространственных характеристик покрытий в новый слой и реляционное соединение их атрибутивных таб-

лиц. Оверлейные операции представляют собой ГИС-операцию наложения друг на друга двух или более слоев, в результате чего образуется графическая композиция (графический оверлей) используемых слоев, либо единственный результирующий слой, несущий в себе набор пространственных объектов исходных слоев, топологию этого набора и атрибуты, которые являются производными от значений атрибутов исходных объектов (рис. 4.15).

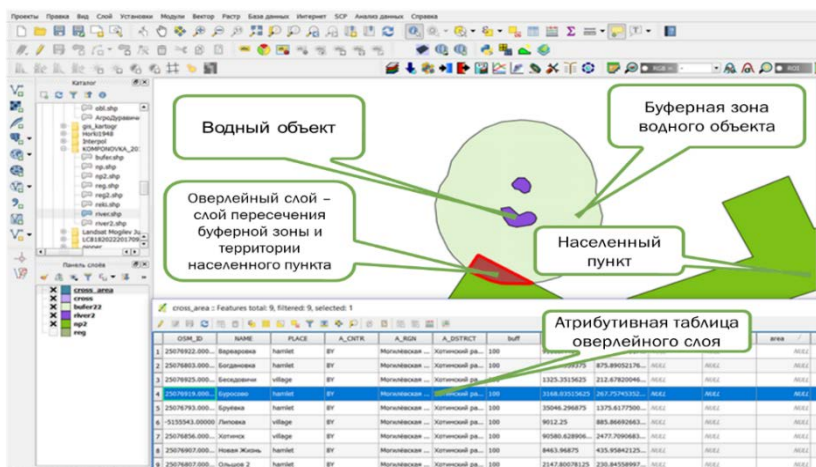


Рис. 4.15. Пример создания оверлейного слоя в QGIS

Они широко используются при агроэкологическом зонировании территории, кадастровой оценке земель населенных пунктов, формировании слоя ограничений ЗИС⁴⁷. С помощью оверлейных операций выполняют формирование нового слоя посредством наложения друг на друга двух слоев. Примерами таких операций могут служить следующие: вырезание объектов одной темы с использованием объектов другой; объединение двух однотипных тем в одну; разрезание объектов одной темы объектами другой. Некоторые оверлейные операции

⁴⁷Земельно-информационная система Республики Беларусь (ЗИС) – комплекс программно-технических средств, баз пространственно-атрибутивных данных, каналов информационного обмена и других ресурсов, обеспечивающий автоматизацию накопления, обработки, хранения и предоставления сведений о состоянии, распределении и использовании земельных ресурсов в электронном виде, в том числе средствами геоинформационных технологий.

можно выполнить только с полигональными объектами, в других могут участвовать также линейные и точечные темы. Целью выполнения оверлейных операций является ответ на вопрос: «Что находится поверх чего?».

Например:

1. Какой тип землепользования соответствует тому или иному типу почвы?

2. Какие земельные участки находятся в пределах зоны, подвергавшейся затоплению в последние 100 лет?

3. Какие дороги расположены в границах определенных административных районов?

4. Какие водоохранные зоны пересекаются с территориями населенных пунктов на определенной территории?

Наиболее широко используемыми инструментами для выполнения оверлейных операций являются следующие: стирание; идентичность; пересечение; симметричная разность; объединение и обновление (рис. 4.16).

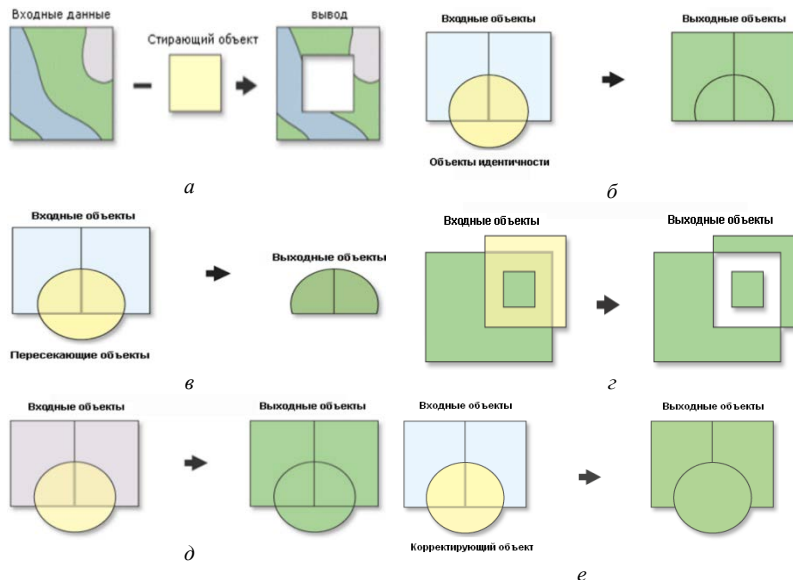


Рис. 4.16. Примеры инструментов для выполнения оверлейных операций: а – стирание; б – идентичность; в – пересечение; г – симметричная разность; д – объединение; е – обновление

Инструмент «стирание» создает класс пространственных объектов путем наложения входных объектов на полигоны стирающих объектов. В выходной класс объектов копируются только те части входных объектов, которые выходят за пределы границ стирающих полигонов.

Инструмент «идентичность» вычисляет геометрическое пересечение между входными объектами и объектами идентичности. К входным объектам или их частям, которые совпадают с объектами идентичности, присоединяются атрибуты соответствующих объектов идентичности.

Инструмент «пересечение» вычисляет геометрическое пересечение между входными объектами, при этом пространственные объекты или части объектов, которые перекрываются во всех слоях и (или) классах пространственных объектов, записываются в выходной класс объектов.

Применение инструмента «симметричная разность» обеспечивает создание выходного класса объектов с неперекрывающимися областями входных пространственных объектов и корректирующих пространственных объектов.

Инструмент «объединение» вычисляет геометрическое объединение входных объектов, при этом в выходной класс объектов записываются все объекты и их атрибуты.

Инструмент «обновление» вычисляет геометрическое пересечение входных объектов и корректирующих объектов, при этом атрибуты и геометрия входных объектов заменяются атрибутами и геометрией корректирующих объектов.

Для реализации топологического наложения в ГИС используется **алгебра логики** – раздел математической логики, в котором изучаются логические операции над высказываниями. Чаще всего предполагается, что высказывания могут быть только истинными или ложными, т. е. используется так называемая бинарная, или двоичная, логика. В алгебре логики истинностные значения высказываний принято обозначать числами 1 (истина – True) и 0 (ложь – False). Каждой логической операции соответствует функция, принимающая значения 1; 0. Такие функции называются функциями алгебры логики, или **булевыми функциями**. Два входных слоя топологического наложения можно рассматривать как два набора данных – набор А и набор В. Для них определяются следующие базовые логические операции (рис. 4.17):

– **конъюнкция** $A \cap B$ – определяет пересечение двух наборов данных, идентифицирующее те сущности, которые принадлежат и набору А, и набору В (истинно А и В);

– **дизъюнкция** $A \cup B$ – определяет объединение двух наборов данных, идентифицирующее те сущности, которые принадлежат или набору А, или набору В (истинно А или В);

– **отрицание** $A \neg B$ – определяет разность двух наборов данных, идентифицирующая те объекты, которые принадлежат А, но не В (истинно не В).

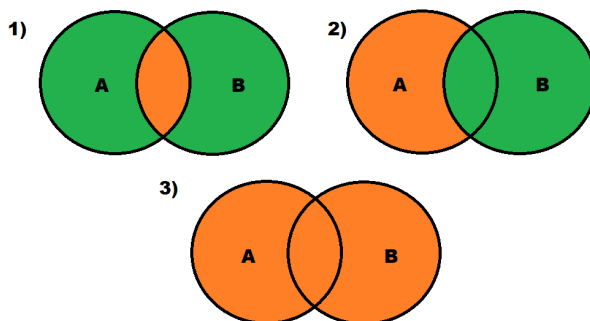


Рис. 4.17. Примеры реализации базовых логических операций:
1 – конъюнкция; 2 – отрицание; 3 – дизъюнкция

Построение буферных зон – это ГИС-операция, позволяющая определить области, окружающие геопространственные объекты. **Буферная зона** представляет собой полигональный слой, образованный путем расчета и построения эквидистант, или эквидистантных линий, равноудаленных относительно множества точечных, линейных или полигональных пространственных объектов. Построение буферной зоны заключается в очерчивании вокруг группы объектов активного слоя контура (или нескольких контуров), все точки которого отстоят от объектов группы на заданную величину. Полученный в результате контур (группа контуров) записывается в редактируемый слой.

Построение буферных зон применяется при ГИС-поддержке решения таких типов задач, как:

- проектирование санитарно-защитных зон предприятий;
- создание водоохранных зон для водных объектов;

- проектирование зон ограничений в радиусе действия линий электропередач;

- создание буферных зон вдоль автомагистралей и др.

Оно выполняется для точечных, линейных или полигональных векторных геопространственных объектов. Ширину буферных зон или **буферное расстояние** можно задавать в различных единицах измерения: метрах, километрах, единицах стоимости, величинах кратности превышения того или иного показателя. Кроме того, возможно объединение буферных зон после их построения, а также построение буферных зон с различной шириной как для одного и того же, так и для различных векторных пространственных объектов (рис. 4.18).

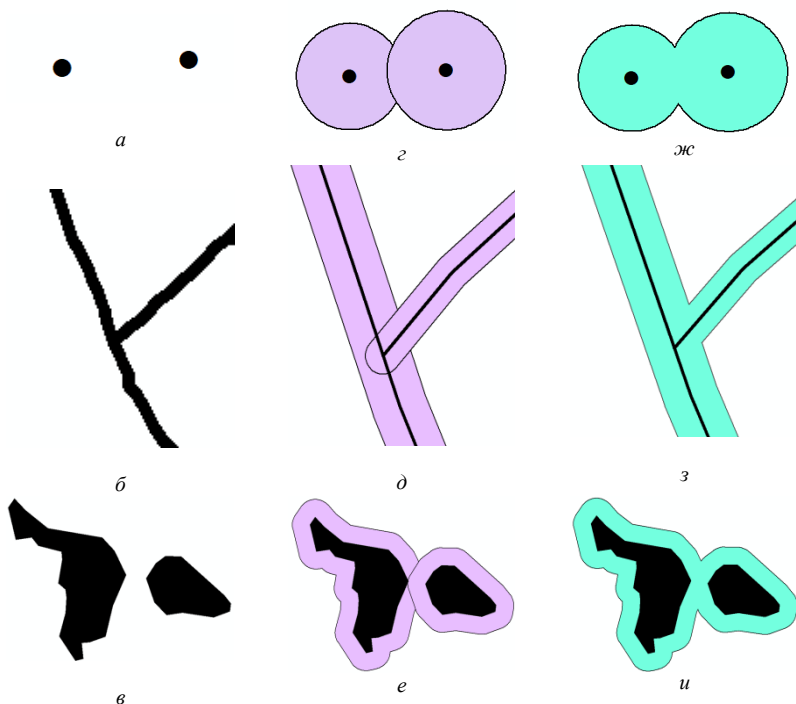


Рис. 4.18. Примеры создания буферных зон для точечных (а), линейных (б) и полигональных (в) объектов со слиянием (г-е) и без слияния (ж-и)

Существует два основных метода построения буферных зон: евклидов и геодезический.

Евклидов буфер измеряет расстояния на двухмерной координатной плоскости, где расстояние между двумя точками на плоскости измеряется по прямой, т. е. как евклидово расстояние. Он является наиболее распространенным и хорошо подходит для анализа расстояний вокруг объектов в системе координат проекции, которые сосредоточены на относительно небольшой территории (такой как одна зона в проекции Гаусса – Крюгера либо UTM).

Геодезические буферы строятся с учетом реальной формы поверхности Земли (например, эллипсоид или еще более точная модель – геоид). Расстояния измеряются между двумя точками на сферической поверхности (геоиде), что отличается от измерения того же расстояния на плоскости. Создание геодезических буферов предпочтительно в следующих случаях: 1) входные объекты располагаются далеко друг от друга (в разных зонах проекции Гаусса – Крюгера либо UTM, разных регионах мира); 2) картографическая проекция карты входных объектов сильно искажает реальные расстояния, сохраняя при этом другие свойства – в первую очередь площади.

4.6. Базы геоданных и их свойства. Проектирование баз геоданных

Современные информационные ресурсы хранятся в базах данных, включающих обычные базы данных, базы геоданных, банки пространственных данных (геоданных), информационные хранилища и фонды инфраструктуры пространственных данных. Информационные хранилища пространственных данных и фонды инфраструктуры пространственных данных, в свою очередь, используют связанные между собой базы геоданных, являющиеся основой хранения пространственной информации.

База географических данных (БГД) – совокупность взаимосвязанных пространственных и непространственных данных, организованная в наборах, классах и отношениях промышленных СУБД⁴⁸ (систем управления базами данных), и предусматривающая возможности хранения, доступа, отображения, обработки и управления данными в

⁴⁸ Система управления базами данных (СУБД) – это совокупность специальных языковых и программных средств, обеспечивающая организацию хранения данных и манипулирования ими.

среде и средствами ГИС. Различают следующие типы баз геоданных: персональные БГД; файловые БГД; многопользовательские БГД (рис. 4.19).



Рис. 4.19. Типы баз геоданных

Персональные базы геоданных используют файловую структуру данных Microsoft Access (файл.mdb). Они поддерживают базы геоданных, которые имеют ограничение по размеру 2 Гб, однако «эффективный» размер базы данных меньше и составляет от 250 и 500 Мб, а при его превышении снижается скорость работы базы данных. Персональные базы геоданных поддерживаются только операционной системой Microsoft Windows.

Файловые базы геоданных – это новая разновидность баз геоданных, назначение которых:

- обеспечение широкодоступного, простого и масштабируемого решения для всех пользователей;
- обеспечение переносимости базы геоданных, которая работает во всех операционных системах;
- постепенное наращивание для управления очень большими наборами данных;
- обеспечение отличной скорости обработки работы и масштабируемости;
- использование эффективной структуры данных, оптимизированной для работы и хранения.

Они оптимальны для работы с наборами данных на базе файлов для ГИС-проектов, для персонального использования и для использования в небольших рабочих группах. У них быстрая скорость работы и хорошая масштабируемость для содержания больших объемов данных

без необходимости использовать СУБД. Файловые базы геоданных используют около одной трети хранимой геометрии объекта по сравнению с шейп-файлами и персональными базами геоданных. Они также позволяют пользователям сжимать векторные данные в доступный только для чтения формат для сокращения дальнейших потребностей в памяти.

Многопользовательские базы геоданных обеспечивают возможность редактирования и использования одновременно многими пользователями, управления распределенной, многопользовательской базой геоданных, а также рядом важных, основанных на версиях, рабочих процессов ГИС. Многопользовательские базы геоданных работают с разнообразными моделями хранения СУБД (IBM DB2, Informix, Oracle, SQL Server и Postgre SQL).

База геоданных состоит из множества элементов и содержит три основных типа наборов данных: классы пространственных объектов; наборы растровых данных и таблицы (рис. 4.20).



Рис. 4.20. Основные элементы базы геоданных

Она способна поддерживать как простые, так и составные объекты и описывать не только статические объекты, но и динамику развития ситуации посредством создания правил отношений, задающих возможность моделирования и построения динамической ситуации.

Файловые и персональные базы геоданных проектируются для поддержки полной информационной модели базы геоданных, включающей топологии, каталоги растров, сетевые наборы данных, наборы данных Terrain, локаторы адресов. Файловые и персональные базы геоданных разрабатываются для редактирования одним пользователем и не поддерживают версиюность. В файловой базе геоданных возможна работа нескольких редактирующих пользователей в одно и то же время, с учетом того ограничения, что они осуществляют редактирование в различных наборах классов пространственных объектов, отдельных классах пространственных объектов или таблицах.

Проектирование баз геоданных предусматривает последовательное прохождение следующих этапов (рис. 4.21).



Рис. 4.21. Основные этапы проектирования базы геоданных

4.7. Подтипы, домены и их характеристика.

Классы отношений геометрической сети и определение отношений между объектами. Создание правил отношений

Подтипами называются подгруппы пространственных объектов класса пространственных объектов или объектов таблицы с одинаковыми атрибутами. Они используются для разделения групп объектов внутри класса. Например, земли одного типа – луговые можно разделить на три подтипа: луговые суходольные чистые; луговые суходоль-

ные закустаренные и луговые суходольные улучшенные. Подтипы могут иметь значения по умолчанию и домены в виде диапазонов, которые отражают эти отличия. Два подтипа внутри объектного класса могут быть связаны с разными топологиями и правилами связности и часто имеют разные значения по умолчанию для определенных атрибутов.

Атрибутивные домены – это правила, описывающие допустимые значения для различных типов полей и обеспечивающие целостность данных. Атрибутивные домены используются для ограничения диапазона значений, допустимых для определенного атрибута таблицы или класса пространственных объектов. Если объекты класса пространственных объектов или объекты таблицы сгруппированы в подтипы, каждому подтипу может быть присвоен атрибутивный домен. **Домен** – это описание допустимых значений атрибутивных данных⁴⁹. Поскольку домен ассоциирован с атрибутивным полем, для этого поля допустимы только входящие в домен значения. Другими словами, поле не примет значение, которое отсутствует в домене. Использование доменов помогает гарантировать целостность данных, ограничивая выбор значений для определенного поля.

Атрибутивные домены могут одновременно использоваться с классами пространственных объектов, таблицами и подтипами базы геоданных. При создании нового домена указывается имя, описывающее параметр, за который он отвечает.

Существуют два типа атрибутивных доменов:

1) **интервальные домены** – определяет корректный диапазон значений для числового атрибута; например, в классе объектов для слоя водотоки можно выделить подтипы для малых, средних и крупных рек: для малых рек площадь водосборного бассейна составляет 10–20 км²; для средних – 20–50 км²; для больших – свыше 50 км²;

2) **домены кодированных значений** – применяются к любым типам атрибутов: текстовым, числовым, дате и определяют валидные (соответствующие правилам, допустимые) значения атрибутов. Они содержат как значение, хранящееся в базе данных, так и описание этого значения. Например: 101 – Пахотные и залежные земли. Контроль данных при использовании доменов кодированных значений осуществляется с помощью ограничения значений, доступных для выбора (рис. 4.22).

⁴⁹ Атрибутивные данные – информация в цифровом или текстовом виде о количественных и качественных характеристиках объектов или явлений.

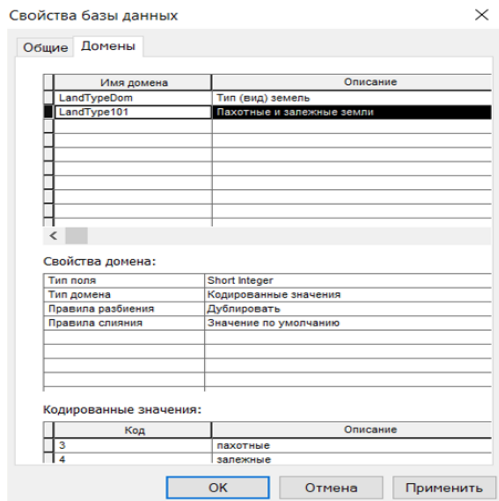


Рис. 4.22. Свойства домена кодированных значений

Классы отношений определяют отношения между объектами в базе геоданных и содержат ряд свойств, определяющих, каким образом объекты в источнике связаны с объектами в назначении. Эти отношения могут быть **простыми** и **составными**.

В *простом отношении* связанные объекты могут существовать независимо друг от друга. Например, в дорожной сети могут быть пешеходные переходы, имеющие один или более связанных с ними светофоров. Однако пешеходный переход может существовать и без светофора, а светофоры могут существовать в дорожной сети и там, где нет пешеходного перехода. Когда удаляется объект-источник в простом отношении, значение поля внешнего ключа для сопоставления объекту-адресату устанавливается как нулевое (рис. 4.23).



Рис. 4.23. Простое отношение: объекты источника и адресата независимы

Когда объект-источник удален, значение в строке внешнего ключа больше не связывается с объектом в источнике и, как результат, значение внешнего ключа устанавливается на нулевое и больше не используется. Основной задачей внешнего ключа является поддержание отношения между объектом-адресатом и связанным с ним объектом-источником. Если отсутствует объект-источник с соответствующим значением первичного ключа, тогда нет необходимости в поддержании значения внешнего ключа.

Сложные отношения также поддерживают целостность на уровне ссылок при удалении объектов, но реализуется это по-другому. В сложном отношении объекты-адресаты не могут существовать независимо от объектов-источников, т. е. при удалении источника связанный объект-адресат тоже удаляется в процессе каскадного удаления (рис. 4.24).



Рис. 4.24. Сложное отношение: объекты источника управляют объектами адресата

При создании правила отношений необходимо выбрать одну из кардинальностей. **Кардинальность отношения** определяет количество объектов в классе-источнике, которые могут быть связаны с объектами в классе-адресате. Отношение может иметь одну из трех кардинальностей: «один-к-одному», «один-к-многим» или «многие-к-многим» (рис. 4.25).

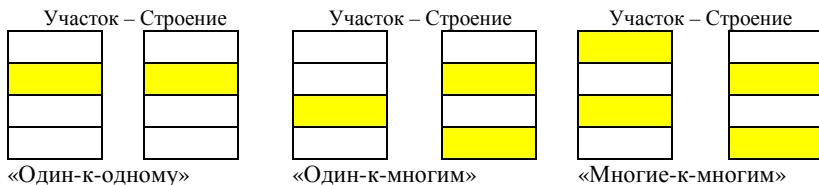


Рис. 4.25. Типы кардинальностей отношений

При кардиальности «один-к-одному» один объект-источник может быть связан только с одним объектом-адресатом. Например, населенный пункт может иметь только одно официальное название.

При кардиальности «один-к-многим» один объект-источник может быть связан с несколькими объектами-адресатами. Например, один земельный участок может иметь несколько зданий.

При кардиальности «многие-к-многим» один объект-источник может быть привязан к нескольким объектам-адресатам и, наоборот, один объект-адресат может быть привязан к нескольким объектам-источникам. Например, у одного объекта собственности может быть несколько владельцев, или у одного владельца может быть несколько объектов собственности.

После создания отношения можно уточнить кардиальность с помощью установки правил для отношения, определяющих, например, количество объектов в источнике, которые могут быть привязаны к объектам в адресате.

Вопросы для самостоятельного изучения

1. Классификация изображения с использованием машинного обучения методом опорных векторов (SVM).
2. Системы управления пространственными базами данных (СУБД).
3. Характеристика наиболее часто используемых СУБД (Oracle, dBASE, Access, Postgree).

Источники информации

1. Воронцов, К. В. Лекции по методу опорных векторов [Электронный ресурс] / К. В. Воронцов, 2007. – Режим доступа: <http://www.ccas.ru/voron/download/SVM.pdf>.
2. Шекхар, Ш. Основы пространственных баз данных / Ш. Шекхар, С. Чаула. – Москва: Кудлиц-Образ, 2004. – 336 с.
3. Richards, J. A. Remote sensing digital image analysis: an introduction / J. A. Richards. – Fifth Edition. – Springer, 2013. – Section 8.14 (p. 276) – 8.17 (p. 287).

Тестовые задания для проверки знаний

1. Дополните определение. Преобразования раstra в слой полигональных, линейных либо точечных объектов называют:

1) конвертация растровых поверхностей в векторные пространственные объекты;

2) конвертация поверхности TIN в векторные пространственные объекты;

3) конвертация набора данных Terrain в векторные пространственные объекты.

2. *Правильным или нет является утверждение?* TIN-модель пространственных данных – это способ моделирования непрерывной поверхности точками и значениями в этих точках, выбранными с постоянной плотностью.

1) да;

2) нет.

3. *Дополните определение.* Процесс пересчета входного растра с высоким разрешением в растр с более низким разрешением называют:

1) переструктуризация данных;

2) агрегирование данных;

3) конвертация данных;

4) переклассификация данных;

5) геокодирование данных.

4. *Дополните определение.* Процесс, преобразующий описание местоположения (например, координаты, адрес или название места) в местоположение на местности, называют:

1) переструктуризация данных;

2) агрегирование данных;

3) конвертация данных;

4) переклассификация данных;

5) геокодирование данных.

5. *Правильным или нет является утверждение?* Локатор адресов содержит свойства и параметры геокодирования, адресные атрибуты базовых данных и запросы для выполнения поиска, а также набор правил разбора и сопоставления адресов, которые указывают подсистеме геокодирования, как именно следует проводить стандартизацию и сопоставление адресов.

1) да;

2) нет.

6. *Дополните определение.* Области изображения, представляющие различные классы для классификации с обучением, называют:

1) обучающая выборка;

2) спектральные кластеры;

3) репрезентативная выборка;

4) функциональная выборка;

5) кластеризованная выборка.

7. *Дополните определение.* Процесс изменения значений ячеек раstra на альтернативные значения с использованием различных методов называют:

- 1) переструктуризация данных;
- 2) агрегирование данных;
- 3) конвертация данных;
- 4) переклассификация данных;
- 5) геокодирование данных.

8. *Дополните определение.* Действия, в результате которых выполняется объединение пространственных характеристик покрытий в новый слой и реляционное соединение их атрибутивных таблиц, называют:

- 1) алгебра логики;
- 2) растровая алгебра;
- 3) оверлейные операции;
- 4) переклассификация данных;
- 5) вычислительная геометрия.

9. *Дополните определение.* Логическая операция, определяющая объединение двух наборов данных, идентифицирующее те сущности, которые принадлежат или набору А, или набору В, называется:

- 1) отрицание;
- 2) конъюнкция;
- 3) дизъюнкция;
- 4) агрегация;
- 5) консолидация.

10. *Дополните определение.* Логическая операция, определяющая пересечение двух наборов данных, идентифицирующее те сущности, которые принадлежат и набору А, и набору В, называется:

- 1) отрицание;
- 2) конъюнкция;
- 3) дизъюнкция;
- 4) агрегация;
- 5) консолидация.

11. *Дополните определение.* Логическая операция, определяющая разность двух наборов данных и идентифицирующая те объекты, которые принадлежат А, но не В, называется:

- 1) отрицание;
- 2) конъюнкция;
- 3) дизъюнкция;
- 4) агрегация;
- 5) консолидация.

12. *Дополните определение.* Полигональный слой, образованный путем расчета и построения эквидистант, или эквидистантных линий, равноудаленных относительно множества точечных, линейных или полигональных пространственных объектов, называют:

- 1) оверлей;
- 2) мультипатч;
- 3) дизъюнкция;
- 4) буферная зона;
- 5) буферное расстояние.

13. *Выберите правильный вариант(ты) ответа.* Персональная база геоданных имеет ограничение по размеру, составляющее:

- 1) 250 Мб;
- 2) 500 Мб;
- 3) 1 Гб;
- 4) 2 Гб;
- 5) 3 Гб.

14. *Правильным или нет является утверждение?* Файловые и персональные базы геоданных разрабатываются для редактирования одним пользователем и не поддерживают версиюность.

- 1) да;
- 2) нет.

15. *Дополните определение.* Подгруппы пространственных объектов класса пространственных объектов или объектов таблицы с одинаковыми атрибутами называют:

- 1) подтипами;
- 2) атрибутивными доменами;
- 3) доменами кодированных значений;
- 4) геоданными;
- 5) интервальными доменами.

16. *Дополните определение.* Правила, описывающие допустимые значения для различных типов полей и обеспечивающие целостность данных, называют:

- 1) подтипами;
- 2) атрибутивными доменами;
- 3) доменами кодированных значений;
- 4) классами отношений;
- 5) интервальными доменами.

17. *Дополните определение.* Домен, определяющий корректный диапазон значений для числового атрибута, называют:

- 1) подтипом;
- 2) атрибутивным;
- 3) доменом кодированных значений;
- 4) классом отношений;
- 5) интервальным доменом.

18. *Дополните определение.* Домен, определяющий валидные (соответствующие правилам, допустимые) значения атрибутов, называют:

- 1) подтипом;
- 2) атрибутивным;
- 3) доменом кодированных значений;
- 4) классом отношений;
- 5) интервальным доменом.

19. *Правильным или нет является утверждение?* Классы отношений определяют отношения между объектами в базе геоданных и содержат ряд свойств, определяющих, каким образом объекты в источнике связаны с объектами в назначении.

- 1) да;
- 2) нет.

20. *Выберите правильный вариант(ты) ответа.* Количество объектов в классе-источнике, которые могут быть связаны с объектами в классе-адресате, определяется:

- 1) классом отношений;
- 2) кардинальностью отношений;
- 3) теснотой отношений;
- 4) иерархией отношений;
- 5) составом отношений.

5. ГЕОПРОСТРАНСТВЕННАЯ СТАТИСТИКА

5.1. Понятие геопространственной статистики и ее прикладные аспекты.

5.2. Усредненный центр данных и индекс ближайшего соседства.

5.3. Анализ и моделирование пространственной корреляции. Анализ кластеризации геопространственных данных.

5.4. Вариография. Меры пространственной корреляции.

5.5. Построение вариограмм. Моделирование вариограмм. Анизотропия вариограмм.

5.6. Пространственный тренд и нестационарность. Моделирование пространственных взаимоотношений.

Количество часов, отведенных на изучение главы, – 6.

Аннотация темы.

Тематика лекции посвящена характеристике геостатистики как метода анализа геопространственных данных и ее прикладному применению. В пункте 5.1 рассмотрены основные принципы интеграции геопространственных и статистических данных, а также функциональные возможности геостатистики и особенности применения ее методов в практической деятельности. В пункте 5.2 изложена информация, относящаяся к расширенным операциям геостатистики, рассмотрен индекс ближайшего соседства и методика его вычисления. В пункте 5.3. приведены особенности пространственной корреляции и такой ее характеристики, как глобальный индекс Морана, а также охарактеризованы концептуальные модели пространственных отношений. Кроме того, подробно освещены вопросы кластеризации геопространственных данных и различных методов ее анализа. В пункте 5.4. рассмотрена сущность вариографии и ее основные цели. Содержание пункта 5.5. посвящено характеристике модели вариограммы и особенностям ее выбора при выполнении подбора к эмпирическим данным. В нем детально рассмотрено явление анизотропии вариограммы и его основные классы. Пункт 5.6. посвящен вопросам установления наличия пространственного тренда в данных и его учета при выполнении геостатистического моделирования. В нем также рассмотрены типы пространственных отношений, их характеристика и особенности использования при моделировании.

Литературные и информационные источники

1. Геоestatистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.

2. Митчелл, Э. Руководство по ГИС-анализу / Э. Митчелл. – ESRI: 2000. – Ч. 1: Пространственные модели и взаимосвязи. – 170 с.

Internet-ресурсы

3. Домашняя страница проекта Quantum GIS [Электронный ресурс]. – Режим доступа: <https://www.qgis.org/en/site/>.

4. Официальный сайт ArcGIS Desktop [Электронный ресурс]. – Режим доступа: <https://desktop.arcgis.com/ru/desktop/>.

5.1. Понятие геопространственной статистики и ее прикладные аспекты

Геопространственная статистика – это статистика с высокой степенью пространственного разрешения, являющаяся результатом тесной увязки статистической и геопространственной информации. Анализ, проводимый в географическом контексте, приобретает все большее значение для понимания не только математического, но и пространственного распределения данных, а также для получения качественно новых данных из уже существующих (рис. 5.1).



Рис. 5.1. Схема интеграции статистических и геопространственных данных и производства геопространственной информации⁵⁰

⁵⁰Источник: <chrome-extension://mhjfbmdgcfjbbpaeojofohoefgiehjai/index.html>.

- Основными операциями геопространственной статистики являются:
- статистическая обработка атрибутивных данных;
 - описательная статистика выборки геопространственных данных;
 - расширенные операции пространственной статистики.

Статистическая обработка атрибутов позволяет выполнять расчет значений новых атрибутивных полей на основе уже существующих. В качестве примера такой операции можно привести расчет соотношений кальция и магния в пределах рабочих участков землепользования сельскохозяйственного предприятия, осуществляемый на основе атрибутивных данных о содержании кальция и магния в почве каждого из рабочих участков, локализованных в пределах пахотных земель (рис. 5.2).

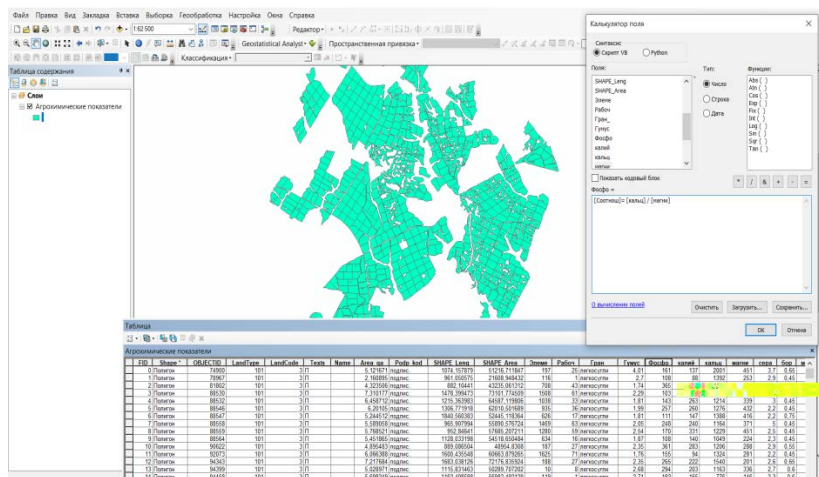


Рис. 5.2. Статистическая обработка атрибутивных данных

Функции описательной статистики выборки геопространственных данных дают возможность выполнить расчет основных элементов описательной статистики массива атрибутивных данных векторных объектов (максимальное, минимальное и среднее значения; медиану, среднеквадратическое отклонение, эксцесс и асимметрию, верхний и нижний квантили), исследовать распределение и суммарную статистику набора геоданных, а также визуализировать данные с тем либо иным набором параметров описательной статистики. На рис. 5.3 пред-

ставлены результаты расчета основных элементов описательной статистики выборки данных о содержании меди в почве рабочих участков в пределах землепользования сельскохозяйственного предприятия. Голубым цветом на схеме подсвечены участки, в пределах которых содержание меди в почве находится в интервале 1,05–1,55 мг/кг.

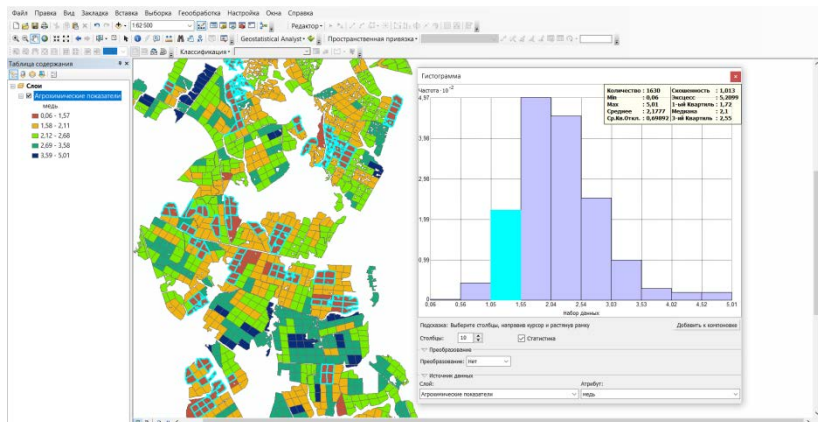


Рис. 5.3. Описательная статистика выборки геопространственных данных

Расширенные операции геопространственной статистики включают определение пространственного распределения геоданных и позволяют выполнить: анализ структурных закономерностей в распределении данных; измерение пространственного распределения данных; моделирование пространственных взаимосвязей между геоданными; расчет и анализ кластеризации данных; установление пространственной зависимости (кластеризацию или дисперсию объектов) по всему диапазону области наблюдения через заданные промежутки расстояний.

Современные возможности геоинформационных систем, в частности методы пространственного анализа, позволяют решать целый ряд различных проблем в экономике, географии, эпидемиологии, криминалистике, демографии, планировании работы экстренных служб, транспортном анализе, археологии, экологии и многих других областях, а потребность в усовершенствованных инструментах пространственного анализа постоянно возрастает (рис. 5.4).

Одним из вариантов применения методов геопространственной статистики для целей сельскохозяйственного производства и землеустройства может стать анализ пространственного распределения агрохимических свойств почв земель сельскохозяйственного назначения.

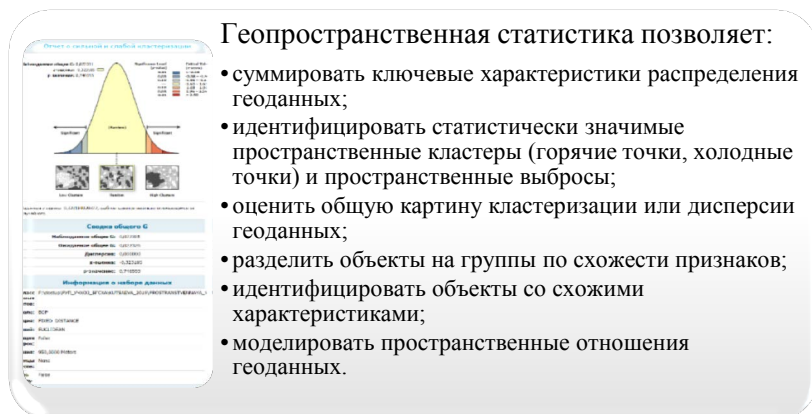


Рис. 5.4. Функциональные возможности геопространственной статистики

При выполнении такого анализа возможно:

- оценить форму и пространственную ориентацию (тренд) распределения агрохимических показателей почвы в пределах исследуемой территории;
- выявить и математически оценить пространственное распределение агрохимических показателей почвы;
- изучить пространственную автокорреляцию данных и определить местоположения в области исследования с аномальными значениями;
- оценить кластеризацию данных об агрохимических свойствах почвы и определить местоположения кластеров в пространстве;
- выполнить визуализацию кластеров путем построения карты локального индикатора пространственной ассоциативности;
- установить наиболее четкие границы между плодородными и малоплодородными землями.

Еще одним прикладным аспектом применения методов геопространственной статистики является выделение однородных по агрофизическим и агрохимическим свойствам почв и агротехнологическим характеристикам участков пахотных земель для дальнейшего внедрения системы точного земледелия.

5.2. Усредненный центр данных и индекс ближайшего соседства

Расчет усредненного центра распределения данных и индекса ближайшего соседства относится к элементам процесса измерения пространственного распределения данных и является расширенной операцией геопространственной статистики.

Усредненный центр распределения данных рассчитывается по атрибутам значений либо по значениям координат x и y . Вычисление координат усредненного центра распределения данных производится по формулам

$$\bar{X}_w = \frac{\sum x_i \cdot w_i}{\sum w_i}; \quad (5.1)$$

$$\bar{Y}_w = \frac{\sum y_i \cdot w_i}{\sum w_i}; \quad (5.2)$$

где \bar{X}_w и \bar{Y}_w – координаты усредненного центра распределения данных;

x_i и y_i – координаты каждой точки, для которой известно значение атрибута;

w_i – значение весового атрибута в точке с координатами x_i и y_i .

В случае анализа пространственного распределения агрохимических показателей почвы целесообразно определять усредненный центр распределения данных по атрибутам значений (одного либо группы агрохимических показателей).

Результаты определения усредненного центра геоданных могут использоваться:

– для проведения криминалистического анализа при определении смещения среднего центра выявления краж в светлое время суток по сравнению с ночью для оптимального распределения ресурсов по предотвращению их осуществления;

– определения среднего центра наблюдаемости видов животных в пределах природоохранного объекта для обеспечения посетителей такого рода информацией;

– оценки среднего центра вызовов доставки еды с целью оптимизации размещения служб доставки;

– оценки среднего центра, взвешенного по количеству жителей младше 25 лет, чтобы определить идеальное местоположение центров по развитию досуга молодежи;

– установления усредненного центра возникновения случаев инфекционных заболеваний, чтобы определить вероятную начальную точку эпидемии (рис. 5.5).

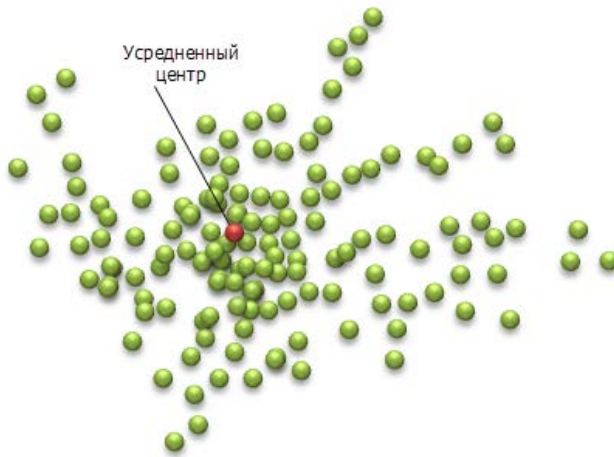


Рис. 5.5. Усредненный центр распределения данных

Пространственной вариацией усредненного центра данных является расчет **эллипса стандартного отклонения** или направленного распределения, выполняемый по формулам:

$$SDx_w = \sqrt{\frac{\sum w_i \cdot (x_i - \bar{x}_w)^2}{\sum w_i}}; \quad (5.3)$$

$$SDy_w = \sqrt{\frac{\sum w_i \cdot (y_i - \bar{y}_w)^2}{\sum w_i}}, \quad (5.4)$$

где SDx_w и SDy_w – среднеквадратичное отклонение по координатам x и y от среднего значения выборки геопространственных данных.

В результате создается новый класс пространственных объектов, содержащий эллиптический полигон с центром в усредненном центре распределения данных. Если данные сгруппированы по какому-либо признаку, создается несколько эллипсов для каждой из групп (рис. 5.6).

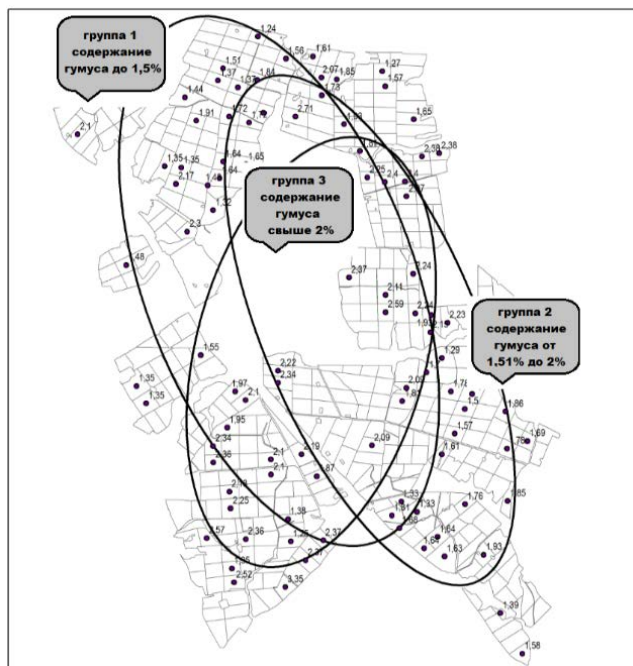


Рис. 5.6. Эллипсы распределения данных о содержании в почве гумуса для каждого из интервалов содержания

Атрибутивная таблица эллипса стандартного отклонения содержит значения координат x и y для центра полигона, величину оси x (короткая) и y (длинная) эллипса направленного распределения и угол поворота длинной оси, измеряемый по часовой стрелке. Если выходных полигонов несколько, атрибутивная таблица содержит значения для каждого из них (рис. 5.7).

При расчете направленного распределения данных можно указать количество стандартных отклонений (1, 2 или 3). Если географические данные имеют нормальное пространственное распределение (это означает, что они сконцентрированы ближе к центру и их становится значительно меньше к периферии), одно стандартное отклонение покрывает приблизительно 68 % всех входных центроидов пространственных объектов; 2 стандартных отклонения покрывают приблизительно 95 % объектов, а 3 стандартных отклонения – 99 % объектов.

FID	Shape *	Id	CenterX	CenterY	XStdDist	YStdDist	Rotation	GRUPPA
0	Полигон	0	3257643,97451	6001185,7578	4342,504442	2115,918411	159,367213	1
1	Полигон	0	3259066,68738	6000418,58562	4264,083734	1674,409307	157,005899	2
2	Полигон	0	3258147,5919	6000078,02039	1803,255013	3435,160617	14,688781	3

Рис. 5.7. Фрагмент атрибутивной таблицы эллипса стандартного отклонения

Для оценки характера распределения данных и уровня их кластеризации вычисляют величину **индекса ближайшего соседства**. Если данный показатель меньше единицы, данные распределены не случайно и в них имеются кластеризированные области. Если данный показатель больше единицы, то данные распределены равномерно и явление кластеризации в них отсутствует. Если индекс ближайшего соседства равен единице – распределение данных является случайным и нулевая гипотеза о том, что данные распределены случайно и пространственно не связаны, не отвергается (рис. 5.8).

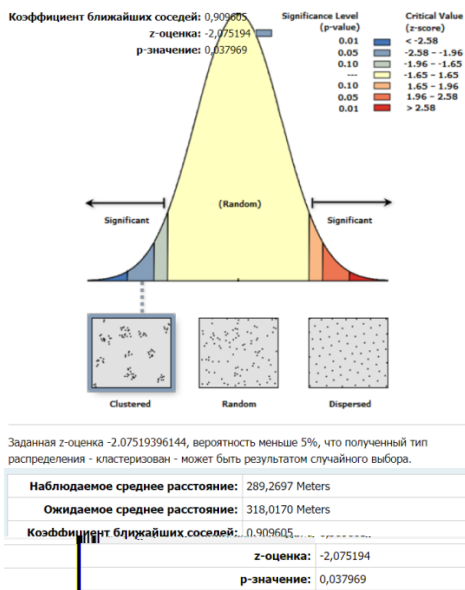


Рис. 5.8. Результат определения индекса ближайшего соседства

5.3. Анализ и моделирование пространственной корреляции. Анализ кластеризации геопространственных данных

Для множества S , состоящего из n географических единиц, пространственная автокорреляция есть соотношение между переменной, наблюдаемой в каждой из n единиц, и мерой географической близости, определенной для всех $n \cdot (n - 1)$ пар единиц из S . Иными словами, пространственная автокорреляция показывает наличие кластеров с учетом атрибутов пространственных данных.

Пространственная корреляция – зависимость между значениями пространственно распределенной функции от взаимного расположения точек, которая оценивает пространственное распределение на основе местоположений пространственных объектов и их атрибутивных значений, используя величину общего или глобального индекса Морана (I). Данный индекс является мерой пространственной автокорреляции, характеризует наличие или отсутствие пространственной автокорреляции геоданных и рассчитывается по формуле

$$I = \frac{n \sum_{i=1}^n \sum_{j=i}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{[\sum_{i=1}^n \sum_{j=i}^n w_{ij}] [\sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}} \quad (5.5)$$

где n – количество единиц в выборке;

w_{ij} – вес пространственной связи между i -й и j -й единицей выборки;

y_i – атрибутивное значение для i -й единицы выборки;

\bar{y} – выборочное среднее значение атрибута.

Глобальный индекс Морана для нормально распределенных данных лежит в диапазоне от -1 до 1 :

1) величина индекса $+1$ означает детерминированную прямую зависимость – группировку (кластеризацию) схожих (низких или высоких) значений и свидетельствует о наличии достоверной кластеризации данных;

2) величина индекса 0 означает абсолютно случайное распределение данных;

3) величина индекса -1 означает детерминированную обратную зависимость – идеальное перемешивание низких и высоких значений, напоминающее шахматную доску, что свидетельствует о равномерном распределении данных и отсутствии их кластеризации (рис. 5.9).

Несмотря на то что при определении глобального индекса Морана в качестве входных используют атрибутивные значения, поскольку выполняется оперирование именно геопространственными данными,

для учета их пространственного расположения используют матрицу величин пространственных весов объектов.

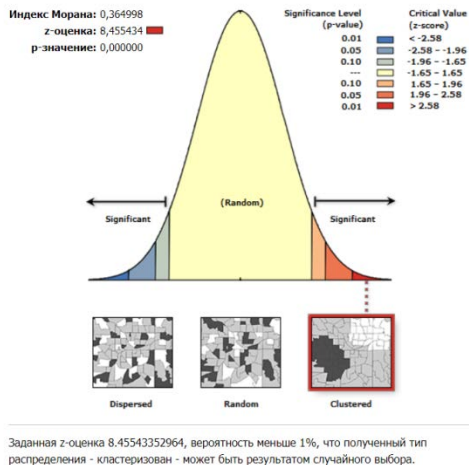


Рис. 5.9. Результат расчета глобального индекса Морана

Пространственные веса характеризуют силу связи между единицами в выборке. Если единицы не являются соседними, то пространственный вес их связи будет равен нулю. Во всех остальных случаях веса будут ненулевыми. Поскольку теоретически каждая единица может быть связана с любой другой единицей, распространена форма представления весов в виде матрицы w размером $n \times n$, где n – число единиц. На пересечении i -й строки и j -го столбца матрицы располагается вес связи между i -й и j -й единицей выборки. Все объекты должны иметь, по крайней мере, одного соседа и ни один объект не должен иметь в качестве соседей все другие объекты. Смежными считаются объекты, границы которых имеют общие точки. При этом возможно два **варианта соседства**: по правилу ферзя (8 соседей) и по правилу ладьи (4 соседа). В первом случае соседними будут считаться все пары территориальных единиц, имеющие хотя бы одну общую точку на границе, т. е. соприкасающиеся сторонами и (или) углами. Соседство по правилу ладьи разрешает только наличие общих сторон вдоль границ, а точечные касания в расчет не берутся (рис. 5.10).

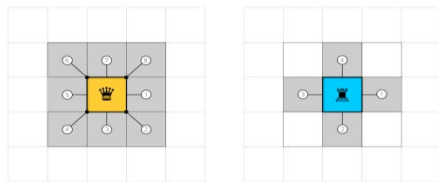


Рис. 5.10. Поиск географических соседей по правилам ферзя и ладьи

Важное отличие между пространственной и традиционной (пространственной или непространственной) статистикой состоит в том, что пространственная статистика интегрирует пространство и пространственные отношения непосредственно в вычисления. В связи с этим необходимо выбрать общую концептуальную модель пространственных отношений, которая будет использоваться для анализа. Различают следующие *концептуальные модели пространственных отношений*:

- модель 1: близко расположенные соседние объекты оказывают большее влияние на вычисления для целевого объекта, чем удаленные объекты;

- модель 2: близко расположенные соседние объекты оказывают большее влияние на вычисления для целевого объекта, чем удаленные объекты, однако угол наклона острее, влияние объектов уменьшается быстрее, и только ближайшие соседи оказывают существенное влияние;

- модель 3: каждый объект анализируется в контексте соседних объектов в пределах указанного порогового расстояния, за пределами которого влияние соседних объектов не учитывается;

- модель 4: объекты в пределах указанного порогового расстояния оказывают влияние на вычисления для целевого объекта, а влияние объектов за его пределами постепенно уменьшается;

- модель 5: пространственные отношения определены путем вычисления матрицы пространственных весов.

Выбор той либо иной концептуальной модели зависит от того, что именно анализируется. Если анализируется кластеризация отдельных видов размножающихся семенами растений, вероятно, наиболее пригодной будет концептуальная модель 1. Для некоторых видов анализа пространство и время могут иметь меньшее значение, чем абстрактные концепции, такие, как знакомство с чем-либо (чем более знакомо что-либо, тем большей функциональностью оно обладает) или простран-

ственное взаимодействие (например, между Минском и Брестом совершается значительно больше телефонных звонков, чем между Минском и небольшим городом рядом с ним, таким как Вилейка, – можно предположить, что Брест и Минск функционально (не территориально) ближе), поэтому оптимальной будет модель 2.

Для определения величины фиксированного расстояния или минимального расстояния окрестности поиска соседства используют **пошаговую пространственную автокорреляцию**, для которой задают величину начального (расстояние, на котором необходимо начать анализ пространственной автокорреляции) и приращенного (расстояние, на которое необходимо увеличивать начальное расстояние при каждой последующей итерации – lag^{51}) расстояний. При выполнении пошаговой пространственной автокорреляции выделяются десять интервалов расстояний, равномерно распределенных по всему экстенду. Для каждого интервала рассчитывается глобальный индекс Морана и интервал, для которого данный индекс будет наибольшим, рекомендуется как оптимальное расстояние для окрестности поиска соседства. В результате строится граф, на котором отмечены минимальное и максимальное расстояния окрестности поиска ближайшего соседства (рис. 5.11).

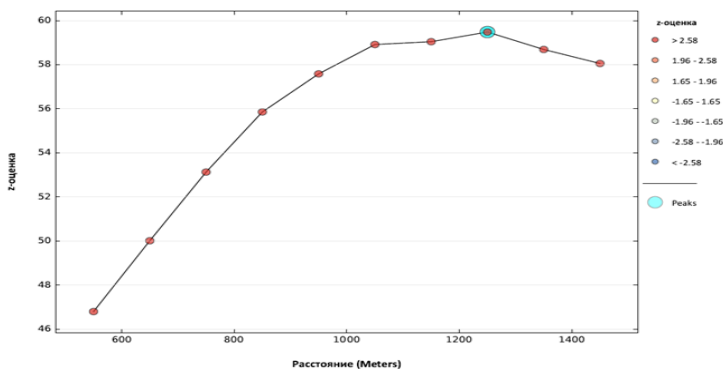


Рис. 5.11. Графическая интерпретация величины минимального расстояния окрестности поиска соседства между значениями содержания гумуса в почве

⁵¹Лag (от англ. *lag* – расстояние, которое выбирается для поиска пар точек при расчете моментов второго порядка (вариограммы, ковариации).

Именно величину лага следует учитывать при подборе шага в процессе создания мониторинговой сети наблюдений за качественным состоянием земель для целей точного земледелия.

В случае установления кластеризации геоданных следующим шагом является установление того, какие именно данные – с высокими или низкими значениями подвержены кластеризации, т. е. выполнить качественное исследование выборки геоданных. Его проводят посредством **анализа горячих точек** (от англ. Hot Spot Analysis, HSA) – инструмента предварительной оценки геоданных, который ориентирован на определение приоритетных областей («горячих точек») и позволяет установить наличие кластеризации данных с высокими и низкими значениями. Его выполняют посредством определения величины **индекса Getis-OrdG***, который рассчитывают по формуле

$$\text{Getis-OrdGi}^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{[n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2]}{n-1}}} \quad (5.6)$$

где x_j – атрибутивное значение объекта наблюдений;

$j, w_{i,j}$ – пространственный вес между объектами i и j ;

n – общее число объектов;

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n}, \quad (5.7)$$

где \bar{X} – выборочное среднее значение атрибута;

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}, \quad (5.8)$$

где S – стандартное отклонение от выборочного среднего значения.

В анализе горячих точек исследуются: 1) все объекты изучаемой области данных; 2) окрестность каждого из объектов; 3) изучаемая область данных. *Целью анализа* является определение наличия у окрестности объекта статистически значимых отличий изучаемого атрибута от всей области значений. Если в окрестности объекта значение изучаемого атрибута выше, чем в изучаемой области, объект является «горячей точкой», если ниже – «холодной».

Статистический показатель Getis-OrdGi^* рассчитывается для каждого пространственного объекта в наборе данных. Однако при его расчете учитываются не атрибутивные значения отдельных объектов, а

атрибутивные значения их окрестностей, которые рассчитываются для каждого объекта и сравниваются со значениями в остальной области исследований. Полученные итоговые z -оценки⁵² и p -значения⁵³ свидетельствуют том, в какой области пространства кластеризуются объекты с высокими или низкими значениями. При этом выполняется анализ каждого объекта в контексте соседних объектов. Чтобы быть статистически значимой «горячей» либо «холодной» точкой, объект должен иметь высокое или низкое значение и быть окружен другими объектами с такими же высокими или низкими значениями. Локальная сумма значений для объекта и его соседей сравнивается пропорционально с суммой всех объектов; когда локальная сумма сильно отличается от ожидаемой локальной суммы и когда это отличие является слишком большим, чтобы быть результатом случайного процесса, получается статистически значимая z -оценка. Высокое значение z -оценки и небольшая величина p -значения для функции указывают на наличие статистически значимой «горячей точки». Низкое отрицательное значение z -оценки и небольшая величина p -значения указывают на наличие статистически значимой «холодной точки». Чем выше (или ниже) показатель z , тем более интенсивная кластеризация данных. Значения z -оценки, близкие к нулю, означают, что пространственная кластеризация данных отсутствует.

В примере, приведенном на рис. 5.12, в результате выполнения анализа горячих точек определены три статистически значимые пространственные кластеры высоких значений «горячих точек» и один статистически значимый пространственный кластер низких значений «холодных точек» с уровнем значимости 99 %.

Если необходимо установить, отличается ли атрибутивное значение объекта от его окрестности, а окрестность объекта от остальной территории исследований, выполняют анализ кластеров и выбросов. Данный анализ позволяет идентифицировать концентрации высоких значений, концентрации низких значений и пространственные выбросы геопространственных данных (рис. 5.13).

⁵² z -оценка (англ. *z-score*) позволяет определить количество стандартных отклонений от среднего значения для рассматриваемого значения из выборки.

⁵³ p -значение (англ. *p-value*) – величина, используемая при тестировании статистических гипотез; вероятность ошибки при отклонении нулевой гипотезы.

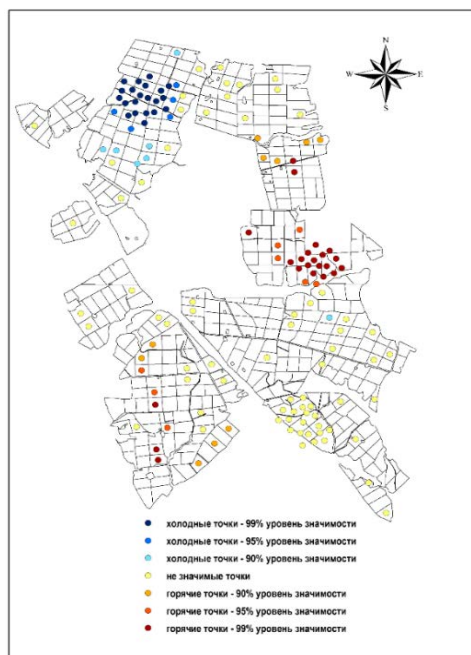


Рис. 5.12. Результаты выполнения анализа горячих точек

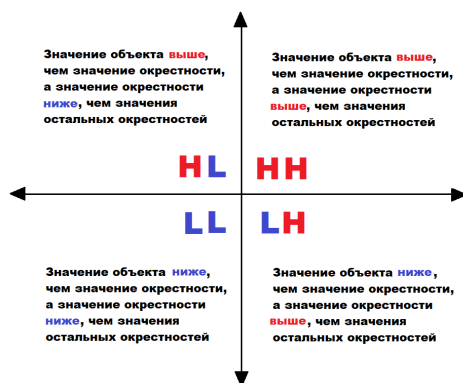


Рис. 5.13. Схема интерпретации результатов выполнения анализа кластеров и выбросов данных

Анализ кластеров и выбросов выполняют посредством определения величины **локального индекса Морана (I_i)**, который рассчитывают по формулам

$$I_i = \frac{x_j - \bar{X}}{S_i^2}, \quad (5.9)$$

$$\sum_{j=1, j \neq i}^n w_{i,j} \cdot (x_j - \bar{X}), \quad (5.10)$$

где x_j – атрибутивное значение объекта наблюдений;
 \bar{X} – выборочное среднее значение атрибута;
 $w_{i,j}$ – пространственный вес между объектами i и j ;

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_i - \bar{X})^2}{n-1}, \quad (5.11)$$

где n – общее число объектов.

С его помощью можно установить, отличаются ли значения изучаемого объекта от значений в изучаемых окрестности и области. Если в результате выполнения анализа установлено наличие кластеров типа HL и LH, то это свидетельствует о том, что в пределах исследуемой области фиксируются пространственные выбросы высоких (HL-кластеры) и низких (LH-кластеры) значений (рис. 5.14).

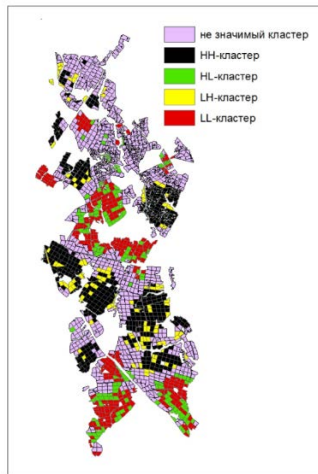


Рис. 5.14. Результат выполнения анализа кластеров и выбросов данных

Пространственные выбросы могут быть обусловлены как субъективными (например, несовершенством методики отбора проб, несовершенством применяемых методов картографирования результатов исследований), так и объективными (например, применением различных доз минеральных удобрений в пределах отдельных участков) причинами. Места их локализации требуют дополнительного изучения, а причины появления – идентификации и логического объяснения.

Важным инструментом пространственной статистики и одним из способов анализа пространственных закономерностей в случайных точечных данных является **пространственный кластерный анализ на основе множественных расстояний**, основанный на К-функции Рипли. Данный вид анализа позволяет установить пространственную зависимость (кластеризацию или дисперсию объектов) по всему диапазону области наблюдения через заданные промежутки расстояний. К-функция Рипли отражает, как центроиды⁵⁴ пространственных кластеров или дисперсий изменяются при изменении размера соседей, причем вычисляется определенное пользователем количество вариаций К-функции Рипли, каждая из которых рассчитывается по формуле

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k_{i,j}}{\pi n(n-1)}}, \quad (5.12)$$

где $L(d)$ – вариация К-функции Рипли;

d – расстояние;

n – общее число объектов;

A – область наблюдений;

$k_{i,j}$ – вес пространственной связи между i -й и j -й единицей выборки.

Результат выполнения анализа на основе множественных расстояний представляется в виде графика (рис. 5.15).

Когда наблюдаемые значения K больше, чем ожидаемые значения K для определенного расстояния, распределение данных более кластеризовано нежели случайно для обозначенного расстояния (масштаба анализа). Когда наблюдаемые значения K меньше, чем ожидаемые значения K , распределение данных более дисперсно нежели случайно для обозначенного расстояния. Когда наблюдаемые значения K больше, чем верхняя граница доверительного интервала, пространственную кластеризацию для данного расстояния можно считать статистически значимой, а когда наблюдаемые значения K меньше, чем нижняя

⁵⁴Центроид фигуры – это среднее арифметическое положений всех ее точек.

граница доверительного интервала, пространственную дисперсию для данного расстояния можно считать статистически незначимой.

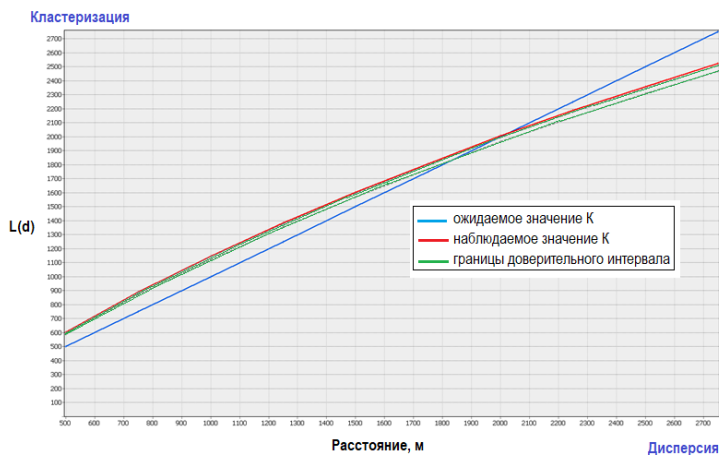


Рис. 5.15. Результат выполнения анализа на основе множественных расстояний

5.4. Вариография. Меры пространственной корреляции

Под **вариографией** понимают анализ и моделирование пространственной корреляционной структуры данных. Анализ пространственной корреляционной структуры данных подразделяется на *два этапа*:

– этап 1: построение и интерпретация мер пространственной непрерывности на основе данных;

– этап 2: моделирование пространственной корреляционной структуры и построение теоретической функции, аппроксимирующей экспериментальные значения мер корреляции аналитической формулой.

Сущность вариографии состоит в выявлении наличия корреляционной структуры в данных и ее описании. Примером может служить проверка данных на наличие или отсутствие крупномасштабного пространственного тренда⁵⁵ (видимой связи значения в точке с ее реальным местоположением), который может быть описан некоторой мате-

⁵⁵Тренд – основная тенденция изменения чего-либо, которая может быть описана различными уравнениями – линейными, логарифмическими, степенными.

матической функцией. Другой пример – проверка зависимости корреляционной структуры⁵⁶ от взаимной пространственной ориентации точек, т. е. наличия или отсутствия пространственной анизотропии⁵⁷. Также определяется *эффективный радиус корреляции данных* (если он существует) – максимальное расстояние, на котором еще наблюдается зависимость между значениями в точках.

Конечной целью вариографии является построение аналитической функции, описывающей пространственную корреляционную структуру данных для использования в геостатистических моделях интерполяции – **модели вариограммы**.

Для *описания пространственной корреляции данных* используют различные моменты второго порядка. Все они характеризуют похожесть (или непохожесть) данных в зависимости от их взаимного расположения (расстояния и направления), тем самым описывая пространственную непрерывность (см. п. 2.4 главы 2). Чтобы сделать статистические выводы о характере распределения при наличии только одной реализации случайной величины (данных измерений), требуется принять дополнительные предположения о пространственной стационарности (см. п. 2.4 главы 2). Принимается предположение о стационарности второго порядка, согласно которому функции корреляции между данными зависят только от взаимного расположения точек измерений, а не от их конкретного местоположения в пространстве. Это означает, что корреляционные функции определяются вектором h . Для изотропного случая, когда корреляция не зависит от направления, а только от расстояния, – вектор h переходит в скаляр (расстояние).

Для проведения пространственного корреляционного анализа наиболее часто используют такие моменты второго порядка, обеспечивающие различное описание пространственной непрерывности на основе двухточечной статистики (пар точек), как ковариация и полувариограмма (вариограмма).

Ковариация (covariance) – статистическая мера корреляции между двумя значениями $Z(x)$ и $Z(x + h)$ в точках, разделенных вектором h . Она *характеризует степень похожести данных* – чем более похожи данные (ближе значения), тем больше значение ковариации.

⁵⁶Корреляция – это один из основных терминов теории вероятности, показывающий меру зависимости между двумя и более случайными величинами. Данная зависимость выражается через коэффициент корреляции, который принимает значения от -1 до $+1$.

⁵⁷Анизотропия – зависимость некоторого свойства функции от ориентации аргумента.

Полувариограмма (*semivariogram*), или вариограмма, – вариация разницы значений переменной в двух точках как функция расстояния и направления между ними. Она *характеризует степень различия данных* в зависимости от расстояния между ними: чем ближе значения данных (меньше разница между ними), тем больше значение вариограммы.

Существуют также статистические моменты, аналогичные вариограмме, но отличающиеся степенью, в которую возводится разница значений пар точек:

– *мадограмма* (*madogram*) – модуль разницы значений переменной в двух точках как функции расстояния и направления между ними, позволяющий уменьшить влияние больших разбросов значений по сравнению с вариограммой

$$|M(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} |Z(x) - Z(x + h)|; \quad (5.13)$$

– *родограмма* (*rodogram*) – квадратный корень разницы значений переменной в двух точках как функции расстояния и направления между ними, позволяющий еще более снизить влияние значений с большим разбросом:

$$R(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} |Z(x) - Z(x + h)|^{0.5}; \quad (5.14)$$

– *дрейф* (*drift*) – зависимость средней разности значений функции точек от вектора расстояния между точками:

$$D(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x + h)]. \quad (5.15)$$

Дрейф служит указателем правомочности предположения о внутренней гипотезе для данных. Такой вывод можно сделать, если значение $D(h)$ колеблется вблизи нуля. Если величина $D(h)$ возрастает или убывает с ростом $|h|$, то это свидетельствует о том, что у данных имеется систематический тренд, т. е. определенная зависимость значения исследуемой функции от пространственного местоположения (координаты).

Для ковариации и вариограммы существуют также стандартизованные варианты, более устойчивые к зашумленным данным и присутствию выбросов высоких и (или) низких значений – *коррелограмма* (5.16) и *стандартизованная вариограмма*:

$$\rho(h) = \frac{C(h)}{\sigma_{-h}\sigma_{+h}}; \quad (5.16)$$

$$\gamma_{st}(h) = \frac{\gamma(h)}{\sigma_{-h}\sigma_{+h}}, \quad (5.17)$$

где σ_{-h} и σ_{+h} – стандартные отклонения для точек, находящихся соответственно в начале и в конце вектора h :

$$\sigma_{-h}^2 = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - m_{-h}]^2; \quad (5.18)$$

$$\sigma_{+h}^2 = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - m_{+h}]^2. \quad (5.19)$$

В данных может присутствовать эффект пропорциональности⁵⁸, когда: 1) среднее и вариабельность постоянны; 2) среднее имеет локальный тренд, в то время как вариабельность остается постоянной; 3) среднее постоянно, но изменяется вариабельность; 4) среднее и вариабельность изменяются вместе пропорционально. При его наличии используют общую и парную относительные вариограммы, которые нивелируют данный эффект.

5.5. Построение вариограмм. Моделирование вариограмм. Анизотропия вариограмм

Для построения вариограммы определяют квадрат разности между значениями для всех пар точек во всех местоположениях. На рис. 5.16 показан пример создания пары одного положения (красная точка) с другими одиннадцатью положениями.

На основании полученных значений квадратов разностей между соседними парами точек строится график, на оси Y которого размещаются половины квадрата разности значений между соседними парами точек, а на оси X – интервал, разделяющий соседние пары точек.

⁵⁸Эффект пропорциональности (*proportional effect*) состоит в наличии явной зависимости между локальными средними значениями и локальной вариабельностью, описываемой локальным стандартным отклонением.

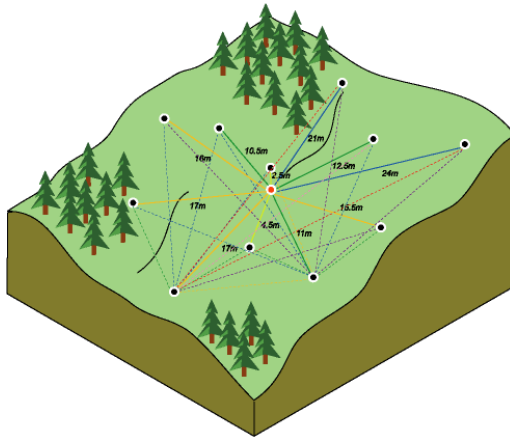


Рис. 5.16. Пример создания пар для точки с определенным местоположением с соседними точками

Полученный график называют *облаком вариограммы*, или *эмпирической вариограммой*⁵⁹ (рис. 5.17).

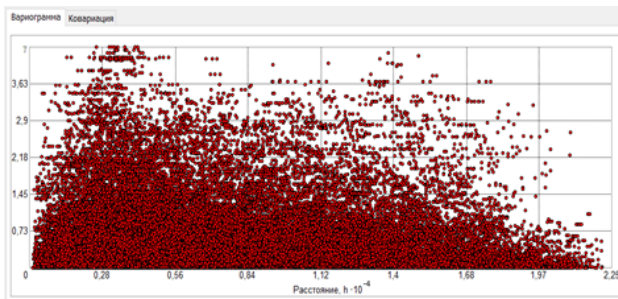


Рис. 5.17. Пример облака вариограммы

Чтобы снизить число точек на эмпирической вариограмме, пары местоположений группируются на основе расстояния, которое отделяет их друг от друга. Процесс такого группирования называется *биннингом*.

⁵⁹ Эмпирическая вариограмма – это диаграмма средних значений вариограммы на оси Y и расстояния (или лага) на оси X.

Моделирование вариограммы – процесс подбора математической модели, удовлетворяющей всем свойствам вариограммы и позволяющей описать ее для любого лага (a) (расстояния) и направления. Иными словами, это процесс выбора модели и подбора параметров, делающих ее подходящей для экспериментальной вариограммы, рассчитанной по фактическим данным. Наиболее часто используемыми типами моделей вариограммы являются сферическая, круговая, экспоненциальная, гауссова и линейная, где c_0 – самородок; c_1 – порог вариограммы (рис. 5.18).

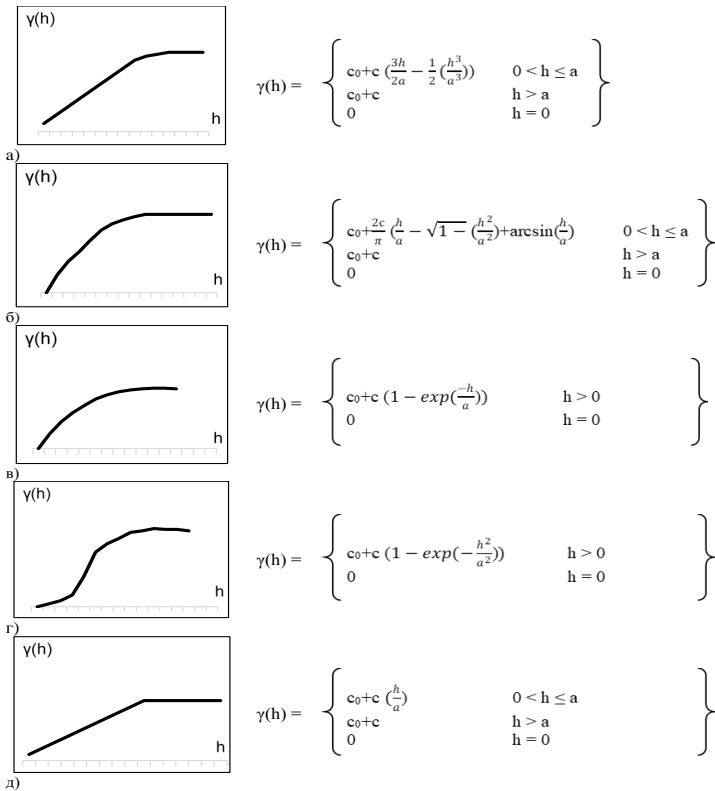


Рис. 5.18. Общие формы и уравнения математических моделей, используемых для описания вариограммы (а – эффективный радиус корреляции): а – сферическая; б – круговая; в – экспоненциальная; г – гауссова; д – линейная

Эффективный радиус корреляции – это порог вариограммы, или расстояние, на котором модель вариограммы выравнивается (рис. 5.19).

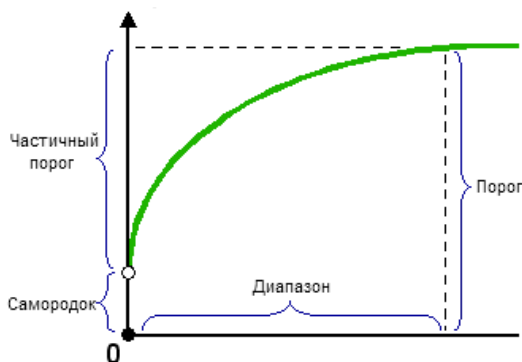


Рис. 5.19. Характеристики модели вариограммы

Исходя из того, что местоположения, близкие друг к другу, должны быть более схожи, на графике вариограммы ближайшим местоположениям (в крайней левой области по оси x) должны соответствовать невысокие значения вариограммы (низкие значения по оси y). По мере увеличения расстояния между парами местоположений (вправо по оси x) значения вариограммы должны также расти (вверх по оси y). Однако по достижении определенного расстояния облако перестает меняться. Это свидетельствует о том, что значения в парах точек, расположенных друг от друга дальше этого расстояния, больше не коррелированы. Значение, в котором модель вариограммы достигает диапазона⁶⁰ (значение на оси y), называется *порогом* (c). *Частичный порог* (c_1) – это порог минус самородок. *Самородок* (c_0) – это погрешность измерения и (или) вариация на микроуровне (изменения в пространственном масштабе, слишком мелкие для обнаружения). Для характеристики пространственной зависимости значений рассчитывается отношение дисперсии самородков к дисперсии порогов: $c_0 / (c_0 + c_1)$. Величина данного отношения менее 25 % указывает на сильную пространственную зависимость, более 75 % – на слабую пространственную зависимость; при $25\% < c_0 / (c_0 + c_1) < 75\%$ пространственная зависимость является умеренной. На практике наиболее часто приме-

⁶⁰Диапазон – это расстояние, при котором кривая модели вариограммы достигает своего предельного значения (порога).

няются сферическая, экспоненциальная и гауссова модели (или их комбинации). Модель вариограммы (ковариации), подобранная к эмпирическим данным, должна: 1) проходить через центр облака объединенных в бины значений (красные точки); 2) проходить как можно ближе к средним значениям (синие перекрестья); 3) располагаться как можно ближе к середине локальных полиномов (зеленые линии) (рис. 5.20).

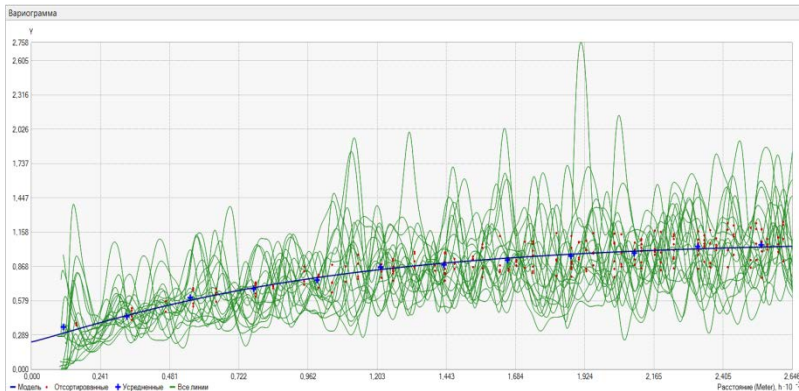


Рис. 5.20. Модель вариограммы, подобранная к эмпирическим данным

Зависимость свойств изучаемых явлений от направлений в пространстве называют *анизотропией*. Анизотропия не является детерминированным процессом, который можно описать одной математической формулой, и не имеет одного источника или влияния, которое предсказуемо воздействует на все измеренные точки. Анизотропия характерна для случайного процесса с более высокой автокорреляцией в одном конкретном направлении. Например, на дне оврага мощность пахотного горизонта почвы значительно меньше, чем на его краях. Если точка наблюдения находится на расстоянии 50 м от точки интерполяции в направлении, параллельном линии дна оврага, то значение в ней скорее будет похоже на значение в точке интерполяции, чем значение в точке наблюдения, расположенной на таком же расстоянии, но в направлении, перпендикулярном к линии дна.

Анизотропия вариограммы – изменение вариограммы и функций ковариации не только с изменением расстояния, но и с изменением

направления. Если модель вариограммы зависит только от расстояния между точками, она называется изотропной. Если же вариограмма зависит и от ориентации пары точек в пространстве, то имеет место наличие анизотропии, что означает существование структур данных с различными пространственными характеристиками в различных направлениях. Изотропная модель вариограммы одинакова во всех направлениях, в то время как анизотропная модель достигает значения порога в одних направлениях быстрее, чем в других. Длина длинной оси при движении к порогу называется *большим радиусом*, длина короткой оси называется *малым радиусом*, также отображается угол поворота линии, которая формирует большой радиус (рис. 5.21).

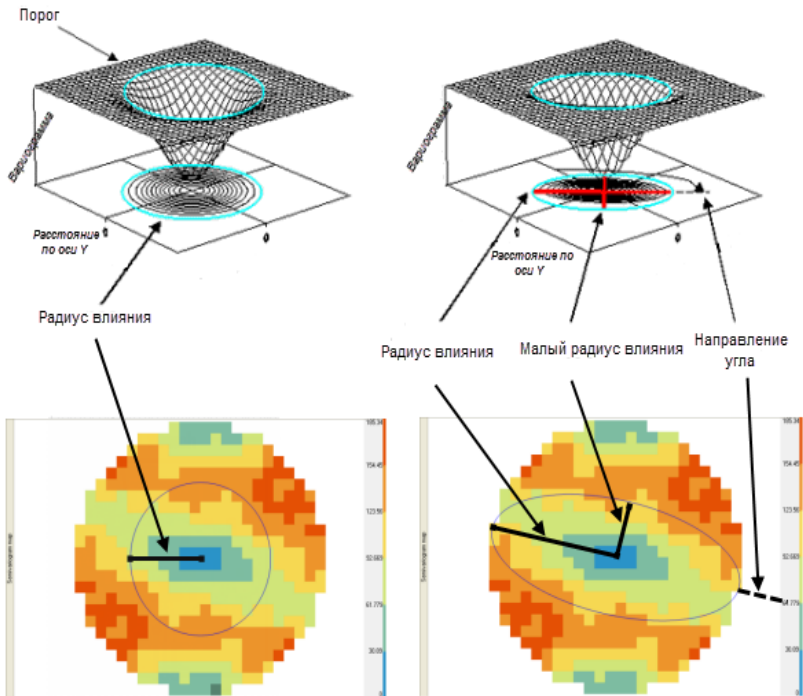


Рис. 5.21. Анизотропия вариограммы

В традиционной геостатистике анизотропию подразделяют на два класса: *геометрическую* и *зонную* (все остальные варианты анизотропии, кроме геометрической). В случае геометрической анизотропии изолинии вариограммы на вариограммной поверхности или вариограммной розе имеют форму эллипса. При негеометрической анизотропии изолинии на вариограммной поверхности или вариограммной розе образуют форму, отличную от эллипса.

Установить наличие анизотропии данных возможно посредством моделирования облака вариограммы в различных направлениях. О наличии данного явления будет свидетельствовать изменение формы облака с изменением направления поиска.

5.6. Пространственный тренд и нестационарность. Моделирование пространственных взаимоотношений

Пространственный тренд – это систематическое изменение наблюдаемой величины либо явления с изменением его местоположения (координаты). В случае присутствия тренда в данных предположение о стационарности наблюдаемой величины⁶¹ не является верным. Например, значение температуры воздуха зависит от высоты над уровнем моря, поэтому в такой местности локальное среднее значение температуры уменьшается с высотой, что нарушает предположение о стационарности среднего.

Существование трендов обычно связано с тем или иным крупномасштабным процессом либо явлением, оказывающим систематическое влияние на наблюдаемую величину. Например, топография и количество осадков влияют на почвообразовательный процесс, а содержание гумуса непосредственно связано с образованными вследствие прохождения почвообразовательного процесса типами почвы (рис. 5.22).

Другим примером может служить анализ биомассы лесного насаждения. Если вернуться назад во времени на 20 лет и выполнить анализ вплоть до сегодняшнего дня, глобальный тренд поверхности биомассы будет неизменным (систематическим), однако он будет изменяться на микроуровне. Причиной неизменности глобального тренда могут быть фиксированные факторы, например, срок жизни лесного насаждения. Вариация же на микроуровне может быть вызвана менее стабильными характеристиками, которые не просматриваются с течением времени,

⁶¹Более подробная информация приведена в п. 2.4 главы 2.

например, кислотностью осадков либо развитием болезней и вредителей леса. Поэтому предполагается, что она является случайной и несистематической.

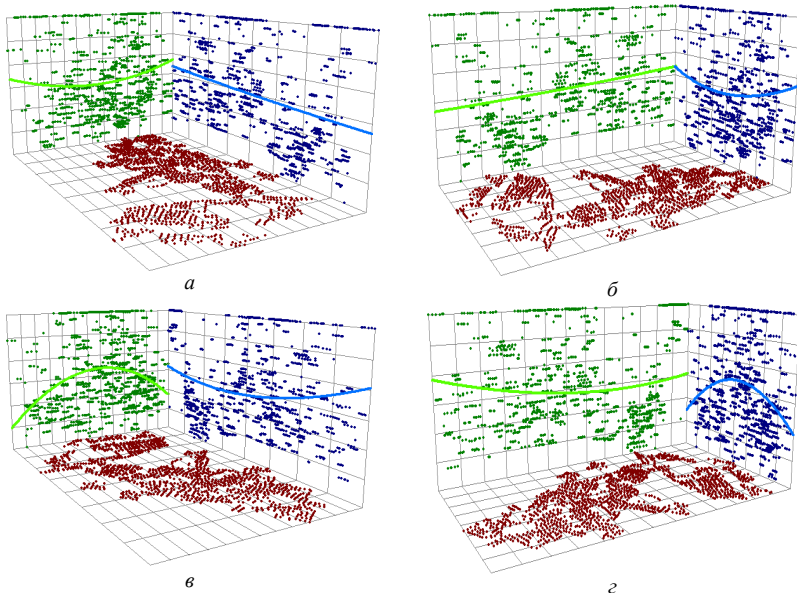


Рис. 5.22. Тренд данных о содержании гумуса в почве в различных направлениях: а – направление 90°; б – направление 180°; в – направление 240°; г – направление 340°; синяя линия – тренд в направлении север – юг, зеленая линия – тренд в направлении восток – запад

Любая поверхность может состоять из двух основных компонентов: 1) фиксированного (систематического) глобального тренда; 2) случайной вариации на микроуровне. Глобальный тренд иногда называют *структурой фиксированного среднего*. Случайная вариация на микроуровне (которую иногда называют случайной ошибкой) может моделироваться в двух частях: пространственная автокорреляция и эффект самородка.

Систематический пространственный тренд должен быть смоделирован и удален из данных измерений до построения вариограммы. В качестве моделей тренда широко используются полиномы, сплайны либо более сложные нелинейные модели. В противном случае вариограмма будет воспроизводить глобальный тренд, что приведет к поте-

ре собственной корреляции наблюдаемой величины при более мелком масштабе. Кроме того, вариограмма данных, имеющих систематический тренд, является нестационарной и не может быть использована в геостатистических моделях кригинга.

Однако существуют геостатистические модели, позволяющие оценивать данные в присутствии пространственного тренда:

– *кригинг с трендом* (или универсальный кригинг⁶²) использует модель тренда как линейную комбинацию набора базисных функций. Он прост в применении, не требует дополнительных настроек параметров, если правильно выбраны базисные функции;

– *кригинг с внешним дрейфом* использует дополнительные данные измерений коррелированной переменной в качестве модели тренда и позволяет достаточно точно оценить тренд при наличии данных дополнительной тренд-переменной во всех точках оценивания;

– *локально меняющееся среднее* использует в качестве модели тренда локальное среднее значение, которое может быть получено с помощью метода движущегося окна⁶³;

– *кригинг невязок с движущимся окном* подобен модели с локально меняющимся средним, однако вычислительно гораздо более сложен, поскольку предполагает подбор модели тренда и модели вариограммы в каждой локальной окрестности (окне);

– *внутренняя случайная функция порядка k* использует моменты более высокого порядка, чем второй, совместно с вариограммой для моделирования трендов;

– моделирование нелинейного тренда на разных масштабах с помощью *искусственной нейронной сети*⁶⁴.

Моделирование пространственных взаимоотношений должно отражать внутренние отношения между пространственными объектами, которые анализируются. Чем точнее будет смоделировано взаимодействие геопространственных объектов в пространстве, тем более точные результаты моделирования будут получены. Внутренние отношения между пространственными объектами, которые анализируются, могут быть описаны с помощью следующих видов отношений: 1) об-

⁶²Более подробная информация приведена в главе 6.

⁶³Чтобы точно отобразить каждое местоположение в исследуемой области, вариограммы вычисляют для каждого интерполируемого местоположения. В каждой окрестности данные предположительно считаются неизменяемыми, поэтому допущения алгоритма кригинга не нарушаются.

⁶⁴Искусственная нейронная сеть – математическая модель, а также ее программная или аппаратная реализация, построенная по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма.

ратно-взвешенное расстояние; 2) обратно-взвешенные расстояние и площадь; 3) фиксированный диапазон расстояний; 4) зона индифферентности; 5) смежность полигонов (по границе и (или) узлу); 6) ближайшая окрестность K ; 7) естественная окрестность (триангуляция Делоне); 8) пространственно-временное окно; 9) матрица пространственных весов (табл. 5.1).

Таблица 5.1. Типы пространственных отношений между анализируемыми объектами

Тип пространственных отношений	Характеристика типа пространственных отношений	Рекомендации по применению
1	2	3
Обратно-взвешенное расстояние	Близко расположенные соседние объекты оказывают большее влияние на вычисления для целевого объекта, чем удаленные объекты	Лучше всего подходят для непрерывных данных или для моделирования процессов, в которых два объекта сближены в пространстве настолько, что с большей вероятностью оказывают влияние друг на друга
Обратно-взвешенные расстояние и площадь	Близко расположенные соседние объекты оказывают большее влияние на вычисления для целевого объекта, чем удаленные объекты, однако угол наклона острее, влияние объектов уменьшается быстрее, и только ближайшие соседи оказывают существенное влияние	
Фиксированный диапазон расстояний	Каждый объект анализируется в контексте соседних объектов в пределах указанного порогового расстояния, за пределами которого влияние соседних объектов не учитывается	Хорошо работает с полигональными данными, имеющими различные размеры полигонов (большие полигоны по краям изучаемой области и очень маленькие полигоны в центре изучаемой области)
Зона индифферентности	Объекты в пределах указанного порогового расстояния оказывают влияние на вычисления для целевого объекта, а влияние объектов за его пределами постепенно уменьшается	Хорошо работает при правильно подобранном фиксированном расстоянии, однако не пригоден для больших наборов данных
Смежность полигонов (по ребру и (или) узлу)	На расчеты для целевого полигонального объекта влияют только полигональные объекты, имеющие смежную границу, общий узел или перекрываются	Эффективен, когда полигоны имеют одинаковые размеры и однотипное распределение, а пространственные отношения являются функцией близости полигонов

1	2	3
Ближайшая окрестность К	Пространственные отношения вычисляются только для определенного числа соседних окрестностей (4 или 8)	Работает, если распределение данных в изучаемом районе неоднородно: некоторые пространственные объекты расположены очень далеко от всех остальных объектов
Естественная окрестность (триангуляция Делоне)	Пространственные отношения вычисляются в окрестностях узлов, соединенных ребрами треугольников, создаваемых методом построения полигонов Вороного по точечным объектам или по центроидам ⁶⁵ пространственных объектов	Используется, когда в гео-данных присутствуют острова и (или) плотность пространственных объектов меняется в довольно широких пределах. Не пригоден, если в данных присутствуют совпадающие объекты
Пространственно-временное окно	Пространственные отношения вычисляются для объектов, которые расположены на определенном расстоянии и в пределах заданного временного интервала целевого объекта	Используется, если требуется создать файл матрицы пространственных весов для анализа «горячих» точек, чтобы выявить «горячие» точки в пространстве-времени
Матрица пространственных весов	Пространственные отношения определены путем вычисления матрицы пространственных весов	Используется при выполнении пространственного моделирования в терминах сети

Если необходимо установить, отличается ли атрибутивное значение объекта от его окрестности, а окрестность объекта от остальной территории исследований, выполняют анализ кластеров и выбросов, важным этапом которого является выбор оптимального типа пространственных отношений. На рис. 5.23 показан результат выполнения кластерного анализа при различных типах пространственных отношений.

⁶⁵Центроид полигона – точка, являющаяся центром его тяжести (геометрическим центром); внутренняя точка полигона со значениями координат.

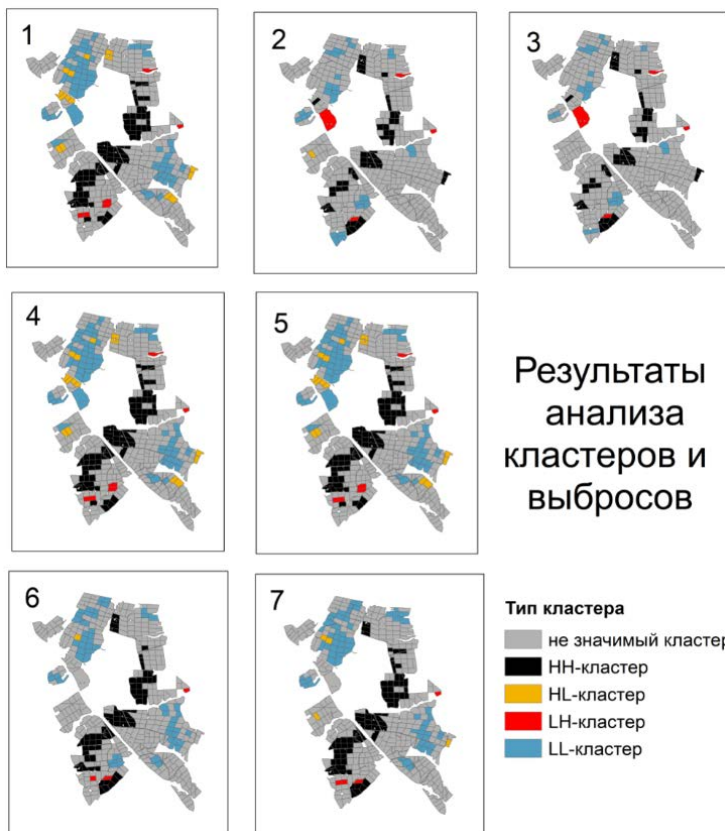


Рис. 5.23. Результаты кластерного анализа при различных типах пространственных отношений:

- 1 – матрица пространственных весов; 2 – смежность полигонов по ребрам и узлам;
 3 – смежность полигонов по ребрам; 4 – зона индифферентности; 5 – фиксированный диапазон расстояний; 6 – обратно-взвешенные расстояние и площадь;
 7 – обратно-взвешенное расстояние

Вопросы для самостоятельного изучения

1. Особенности поведения вариограмм на больших расстояниях и вблизи нуля. Анизотропия плато.
2. Анализ группирования в геопространственной статистике.

Источники информации

1. Геостатистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.
2. Kovalevskiy, E. Geological modelling on the base of geostatistics / E. Kovalevskiy. – EAGE, 2011. – 122 с.
3. Официальный сайт компании StatSoft [Электронный ресурс]. – Режим доступа: http://statsoft.ru/solutions/ExamplesBase/branches/detail.php?ELEMENT_ID=1573.

Тестовые задания для проверки знаний

1. *Правильным или нет является утверждение?* Геопространственная статистика – это статистика с высокой степенью пространственного разрешения, являющаяся результатом тесной увязки статистической и геопространственной информации.

- 1) да;
- 2) нет.

2. *Дополните определение.* Анализ структурных закономерностей в распределении данных и измерение их пространственного распределения; моделирование пространственных взаимосвязей между геоданными; расчет и анализ кластеризации данных; установление пространственной зависимости (кластеризации или дисперсии объектов) называют:

- 1) статистическая обработка атрибутов;
- 2) кардиальность отношений;
- 3) расширенные операции геостатистики;
- 4) статистические операции;
- 5) интеграция геопространственных данных.

3. *Правильным или нет является утверждение?* Усредненный центр распределения данных и индекс ближайшего соседства относятся к элементам процесса измерения пространственного распределения данных и не являются расширенными операциями геопространственной статистики.

- 1) да;
- 2) нет.

4. *Выберите правильный вариант(ты) ответа.* Пространственной вариацией усредненного центра данных является:

- 1) эллипс стандартного отклонения;
- 2) усредненный центр распределения данных;
- 3) индекс ближайшего соседства;
- 4) стандартное отклонение;
- 5) медианное значение.

5. Выберите правильный вариант(ты) ответа. Для оценки характера распределения данных и уровня их кластеризации вычисляют величину:

- 1) эллипса стандартного отклонения;
- 2) усредненного центра распределения данных;
- 3) индекса ближайшего соседства;
- 4) стандартного отклонения;
- 5) медианного значения.

6. Выберите правильный вариант(ты) ответа. Если геопространственные данные распределены не случайно и в них имеются кластеризованные области, то индекс ближайшего соседства имеет значение:

- 1) >1 ;
- 2) <1 ;
- 3) 1.

7. Выберите правильный вариант(ты) ответа. Если геопространственные данные распределены равномерно и явление кластеризации в них отсутствует, то индекс ближайшего соседства имеет значение:

- 1) >1 ;
- 2) <1 ;
- 3) 1.

8. Выберите правильный вариант(ты) ответа. Если геопространственные данные распределены случайно, то индекс ближайшего соседства имеет значение:

- 1) >1 ;
- 2) <1 ;
- 3) 1.

9. Дополните определение. Мету пространственной автокорреляции, характеризующую наличие или отсутствие пространственной автокорреляции геоданных, называют:

- 1) индекс Морисита;
- 2) индекс Getis-OrdGi*;
- 3) глобальный индекс Морана;
- 4) локальный индекс Морана;
- 5) геопространственный индекс.

10. Выберите правильный вариант(ты) ответа. Если геопространственные данные демонстрируют кластеризацию в распределении, то глобальный индекс Морана имеет значение:

- 1) >1 ;
- 2) <1 ;
- 3) 1.

11. Выберите правильный вариант(ты) ответа. Если геопространственные данные распределены равномерно, то глобальный индекс Морана имеет значение:

- 1) >1 ;
- 2) <1 ;
- 3) 1.

12. Выберите правильный вариант(ты) ответа. Если геопространственные данные распределены случайно, то глобальный индекс Морана имеет значение:

- 1) >1 ;
- 2) <1 ;
- 3) 1.

13. Правильным или нет является утверждение? При определении глобального индекса Морана в качестве входных используют атрибутивные значения окрестностей каждого объекта, которые сравниваются со значениями в остальной области исследований.

- 1) да;
- 2) нет.

14. Дополните определение. Вариант соседства, при котором соседними считаются все пары территориальных единиц, имеющие хотя бы одну общую точку на границе, т. е. соприкасающиеся сторонами и (или) углами, называют:

- 1) соседство один-к-многим;
- 2) соседство по правилу ладьи;
- 3) соседство многие-к-многим;
- 4) соседство по правилу ферзя;
- 5) соседство по правилу туры.

15. Дополните определение. Вариант соседства, при котором соседними считаются все пары территориальных единиц, имеющие общие стороны вдоль границ, а точечные касания в расчет не берутся, называют:

- 1) соседство «один-к-многим»;
- 2) соседство по правилу ладьи;
- 3) соседство «многие-к-многим»;
- 4) соседство по правилу ферзя;
- 5) соседство по правилу туры.

16. Дополните определение. Анализ, целью которого является определение наличия у окрестности объекта статистически значимых отличий изучаемого атрибута от всей области значений, называют:

- 1) геопространственным;
- 2) геостатистическим;
- 3) анализом «горячих» точек;
- 4) геоинформационным;
- 5) пространственным.

17. *Правильным или нет является утверждение?* При определении индекса Getis-OrdGⁱ* в качестве входных используют атрибутивные значения окрестностей каждого объекта, которые сравниваются со значениями в остальной области исследований.

- 1) да;
- 2) нет.

18. *Вставьте пропущенное словосочетание.* Анализ кластеров и выбросов выполняют с помощью определения величины _____, позволяющего идентифицировать концентрации высоких значений, концентрации низких значений и пространственные выбросы геопространственных данных.

- 1) индекса Морисита;
- 2) индекса Getis-OrdGⁱ*;
- 3) глобального индекса Морана;
- 4) локального индекса Морана;
- 5) геопространственного индекса.

19. *Дополните определение.* Анализ и моделирование пространственной корреляционной структуры данных, целью которого является построение аналитической функции, описывающей пространственную корреляционную структуру данных для использования в геостатистических моделях интерполяции, называют:

- 1) вариография;
- 2) моделирование;
- 3) геостатистика;
- 4) корреляция;
- 5) регрессия.

20. *Правильным или нет является утверждение?* Вариограмма характеризует степень похожести данных: чем более похожи данные (ближе значения), тем больше значение вариограммы.

- 1) да;
- 2) нет.

21. *Правильным или нет является утверждение?* Ковариация характеризует степень различия данных в зависимости от расстояния

между ними: чем ближе значения данных (меньше разница между ними), тем больше значение ковариации.

- 1) да;
- 2) нет.

22. *Дополните определение.* Процесс подбора математической модели, удовлетворяющей всем свойствам вариограммы и позволяющей описать ее для любого расстояния и направления, называют:

- 1) биннинг вариограммы;
- 2) дрейф вариограммы;
- 3) тренд вариограммы;
- 4) анизотропия вариограммы;
- 5) моделирование вариограммы.

23. *Дополните определение.* Порог вариограммы или расстояние, на котором модель вариограммы выравнивается, называют:

- 1) эффективный радиус корреляции;
- 2) диапазон вариограммы;
- 3) самородок вариограммы;
- 4) частичный порог вариограммы;
- 5) плато вариограммы.

24. *Дополните определение.* Изменение вариограммы и функций ковариации не только с изменением расстояния, но и с изменением направления, называют:

- 1) биннинг вариограммы;
- 2) дрейф вариограммы;
- 3) построение вариограммы;
- 4) анизотропия вариограммы;
- 5) моделирование вариограммы.

25. *Дополните определение.* Систематическое изменение наблюдаемой величины либо явления с изменением его местоположения (координаты) называют:

- 1) биннинг;
- 2) дрейф;
- 3) пространственный тренд;
- 4) анизотропия;
- 5) динамика.

6. ГЕОПРОСТРАНСТВЕННОЕ МОДЕЛИРОВАНИЕ

6.1. Детерминированные методы геопространственного моделирования.

6.2. Геостатистические методы пространственного моделирования.

6.3. Автоматизация геопространственного моделирования.

Количество часов, отведенных на изучение главы, – 8.

Аннотация темы.

Тематика лекции посвящена вопросам геопространственного моделирования и особенностям автоматизации процесса его выполнения. В пункте 6.1 подробно рассматриваются глобальные и локальные методы детерминированной интерполяции: метод обратных взвешенных расстояний, интерполяция по методу глобального и локального полиномов, по методу диффузного ядра и ядра с барьерами, радиальные базисные функции. В пункте 6.2 охарактеризованы геостатистические методы пространственного моделирования, рассмотрена базовая модель геостатистики – кригинг и специфические особенности простого, ординарного, универсального и эмпирического байесовского кригинга. В пункте 6.3. приведены сведения об особенностях геообработки и ее основных инструментах, рассмотрены возможности геоинформационного моделирования на примере функционала модуля Model Builder-ArcGIS.

Литературные и информационные источники

1. Геоистатистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.
2. Курлович, Д. М. Геоинформационные методы анализа и прогнозирования погоды: учеб.-метод. пособие / Д. М. Курлович. – Минск: БГУ, 2013. – 191 с.
3. Krivoruchko, K. Pragmatic Bayesian kriging for non-stationary and moderately non-Gaussian data / K. Krivoruchko, A. Gribov // Mathematics of Planet Earth, 61 Lecture Notes in Earth System Sciences. – Springer-Verlag Berlin Heidelberg, 2014. – P. 61–64.

Internet-ресурсы

4. Домашняя страница проекта Quantum GIS [Электронный ресурс]. – Режим доступа: <https://www.qgis.org/en/site/>.
5. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.

6.1. Детерминированные методы геопространственного моделирования

Чтобы создать модель непрерывной поверхности, необходимо наличие множества регулярно расположенных точек со значениями. Однако выполнить измерения в каждой точке невозможно, поэтому на практике используют выборку точек, для которых определяются значения (высота, концентрация вещества), а затем используются различные методы интерполяции, чтобы предсказать значения во всех точках по ограниченному набору опорных точек с известными значениями.

Существует две основные группы методов интерполяции: *детерминированные* и *геостатистические*. Среди детерминированных методов интерполяции наиболее широко применяются:

- метод обратных взвешенных расстояний;
- интерполяция по методу глобального полинома;
- интерполяция по методу локального полинома;
- радиальные базисные функции;
- интерполяция с барьерами, использующая проницаемые или полупроницаемые барьеры.

Методы детерминированной интерполяции создают поверхность из измеренных точек, основываясь или на степени схожести (обратные взвешенные расстояния), или на уровне сглаживания (радиальные базисные функции). Детерминированные методы интерполяции можно разделить на две группы: *глобальные* и *локальные*. **Глобальные методы** вычисляют проинтерполированные значения на основании всего набора данных. **Локальные методы** вычисляют проинтерполированные значения на основании измеренных точек в пределах окрестностей, которые являются меньшими пространственными областями внутри большей изучаемой территории.

Детерминированная интерполяция создает результирующую поверхность, как проходящую, так и не проходящую через все значения данных. Метод интерполяции, который вычисляет значение, идентичное измеренному в опорном местоположении, называется **жестким интерполятором**. **Нежесткий интерполятор** вычисляет значение, которое отличается от измеренного. Последний можно использовать с целью избежать возникновения острых вершин или углублений на выходной поверхности. Обратные взвешенные расстояния и радиальные базисные функции являются точными интерполяторами, тогда как глобальный полином, локальный полином, интерполяция ядра с барьерами и интерполяция диффузии с барьерами – это нежесткие интерполяторы.

Метод обратных взвешенных расстояний (Inverse Distance-Weights, IDW) является одним из наиболее часто применяемых детерминированных методов интерполяции. IDW однозначно предполагает, что объекты, которые находятся поблизости, более подобны друг другу, чем объекты, удаленные друг от друга. Чтобы проинтерполировать значение для неизмеренного положения, IDW использует измеренные значения вокруг интерполируемого местоположения. Наиболее близкие к проинтерполированному местоположению измеренные значения оказывают большее влияние на прогнозируемое значение, чем удаленные от него на значительное расстояние. При этом предполагается, что каждая измеренная точка оказывает локальное влияние, которое уменьшается с увеличением расстояния. Это придает больший вес точкам, расположенным ближе всего к интерполируемому местоположению. Вес точки уменьшается как функция от расстояния, поэтому метод носит название обратных взвешенных расстояний. Веса, назначенные точкам данных, представлены на рис. 6.1.

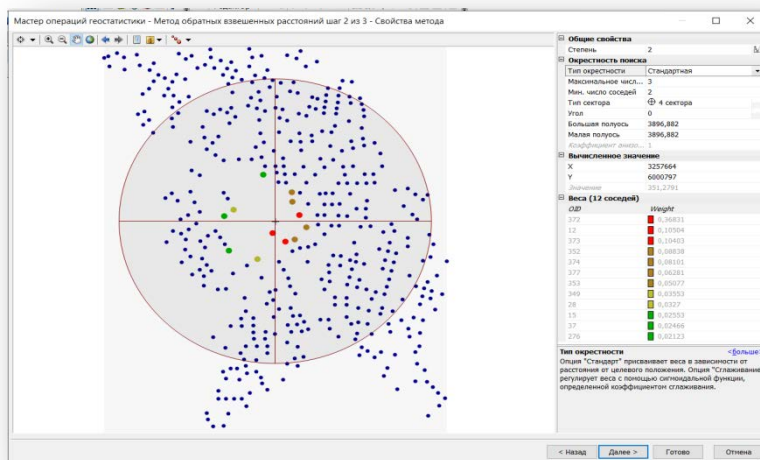


Рис. 6.1. Иллюстрация присвоения весов точкам в окрестности поиска

Каждой точке данных, которая используется для интерполяции значения в местоположении, отмеченном перекрестием, присваивается определенный вес, значение которого тем больше, чем ближе находит-

ся точка к местоположению интерполируемого значения. Список весов со значениями показан в окне веса. Поскольку была задана стандартная окрестность поиска в четырех секторах, а максимальное число соседних точек в каждом секторе ограничено тремя, общее число соседей составило 12.

Поскольку объекты, находящиеся поблизости друг от друга, более сходны, чем объекты, находящиеся на большом расстоянии, то по мере удаления от местоположений измеренные значения будут находиться в несущественных отношениях со значением проинтерполированного местоположения. В этой связи для увеличения скорости вычислений следует исключить большинство удаленных точек, которые оказывают незначительное влияние на интерполяцию, с помощью определения **окрестности поиска**. На рис. 6.2 представлены пять измеренных точек (соседи), значения в которых используются для прогнозирования значения в точке, где измерения не проводились (показана желтым цветом).

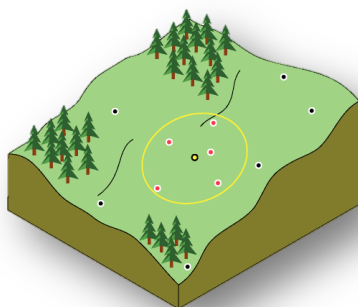


Рис. 6.2. Иллюстрация окрестности поиска

Форма окрестности накладывает ограничения на дальность и место поиска измеренных значений, используемых для интерполяции, а также оказывает влияние на входные данные и поверхность, которую необходимо создать. При отсутствии направленных воздействий на данные можно рассматривать точки одновременно во всех направлениях и определять окрестность поиска как окружность. Однако если присутствует направленное воздействие на данные, например преоб-

терполяции является проблематичным. Интерполированное значение при применении метода IDW определяется по формуле

$$z(x_0) = \frac{\sum_{i=1}^n \frac{x_i}{h_{ij}^\beta}}{\sum_{i=1}^n \frac{1}{h_{ij}^\beta}}, \quad (6.1)$$

где $z(x_0)$ – интерполируемое значение;

n – общее количество значений данных выборки;

x_i – i -е значение данных;

h_{ij} – расстояние между интерполированным значением и значением данных выборки;

β – весовое значение.

Интерполяция по методу глобального полинома (тренд-интерполяция) подбирает к входным опорным точкам сглаженную поверхность, определяемую математической функцией (полиномом). Поверхность глобального полинома постепенно изменяется и характеризует грубую структуру в данных. Концептуально интерполяция по методу глобального полинома схожа с установкой листа бумаги между поднятыми до высоты значения точки. Данное положение иллюстрирует рис. 6.4, где отображена поверхность, созданная для набора опорных точек высот, взятых на холме с небольшим уклоном (лист бумаги розового цвета).

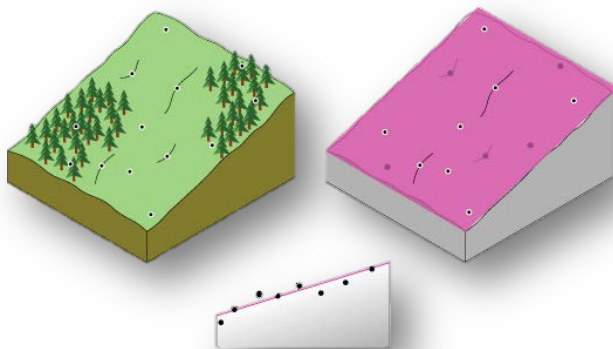


Рис. 6.4. Интерполяция по методу глобального полинома (полином первого порядка)

Очевидно, что плоский лист бумаги, имитирующий интерполяционную поверхность, не будет точно характеризовать ландшафт, содержащий долину либо несколько разновысотных возвышенностей. Однако если согнуть условный лист бумаги один раз, соответствие реальной ситуации станет значительно больше. Математически это реализуется посредством добавления члена в математическую формулу. Плоская поверхность (нет сгиба на листе бумаги) – это полином первого порядка (линейный). При допущении одного изгиба получается полином второго порядка (квадратичный), двух изгибов – полином третьего порядка (кубический) и так далее (допускается до 10 изгибов).

Однако на практике довольно редко возможен случай, когда лист бумаги (аналог интерполяционной поверхности) проходит через все фактически измеренные точки. В этой связи интерполяция по методу глобального полинома относится к нежестким методам интерполяции. Некоторые точки будут расположены выше листа бумаги, другие – ниже.

На рис. 6.5 концептуально показан полином второго порядка, позволяющий создать интерполированную поверхность, тождественную поверхности с изгибом.

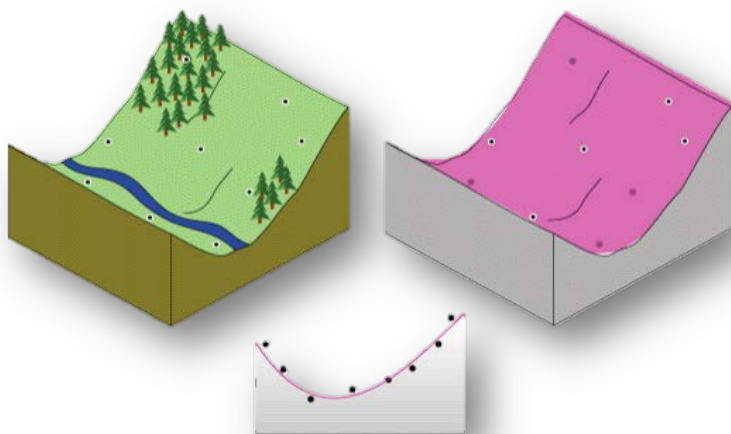


Рис. 6.5. Интерполяция по методу глобального полинома (полином второго порядка)

Однако можно добавить высоту до высоты точек выше и ниже листа бумаги (эти две суммы должны быть равны). В таком случае результирующая поверхность получается путем подбора подходящего уравнения регрессии по методу наименьших квадратов, а различия между квадратами поднятых значений и листом бумаги будут сведены к минимуму. Интерполяционные поверхности, полученные методом глобального полинома, весьма чувствительны к выпадающим значениям (чрезвычайно высоким и низким значениям), особенно на ребрах. В результате интерполяции по методу глобального полинома формируется сглаженная поверхность, которая представляет постепенные тренды поверхности в исследуемой области (рис. 6.6).

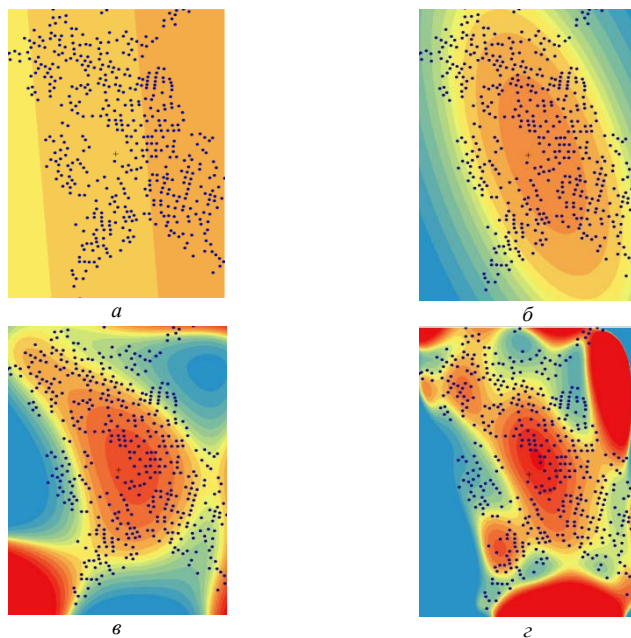


Рис. 6.6. Интерполяционная поверхность, созданная методом глобального полинома:
а – полином 1-го порядка; *б* – полином 2-го порядка; *в* – полином 3-го порядка;
г – полином 4-го порядка

Интерполяция по методу глобального полинома используется в случае:

– необходимости подбора поверхности к опорным точкам, если поверхность в пределах исследуемой области постепенно варьирует от участка к участку, например, загрязнение окружающей среды в пределах санитарно-защитной зоны промышленного предприятия;

– исследования и (или) удаления эффектов трендов с большим диапазоном, или глобальных трендов.

Метод локальных полиномов (Local Polynomial Interpolation, LPI) – это умеренно быстрый интерполятор, который является сглаженным (неточным). Данный метод интерполяции используется в случае необходимости создания сложных полиномиальных плоскостей. Если интерполяция по методу локального полинома согласовывает полином со всей поверхностью, то интерполяция по методу локальных полиномов подбирает множество полиномов, каждый из которых подбирается к определенной перекрывающейся окрестности (рис. 6.7).

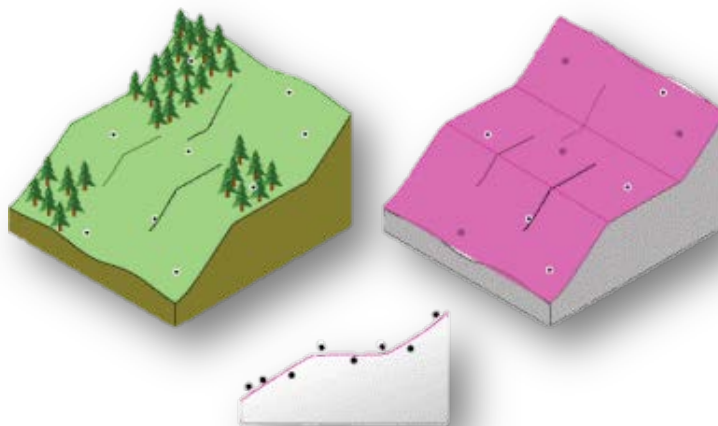


Рис. 6.7. Интерполяция по методу глобального полинома

Обычно на практике для двумерного случая используют один из четырех типов полиномов:

- плоскость: $P_1(x, y) = a + bx + cy$;
- квадратичный: $P_2(x, y) = a + bx + cy + dx + ex^2 + fy^2$;
- кубический: $P_3(x, y) = a + bx + cy + dx + ex^2 + fy^2 + gx^2y + hxy^2 + ix^3 + jy^3$.

Задача полиномиальной интерполяции сводится к тому, чтобы определить неизвестные коэффициенты a_i так, чтобы полиномы максимально хорошо соответствовали данным в заданных точках. Для этого находят минимум по всем коэффициентам (a, b, c, d и т. д.) функции χ^2 , задающей интегральную ошибку интерполяции и определяемую следующим образом:

$$\chi^2 = \sum_{i=1}^N [Z(x_i, y_i) - P_n(x_i, y_i)]^2, \quad (6.2)$$

где Z – известное значение в точке измерений;

P_n – полином n -й степени.

Полиномиальный метод обеспечивает построение поверхности прогнозируемых значений, поверхности стандартной ошибки прогнозирования и числа обусловленности, которые сравнимы с обычным кригингом по погрешностям измерения. *Пространственное число обусловленности* является мерой стабильности или нестабильности решения уравнений интерполяции для конкретного положения. Иными словами, оно позволяет определить, достоверны ли результаты, полученные посредством интерполяции по методу LPI, в определенной области или нет. Число обусловленности зависит только от местоположений входных точек, а не их текущих значений. Значения ниже критического порога пространственного числа обусловленности указывают на местоположения, в которых решения достоверны, а значения выше критических пределов – недостоверны.

Метод локальных полиномов не позволяет исследовать автокорреляцию данных, что делает его менее гибким и более автоматизированным, чем кригинг. На данные не накладываются никакие ограничения, однако более целесообразно использовать этот метод при наличии данных, полученных при отборе образцов по равномерной сетке опробования. Применение метода LPI целесообразно и в случае, когда присутствует территория с изменчивой формой рельефа, а величина интерполируемого показателя прямо пропорционально связана с характером рельефа, изменяясь в соответствии с его изменением.

Интерполяция по методу локальных полиномов чувствительна к размеру окрестности поиска, а небольшие окрестности поиска могут создать на интерполируемой поверхности пустые области. На рис. 6.8. представлена поверхность пространственного числа обусловленности, созданная для полиномов второго порядка, и интерполяционная поверхность, созданная методом локальных полиномов.

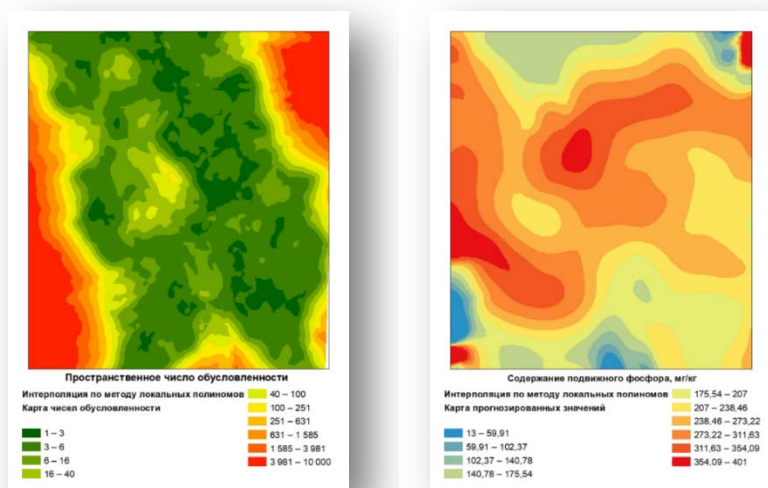


Рис. 6.8. Поверхность пространственного числа обусловленности (а) и интерполяционная поверхность, созданная методом обратных взвешенных расстояний (б)

Интерполяция ядра с барьерами – это детерминированный интерполятор с движущимся окном, который использует кратчайшее расстояние между точками, соединя точки по обе стороны от линейных барьеров так, что точки по разным сторонам определенного непроходимого (абсолютного) барьера соединятся серией прямых линий. Данный метод является вариантом метода локальных полиномов и используется в случае, если в пределах интерполируемой поверхности находятся барьеры – векторный слой линейных объектов. Модель интерполяции ядра с барьерами использует следующие радиально-симметричные ядра (r – радиус окрестности поиска с центром в точке s ; h – ширина канала):

- экспоненциальное: $e^{-3(\frac{r}{h})}$;
- Гауссово: $e^{-3(\frac{r}{h})^2}$;
- биквадратное: $(1 - (\frac{r}{h})^2)^2$, для $\frac{r}{h} < 1$;
- Епанечникова: $1 - (\frac{r}{h})^2$, для $\frac{r}{h} < 1$;

– полиномиальное 5-го порядка: $1 - \left(\frac{r}{h}\right)^3(10 - \frac{r}{h})(15 - 6\left(\frac{r}{h}\right))$,
 для $\frac{r}{h} < 1$;

– константу: $I(s - h < s_i < s + h)$, где индикатор функции I принимает значение 1, если $I = true$, и значение 0, если $I = false$.

На рис. 6.9 показана интерполированная поверхность, полученная посредством выполнения интерполяции с барьерами, где в качестве барьеров использован векторный слой мелиоративных каналов.

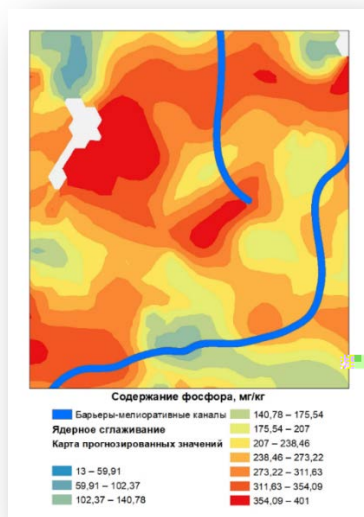


Рис. 6.9. Интерполяционная поверхность, созданная методом интерполяции ядра с барьерами (функция ядра – полиномиальная 5-го порядка)

Интерполяция по методу диффузного ядра применяет ядро, основанное на уравнении теплового баланса, которое описывает, как тепло или частицы со временем распределяются в однородной среде, и позволяет использовать в качестве барьера комбинацию растровых и векторных наборов геопространственных данных. Интерполяция диффузии может использовать сложную метрику расстояния, определяемую стоимостной поверхностью, которая является обычной функцией растра, вычисляющей стоимость пути от одной ячейки растра к другой. В случае интерполяции по методу диффузного ядра форма ядра

изменяется рядом с барьером в соответствии с уравнением диффузии (рис. 6.10, *а*), тогда как без ее применения этого не происходит (рис. 6.10, *б*).

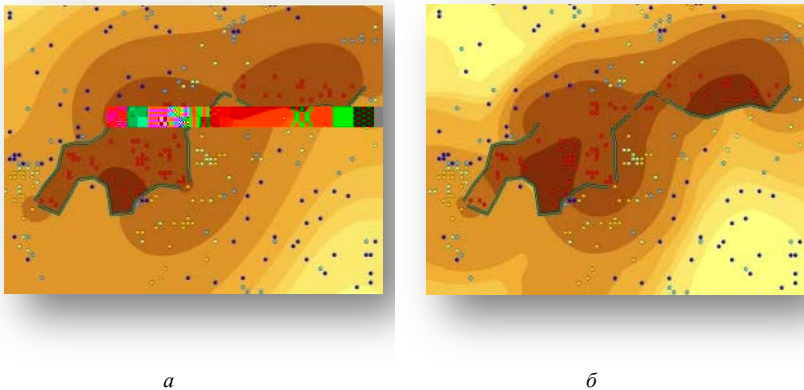


Рис. 6.10. Сравнение интерполяционных поверхностей, созданных методом диффузного ядра (*а*) и интерполяции ядра с барьерами (*б*); функция ядра – гауссова

Радиальные базисные функции (Radial Basis/Bias Functions, RBF) представляют собой жесткие интерполяторы, которые создают гладкие поверхности, и, в отличие от интерполяции по методу глобального и локального полиномов, предусматривают прохождение интерполированной поверхности через все точки с измеренными значениями. Кроме того, в отличие от интерполяции по методу обратных взвешенных расстояний, также являющейся жестким интерполятором, радиальные базисные функции интерполируют значения выше максимального или ниже минимального измеренного значения (рис. 6.11).

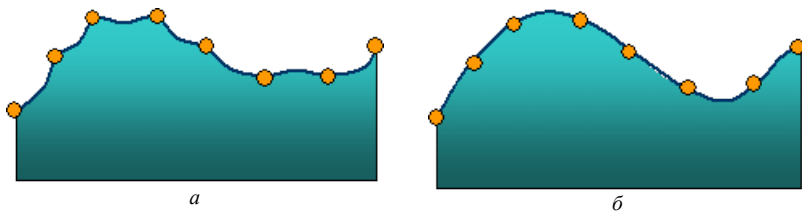


Рис. 6.11. Сравнение профиля поперечного сечения опорных данных: *а* – метод обратных взвешенных расстояний; *б* – метод радиальных базисных функций

Поскольку интерполяторы являются жесткими, радиальные базисные функции могут быть локально чувствительны к выпадающим значениям (т. е. поверхность будет содержать локально высокие или низкие значения).

Интерполяция методом радиальных базисных функций строится как линейная комбинация из базисных функций:

$$Z^*(x_0) = \sum_{i=1}^n c_i B(h_{0i}), \quad (6.3)$$

где h_{0i} – расстояние от точки x_0 до точки x_i ;

$B(h_{0i})$ – базисная функция, определяемая от расстояния;

c_i – весовые коэффициенты (определяет алгебраический знак вхождения соответствующего члена в уравнение и степень его влияния).

Метод RBF представляет собой семейство из пяти методов детерминированной точной интерполяции (h – расстояние; δ – сглаживающий параметр):

– тонкий сплайн (TPS): $B(h) = (h^2 + \delta^2) \lg(h^2 + \delta^2)$;

– сплайн с напряжением (SPT): $B(h) = (h^2 + \delta^2)^{1/2}$;

– полностью регуляризованный сплайн (CRS): $B(h) = \lg(h^2 + \delta^2)$;

– мультиквадратичная функция (MQ): $B(h) = \sqrt{h^2 + \delta^2}$;

– обратная мультиквадратичная функция (IMQ): $B(h) = \frac{1}{\sqrt{h^2 + \delta^2}}$.

Метод RBF подходит к поверхности через измеренные значения при минимизации общей кривизны поверхности и неэффективен, когда происходит резкое изменение значений поверхности на коротких расстояниях. Радиальные базисные функции гораздо более гибкие, чем IDW, поскольку позволяют прогнозировать значения выше и ниже максимального и минимального соответственно, но требуют принятия большего количества решений относительно параметров и не позволяют оценивать прогнозируемые ошибки. Интерполяция по методу RBF приближается к построению поверхности по измеренным значениям, минимизируя общую кривизну поверхности, и неэффективна, когда происходит резкое изменение значений показателя на коротких расстояниях. Радиальные базисные функции используют для создания сглаженных поверхностей из большого (несколько сотен) количества исходных данных, которые не подвержены значительным изменениям в пределах коротких расстояний. Наиболее оптимальным при интерполировании является выбор радиальной базисной функции мультиквадратов, однако возможно использовать и другие функции, опреде-

лив наиболее оптимальную посредством выполнения перекрестной проверки. Каждая радиальная базисная функция содержит параметр, управляющий сглаженностью поверхности – параметр ядра (рис. 6.12).

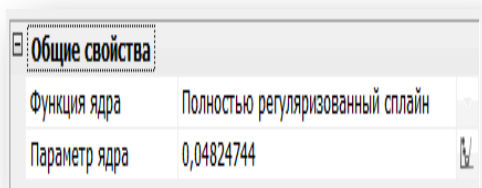


Рис. 6.12. Иллюстрация параметра, управляющего сглаженностью интерполяционной поверхности

Для всех радиальных базисных функций, кроме обратных мультиквадриков, чем выше значение параметра, тем более сглаженной будет поверхность; обратное верно для метода мультиквадриков. Каждая радиальная базисная функция имеет различную форму и позволяет получать разные интерполированные поверхности, а метод RBF – это специальный случай сплайнов. На рис. 6.13 приведена интерполяционная поверхность, полученная методом RBF (а) и методом глобального полинома третьей степени (б).

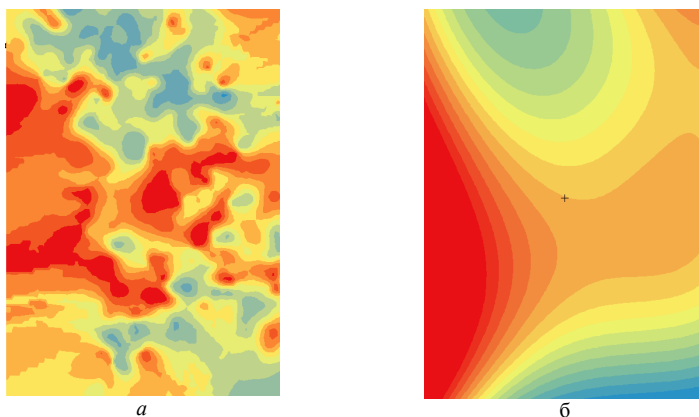


Рис. 6.13. Интерполяционная поверхность, созданная методом RBF (функция ядра – мультиквадратичная) (а) и методом глобального полинома третьей степени (б)

6.2. Геоestatистические методы пространственного моделирования

Геоestatистические методы интерполяции используют статистические свойства значений в точках, а не их абсолютные значения. С их помощью измеряется пространственная автокорреляция в измеренных точках и рассчитывается пространственная конфигурация опорных точек вокруг интерполируемого местоположения.

Кригинг является базовой интерполяционной моделью геоestatистики и основой всех методов, связанных с геоestatистикой, – интерполяции, вероятностного картирования, стохастического моделирования. Он основывается на гипотезе о пространственной однородности («стационарность 2-го порядка»⁶⁶), которая предполагает, что:

- пространственная изменчивость, представленная значениями Z в исходных опорных точках x , принадлежащих области пространства S , статистически однородна по всей поверхности;
- вариации в значениях исходных опорных точек зависят от расстояния между ними, но не зависят от их местоположения.

Исходя из этого, исходные данные в виде наборов точек (значений переменных), включающих аномальные впадины, высоты или другие какие-либо резкие изменения, не предназначены для кригинг-метода. На рис. 6.14. приведена иллюстрация основных элементов кригинга.

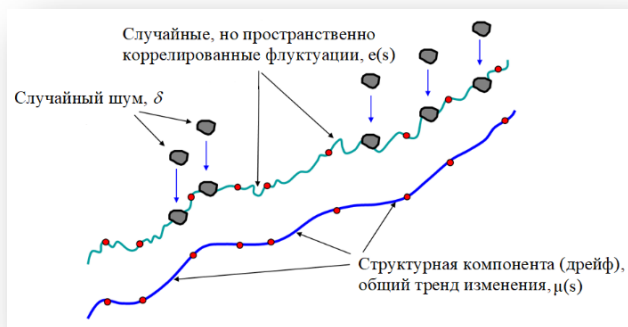


Рис. 6.14. Основные элементы кригинга

⁶⁶Более подробная информация приведена в главе 2.

Кригинг обрабатывает поверхность, считая пространственные изменения значений $Z(s)$ в точках s в общем случае суммой трех компонентов:

$$Z(s) = \mu(s) + e(s) + \delta, \quad (6.4)$$

где $\mu(s)$ – структурная (неслучайная) компонента, которая представляет поверхность как общий тренд в определенном направлении и может быть описана какой-либо математической функцией. Например, пологий склон (наклонная плоскость) может быть представлен полиномом 1-го порядка, долина U -образной формы – полиномом 2-го порядка. В большинстве случаев использования интерполяции по методу кригинга считается, что исходные данные не содержат никакой тенденции (тренда);

$e(s)$ – случайная, но пространственно-коррелированная компонента – некоторые отклонения от общей тенденции (тренда), которые являются случайными, но связанными друг с другом пространственно;

δ – случайный шум, не связанный с общей тенденцией и не имеющий пространственной автокорреляции (остаточная ошибка).

Предположения, которые делаются относительно компоненты δ :

– среднее всех $e(s) = 0$;

– вариации значений $e(s)$ и $e(s + h)$ в любых точках s и $(s + h)$ зависят от смещения h (расстояния между точками), но не зависят от местоположения точек.

В кригинге для определения неизвестного значения в некоторой точке значения исходных опорных точек, попавших в некую окрестность вокруг интерполируемой точки, взвешиваются:

$$Z_0 = \sum_{i=1}^n \lambda_i Z_i, \quad (6.5)$$

где Z_0 – прогнозируемое значение в некоторой точке (s_0);

Z_i – известное значение i -й опорной точки (s_i);

n – число опорных точек, попавших в окрестность искомой точки;

λ_i – неизвестный вес i -й точки.

Если поверхность имеет тренд (компоненту m), тогда из анализа значений в опорных точках тренд вычитается и моделируется только

случайная компонента $e(s)$, а перед окончательным интерполированием поверхности тренд добавляется обратно.

В кригинге (в отличие от интерполяции по методу обратных взвешенных расстояний) веса опорных точек в окрестности искомой точки зависят не только от расстояния между опорной и искомой точкой, но и от пространственной структуры данных в целом. Поэтому для установления весов λ_i в кригинге сначала необходимо определить пространственную структуру данных, т. е. количественно установить пространственную автокорреляцию данных⁶⁷. Исходя из этого, кригинг предусматривает *решение двух основных задач*:

1) установить пространственную структуру данных – подобрать к данным модель пространственной изменчивости (вариограмму);

2) построить поверхность, используя для расчета (прогноза) неизвестных значений подобранную вариограмму, а также расположение и известные значения опорных точек, находящихся в пределах заданного радиуса поиска вокруг точки с искомым значением.

Вариограмма – это функция, которая связывает различие в значениях опорных точек и расстояние, на которое они отстоят друг от друга. Она служит инструментом для исследования пространственной автокорреляции (связей) между опорными точками.

Интерполяция по методу кригинга осуществляется в *четыре этапа*.

Этап 1. Расчет эмпирической семивариограммы по набору исходных опорных точек («semi» – половина), где (рис. 6.15):

– по оси X откладывается расстояние h между парами точек;

– по оси Y – значение $\gamma(h)$, равное $1/2$ среднеквадратической разности значений Z между всеми парами исходных опорных точек, расположенных на расстоянии h друг от друга (полудисперсия):

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n (z_i - z_{i+h})^2, \quad (6.6)$$

где z_i, z_{i+h} – значения в опорных точках $i, (i+h)$, находящихся на расстоянии h друг от друга;

n – число пар таких опорных точек.

⁶⁷Более подробная информация приведена в главе 5.

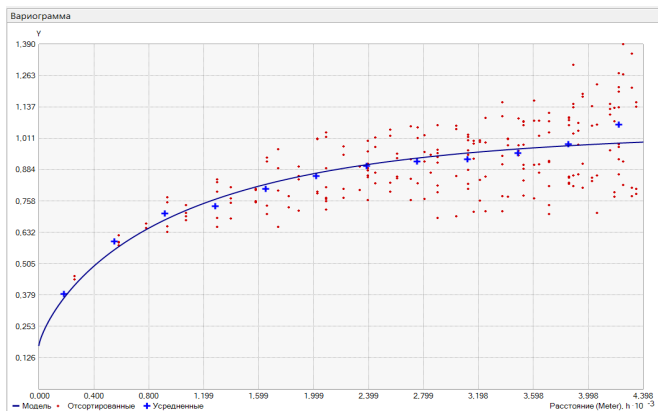


Рис. 6.15. Эмпирическая семивариограмма

Этап 2. Подбор теоретической функции к полученной эмпирической семивариограмме (по методу наименьших квадратов) (рис. 6.16).

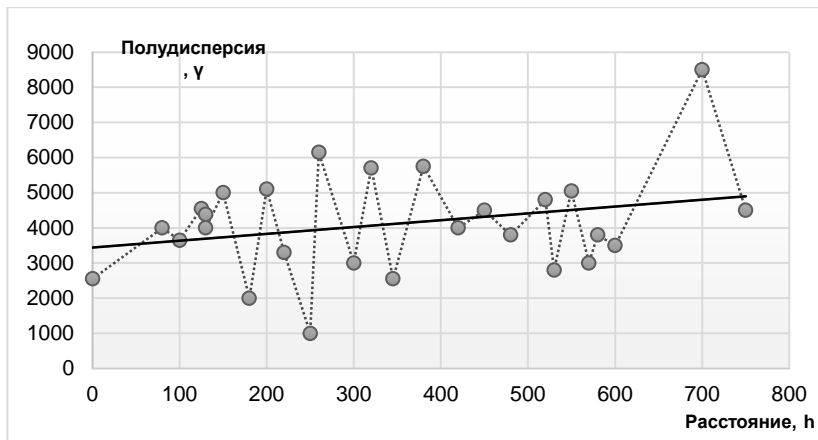


Рис. 6.16. Графики семивариограмм (пунктирная линия – эмпирическая вариограмма; сплошная линия – подобранная вариограмма)

Варианты возможных функций для подбора модели вариограммы: круговая, сферическая, тетрасферическая, пентасферическая, экспоненциальная, гауссова, рациональная квадратическая, эффекта дыры,

k -Бесселя, j -Бесселя, устойчивая⁶⁸. Выбранная модель влияет на интерполяцию неизвестных значений, особенно если форма кривой вблизи исходных данных значительно отличается. Чем круче кривая вблизи исходных данных, тем больше влияния на интерполяцию будет иметь ближайшая окрестность. В результате выходная поверхность будет менее сглаженной.

Этап 3. Нахождение весов λ_i на основе полученной модели вариограммы:

$$w_i(x) = \begin{cases} 1, & x \in A(x_i), \\ 0, & x \notin A(x_i), \end{cases} \quad (6.7)$$

где $A(x_i)$ – область влияния точки x_i .

Этап 4. Выполнение интерполяции, когда на основе рассчитанных весов кригинга λ_i и известных значений Z_i в опорных точках в пределах заданного радиуса поиска вычисляется значение в искомой точке Z_0 :

$$Z_0 = \sum_{i=1}^n \lambda_i Z_i, \quad (6.8)$$

где Z_0 – прогнозируемое значение в некоторой точке (s_0);

Z_i – известное значение i -й опорной точки (s_i);

n – число опорных точек, попавших в окрестность искомой точки;

λ_i – неизвестный вес i -й точки.

Существует несколько различных видов кригинга: простой, ординарный, универсальный, эмпирический байесовский кригинг.

Простой кригинг (Simple Kriging, SK) предполагает использование для интерполирования значений модели следующего вида:

$$Z(s) = \mu + e(s), \quad (6.9)$$

где μ – известная и постоянная модель тренда.

Простой кригинг предусматривает также наличие предположения о стационарности второго порядка случайной переменной $Z(s)$ (рис. 6.17).

⁶⁸Более подробная информация приведена в главе 5.

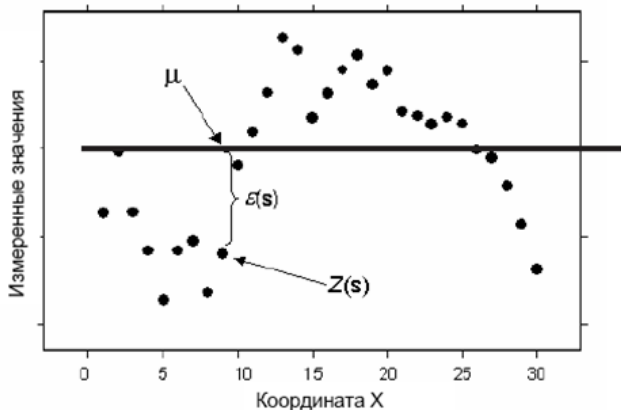


Рис. 6.17. Простой кригинг

Оценка простого кригинга имеет следующий вид:

$$Z(s) = \mu + \sum_{i=1}^n \lambda_i(x) Z(x). \quad (6.10)$$

Простой кригинг обладает рядом свойств:

1) оценка простого кригинга является точной: если координата оцениваемой точки x_0 совпадает с какой-либо координатой из исходного набора данных ($x_0 = x_i, i = 1, \dots, n$), то полученная оценка также будет совпадать с исходным значением;

2) веса простого кригинга не зависят от значений исходного набора данных, а зависят только от пространственной корреляции поля, построенного на основе этих данных;

3) оценка простого кригинга является сглаженной по сравнению с распределением исходных данных.

Основным недостатком простого кригинга как метода интерполяции является предположение о знании среднего (существовании постоянного тренда с известным значением). Поэтому простой кригинг редко применяется как самостоятельный метод оценивания, обычно его использование связано с искусственными комбинациями, где среднее известно вследствие предварительных манипуляций с исходными данными.

Ординарный, или обычный, кригинг (Ordinary Kriging, ОК) в отличие от простого кригинга предусматривает, что модель тренда μ является постоянной, но неизвестной (рис. 6.18).

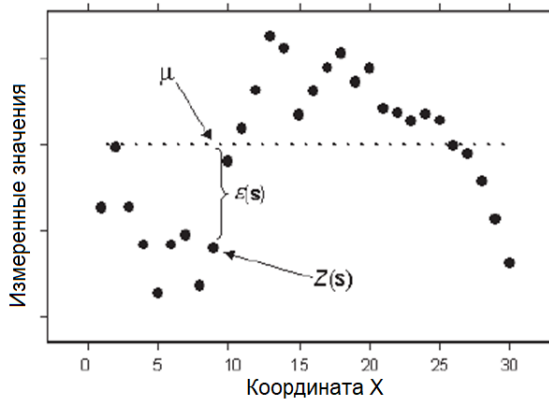


Рис. 6.18. Ординарный кригинг

Оценка ординарного кригинга $Z(s)$ и дисперсия оценки ошибок $\sigma_k^2(x_0)$ в любой точке x_0 вычисляются следующим образом:

$$Z(s) = \mu + \sum_{i=1}^n \lambda_i(x) Z(x_i); \quad (6.11)$$

$$\sigma_k^2(x_0) = \mu + \sum_{i=1}^n \lambda_i \gamma(x_0 - x_i), \quad (6.12)$$

где λ_i — вес i -й точки;

μ — постоянная Лагранжа;

$\gamma(x_0 - x_i)$ — значение вариограммы, соответствующее расстоянию между x_0 и x_i .

При интерполяции по методу ординарного кригинга возможно использование вариограммы либо ковариации (математические формы, используемые для выражения автокорреляции), а также применение преобразований (по методу Бокса-Кокса, логарифмическое либо по методу арксинуса) и удаление тренда. На рис. 6.19 приведена оценка

ординарного кригинга для случая изотропной модели вариограммы (т. е. радиус корреляции не зависит от направления).

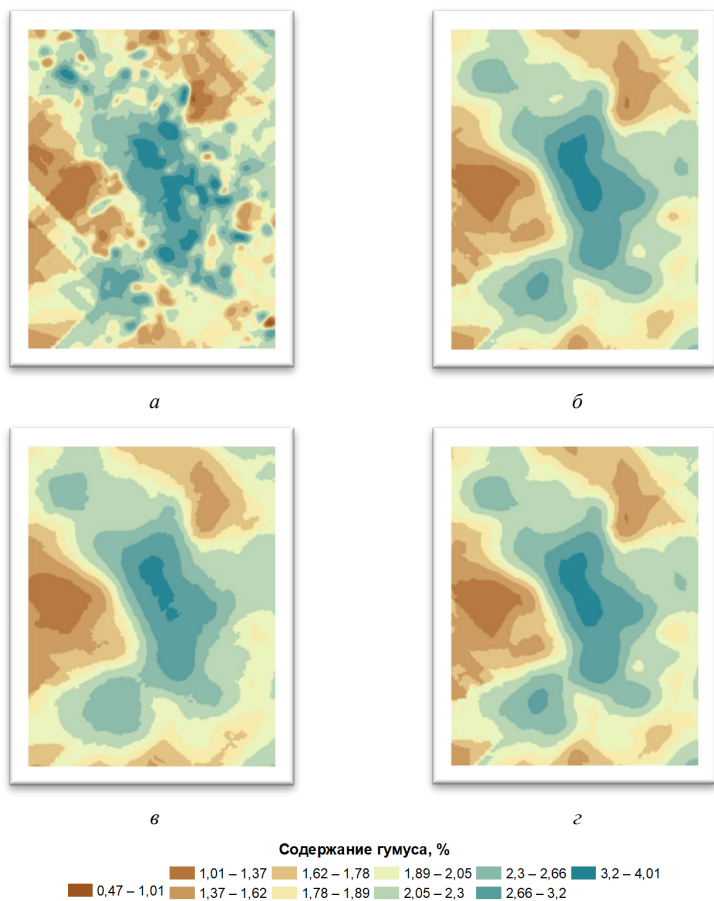


Рис. 6.19. Оценка обычного кригинга при изотропной модели пространственной корреляционной структуры (вариограмме) и различных значениях радиуса корреляции: *а* – 50 м; *б* – 150 м; *в* – 350 м; *г* – 700 м

Увеличение радиуса корреляции приводит к использованию большего числа опорных точек и вследствие этого к их заметному влиянию

на результат интерполяции. У оценки кригинга увеличиваются зоны высоких и низких значений, а зона, соответствующая среднему значению, уменьшается.

Универсальный кригинг, или **кригинг с трендом (Universal Kriging, UK)**, используется тогда, когда предполагается, что в данных имеется какая-либо доминирующая тенденция (тренд), которую можно смоделировать с помощью детерминистической полиномиальной функции. Он может использовать либо вариограммы, либо ковариации (математические формы, используемые для выражения автокорреляции), применять преобразования и учитывать погрешность измерения. Интерполированное значение при применении универсального кригинга определяется по формуле

$$Z(s) = \mu(s) + \varepsilon(s), \quad (6.13)$$

где $\mu(s)$ – это некоторая детерминированная функция, описываемая полиномом второго порядка;

$\varepsilon(s)$ – случайная ошибка, которая вычисляется путем вычитания полинома второго порядка из оригинальных данных (рис. 6.20).

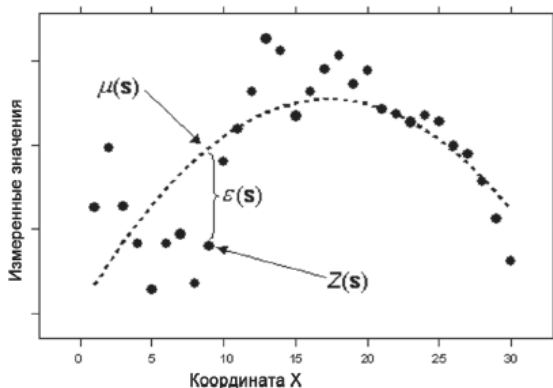


Рис. 6.20. Универсальный кригинг

Эмпирический байесовский кригинг (Empirical Bayesian Kriging, EBK) отличается от других методов кригинга использованием внутренней случайной функции в качестве модели кригинга. В EBK

можно анализировать эмпирическое распределение оценок параметров, поскольку в каждом местоположении рассчитывается множество вариограмм. Процесс ЕВК неявно предполагает, что оценочная полувариантность является истинной вариограммой для интерполяционной области и линейным предсказанием, которое включает в себя переменное пространственное затухание. В результате получается надежный нестационарный алгоритм пространственных интерполяционных геофизических поправок. Для каждого местоположения интерполяция рассчитывается с использованием новой вариограммы распределения, которая сгенерирована посредством базирующейся на подобию выборке из индивидуальных вариограмм спектров вариограмм в окрестности точки.

Среди *достоинств* эмпирического байесовского кригинга – минимум требуемого интерактивного моделирования, меньшие по сравнению с другими методами кригинга стандартные ошибки интерполяции, возможность точной интерполяции умеренно нестационарных данных и более высокая точность для небольших наборов данных.

Недостатками данного метода интерполяции является то, что обработка данных происходит медленнее, чем в других методах кригинга, особенно при выводе результатов в растр, а также недоступность кокригинга и анизотропной коррекции.

В отличие от других методов кригинга, которые используют метод взвешенных наименьших квадратов, параметры вариограммы при выполнении ЕВК оцениваются с использованием ограниченной максимальной вероятности. В связи с вычислительными ограничениями для крупных наборов данных входные данные сначала делятся на перекрывающиеся поднаборы определенного размера (по умолчанию 100 точек на каждый поднабор), и в каждом поднаборе выполняется расчет вариограммы, предусматривающий такую последовательность действий:

- 1) на основе данных в поднаборе рассчитывается вариограмма;
- 2) рассчитанная вариограмма используется в качестве модели. с помощью которой выполняется безусловное моделирование новых данных в каждой входной точке в поднаборе;
- 3) на основе смоделированных данных рассчитывается новая вариограмма;
- 4) шаги 2 и 3 повторяются заданное число раз, при этом при каждом повторении вариограмма, рассчитанная в шаге 1, используется для

моделирования нового набора данных во входных точках, а на основе смоделированных данных выполняется оценка новой вариограммы.

В результате последовательного выполнения всех шагов для каждого поднабора данных создается множество вариограмм. Если отобразить их на одной диаграмме, получится распределение вариограмм, окрашенных в зависимости от плотности (чем темнее синий цвет, тем больше вариограмм проходит через данный участок) (рис. 6.21).

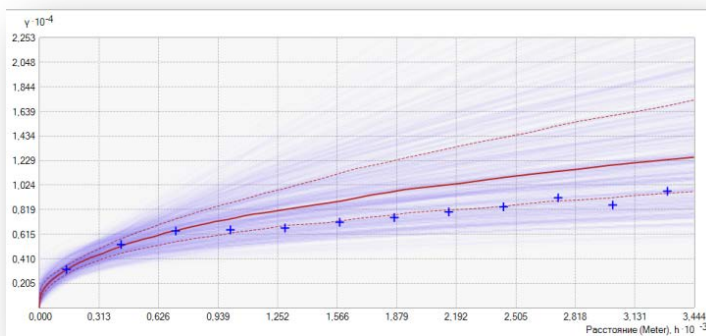


Рис. 6.21. Вариограммы, смоделированные в процессе реализации эмпирического байесовского кригинга

Эмпирические вариации на рисунке представлены синими крестиками, медиана распределения обозначена сплошной красной линией, а 25-й и 75-й процентили – пунктирными красными линиями. Для каждого местоположения интерполяция рассчитывается с использованием новой вариограммы распределения, которая сгенерирована посредством базирующейся на подобию выборке из индивидуальных вариограмм спектров вариограмм в окрестности точки.

Эмпирический байесовский кригинг отличается от других методов кригинга использованием внутренней случайной функции в качестве модели кригинга. Другие методы кригинга предполагают, что процесс следует общему среднему (или заданному тренду) с отдельными отклонениями относительно этого среднего, а большие отклонения при этом смещаются к среднему, в связи с чем значения никогда не отклоняются слишком сильно. В отличие от них, метод ЕВК не предполагает тенденции к общему среднему, поэтому большие отклонения могут

с равной вероятностью стать как меньше, так и больше. ЕВК поддерживает следующие вариограммы:

- степенная: $\gamma(h) = \text{Nugget} + b|h|^a$;
- линейная: $\gamma(h) = \text{Nugget} + b|h|$;
- плоский сплайн: $\gamma(h) = \text{Nugget} + b|h^2| \cdot \ln(|h|)$.

При этом самородок (Nugget) и уклон (b) должны быть положительными, а a (степень) принимать значения в диапазоне от 0,25 до 1,75.

ЕВК позволяет визуализировать и проанализировать эмпирическое распределение оценок параметров моделированных вариограмм, поскольку в каждом местоположении рассчитывается их множество (рис. 6.22).

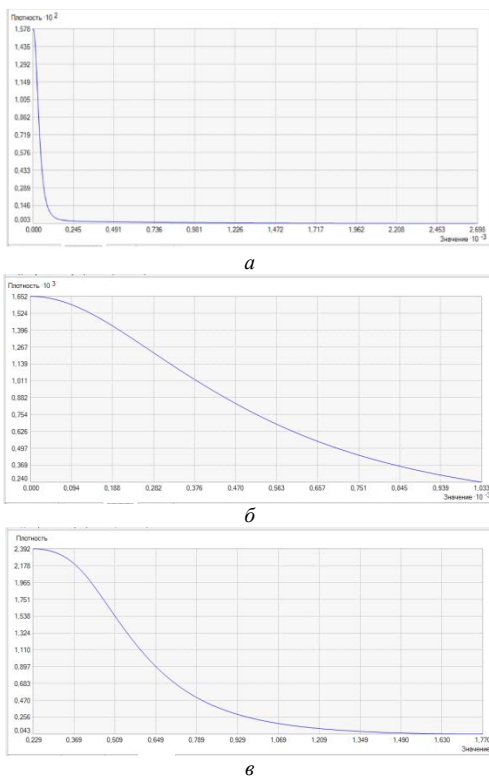


Рис. 6.22. Распределение параметров моделированных вариограмм: a – самородок; b – уклон; v – степень

Интерполяция методом эмпирического байесовского кригинга требует наличия нормального распределения входных данных. С этой целью выполняется их *преобразование по методу нормальных меток*, вследствие чего набор данных изменяется так, чтобы приблизиться к стандартному нормальному распределению. Это достигается посредством расположения значений в наборе данных по возрастанию и связывания этих рядов с эквивалентными рядами, созданными с помощью нормального распределения. Преобразование по методу нормальных меток подстраивается к каждому отдельному набору данных, тогда как другие виды преобразования этого не предусматривают.

Эмпирический байесовский кригинг выполняет преобразование по методу нормальных меток для мультипликативного искажения с двумя вариантами базовых распределений: *эмпирическим* и *логарифмическим эмпирическим*. Для выполнения логарифмического эмпирического преобразования данных необходимо, чтобы все их значения были положительными – в этом случае и результаты интерполяции будут также положительные. Этот вариант преобразования подходит для данных, которые не могут быть отрицательными.

В случае выполнения преобразования данных поддерживаются следующие типы вариограммы (табл. 6.1).

Таблица 6.1. Преимущества и недостатки моделей вариограммы

Модель вариограммы	Преимущества модели	Недостатки модели
Степень	Относительно быстрая и гибкая, обеспечивает оптимальный баланс между производительностью и точностью	Менее гибкая и работает значительно медленнее, чем другие модели
Линейная	Очень быстрая	Наименее гибкая
Плоский сплайн	Очень быстрая, отлично работает в случае наличия значительного тренда	Менее гибкая, особенно в случае отсутствия тренда данных
Экспоненциальная	Предлагает гибкое преобразование и работает значительно быстрее, чем вариограммы К-Бесселя или К-Бесселя с исключенным трендом	Форма вариограммы не является гибкой; более медленная по сравнению со степенной и линейной вариограммами, а также с плоским сплайном
Распределение Уиттла	Предлагает гибкое преобразование и работает значительно быстрее, чем вариограммы К-Бесселя или К-Бесселя с исключенным трендом	
Экспоненциальная с исключенным трендом	Предлагает гибкое преобразование и работает значительно быстрее, чем вариограммы К-Бесселя или К-Бесселя с исключенным трендом; удаляет тренд первого порядка	
Распределение Уиттла с исключенным трендом	Предлагает гибкое преобразование и работает значительно быстрее, чем вариограммы К-Бесселя или К-Бесселя с исключенным трендом; удаляет тренд первого порядка	
К-Бесселя	Наиболее гибкая и точная	Требует значительного времени для выполнения вычислений
Распределение К-Бесселя с исключенным трендом	Наиболее гибкая и точная, удаляет тренд первого порядка	

6.3. Автоматизация геопространственного моделирования

Автоматизация – одно из направлений научно-технического прогресса, использующее саморегулирующиеся технические средства и математические методы с целью освобождения человека от участия в процессах получения, преобразования, передачи и использования энергии, материалов, изделий или информации либо существенного уменьшения степени этого участия или трудоемкости выполняемых операций. Основным инструментом для автоматизации геопространственного моделирования является геообработка – методичное выполнение последовательности операций с географическими данными, в результате которого создается новая информация. Она позволяет экономить время при выполнении однотипных повторяющихся задач, сократить ошибки и выполнять анализ более эффективно. В частности, современные ГИС имеют необходимый функционал для создания сквозных рабочих процессов посредством объединения моделей и пространственных алгоритмов в единую цепочку. Они также позволяют создавать полнофункциональные модели геообработки без программирования с помощью конструктора визуальных моделей или использовать язык программирования Python⁶⁹ для написания соответствующих скриптов и создания готовых для совместного использования моделей.

Эффективным примером приложения, используемого для создания, редактирования и управления моделями⁷⁰, является модуль Model Builder программного продукта ArcGIS. Модуль Model Builder можно также рассматривать как визуальный язык программирования для построения рабочих потоков без применения традиционного программирования.

⁶⁹Python – высокоуровневый скриптовый объектно ориентированный язык программирования общего назначения. Синтаксис ядра Python минималистичен, а его стандартная библиотека включает широкий набор полезных функций, ориентированных на повышение производительности разработчика и читаемости кода. Данный язык содержит ряд библиотек, созданных для проведения научных исследований и вычислений: SciPy – библиотека с научными инструментами; NumPy – расширение, которое добавляет поддержку матриц и многомерных массивов, а также математические функции для работы с ними; Matplotlib – библиотека для работы с 2D- и 3D-графикой. На нем пишут алгоритмы программ с машинным обучением и аналитические приложения, а также обслуживают хранилища данных и облачные сервисы.

⁷⁰Модели – это рабочие процессы, соединенные друг с другом в последовательности инструментов геообработки и направляющие результат реализации функционала одного инструмента в качестве входных данных для другого.

Преимущества Model Builder в обобщенном виде можно представить следующим образом:

- удобное в работе приложение для создания и запуска рабочих потоков, содержащих последовательность инструментов;
- позволяет создавать собственные инструменты, которые также могут использоваться в средствах поддержки скриптов Python и в других моделях;
- предоставляет дополнительные методы для функциональных возможностей ArcGIS, позволяя создавать и совместно использовать конкретные модели в качестве инструмента в других, более сложных моделях;
- позволяет интегрировать программный продукт ArcGIS с другими приложениями, например, интеграция внешнего статистического пакета R.

Различают три основных *типа элементов модели* – инструменты, переменные и соединители (рис. 6.23).

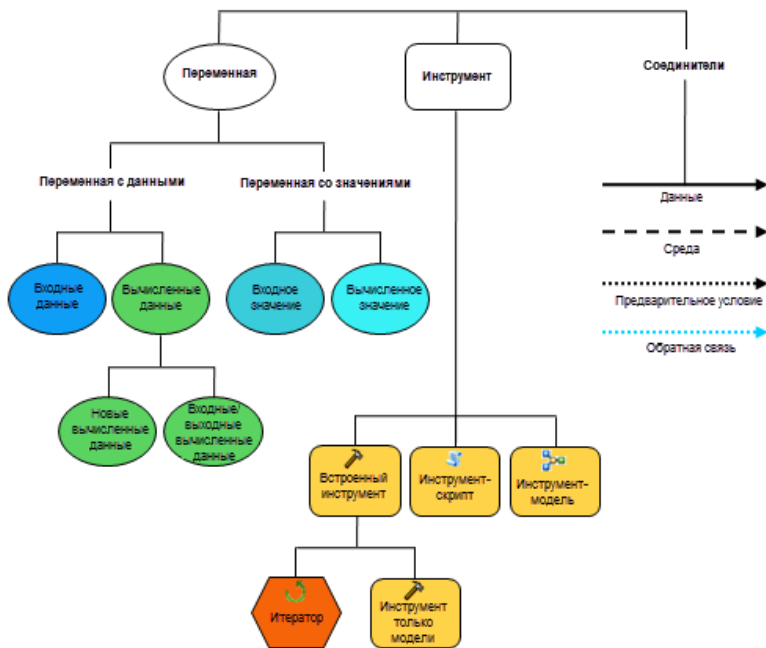


Рис. 6.23. Элементы геоинформационной модели

Инструменты геообработки являются основными блоками, из которых строятся рабочие процессы в модели. С их помощью выполняются различные операции с геопространственными или табличными данными. *Параметры инструмента* – это его входные и выходные значения, а также спецификации, которые влияют на обработку геоданных; одни параметры являются обязательными, другие – опциональными.

Переменные являются элементами модели, в которых хранятся значения или ссылки на данные, хранящиеся в атрибутивных таблицах. Существует два типа переменных:

1) *переменные данных*, в которых содержатся описательные сведения о геоданных (сведения о поле, пространственная привязка и путь доступа);

2) *переменные значений*, представляющие собой значения, такие как строки, числа, логические значения (true/false), пространственные привязки, линейные единицы или экстеннты.

Соединители соединяют данные и значения с инструментами модели. Стрелки соединителя указывают направление обработки. Существует *четыре типа соединителей*:

1) *соединители данных* – соединяют переменные данных и значений с инструментами модели;

2) *соединители среды* – соединяют переменную, содержащую параметр среды (данные или значение), с инструментом модели;

3) *соединители предварительного условия* – соединяют переменную с инструментом модели;

4) *соединители обратной связи* – направляют выходные данные инструмента обратно в тот же инструмент в качестве входных данных.

Основной структурной единицей геоинформационной модели является *процесс*, состоящий из инструмента и всех переменных, соединенных с ним (рис. 6.24). Несколько процессов могут быть связаны вместе, образуя более крупную модель, а несколько простых моделей могут образовывать более сложную. Если возникает необходимость регулярного выполнения одного и того же процесса, соответствующий инструмент модели заменяют на итератор⁷¹. Итерация, иногда называемая циклической организацией, обеспечивает неоднократное повторение процесса с определенной степенью автоматизации.

⁷¹Итератор – это инструмент, который повторяет одну операцию или цикл для набора входных данных или значений.

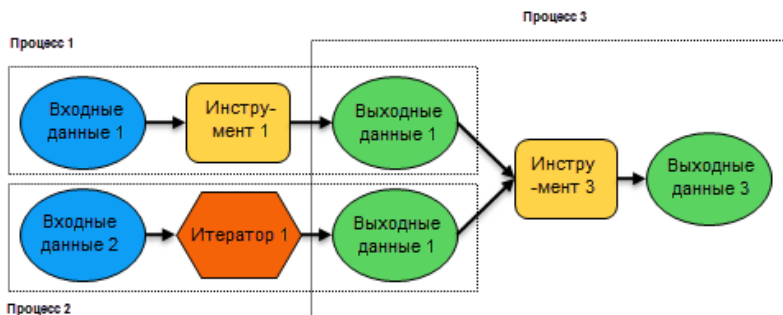


Рис. 6.24. Структура процесса геоинформационной модели

Соединительные линии на рисунке показывают последовательность выполнения процессов. Каждый из процессов в геоинформационной модели может находиться в одном из четырех состояний:

- 1) не готов к выполнению – первоначальное состояние процесса, когда не указаны значения параметров инструмента;
- 2) готов к выполнению – для инструмента указаны все необходимые значения параметров;
- 3) выполняется – происходит создание результатов с выведением сообщений в соответствующем окне;
- 4) выполнен – процесс уже выполнялся и производные данные уже были сгенерированы.

Параметры геоинформационной модели – это параметры, отображающиеся в диалоговом окне инструмента модели. Любую переменную модели можно преобразовать в параметр модели (рис. 6.25).



Рис. 6.25. Геоинформационная модель для определения степени расчленения рельефа территории по данным дистанционного зондирования

Параметром модели могут быть либо переменные, значения которых пользователь модели должен указывать при каждом выполнении модели, либо результат выполнения процесса или целой модели, который следует сохранить как постоянный объект.

Параметры среды модели – это дополнительные параметры, которые оказывают воздействие на результаты инструмента. Различают четыре уровня параметров среды: приложение, инструмент, модель и обработка модели. Все уровни содержат одинаковые переменные среды и одинаково влияют на выходные результаты реализации процесса либо модели, формируя иерархию, возрастающую при переходе на уровень ниже, поскольку настройки, сделанные для каждого уровня, являются приоритетными по сравнению с уровнем выше (рис. 6.26).

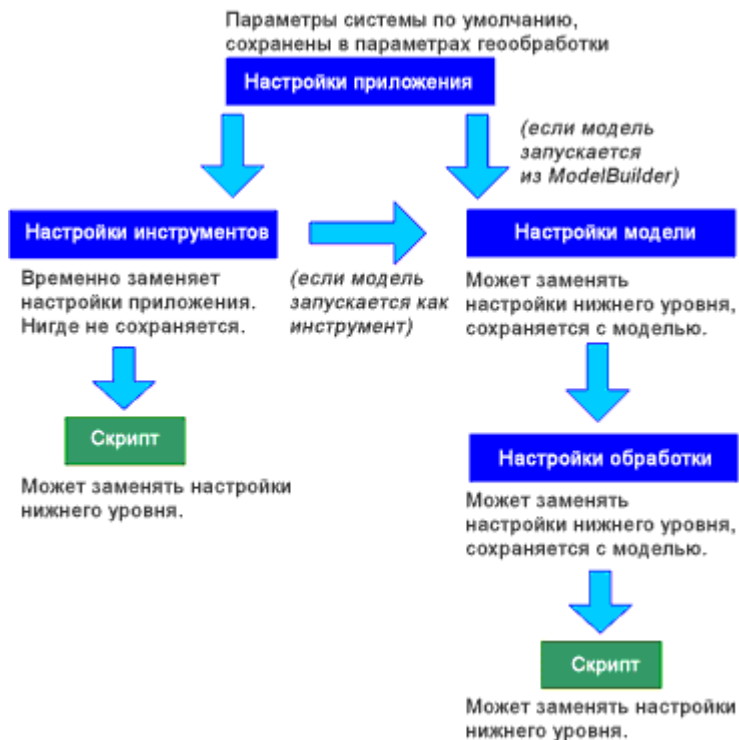


Рис. 6.26. Иерархия параметров среды геоинформационной модели

Для автоматизации и упрощения управления данными модели используют параметры среды рабочей области, которая определяется как местоположение по умолчанию для входных и выходных данных. Различают четыре варианта рабочей среды:

1) *текущая* – рабочая область, определенная как местоположение по умолчанию для входных и выходных данных инструментов геообработки;

2) *временная* – предназначена для выходных данных, которые нет необходимости сохранять;

3) *временная папка* – указывает местоположение папки, которую можно использовать для записи файловых данных, например, шейп-файлов, текстовых файлов и файлов слоев;

4) *временная база геоданных* – параметр, определяющий расположение файловой базы геоданных, которую можно использовать для записи временных данных.

Параметрические фильтры используются для ограничения или запрета входных значений или данных, которые можно указать для параметра инструмента модели. К параметрам модели можно применять шесть типов фильтров:

- список значений;
- диапазон и класс объекта;
- файл и поле;
- рабочая область.

Добавление и запуск одной модели в качестве инструмента в другой модели называется *подмоделью*, вложенной моделью или моделью внутри модели. Данный тип иерархии моделей используется в случае необходимости упрощения крупной, конструктивно сложной модели или для расширенного использования итераторов модели. Внедрение подмодели позволяет разбить крупную, сложную модель на мелкие элементы, которые более просты в управлении и повтором использовании в других моделях и скриптах.

Внедрение подмодели в другую модель также обеспечивает возможность расширенного использования итераторов модели. Когда требуется итерирование модели, особые инструменты или процессы, которые необходимо запускать многократно, нужно отделить от инструментов и процессов, запускаемых один раз для каждого рабочего потока. В этой ситуации все инструменты, которые нужно запускать многократно, помещают в одну модель с итератором и используют в качестве подмодели (рис. 6.27).

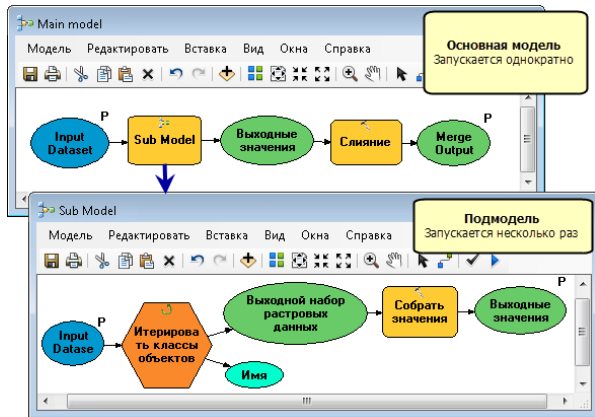


Рис. 6.27. Расширенное использование итераторов геоинформационной модели

На рис. 6.28 представлена универсальная модель, позволяющая автоматизировать процесс выполнения морфометрического анализа цифровой модели рельефа, генерацию системы водотоков и водосборных бассейнов, а также получать таблицы площадей с различной степенью уклона, рассчитывать площади водосборных бассейнов и определять порядок водотоков.

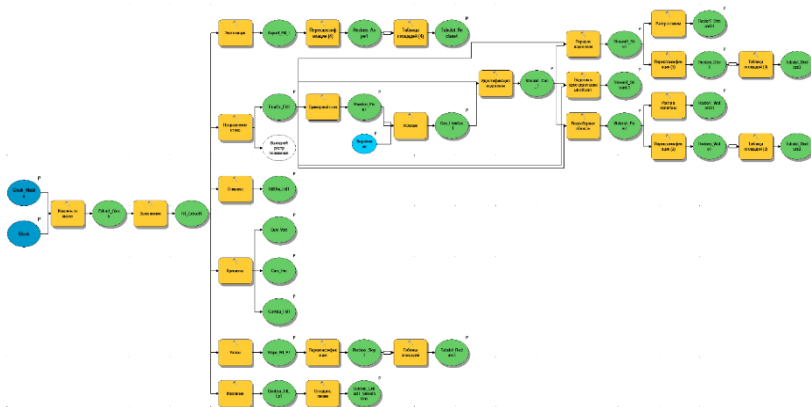


Рис. 6.28. Геоинформационная модель для автоматизации морфометрического анализа, извлечения системы водотоков и водосборных бассейнов

Созданная модель структурно состоит из трех исходных параметров, 25 инструментов и 27 результирующих растровых, векторных слоев и слоев в виде таблиц. Использование модели позволяет ускорить процесс анализа более чем в 30 раз по сравнению с процессом анализа, выполняемого посредством последовательного применения отдельных инструментов, а время, затраченное на ее реализацию, составляет 0,54 мин.

Скрипты, написанные с использованием языка программирования Python или другого языка, можно интегрировать в геоинформационную модель для ее улучшения, преобразуя скрипт в инструмент-скрипт. Процесс программирования в среде Python является универсальным подходом к решению разнообразных задач и повышению производительности труда при работе с пространственными данными. Скрипты, созданные на Python, используются для автоматизации выполнения задач геообработки, исследования и манипулирования пространственными и табличными данными, работы с геометриями с помощью курсоров, работы с растрами и алгеброй карт, автоматизации картографических задач и других рабочих процессов.

Вопросы для самостоятельного изучения

1. Индикаторный, вероятностный и дизъюнктивный кригинг.
2. Характеристика и особенности применения площадной интерполяции.

Источники информации

1. Геостатистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.
2. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.

Тестовые задания для проверки знаний

1. *Вставьте пропущенное слово либо словосочетание.* Методы _____ интерполяции создают поверхности из измеренных точек, основываясь или на степени схожести (обратные взвешенные расстояния), или на уровне сглаживания (радиальные базисные функции).

- 1) детерминированной;

- 2) геостатистический;
- 3) глобальной;
- 4) локальной;
- 5) жесткой.

2. *Вставьте пропущенное слово либо словосочетание.*

_____ методы вычисляют проинтерполированные значения на основании всего набора данных, тогда как _____ методы вычисляют проинтерполированные значения на основании измеренных точек в пределах окрестностей, которые являются меньшими пространственными областями внутри большей изучаемой территории.

- 1) локальные и глобальные;
- 2) жесткие и нежесткие;
- 3) глобальные и локальные;
- 4) детерминированные и геостатистические;
- 5) геостатистические и детерминированные;
- 6) точные и неточные.

3. *Правильным или нет является утверждение?* Метод интерполяции, который вычисляет значение, идентичное измеренному в опорном местоположении, называется нежестким интерполятором.

- 1) да;
- 2) нет.

4. *Выберите правильный вариант(ты) ответа.* Какой детерминированный метод интерполяции целесообразно использовать, когда набор точек достаточно плотный, чтобы фиксировать степень локального изменения поверхности, необходимую для анализа, а также если не предъявляются особые требования к точности интерполированной поверхности, а необходимо получить общее представление о характере распространения того либо иного показателя без оценки ошибки прогнозирования?

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяцию ядра с барьерами;
- 5) радиальные базисные функции;
- 6) интерполяцию по методу диффузного ядра;
- 7) кригинг.

5. *Дополните определение.* Нежесткий метод интерполяции, подбирающий к входным опорным точкам сглаженную поверхность, определяемую линейным полиномом, поверхность которого постепенно изменяется и характеризует грубую структуру в данных, называют:

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяция ядра с барьерами;
- 5) радиальные базисные функции;
- 6) интерполяция по методу диффузного ядра;
- 7) кригинг.

6. *Дополните определение.* Метод интерполяции, используемый в случае необходимости создания сложных полиномиальных плоскостей и обеспечивающий построение поверхности прогнозируемых значений, поверхности стандартной ошибки прогнозирования и числа обусловленности, однако не позволяющий исследовать автокорреляцию данных, называют:

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяция ядра с барьерами;
- 5) радиальные базисные функции;
- 6) интерполяция по методу диффузного ядра;
- 7) кригинг.

7. *Дополните определение.* Метод интерполяции, являющийся вариантом метода локальных полиномов и используемый в случае, если в пределах интерполируемой поверхности находятся барьеры – векторный слой линейных объектов, называют:

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяция ядра с барьерами;
- 5) радиальные базисные функции;
- 6) интерполяция по методу диффузного ядра;
- 7) кригинг.

8. *Дополните определение.* Метод интерполяции, основанный на уравнении теплового баланса, и позволяющий использовать в качестве барьера комбинацию растровых и векторных наборов данных, называют:

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяция ядра с барьерами;

- 5) радиальные базисные функции;
- 6) интерполяция по методу диффузного ядра;
- 7) кригинг.

9. *Дополните определение.* Жесткий метод интерполяции, создающий сглаженные поверхности и позволяющий интерполировать значения выше максимального или ниже минимального измеренного значения, называют:

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяция ядра с барьерами;
- 5) радиальные базисные функции;
- 6) интерполяция по методу диффузного ядра;
- 7) кригинг.

10. *Правильным или нет является утверждение?* Геоestatистические методы интерполяции используют статистические свойства значений в точках, а не их абсолютные значения.

- 1) да;
- 2) нет.

11. *Дополните определение.* Метод интерполяции, при котором веса опорных точек в окрестности искомой точки зависят не только от расстояния между опорной и искомой точкой, но и от пространственной структуры данных в целом, требующий количественно установить пространственную автокорреляцию данных, называют:

- 1) метод обратных взвешенных расстояний;
- 2) метод локальных полиномов;
- 3) метод глобального полинома;
- 4) интерполяция ядра с барьерами;
- 5) радиальные базисные функции;
- 6) интерполяция по методу диффузного ядра;
- 7) кригинг.

12. *Выберите правильную последовательность.* Интерполяция по методу кригинга осуществляется посредством последовательного прохождения следующих этапов:

- 1) выполнение интерполяции по методу кригинга;
- 2) расчет эмпирической семивариограммы по набору исходных опорных точек;
- 3) нахождение весов на основе полученной модели вариограммы;
- 4) подбор теоретической функции к полученной эмпирической семивариограмме.

13. Дополните определение. Метод кригинга, предусматривающий, что модель тренда является известной и постоянной, называют:

- 1) универсальный кригинг;
- 2) простой кригинг;
- 3) обычный кригинг;
- 4) эмпирический байесовский кригинг.

14. Дополните определение. Метод кригинга, предусматривающий, что модель тренда является неизвестной и постоянной, называют:

- 1) универсальный кригинг;
- 2) простой кригинг;
- 3) обычный кригинг;
- 4) эмпирический байесовский кригинг.

15. Дополните определение. Метод кригинга, используемый тогда, когда предполагается, что в данных имеется какая-либо доминирующая тенденция (тренд), которую можно смоделировать с помощью детерминистической полиномиальной функции, называют:

- 1) универсальный кригинг;
- 2) простой кригинг;
- 3) обычный кригинг;
- 4) эмпирический байесовский кригинг.

16. Дополните определение. Метод кригинга, использующий внутреннюю случайную функцию в качестве модели кривой N и e

- 1) универсальный кригинг;
- 2) простой кригинг;
- 3) обычный кригинг;
- 4) эмпирический байесовский кригинг.

16.6

1) г.

16.

1) ст

1). жу

1) м

19. *Дополните определение.* Элементы модели геообработки, в которых хранятся значения или ссылки на данные, хранящиеся в атрибутивных таблицах, называют:

- 1) соединители;
- 2) переменные;
- 3) инструменты;
- 4) блоки;
- 5) компоненты.

20. *Дополните определение.* Элементы модели геообработки, соединяющие данные и значения с инструментами, называют:

- 1) соединители;
- 2) переменные;
- 3) инструменты;
- 4) блоки;
- 5) компоненты.

21. *Вставьте пропущенное слово либо словосочетание.* Основной структурной единицей геоинформационной модели является _____, состоящий из инструмента и всех переменных, соединенных с ним.

- 1) результат;
- 2) компонент;
- 3) процесс;
- 4) блок;
- 5) элемент;
- 6) итератор.

22. *Дополните определение.* Инструмент геоинформационной модели, обеспечивающий неоднократное повторение процесса с определенной степенью автоматизации, называют:

- 1) результат;
- 2) компонент;
- 3) процесс;
- 4) блок;
- 5) элемент;
- 6) итератор.

7. ТРЕХМЕРНОЕ МОДЕЛИРОВАНИЕ

7.1. Пространственно-временное моделирование и визуализация куба «пространство – время».

7.2. Особенности создания трехмерных моделей поверхности и их геообработки.

7.3. 3D-визуализация и 3D-анимация в ГИС.

Количество часов, отведенных на изучение главы, – 4.

Аннотация темы.

Тематика лекции посвящена особенностям создания и анализа трехмерных моделей в ГИС. В пункте 7.1 дано определение понятия «пространственно-временная модель», рассмотрены особенности выполнения визуализации куба «пространство-время» и анализа тренда по алгоритму Манна-Кендалла. В этом пункте приведена также подробная характеристика методов интеллектуального анализа пространственно-временных данных. В пункте 7.2 раскрыты особенности выполнения трехмерного моделирования поверхности и рассмотрены операции геообработки 3D-моделей. В пункте 7.3. приведено описание процесса выполнения визуализации данных в трех измерениях и 3D-анимации.

Литературные и информационные источники

1. Домрачева, А. Б. Пространственно-временное моделирование / А. Б. Домрачева. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2010. – 56 с.
2. Курлович, Д. М. Геоинформационные технологии. Лабораторный практикум / Д. М. Курлович, Н. В. Жуковская, О. М. Ковалевская. – Минск: БГУ, 2015. – 160 с.
3. Цветков, В. Я. Модели в информационных технологиях / В. Я. Цветков. – М.: Макс Пресс, 2006. – 104 с.
4. Classification and regression trees / L. Breiman, [et al.]. – New York: Academic, 2017. – 368 p.

Internet-ресурсы

5. Сайт компании ESRI [Электронный ресурс]. – Режим доступа: <https://www.esri.com/ru-ru/home>.
6. Концепции инструментов углубленного анализа пространственно-временных закономерностей [Электронный ресурс] – Режим доступа: <https://pro.arcgis.com/ru/proapp/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm>.

7.1. Пространственно-временное моделирование и визуализация куба «пространство – время»

Пространственно-временные данные – это данные, относящиеся как к пространству, так и ко времени, содержащие сведения о геопространственных объектах за определенный промежуток времени. **Пространственно-временная модель** – модель, в которой входные и выходные данные являются пространственно-временными. Такая модель может классифицироваться как детерминированная, и как стохастическая, а для обработки пространственно-временных данных используют методы пространственной статистики. Пространственно-временные модели применяют для проверки гипотез, а также при имитационном моделировании. Под **пространственно-временным моделированием** подразумевают построение пространственно-временных моделей исследуемых объектов на основе результатов их дистанционного мониторинга или проектирования.

Пространственно-временные данные представляются в виде куба «пространство-время», позволяющего визуализировать и анализировать их посредством анализа временных рядов и выполнения интеллектуального анализа пространственных и временных закономерностей (рис. 7.1).

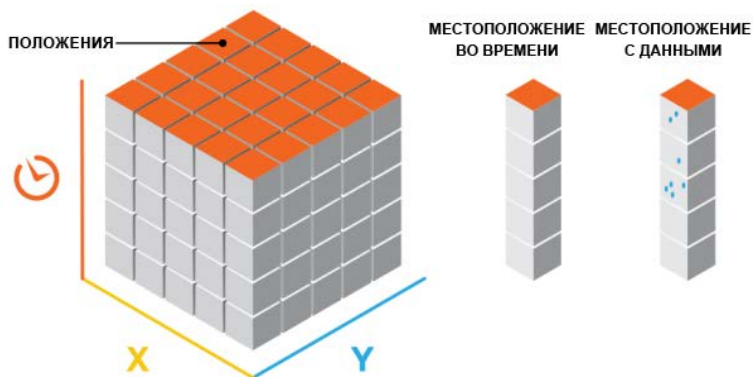


Рис. 7.1. Визуализация куба «пространство – время»

Куб «пространство – время» создается при наличии точечных геопространственных объектов с временной меткой, которые необходимо агрегировать в пространстве, чтобы определить пространствен-

но-временные закономерности в изучаемых местоположениях данных объектов. В результате его построения получают куб сетки (с прямоугольными или гексагональными ячейками), либо куб, структурированный по заданным местоположениям, использованным как полигоны агрегации. В каждом бине этого куба, где x , y -измерения представляют пространство, а t -измерение – время, выполняется подсчет количества точек, вычисляется статистика поля суммирования и определяется наличие трендов во времени в каждом местоположении с использованием **тенденционного теста Манна-Кендалла** – непараметрического теста для определения наличия монотонной, статистически значимой тенденции, основанного на S -статистике (при наличии временных рядов из менее чем 10 значений) и Z -статистике (при наличии временных рядов из более чем 10 значений), рассчитываемых как разность между возрастающими и уменьшающимися парами значений во временном ряду. Каждый бин имеет определенное положение в пространстве (x , y) и во времени (t), а бины, расположенные в одном и том же фрагменте пространства (x , y), имеют один и тот же идентификационный номер местоположения. Бины, охватывающие одинаковый временной период, имеют один и тот же идентификатор временного шага (рис. 7.2).

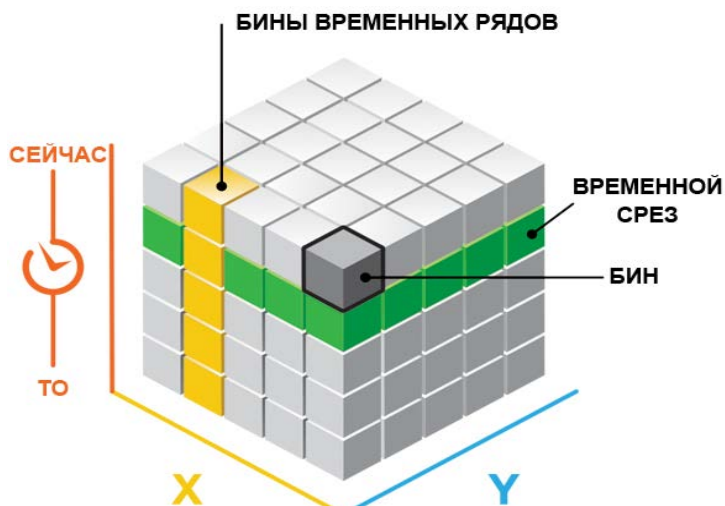


Рис. 7.2. Бины в пространственно-временном кубе

Статистика Манна-Кендалла подразумевает ранговый корреляционный анализ количества точек или значений и их временных периодов. Значение бина первого интервала сравнивается со значением бина в следующем, и если первое значение меньше второго – результат равен +1, если больше – результат равен –1, если два значения совпадают – результат равен 0. Результаты каждой пары временных периодов сравниваются и суммируются. Ожидаемое значение суммы равно 0 ($S = 0$, нулевая гипотеза) и свидетельствует об отсутствии тренда в значениях во времени. На основе вариации значений во временных рядах бинов, количества совпадений и общего числа временных периодов общая сумма сравнивается с ожидаемой суммой для оценки того, статистически значима ли разница. Тренд для каждого временного ряда бинов определяется z -оценкой и p -значением. Низкие значения p свидетельствуют о наличии статистически значимого тренда в пространственно-временных данных, тогда как величина z -оценки показывает, свидетельствует ли тренд о возрастании значений в бинах (положительное значение z -оценки) или об их убывании (отрицательное значение z -оценки).

При создании куба «пространство – время» наборы данных структурируются и суммируются в формат данных netCDF посредством создания бинов пространства-времени. NetCDF (Network Common Data Form) – это набор программных библиотек и машинно-независимых форматов данных, которые поддерживают создание, доступ и совместное использование научных данных, ориентированных на массивы.

Интеллектуальный анализ пространственно-временных данных относится к категории задач data mining⁷² и представляет собой процесс получения новых знаний посредством обнаружения закономерностей в существующих данных об изменении характеристик геопространственных объектов во времени (рис. 7.3).

Существуют следующие подходы к интеллектуальному анализу пространственно-временных данных:

- 1) пространственная кластеризация;
- 2) анализ пространственных шаблонов;
- 3) обнаружение пространственно-временных аномалий;
- 4) анализ пространственной автокорреляции.

⁷²Data mining – собирательное название, используемое для обозначения совокупности методов обнаружения в данных, ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.



Рис. 7.3. Методы интеллектуального анализа пространственно-временных данных

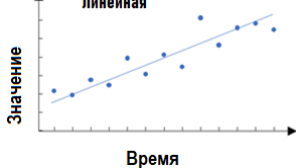
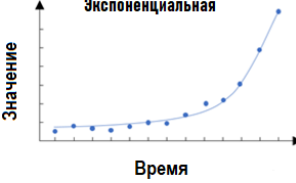
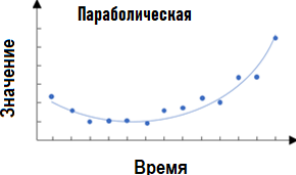
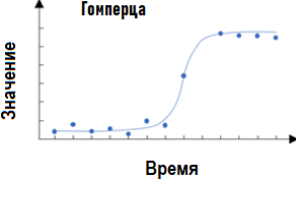
Примерами интеллектуального анализа пространственно-временных данных могут служить:

- обнаружение эволюционной истории городов и земель;
- поиск оптимальных мест для размещения торговых и социально-культурных объектов;
- выявление неблагоприятных погодных условий;
- прогнозирование землетрясений и ураганов и определение тенденций глобального потепления;
- прогнозирование спроса на товары и услуги на заданный временной период (например, день следующей недели).

Метод подгонки кривой для моделирования временных рядов и прогнозирования будущих значений в каждом местоположении в кубе пространства-времени позволяет спрогнозировать величину значения на будущий период. Например, используя куб пространства-времени с ежегодными данными о количестве того либо иного вида земель в пределах административно-территориальной единицы, возможно спрогнозировать динамику их площади в предстоящие годы. По результатам выполнения такого прогнозирования возможно создать новый куб «пространство-время», содержащий данные из исходного куба и добавленные прогнозируемые значения. В процессе моделирования в

каждый бин входного куба «пространство-время» вписывается кривая, на основании которой прогнозируется временной ряд, и которая затем экстраполируется на будущие временные шаги. Кривые могут быть линейными, параболическими, S-образными (функция Гомперца) или экспоненциальными (табл. 7.1).

Таблица 7.1. Типы кривых и их возможное применение при прогнозировании

Тип кривой	Применение при прогнозировании
<p data-bbox="252 408 319 424">Линейная</p>  <p data-bbox="296 592 352 608">Время</p>	<p data-bbox="496 440 957 608">Линейный тип кривой подходит для данных, которые равномерно возрастают или уменьшаются с течением времени. Например, он может быть использован для прогнозирования численности популяции на том уровне развития сообществ, когда ее прирост более или менее линейный</p>
<p data-bbox="252 616 386 632">Экспоненциальная</p>  <p data-bbox="296 783 352 799">Время</p>	<p data-bbox="496 647 957 767">Экспоненциальная кривая подходит для данных, которые резко возрастают или уменьшаются с течением времени. Например, с ее помощью возможно спрогнозировать периоды быстрого роста спроса на тот либо иной товар</p>
<p data-bbox="252 807 364 823">Параболическая</p>  <p data-bbox="296 967 352 983">Время</p>	<p data-bbox="496 807 957 975">Параболические кривые подходят для данных, направление изменений которых меняется с течением времени, не важно, с увеличения на уменьшение, или наоборот. Остальные типы кривых предполагают, что значения в данных последовательно возрастают или уменьшаются во времени</p>
<p data-bbox="252 991 319 1007">Гомперца</p>  <p data-bbox="296 1158 352 1174">Время</p>	<p data-bbox="496 991 957 1198">Кривые Гомперца используются для моделирования роста с ограничением емкости. Например, численность населения, как правило, сначала растет довольно медленно, но как только достигает определенного порога, достаточного для поддержки промышленности, сразу же существенно ускоряется. Затем рост снова замедляется, так как численность достигает предела возможностей региона</p>

Еще одним методом анализ куба «пространство – время», используемого как входные данные, является **анализ горячих точек**, в ходе выполнения которого идентифицируются тренды в агрегированных в бинах точечных данных, полях суммирования или переменных, а также вычисляется локальный индекс Морана для идентификации стати-



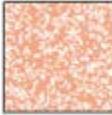
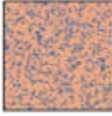



стически важных кластеров и выбросов в контексте пространства и времени. Иными словами, по аналогии с двумерными пространственными данными в пределах куба производится анализ горячих и холодных точек, а также выполняется кластерный анализ, в ходе которого определяются бины со статистически значимыми аномально высокими либо аномально низкими значениями, а также выбросы высоких и низких значений⁷³. Однако наличие третьей координаты – времени позволяет, в отличие от двумерных геопространственных данных, идентифицировать также новые и постоянные, последовательные и спорадические, колеблющиеся и исторические, возрастающие и убывающие горячие и холодные точки (табл. 7.2).

Таблица 7.2. Категории местоположений, идентифицируемые при анализе горячих и холодных точек

Обозначение	Закономерность	Определение
1	2	3
	Закономерность не обнаружена	Не соответствует ни одной закономерности горячей или холодной точки
	Новая горячая точка	Местоположение, являющееся статистически значимой горячей точкой в последнем временном шаге, которое ранее не являлось статистически значимой горячей точкой
	Последовательная горячая точка	Местоположение, характеризующееся одной непрерывной серией бинов, которые являются статистически значимыми горячими точками в последних интервалах временных шагов; не являлось статистически значимой горячей точкой до возникновения закономерности в последних временных шагах, кроме того, в этом местоположении менее 90 % всех бинов являются статистически значимыми горячими точками
	Возрастающая горячая точка	Местоположение, являющееся статистически значимой горячей точкой для 90 % интервалов временных шагов, включая последний временной шаг; интенсивность кластеризации высоких значений в каждом временном шаге с течением времени возрастает, одновременно с ростом статистической значимости

⁷³ Более подробная информация содержится в главе 5.

Продолжение табл. 7.2

1	2	3
	Постоянная горячая точка	Местоположение, являющееся статистически значимой горячей точкой для 90 % интервалов временных шагов, в котором не обнаруживается трендов, свидетельствующих о возрастании или убывании кластеризации с течением времени
	Убывающая горячая точка	Местоположение, являющееся статистически значимой горячей точкой для 90 % интервалов временных шагов, включая последний временной шаг; интенсивность кластеризации в каждом временном шаге с течением времени убывает одновременно со снижением статистической значимости
	Спорадическая горячая точка	Местоположение, в котором горячая точка то появляется, то исчезает. Менее 90 % всех интервалов временных шагов являются статистически значимыми горячими точками и ни один из интервалов не является статистически значимой холодной точкой
	Колеблущаяся горячая точка	Статистически значимая горячая точка в последнем временном шаге, которая в предыдущем временном шаге регистрировалась как статистически значимая холодная точка; статистически значимыми горячими точками являются менее 90 % интервалов
	Историческая горячая точка	Местоположение в последнем временном периоде не является горячей точкой, но ранее, не менее чем в 90 % интервалов временных шагов, являлось статистически значимой горячей точкой
	Новая холодная точка	Местоположение, являющееся статистически значимой холодной точкой в последнем временном шаге, которое ранее не являлось статистически значимой холодной точкой
	Последовательная холодная точка	Местоположение, характеризующееся одной непрерывной серией бинов, которые являются статистически значимыми холодными точками в последних интервалах временных шагов; не являлось статистически значимой холодной точкой до возникновения закономерности в последних временных шагах, кроме того, в этом местоположении менее 90 % всех бинов являются статистически значимыми холодными точками

Окончание табл.7.2

1	2	3
	Возрастающая холодная точка	Местоположение, являющееся статистически значимой холодной точкой для 90 % интервалов временных шагов, включая последний временной шаг; интенсивность кластеризации низких значений в каждом временном шаге с течением времени возрастает одновременно с ростом статистической значимости
	Постоянная холодная точка	Местоположение, являющееся статистически значимой холодной точкой для 90 % интервалов временных шагов, в котором не обнаруживается трендов, свидетельствующих о возрастании или убывании кластеризации с течением времени
	Убывающая холодная точка	Местоположение, являющееся статистически значимой холодной точкой для 90 % интервалов временных шагов, включая последний временной шаг; интенсивность кластеризации низких значений в каждом временном шаге с течением времени убывает одновременно со снижением статистической значимости
	Спорадическая холодная точка	Местоположение, в котором холодная точка то появляется, то исчезает. Менее 90 % всех интервалов временных шагов являются статистически значимыми холодными точками, и ни один из интервалов не является статистически значимой горячей точкой

Колеблущаяся

х х х х

кл

С

амз



Рис. 7.4. Результат визуализации куба «пространство – время» после выполнения анализа горячих точек

Еще одним способом интеллектуального анализа куба «пространство – время» является выполнение **прогноза в каждом местоположении** выбор наиболее точного метода для представления прогнозных данных для этого местоположения, позволяющий попробовать несколько методов прогнозирования и выбрать наиболее точное местоположение по исходным данным местоположения. При этом для каждого анализируемого положения создается интерактивная диаграмма, показывающая установленные значения, значения прогноза и доверительный интервал (если метод поддерживает доверительные интервалы) выбранного метода в местоположении (рис. 7.5).

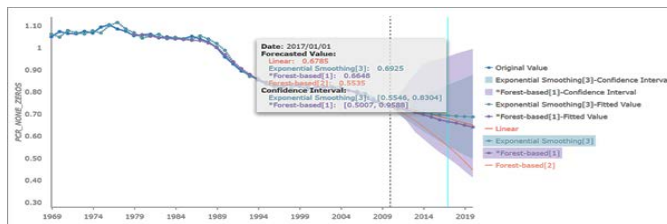


Рис. 7.5. Диаграмма прогноза, построенная по результатам анализа местоположения в кубе «пространство – время»: вертикальная пунктирная линия – время начала прогноза; окрашенные области – доверительные интервалы для выбранных методов прогноза; вертикальная голубая линия – отображает все значения диаграммы на заданном временном шаге

Для эффективного **прогнозирования будущих временных шагов** в каждом местоположении применяют метод экспоненциального

сглаживания Хольта – Винтерса⁷⁴, с помощью которого разделяют временные ряды в каждом местоположении куба «пространство-время» на сезонные и трендовые компоненты. При этом строятся две модели:

- первая – модель прогнозирования, которая используется для прогнозирования значений будущих временных шагов;
- вторая – модель проверки, используемая для проверки прогнозируемых значений (рис. 7.6).

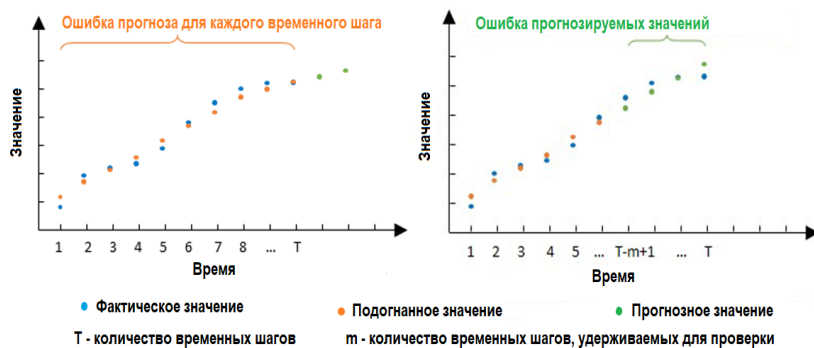


Рис. 7.6. Модель прогноза и модель проверки для временных шагов

По результатам выполнения такого прогнозирования возможно создать новый куб «пространство – время», содержащий данные из исходного куба и добавленные прогнозируемые значения.

Для прогнозирования будущих временных срезов в кубе пространства-времени используется **регрессия на основе леса решений**. Обучающие данные, использованные для построения модели регрессии леса, создаются путем использования временных окон в каждом местоположении в кубе пространства – времени (рис. 7.7).

⁷⁴Метод Хольта – Винтерса используется для прогнозирования временных рядов, когда в структуре данных есть сложившийся тренд и сезонность или тенденция к росту или падению значений временного ряда. Также применяется для прогнозирования временных рядов, когда имеются данные за неполный цикл (например, за неполный год для прогноза по месяцам) и выделить сезонность не представляется возможным.

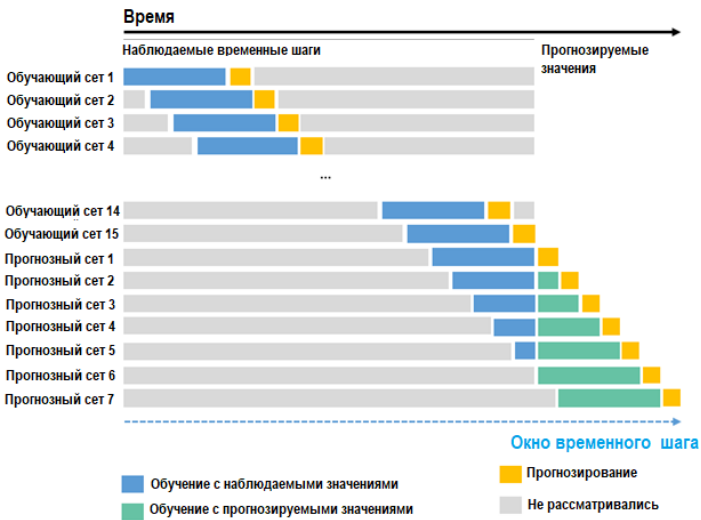


Рис. 7.7. Процесс обучения модели леса и прогнозирования первых семи будущих временных шагов (обучающий сет – набор независимых переменных; прогнозный сет – набор зависимых переменных)

Классификация пространственно-временных данных на основе регрессионных лесов обучает модель на основании известных значений, предоставленных как часть обучающего набора данных. Затем обученная модель прогнозирования используется для прогноза неизвестных значений в наборе данных прогнозирования, который содержит такие же связанные независимые переменные. Например:

- при наличии пространственно-временных данных об урожайности культур на сотнях сельскохозяйственных предприятий по всей стране, а также других атрибутов для каждого из этих предприятий (количество работников, площадь пахотных земель), а также ряда растров, которые представляют пространственное распределение агрохимических и физико-химических свойств почв, возможно спрогнозировать во времени урожайность для тех предприятий, данные по которым отсутствуют, однако имеются в наличии все остальные переменные;

- на основе сведений о ценах на недвижимость за предыдущий период, а также при наличии информации о числе комнат, расстоянии до школ, близости автомагистралей, объектов торговли и социальной ин-

фраструктуры может быть спрогнозирована продажная стоимость аналогичной недвижимости на перспективу.

Модели регрессии на основе леса особенно эффективны, когда данные содержат сложные тренды или являются сезонными, или изменяются способами, отличными от стандартных математических функций, таких как полиномы, экспоненциальные кривые или синусоидальные волны.

7.2. Особенности создания трехмерных моделей поверхности и их геообработки

Трехмерное моделирование поверхности – это процесс разработки пространственной формы поверхности в трех измерениях с учетом геоференции⁷⁵, являющийся одной из разновидностей геоинформационного моделирования. Его цель – информационное управление, получение качественно новых знаний и формирование информационных геопространственных ресурсов. Трехмерное моделирование играет важную роль в научных исследованиях и практической деятельности, позволяя выполнять:

- изучение рельефа и его морфометрических характеристик;
- моделирование полей рассеивания химических веществ в гидросфере, атмосфере и педосфере;
- моделирование городской среды и транспортных коммуникаций;
- имитацию чрезвычайных ситуаций (наводнения, сход лавин и селевых потоков).

В процессе создания трехмерной модели поверхности выполняются следующие *операции*:

- конструирование трехмерной сцены;
- назначение поверхности 3D-сцены растровой текстуры, имитирующей реальные объекты;
- анализ трехмерной поверхности;
- установка и настройка освещения 3D-сцены;
- анимация 3D-сцены для передачи динамики процесса;
- вывод 3D-сцены на дисплей или принтер.

Трехмерное моделирование является важным направлением в развитии ГИС-технологий и позволяет представлять реальную местность,

⁷⁵Геоференция – соотнесение информации об объектах на земной поверхности, в околоземном пространстве, в подземном пространстве с метрической составляющей гео-данных.

объекты окружающего мира и их взаимное расположение, а также рассматривать объекты в виртуальной реальности. Виртуальная реальность⁷⁶ в свою очередь позволяет осуществлять имитационное моделирование и выбирать оптимальное расположение проектируемого объекта.

3D-модель поверхности является цифровым отображением пространственных объектов, как реальных, так и гипотетических, в трехмерном пространстве. Поверхности могут быть как реальными (ландшафт, городские улицы, цифровая модель местности), так и вычисленными или гипотетическими (пример вычисленной поверхности – уровень загрязнения почвы тяжелыми металлами в пределах каждого поля; вымышленные поверхности используются в видеоиграх или компьютерных симуляциях).

3D-поверхности создаются посредством вычисления с использованием специальных алгоритмов, которые обрабатывают исходные точечные, линейные или полигональные геопространственные данные и конвертируют их в цифровую 3D-поверхность. В ГИС могут создаваться и храниться четыре типа моделей поверхности: растровые, наборы данных TIN и terrain⁷⁷ и наборы данных LAS⁷⁸. Эти модели поверхностей создаются с помощью двух основных методов – интерполяции и триангуляции.

Растры, TIN, наборы данных terrain и наборы данных LAS являются различными типами функциональных поверхностей – непрерывных полей значений, которые могут меняться в бесконечном количестве точек.

Модели поверхностей позволяют хранить информацию о поверхности в ГИС. Поскольку поверхность может содержать бесконечное количество точек, значения в которых невозможно измерить, в модели поверхности используется аппроксимация поверхности по отдельным значениям, взятым в различных точках поверхности, с последующей интерполяцией значений между этими точками.

⁷⁶Виртуальная (искусственная) реальность – созданный техническими средствами мир, передаваемый человеку через его ощущения: зрение, слух, осязание и имитирующий как воздействие, так и реакции на воздействие.

⁷⁷Наборы данных terrain – эффективный способ управления большими объемами точечных данных дистанционного зондирования, получаемых с помощью LIDAR и SONAR.

⁷⁸Набор данных LAS – это отраслевой стандартный двоичный формат для хранения бортовых лазерных данных, позволяющий быстро и легко изучать файлы LAS в их исходном формате и предоставляющий детализированную статистику и область покрытия лидарных данных, содержащихся в файлах LAS.

На рис. 7.8 показана модель поверхности содержания гумуса в почве в пределах землепользования сельскохозяйственного предприятия. Точками указаны места, где были получены значения отображаемого показателя.

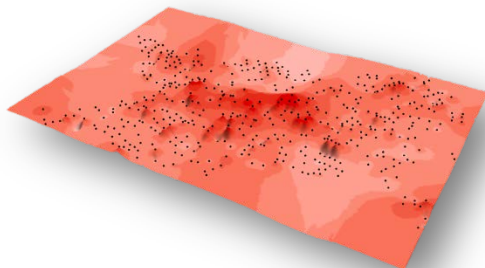


Рис. 7.8. 3D-модель поверхности содержания гумуса, созданная посредством интерполяции методом обратных взвешенных расстояний

Геообработка – это обработка пространственных данных, которая обеспечивает возможность создания новых данных путем обработки существующих. Любое преобразование данных или извлечение информации требует выполнения задач геообработки, например, конвертация географических данных в другой формат или сложный процесс, состоящий из нескольких последовательных операций, таких как вырезание, выборка и пересечение наборов данных.

Операции геообработки 3D-моделей позволяют:

- создавать и редактировать растровые, TIN и terrain-поверхности и извлекать из них различную информацию;
- конвертировать TIN в векторные или растровые данные;
- создавать 3D пространственные объекты на основе функциональных поверхностей путем извлечения данных высот;
- интерполировать данные растров;
- выполнять математическую обработку и переклассификацию растров;
- получать из TIN и растров данные о высотах, уклонах, экспозициях и объемах.

Посредством геообработки возможно выполнять **количественное выражение и отображение форм рельефа** земной поверхности, пред-

ставленного цифровой 3D-моделью рельефа, и выявлять закономерности, которые не видны на исходной поверхности:

- создать класс линейных объектов изолиний из растровой 3D-поверхности, класс пространственных объектов выбранных значений изолиний или построить изолинии по растровой поверхности с учетом включения объектов-барьеров;

- вычислить профильную и плановую кривизну растровой поверхности, объем насыпей и выемок, разницу в объеме между двумя поверхностями;

- создать рельеф с тенью из растровой поверхности, принимая во внимание угол источника освещения и тени;

- определить уклон (градиент, или скорость максимального изменения значения z) для каждой ячейки поверхности растра;

- создать полигональные объекты, которые представляют диапазоны значений уклона для триангулированных поверхностей (рис. 7.9).

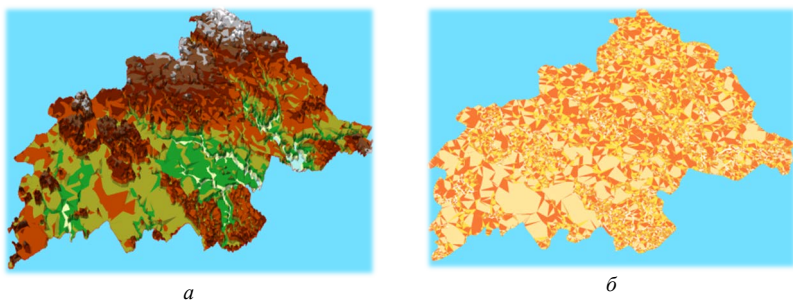


Рис. 7.9. Уклон растровой 3D-поверхности:
а – TIN-модель поверхности с применением вытягивания; *б* – полигональные объекты, которые представляют диапазоны значений уклона

Посредством **определения объема насыпей и выемок** вычисляются площади и объемы, которые изменяются в два различных момента времени, и определяются участки, с которых материал был удален, участки, на которые материал был добавлен, а также участки, где поверхность не претерпела изменений (рис. 7.10).

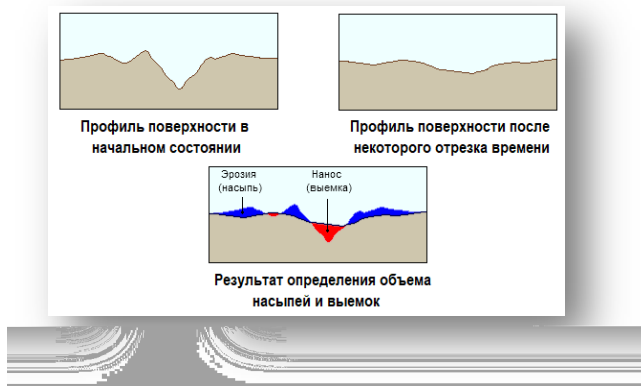


Рис. 7.10. Определение объема насыпей и выемок

Данная операция *позволяет*:

- определять участки проявления эрозии и накопления осадочного материала в долине реки;
- вычислять объемы и площади поверхностного материала, который должен быть удален, а также площади и объемы участков, которые должны быть выровнены до уровня площадки под строительство;
- вычислять объемы и площади поверхностного материала, необходимого при выполнении рекультивационных работ на территории, нарушенной вследствие добычи полезных ископаемых открытым способом;
- выявлять участки, которые часто заполняются поверхностным материалом во время грязевых оползней, а также определять безопасные для строительства домов участки со стабильным грунтом.

Анализ зон видимости⁷⁹ – функция, использующая значение высоты каждой ячейки раstra цифровой модели рельефа (ЦМР) для определения видимости конкретной ячейки или из нее. Местоположение этой конкретной ячейки варьируется в зависимости от потребностей анализа. Каждой ячейке выходного раstra присваивается значение, указывающее, из скольких точек наблюдения видно данное местоположение. Если точка наблюдения только одна, каждой ячейке, видимой из этой точки, присваивается значение 1. Всем ячейкам, которые не видны из этой точки наблюдения, присваивается значение 0.

⁷⁹Зона видимости – область, которая видна из конкретного местоположения.

Анализ видимости определяет местоположения растровой поверхности, видимые для набора объектов-наблюдателей, или какие точки наблюдения видны из каждого положения на поверхности раstra. Он обычно используется при определении местоположения телекоммуникационных вышек; оценке стоимости недвижимости; определении вида в пределах застройки; оценке видимости объекта, который необходимо скрыть (полигон твердых бытовых отходов, место захоронения). При этом ширина обзора может быть рассчитана с использованием как отдельной точки, например, высотного жилого здания, так и нескольких точек, например линии, обозначающей дорогу, а также оценена как для отдельного наблюдателя, так и для нескольких (рис. 7.11).

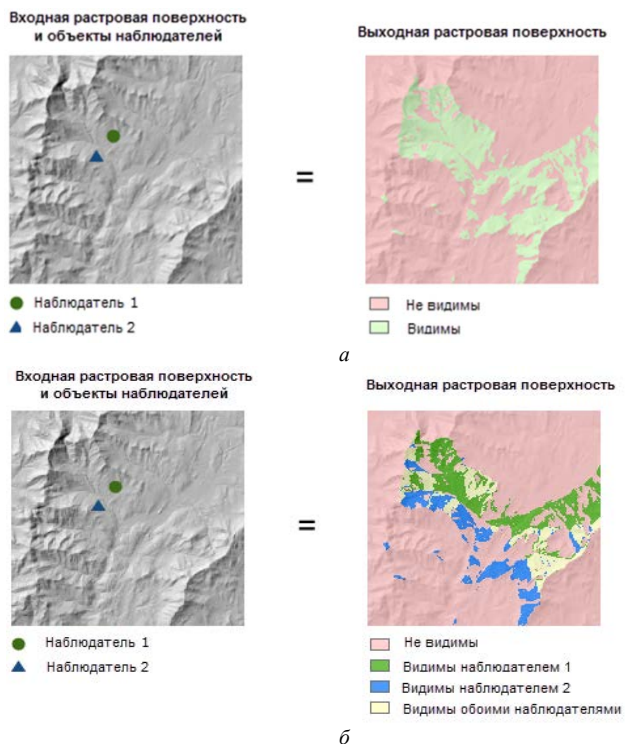


Рис. 7.11. Анализ видимости в пределах растровой поверхности:
a – анализ обоюдной видимости; *б* – анализ видимости для отдельных наблюдателей

7.3. 3D-визуализация и 3D-анимация в ГИС

Визуализация⁸⁰ данных в трех измерениях дает новые перспективы и позволяет выявить особенности, которые не видны на плоской карте с теми же данными в 2D-формате. Например, вместо того, чтобы предполагать наличие долины по конфигурации изолиний, в процессе выполнения ее 3D-визуализации возможно реально увидеть эту долину и оценить разницу высот между ложем долины и ее хребтом.

Просмотр 3D-данных в ГИС возможен в *глобальном масштабе* (используется для просмотра больших объемов 3D-данных) и *локальном масштабе*, когда осуществляется просмотр данных небольшого участка в местной системе координат (рис. 7.12).

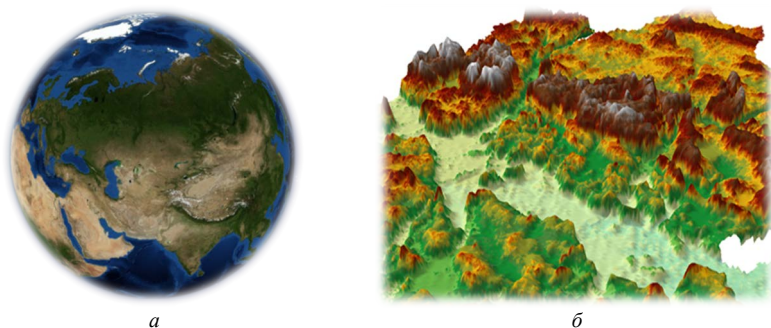


Рис. 7.12. Просмотр 3D-данных в ГИС:
а – глобальный масштаб; б – локальный масштаб

В процессе визуализации *возможно*:

- создавать многослойные 3D-ландшафты в 3D-пространстве и для каждого слоя управлять особенностями его отображения и размещения в пространстве;
- управлять глобальными характеристиками 3D-просмотра, такими как освещение или вертикальное преувеличение;
- выбирать пространственные объекты по их атрибутам или по относительному положению;

⁸⁰Визуализация – это метод представления воспринимаемой в виде оптического изображения (рисунков, диаграмм, графиков, структурных схем, карт, таблиц) информации.

– осуществлять интерактивную навигацию по 3D-просмотру или задавать координаты наблюдателя и цели для средства просмотра.

В *прикладном аспекте* 3D-визуализация *позволяет*:

– наглядно оценить воздействие проектируемых зданий на городской пейзаж при проведении проектных работ в строительстве и на транспорте;

– определить видимость и безопасность траекторий полета, в частности для беспилотных летательных аппаратов;

– изучать взаимосвязи между трехмерными объектами над и под земной поверхностью.

Одним из способов визуализация 3D-данных является выполнение драпировки изображения или векторных данных на поверхность, а также вытягивание векторных объектов в направлении от поверхности, создавая таким образом линии, стены и объемные фигуры. Использование 3D-символов позволяет делать отображение данных более реалистичным и создавать высококачественную анимацию для демонстрации или распространения. Кроме того, можно просматривать сцену с различных точек обзора с помощью разных вьюеров и менять параметры 3D-слоев, чтобы использовать прозрачность и оттенение, а также задавать систему координат и экстенг сцены, моделировать ее освещение, выполнять вертикальное вытягивание рельефа (рис. 7.13).

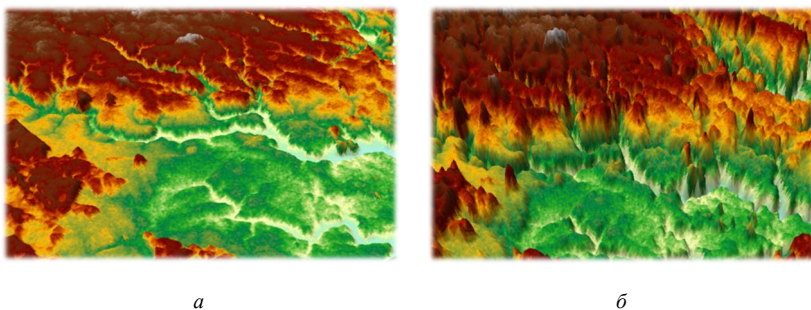


Рис. 7.13. Вытягивание вертикального размера 3D-сцены:

а – коэффициент конвертации значений высот 1,5; *б* – коэффициент конвертации значений высот 5,5

Возможна также одновременная визуализация нескольких слоев, в результате чего отображаются 3D-поверхность и трехмерные точечные, линейные либо полигональные объекты. На рис. 7.14 представлен фрагмент тепловой карты, на которой в форме 3D-поверхности отоб-

ражено содержание гумуса в почве, а в форме вытянутых 3D-точек – содержание подвижного фосфора в определенных точках поверхности.

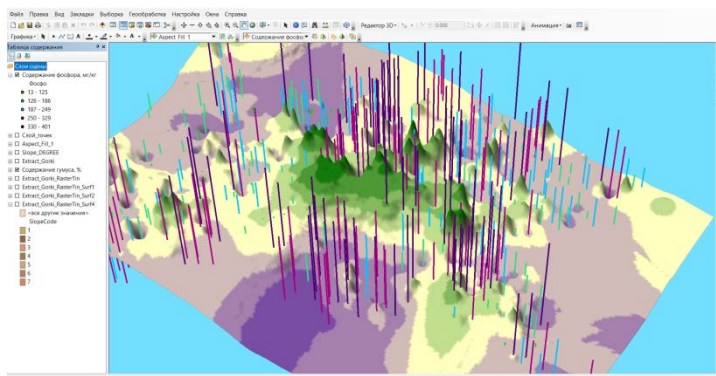


Рис. 7.14. 3D-визуализация тепловой карты

Примером реализации 3D-визуализации служат функциональные возможности модуля ArcScene ГИС ArcGIS, используемые для трехмерной визуализации 3D-сцен и их участков, а также для создания 3D-моделей рельефа и местности. С его помощью создается трехмерный вид сцены, в которой возможно выполнять различную навигацию. Данный модуль является также средством просмотра 3D-изображений и хорошо подходит для генерации перспектив, которые можно использовать для навигации и работы с 3D-векторными и растровыми данными. Основанный на технологии OpenGL⁸¹ модуль ArcScene поддерживает сложные трехмерные линейные символы и наложение текстур, а также создание поверхностей и отображение TIN. Все данные загружаются в память, что обеспечивает довольно быструю навигацию, панорамирование и изменение масштаба сцен.

3D-анимация – это визуальное отображение изменений свойств одного объекта (например, сцены или слоя) или набора объектов (например, нескольких слоев). Данный прием позволяет «оживить» визуализацию объекта благодаря сохранению действий, которые могут повторяться, создавая таким образом иллюзию непрерывности. С по-

⁸¹OpenGL (Open Graphics Library) – спецификация, определяющая независимый от языка программирования программный интерфейс для написания приложений, использующих двумерную и трехмерную компьютерную графику.

мощью 3D-анимации возможно визуализировать изменения в перспективе, изменения в свойствах объектов и географическую динамику. Она также используется для понимания закономерностей, проявляющихся в данных с течением времени, и для автоматизации процессов, которые должны быть произведены при показе точек, создаваемых только посредством наглядной динамики. Различают следующие типы 3D-анимации в ГИС:

- перемещение по изображению посредством изменения места положения сцены или объекта над ландшафтом;
- изменение прозрачности и видимости слоев сцены, позволяющее просматривать информацию, содержащуюся в различных слоях;
- перемещение камеры или вида сцены вдоль маршрута;
- анимация данных во времени при перемещении камеры;
- изменение цвета фона, освещения или вертикального масштаба сцены.

3D-анимация позволяет выполнить имитацию процесса или явления. На рис. 7.15 отображены фрагменты 3D-сцен с имитацией наводнения при различном уровне поднятия воды в поверхностном водном объекте.

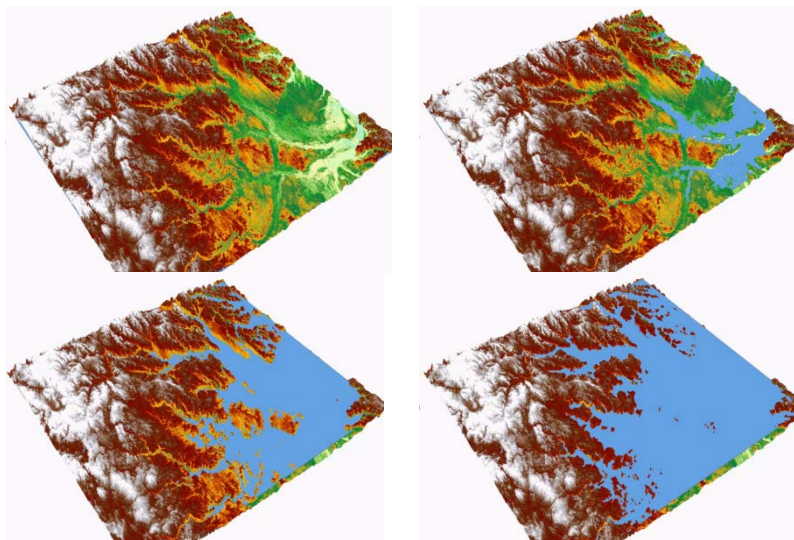


Рис. 7.15. Фрагменты анимации симуляции наводнения

Чтобы анимировать свойства объекта (слоя, таблицы, камеры, вида или сцены) или нескольких объектов, необходимо создать связанный с объектом трек⁸². Анимация состоит из одного или нескольких треков, а каждый трек состоит из нескольких кадров⁸³. Существует три типа кадров, которые могут использоваться для построения анимации 3D-сцен:

- первый тип – это кадр камеры, являющийся мгновенным снимком вида, который отображается в текущем экстенде сцены;
- второй тип – кадр слоя, являющийся снимком с отображением свойств слоя;
- третий тип – кадр сцены, хранящий свойства сцены.

Объектом, к которому привязывается трек, может быть сцена, слой, таблица, камера или вид, а контролируемые треком свойства объекта – это фоновый цвет документа, видимость слоя, местоположение камеры или временная метка слоя.

При воспроизведении анимации изменения свойств объекта или объектов в каждом кадре в составе трека отображаются на экране. Для трека расположение камеры может быть различным от кадра к кадру. Для трека слоя времени время, интервал и единицы времени (часы или дни) могут изменяться в каждом кадре, а попадающие в интервал между кадрами временные метки отображаются последовательно.

Вопросы для самостоятельного изучения

1. 3D-символы, графика и текст в ГИС.
2. 3D-картографирование и виртуальное картографирование.

Источники информации

1. Лурье, И. К. Основы геоинформатики / И. К. Лурье // Информатика с основами геоинформатики: учеб. пособие: в 2 ч. / И. К. Лурье, Т. Е. Самсонов – М.: Географический факультет МГУ, 2016. – Ч. 2: Основы геоинформатики. – 200 с.

2. Самсонов, Т. Е. Основы геоинформатики: практикум [Электронный ресурс] / Т. Е. Самсонов. – Режим доступа: <https://tsamsonov.github.io/arcgis-course/index.html#%D0%BE%D0%BF%D1%80%D0%B0%D0%BA%D1%82%D0%B8%D0%BA%D1%8%D0%BC%D0%B5>.

⁸²Трек – упорядоченный набор однородных кадров, которые при проигрывании анимации показывают изменения в динамике.

⁸³Кадр – моментальный снимок свойств объекта в определенный момент времени при анимации.

Тестовые задания для проверки знаний

1. *Правильным или нет является утверждение?* Процесс получения новых знаний посредством обнаружения закономерностей в существующих данных об изменении характеристик геопространственных объектов во времени называют data mining.

- 1) да;
- 2) нет.

2. *Вставьте пропущенное слово либо словосочетание.* Куб «пространство – время» создается при наличии _____ геопространственных объектов с временной меткой, которые необходимо агрегировать в пространстве, чтобы определить пространственно-временные закономерности в изучаемых местоположениях данных объектов.

- 1) точечных;
- 2) полигональных;
- 3) линейных;
- 4) мультиточечных;
- 5) полилинейных.

3. *Дополните определение.* Метод интеллектуального анализа пространственно-временных данных, позволяющий спрогнозировать величину значения на будущий период, называют:

- 1) анализ горячих точек;
- 2) метод прогноза в каждом местоположении;
- 3) метод прогнозирования будущих временных шагов;
- 4) регрессия на основе леса решений;
- 5) метод подгонки кривой.

4. *Дополните определение.* Метод интеллектуального анализа пространственно-временных данных, позволяющий попробовать несколько методов прогнозирования и выбрать наиболее точное местоположение по исходным данным местоположения, называют:

- 1) анализ горячих точек;
- 2) метод прогноза в каждом местоположении;
- 3) метод прогнозирования будущих временных шагов;
- 4) регрессия на основе леса решений;
- 5) метод подгонки кривой.

5. *Дополните определение.* Метод интеллектуального анализа пространственно-временных данных, позволяющий с помощью экспоненциального сглаживания Хольта – Винтерса разделить временные ряды в каждом местоположении куба «пространство – время» на сезонные и трендовые компоненты, называют:

- 1) анализ горячих точек;
- 2) метод прогноза в каждом местоположении;
- 3) метод прогнозирования будущих временных шагов;
- 4) регрессия на основе леса решений;
- 5) метод подгонки кривой.

6. *Дополните определение.* Метод интеллектуального анализа пространственно-временных данных, позволяющий выполнить прогнозирование будущих временных срезов в кубе «пространство – время», называют:

- 1) анализ горячих точек;
- 2) метод прогноза в каждом местоположении;
- 3) метод прогнозирования будущих временных шагов;
- 4) регрессия на основе леса решений;
- 5) метод подгонки кривой.

7. *Правильным или нет является утверждение?* В ГИС могут создаваться и храниться четыре типа моделей поверхности: растровые, наборы данных TIN и terrain и наборы данных LAS, создаваемые с помощью двух основных методов – интерполяции и триангуляции.

- 1) да;
- 2) нет.

8. *Дополните определение.* Упорядоченный набор однородных кадров, которые при проигрывании анимации показывают изменения в динамике, называют:

- 1) кадр;
- 2) сцена;
- 3) трек;
- 4) блок;
- 5) бин;
- 6) итератор.

9. *Правильным или нет является утверждение?* 3D-модель поверхности является цифровым отображением только реально существующих пространственных объектов в трехмерном пространстве.

- 1) да;
- 2) нет.

10. *Правильным или нет является утверждение?* 3D-поверхности создаются посредством вычисления с использованием специальных алгоритмов, которые обрабатывают только исходные точечные геопространственные данные и конвертируют их в цифровую 3D-поверхность.

- 1) да;
- 2) нет.

ОСНОВНЫЕ ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Атрибут пространственного объекта – непозиционная характеристика пространственного объекта с ее качественным или количественным значением.

Адресные данные (пространственного объекта) – минимальный набор атрибутов пространственного объекта, позволяющий идентифицировать пространственный объект как уникальный среди других пространственных объектов, включающий наименование пространственного объекта и его характеристики, используемые для обмена данными.

Аналитическая модель рельефа – модель рельефа, предполагающая использование нелинейных методов интерполяции высот или глубин. В аналитических моделях рельефа используются сплайны порядка, полиномиальные (полиномы 2-го и более высокого порядка), мультиквадратические, тригонометрические и другие функции.

Атрибутивная точность пространственных данных – близость фактических атрибутивных данных пространственных объектов к истинным.

Атрибутивные данные пространственного объекта (атрибутика пространственного объекта) – набор имен и значений атрибутов пространственного объекта.

Атрибутирование пространственного объекта – присвоение пространственному объекту атрибутов.

База пространственных данных (база геоанных, пространственная база данных, база данных ГИС) – совокупность пространственных данных, организованных по определенным правилам, устанавливающим общие принципы описания, хранения и манипулирования данными, предназначенная для удовлетворения информационных потребностей пользователя.

Базовые пространственные данные, БПД – общедоступная часть ресурсов пространственных данных, включающая информацию об их координатной основе и избранных пространственных объектах, необходимых для позиционирования пространственных данных.

Векторизация – преобразование растровой модели пространственных данных в векторную модель.

Векторизация цифровой картографической информации – преобразование цифровой картографической информации из растровой формы представления в векторную.

Векторная модель пространственных данных – модель пространственных данных, включающая описание координатных данных про-

пространственных объектов и, возможно, топологических отношений между ними.

Векторная нетопологическая модель пространственных данных – векторная модель пространственных данных, не включающая в себя описание топологических отношений между пространственными объектами.

Векторная топологическая модель пространственных данных – векторная модель пространственных данных, включающая в себя описание топологических отношений между пространственными объектами.

Векторная форма представления цифровой картографической информации – способ представления метрической картографической информации в виде последовательности векторов.

Визуализация данных – преобразование цифровых данных в изображение, доступное для восприятия человеком или специальным устройством. Программные средства ГИС обеспечивают визуализацию данных в форме картографических, графических, виртуально-реальностных и других геоизображений, выводимых на монитор компьютера, принтер, плоттер или иное устройство отображения.

Временная точность пространственных данных – близость фиксируемого времени существования пространственных данных к фактическому.

Генерализация данных – обобщение координатных и (или) атрибутивных данных пространственных объектов. Главные операции генерализации: упрощение, сглаживание, прерывание и утоньшение линейных объектов, объединение смежных полигональных объектов, отбор, агрегация, утрирование размеров и форм объектов, изменение их мерности, переклассификация данных.

Геоид – поверхность гравитационного поля, которая совпадает со средним уровнем моря и перпендикулярна вектору силы гравитации.

Геоинформатика – научно-техническое направление, объединяющее теорию цифрового моделирования предметной области с использованием пространственных данных, технологии создания и использования геоинформационных систем, производство геоинформационной продукции и оказание геоинформационных услуг.

Геоинформационная система, ГИС – информационная система, опирающаяся пространственными данными. По пространственному охвату различают глобальные, субконтинентальные, национальные, межнациональные, региональные, субрегиональные и локальные ГИС.

Геоинформационная технология – совокупность приемов, способов и методов применения программно-технических средств обработки и передачи информации, позволяющая реализовать функциональные возможности геоинформационных систем.

Геоинформационное картографирование – автоматизированное составление и использование карт на основе геоинформационных технологий и баз географических (геологических, экологических, социально-экономических и др.) данных.

Геоинформационное картографирование земель – автоматизированное составление и использование карт земельно-кадастровой системы на основе геоинформационных технологий и земельно-информационных баз геоданных (БГД).

Геокодирование пространственного объекта – косвенное описание местоположения пространственного объекта путем его соотнесения с позиционированным объектом. Местоположение геокодированного объекта обычно описывается через географическое название, почтовый адрес, почтовый код и другие идентификационные и адресные характеристики какого-либо позиционированного объекта.

Геоматика – научно-техническое направление, объединяющее методы и средства интеграции информационных технологий сбора, обработки и использования пространственных данных, включая геоинформационные технологии.

Геометрическая генерализация – обобщение плановых очертаний объектов с выделением типичных форм этих объектов.

Геометрический примитив – тип пространственного объекта с присущими ему геометрическими свойствами и размерностью, рассматриваемый как неделимый.

Геореляционная модель данных – хранимые отдельно позиционные характеристики и атрибутивные данные, последние из которых хранятся и управляются средствами реляционной системы управления базами данных.

Графические переменные – элементарные графические средства, используемые для построения картографических знаков и знаковых систем.

Графический оверлей – графическая композиция, получаемая наложением двух или более слоев.

Данные – информация, представленная в виде, пригодном для обработки автоматическими средствами при возможном участии человека.

Земельно-информационная система Республики Беларусь – комплекс программно-технических средств, баз пространственно-атрибутивных данных, каналов информационного обмена и других

ресурсов, обеспечивающий автоматизацию накопления, обработки, хранения и предоставления сведений о состоянии, распределении и использовании земельных ресурсов в электронном виде, в том числе средствами геоинформационных технологий.

Идентификатор пространственного объекта – уникальная характеристика пространственного объекта, присваиваемая ему пользователем или назначаемая информационной системой, которая используется для фиксации связи координатных и адресных данных пространственных объектов.

Импорт данных – прием данных из внешней среды путем их конвертирования для использования в данной геоинформационной системе в ее собственном формате.

Информационная система – система, предназначенная для хранения, обработки, поиска, распространения, передачи и представления информации.

Информационное обеспечение геоинформационной системы – совокупность знаний о предметной области информационных ресурсов, информационных услуг, классификаторов, правил цифрового описания, форматов данных и соответствующей документации, предоставляемых пользователю и (или) разработчику геоинформационных систем для решения задач ее создания, эксплуатации и использования.

Информация – сведения, воспринимаемые человеком и (или) специальными устройствами как отражение фактов материального или духовного мира в процессе коммуникации.

Инфраструктура пространственных данных, ИПД – информационно-телекоммуникационная система, обеспечивающая доступ граждан, хозяйствующих субъектов, органов государственной власти к распределенным ресурсам пространственных данных, а также распространение и обмен данными в общедоступной глобальной информационной сети в целях повышения эффективности их производства и использования.

Конвертирование данных – преобразование пространственных данных из одного формата в другой в рамках одной модели данных.

Координатные данные пространственного объекта – позиционная характеристика пространственного объекта, описывающая его местоположение в установленной системе координат в виде последовательности наборов координат точек.

Линейный объект – одномерный пространственный объект, координатные данные которого состоят из двух или более пар плановых координат, образуя последовательность из одного или более сегментов.

Логическая согласованность пространственных данных – соблюдение ограничений на координатные, атрибутивные данные и топологические отношения пространственных объектов и их наборов.

Модель геометрической сети – модель пространственных данных, описывающая пространственные объекты в виде структуры из позиционированных узлов и соединяющих их ребер.

Модель пространственных данных – набор пространственных объектов и межобъектных связей, сформированных с учетом общих для этих объектов правил цифрового описания.

Направление цифрования объекта – последовательность цифрования линейного объекта или контура площадного объекта.

Организационное обеспечение геоинформационной системы – совокупность стратегий, регламентированных мер, научно-технической документации, нормативных документов, организационно-правовых норм, материальных и финансовых ресурсов и квалифицированных кадров, поддерживающих функционирование геоинформационной системы на всех стадиях ее жизненного цикла.

Операции с координатами – изменение координат пространственных объектов с использованием их математической связи при переходе от одной системы координат к другой.

Операция «линия в полигоне» – алгоритм определения принадлежности линейного пространственного объекта полигональному объекту.

Операция «точка в полигоне» – алгоритм определения принадлежности точечного пространственного объекта полигональному объекту.

Перевычисление координат – операция с координатами пространственных объектов, основанная на математически строго определенной связи, при переходе из одной системы координат в другую, используя одни и те же исходные геодезические даты.

Переклассификация – логическое или математическое преобразование значений атрибутов пространственных объектов, принадлежащих к одному слою данных, в результате которого появляются новые атрибуты или замена исходных значений атрибутов на производные.

Поверхность – двумерный пространственный объект, образованный в своих границах набором значений функции двумерных координат в виде непрерывного поля.

Позиционирование пространственного объекта – описание координатных данных пространственного объекта в системах координат двумерного или трехмерного пространства и системах координат времени в явной форме или путем геокодирования.

Позиционная точность пространственных данных – близость к истинным результатам позиционирования пространственного объекта в пространстве. Различают точность планового положения пространственного объекта и его положения по высоте.

Полигональный объект – двухмерный пространственный объект, ограниченный замкнутым линейным объектом и обычно идентифицированный своим центроидом.

Полнота пространственных данных – необходимая достаточность и отсутствие избыточности пространственных данных.

Построение буферной зоны – порождение полигонального объекта, граница которого образована линией, равноудаленной от точечного, линейного или полигонального объекта.

Правовое обеспечение геоинформационной системы – совокупность правовых норм, регламентирующих правовые отношения на всех стадиях жизненного цикла геоинформационной системы.

Программное обеспечение геоинформационной системы – совокупность программ, в которых реализованы функциональные возможности геоинформационных систем и сопровождающей программной документации.

Происхождение пространственных данных – сведения об источниках пространственных данных и описание их жизненного цикла, используемые для косвенной оценки пространственных данных.

Простой пространственный объект – пространственный объект, описываемый одним идентификатором и одним набором атрибутивных данных.

Пространственные данные (геоинформационные, геопространственные данные, геоданные) – данные о пространственных объектах и их наборах.

Пространственные метаданные – данные о пространственных данных. Пространственные метаданные, описывающие набор пространственных данных, в общем случае могут содержать сведения о составе, статусе (актуальности и обновляемости), происхождении, местонахождении, качестве, форматах представления, условиях доступа, приобретения и использования, авторских правах на данные, применяемых системах координат, позиционной точности, масштабах и других характеристиках.

Пространственный запрос «объект в полигоне» – запрос на поиск пространственных объектов внутри или вне области, образованной кругом, прямоугольником или фигурой произвольной формы.

Пространственный объект (геообъект, геоинформационный объект) – цифровая модель материального или абстрактного объекта реального или виртуального мира с указанием его идентификатора, координатных и атрибутивных данных.

Пространственный поиск «объект в базе» – поиск объектов в базе пространственных данных по их координатам или функциям от них.

Растрезация – преобразование векторной модели пространственных данных в растровую модель.

Растровая модель пространственных данных – модель пространственных данных, описывающая пространственные объекты в виде набора пикселей с присвоенными им значениями.

Регулярная модель пространственных данных – модель пространственных данных, описывающая пространственные объекты в виде набора регулярных ячеек с присвоенными им значениями, где ячейка – минимальный адресуемый элемент дискретизации земной поверхности.

Сетевой анализ – решение оптимизационных задач с использованием модели геометрической сети: поиск кратчайшего или оптимального пути, в том числе в условиях ограничений на передвижение; расчет зон обслуживания и оптимизация пунктов обслуживания; решение транспортной задачи и задачи коммивояжера.

Сеточная модель рельефа – описание рельефа в виде набора высотных отметок в узлах прямоугольной регулярной сети в виде матрицы высот или глубин. Регулярная сеть может быть составлена из равносторонних (равнобедренных) треугольников, квадратов или прямоугольников.

Сложный пространственный объект – пространственный объект, состоящий из нескольких простых и (или) сложных пространственных объектов.

Слой пространственных данных – подмножество пространственных объектов предметной области, обладающих тематической общностью и единой для всех слоев системой координат.

Структурная модель рельефа – описание рельефа в виде множества координат, состоящего из набора подмножеств, каждое из которых описывает структурную линию рельефа. Любая зафиксированная на поверхности рельефа местности ломаная линия, которая допускает с требуемой точностью линейное интерполирование высот или глубин между смежными вершинами, может быть использована в качестве структурной. Наиболее ярко выраженными структурными линиями рельефа являются линии водоразделов и тальвегов.

Сфероид – трехмерное тело, созданное из двумерного эллипса путем вращения его вокруг малой оси.

Тело – трехмерный пространственный объект, ограниченный набором поверхностей. При моделировании тел используются трехмерные расширения традиционных двумерных моделей пространственных данных, например, в виде набора вокселей или специальные модели (например, модели конструктивной геометрии).

Техническое обеспечение геоинформационной системы (аппаратное обеспечение геоинформационной системы) – комплекс технических средств, используемых для реализации функциональных возможностей геоинформационных систем, включая устройства ввода, обработки, хранения и передачи данных.

Топологизация – внесение изменений в векторную модель пространственных данных, которые превращают ее в векторную топологическую модель.

Топологические отношения пространственных объектов (топология пространственных объектов) – свойства пространственных объектов, не нарушающиеся при взаимно-однозначных и взаимно-непрерывных преобразованиях. К топологическим отношениям относятся такие свойства, как связность, соседство, совпадение, пересечение, вложенность и т. п., используемые в векторной топологической модели пространственных данных и в операциях пространственного анализа.

Топологический оверлей – наложение двух или более полигональных объектов, в результате которого образуется новый слой, состоящий из фрагментов исходных полигональных объектов и наследующий их координатные, атрибутивные данные и топологические отношения.

Точечный объект – нульмерный пространственный объект, координатные данные которого состоят из единственной пары плановых координат. Плановые координаты описывают положение точечного объекта в двумерном пространстве. Точечный объект может иметь третью координату, определяющую его положение по высоте (глубине).

Трансформирование координат – операция с координатами пространственных объектов при переходе от одной координатной системы отсчета к координатной системе отсчета, основанной на других данных.

Триангуляционная модель рельефа – описание рельефа в виде набора высотных отметок или отметок глубин в узлах треугольников – элементов триангуляции Делоне и ее обобщений.

Цифрование (оцифровка, дигитализация) – преобразование картографических материалов в цифровые модели пространственных данных с использованием полуавтоматических и автоматических технологий и устройств ввода данных.

Цифрование картографического материала – преобразователь картографической информации в цифровую форму.

Цифровая карта, ЦК – цифровая картографическая модель, содержание которой соответствует содержанию карты определенного вида и масштаба. Классификация цифровых карт соответствует общей классификации карт: цифровая топографическая карта, цифровая авиационная карта, цифровая геологическая карта, цифровая кадастровая карта.

Цифровая картографическая модель – логико-математическое представление в цифровой форме объектов картографирования и отношений между ними.

Цифровое моделирование рельефа – создание цифровой модели рельефа и ее использование. Обработка цифровой модели рельефа служит для получения производных морфометрических показателей; расчета и построения линий тока; экстракции структурных линий и линий перегиба склонов; оконтуривания водосборных бассейнов; интерполяции высот; построения горизонталей и иных изолиний по множеству значений отметок высот (глубин); анализа видимости (невидимости); построения вертикальных профилей сечения рельефа, трехмерных изображений, в том числе блок-диаграмм; автоматизации отмывки рельефа; цифрового ортотрансформирования снимков и других вычислительных операций и графоаналитических построений. Методы и алгоритмы создания и обработки цифровой модели рельефа применимы к иным физическим или статистическим рельефам и полям.

Экспорт (данных) – передача данных в собственном внутреннем формате геоинформационной системы во внешнюю среду для использования вне данной геоинформационной системы.

Эллипс искажения (индикатриса Тиссо) – бесконечно малый эллипс в каждой точке на карте, являющийся изображением бесконечно малой окружности на поверхности эллипсоида или шара, с помощью которого производится обобщенная характеристика искажений картографических проекций.

ОТВЕТЫ К ТЕСТОВЫМ ЗАДАНИЯМ

Глава 1		Глава 2		Глава 3		Глава 4	
№ задания	Вариант ответа	№ задания	Вариант ответа	№ задания	Вариант ответа	№ задания	Вариант ответа
1	4	1	2	1	5	1	1
2	5-3-2	2	4	2	2	2	2
3	1	3	1	3	2	3	2
4	1	4	5	4	1, 2, 3, 5	4	5
5	2	5	1, 3, 5	5	2	5	1
6	2	6	2, 4, 6	6	5	6	1
7	3	7	1	7	1	7	4
8	2	8	2, 4, 5	8	2	8	3
9	2	9	1	9	6	9	3
10	4	10	2	10	3	10	2
		11	4	11	7	11	1
		12	1, 3, 4	12	8	12	4
		13	3	13	4	13	4
		14	1, 3, 4	14	1	14	1
		15	2	15	4	15	1
		16	2	16	1-А, 2-Б, 3-В, 4-Г, 5-Д	16	2
		17	1	17	2	17	5
		18	4	18	3	18	3
		19	1, 3, 4	19	1	19	1
		20	3	20	2	20	2
				21	4		
				22	1		
				23	2, 4, 5		
				24	2		
				25	5		
				26	4		
				27	1, 3, 6		
				28	3		
				29	6		
				30	1		
				31	1, 4, 5		
				32	4		
				33	5		
				34	1		
				35	1		
				36	2		
				37	1		
				38	2		
				39	3		
				40	4		
				41	2		
				42	2		
				43	2		
				44	1		
				45	1		
				46	1		
				47	4		
				48	2		
				49	5		
				50	1		
				51	1-6-4-3-5-2		
				52	1-А, 2-Б, 3-В, 4-Г		
				53	4		
				54	5		

Глава 5		Глава 6		Глава 7	
№ задания	Вариант ответа	№ задания	Вариант ответа	№ задания	Вариант ответа
1	1	1	1	1	1
2	3	2	3	2	1
3	2	3	2	3	5
4	1	4	1	4	2
5	3	5	3	5	3
6	2	6	2	6	4
7	1	7	4	7	1
8	3	8	6	8	3
9	3	9	5	9	2
10	1	10	1	10	2
11	2	11	7		
12	3	12	2-4-3-1		
13	2	13	2		
14	4	14	3		
15	2	15	1		
16	3	16	4		
17	1	17	1		
18	4	18	3		
19	1	19	2		
20	2	20	1		
21	2	21	3		
22	5	22	6		
23	1				
24	4				
25	3				

ПЕРЕЧЕНЬ ТЕМ ЛАБОРАТОРНЫХ ЗАНЯТИЙ

Тема лабораторного занятия	Цель лабораторного занятия	Рекомендуемое количество часов
1. Исследование геопространственных данных с помощью набора графиков для исследовательского анализа (ESDA)	Освоение методики выполнения исследовательского анализа геопространственных данных с использованием функциональных возможностей модуля Geostatistical Analyst	2
2. Создание базы геоданных в среде ArcCatalog ГИС ArcGIS	Освоение алгоритма формирования баз геоданных в среде ArcCatalog ГИС ArcGIS	4
3. Загрузка данных в кадастровую фабрику в среде ГИС ArcGIS	Освоение методики загрузки геопространственных данных в кадастровую фабрику в среде ArcCatalog ГИС ArcGIS	4
4. Выполнение сетевого анализа в ArcGIS с помощью модуля Network Analyst	Приобретение навыков создания графа и выполнения операций сетевого анализа, используя функциональные возможности ГИС ArcGIS	6
5. Моделирование пространственных отношений и расчет кластеризации геоданных	Освоение методики выполнения анализа геопространственных данных с использованием функциональных возможностей модуля Special Analyst	4
6. Моделирование пространственного распределения данных с помощью детерминированных методов интерполяции	Освоение алгоритма создания интерполированных поверхностей с использованием функциональных возможностей модуля Geostatistical Analyst	2
7. Моделирование пространственного распределения данных с помощью геостатистических методов интерполяции	Освоение алгоритма создания интерполированных поверхностей с использованием функциональных возможностей модуля Geostatistical Analyst	4
8. Комплексный морфометрический анализ ЦМР с использованием функциональных возможностей ГИС ArcGIS и его автоматизация	Ознакомление с функциональными возможностями набора инструментов SpatialAnalyst и освоение методики выполнения морфометрического анализа территории с использованием цифровой модели рельефа в среде ГИС ArcGIS. Ознакомление с функциональными возможностями модуля Model Builder и освоение методики создания геоинформационных моделей	6
9. Создание 3D-изображений и их анимация	Освоение алгоритма создания 3D-изображений и настройки их анимации	4

ПЕРЕЧЕНЬ БЕСПЛАТНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ ДЛЯ ПОЛУЧЕНИЯ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ

1. Copernicus Open Access Hub [Электронный ресурс]. – Режим доступа: <https://scihub.copernicus.eu/>.
2. FAO soils portal [Электронный ресурс]. – Режим доступа: <https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>.
3. SentinelHub, сервис EO Browser [Электронный ресурс]. – Режим доступа: <https://apps.sentinel-hub.com/eo-browser>.
4. Бесплатный сервис www.extract.bbbike.org.
5. Официальный сайт Геологической службы США [Электронный ресурс]. – Режим доступа: <https://earthexplorer.usgs.gov/>.
6. Сайт GisLab [Электронный ресурс]. – Режим доступа: <https://gis-lab.info/>.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1. ВВЕДЕНИЕ В ГЕОМАТИКУ	4
1.1. Геомастика как научная дисциплина, технология и сфера производственной деятельности. Цель и задачи изучения дисциплины.....	5
1.2. Место геомастики в системе наук. Понятие и функции геоинформационных систем и технологий	7
1.3. Возможности применения геоинформационных систем и технологий	12
2. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ ГЕОМАТИКИ.....	20
2.1. Основные понятия и термины геомастики. Пространственно-распределенные данные.....	21
2.2. Метрика в пространстве. Пространственное разрешение.....	26
2.3. Сеть мониторинга и кластерность. Декластеризация данных	29
2.4. Пространственная непрерывность данных. Стационарность данных.....	37
2.5. Геостатистическое оценивание данных. Проверка качества модели методом кросс-валидации	39
3. ВВЕДЕНИЕ В ГЕОПРОСТРАНСТВЕННЫЙ АНАЛИЗ	47
3.1. Пространственные объекты и пространственные атрибуты. Анализ местоположения объектов	48
3.2. Анализ распределения числовых показателей. Карты плотности. Поиск объектов внутри области. Анализ окружения и пространственных изменений	58
3.3. Основные виды векторного ГИС-анализа.....	78
3.4. Основные виды растрового ГИС-анализа	86
3.5. Основные операции моделирования в ГИС.....	95
4. РАБОТА С ГЕОПРОСТРАНСТВЕННЫМИ ДАННЫМИ	114
4.1. Операции реструктуризации данных. Агрегирование данных	115
4.2. Геокодирование данных	122
4.3. Классификация и переклассификация изображений.....	127
4.4. Локальные, зональные и фокальные операции.....	129
4.5. Операции вычислительной геометрии. Оверлейные операции. Построение буферных зон.....	133
4.6. Базы геоданных и их свойства. Проектирование баз геоданных	139
4.7. Подтипы, домены и их характеристика. Классы отношений геометрической сети и определение отношений между объектами. Создание правил отношений... ..	142
5. ГЕОПРОСТРАНСТВЕННАЯ СТАТИСТИКА	151
5.1. Понятие геопространственной статистики и ее прикладные аспекты	152
5.2. Усредненный центр данных и индекс ближайшего соседства	156
5.3. Анализ и моделирование пространственной корреляции. Анализ кластеризации геопространственных данных	160
5.4. Вариография. Меры пространственной корреляции.....	169
5.5. Построение вариограмм. Моделирование вариограмм. Анизотропия вариограмм.....	172
5.6. Пространственный тренд и нестационарность. Моделирование пространственных взаимоотношений	178
6. ГЕОПРОСТРАНСТВЕННОЕ МОДЕЛИРОВАНИЕ	189
6.1. Детерминированные методы геопространственного моделирования.....	190

6.2. Геостатистические методы пространственного моделирования	204
6.3. Автоматизация геопро пространственного моделирования	217
7. ТРЕХМЕРНОЕ МОДЕЛИРОВАНИЕ	230
7.1. Пространственно-временное моделирование и визуализация куба «пространство – время»	231
7.2. Особенности создания трехмерных моделей поверхности и их геообработки	242
7.3. 3D-визуализация и 3D-анимация в ГИС	248
ОСНОВНЫЕ ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ	255
ОТВЕТЫ К ТЕСТОВЫМ ЗАДАНИЯМ	264
ПЕРЕЧЕНЬ ТЕМ ЛАБОРАТОРНЫХ ЗАНЯТИЙ	266
ПЕРЕЧЕНЬ БЕСПЛАТНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ ДЛЯ ПОЛУЧЕНИЯ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ	267

Учебное издание

Мыслыва Тамара Николаевна
Куцаева Олеся Алексеевна

МЕТОДОЛОГИЯ И СОВРЕМЕННЫЕ
ПРОБЛЕМЫ ГЕОМАТИКИ

Учебное пособие

Редактор *С. Н. Кириленко*
Технический редактор *Н. Л. Якубовская*
Корректор *Н. П. Лаходанова*

Подписано в печать 13.05.2022. Формат 60×84 ¹/₁₆. Бумага офсетная.
Ризография. Гарнитура «Таймс». Усл. печ. л. 15,81. Уч.-изд. л. 13,66.
Тираж 20 экз. Заказ .

УО «Белорусская государственная сельскохозяйственная академия».
Свидетельство о ГРИИРПИ № 1/52 от 09.10.2013.
Ул. Мичурина, 13, 213407, г. Горки.

Отпечатано в УО «Белорусская государственная сельскохозяйственная академия».
Ул. Мичурина, 5, 213407, г. Горки.