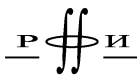


УДК 519.7  
ББК 32.81  
Г 51



Издание осуществлено при поддержке  
Российского фонда фундаментальных  
исследований по проекту 04-07-95003д

Гитис В. Г., Ермаков Б. В. **Основы пространственно-временного прогнозирования в геоинформатике.** — М.: ФИЗМАТЛИТ, 2004. — 256 с. — ISBN 5-9221-0512-4.

В книге изложены принципы геоинформационного подхода к исследованию пространственно-временных процессов и явлений. Цель книги состоит в том, чтобы снабдить специалистов руководством как для разработки аналитических геоинформационных систем (ГИС), так и для создания предметно-ориентированных ГИС-приложений, предназначенных для решения задач пространственно-временного прогноза в фундаментальных исследованиях и прикладных областях. Книга объединяет в себе методологию, методы и алгоритмы решения задач геоинформационного анализа и прогнозирования, а также примеры решения практических задач сейсмологии, геофизики, геологии, экологии и экономики. В основу книги положены оригинальные результаты, полученные в секторе Геоинформационных технологий и систем Института проблем передачи информации РАН.

Для научных работников и специалистов в области геоинформатики, анализа пространственно-временных данных и наук о Земле. Книга может быть использована в качестве учебного пособия для высших учебных заведений.

---

Научное издание

*ГИТИС Валерий Григорьевич*  
*ЕРМАКОВ Борис Владимирович*

## **ОСНОВЫ ПРОСТРАНСТВЕННО-ВРЕМЕННОГО ПРОГНОЗИРОВАНИЯ В ГЕОИНФОРМАТИКЕ**

Редактор *В.Д. Боксер*  
Оригинал-макет: *В.В. Худяков*

ЛР № 071930 от 06.07.99. Подписано в печать 21.06.04. Формат 60×90/16.  
Бумага офсетная. Печать офсетная. Усл. печ. л. 16. Уч.-изд. л. 17,6.  
Тираж 400 экз. Заказ №

Издательская фирма «Физико-математическая литература»  
МАИК «Наука/Интерпериодика»  
117997, Москва, ул. Профсоюзная, 90  
E-mail: fizmat@maik.ru, fmlsale@maik.ru; <http://www.fml.ru>

Отпечатано с диапозитивов  
в ОАО «Чебоксарская типография № 1»  
428019, г. Чебоксары, пр. И. Яковлева, 15

ISBN 5-9221-0512-4



9 785922 105125

ISBN 5-9221-0512-4

© ФИЗМАТЛИТ, 2004

© В. Г. Гитис, Б. В. Ермаков, 2004



## ОГЛАВЛЕНИЕ

Предисловие . . . . .	5
-----------------------	---

### I. Основные понятия геоинформационного прогноза

Глава 1. Введение в геоинформационный анализ . . . . .	7
1.1. Географическая информация . . . . .	7
1.2. Типы задач . . . . .	11
1.3. Средства аналитических ГИС . . . . .	12
Глава 2. Методология пространственно-временного прогноза . . . . .	17
2.1. Введение . . . . .	17
2.2. Критерии качества решения задач прогноза с неполной информацией . . . . .	18
2.3. Информационная модель как версия решения в условиях неопределённости . . . . .	20
2.4. Элементы решения задач пространственно-временного прогноза . . . . .	22
2.5. Компоненты среды информационного моделирования . . . . .	23

### II. Пространственный прогноз

Глава 3. Прогноз свойств геологической среды, не зависящих от времени . . . . .	29
3.1. Введение . . . . .	29
3.2. Технология построения информационной модели пространственного прогноза . . . . .	30
3.3. Аналитическая ГИС ГЕО 2.5 . . . . .	40
Глава 4. Вывод эмпирических закономерностей в условиях неопределённости . . . . .	42
4.1. Введение . . . . .	42
4.2. Метод интервальных экспертных оценок . . . . .	43
4.3. Метод балльных экспертных оценок . . . . .	49
4.4. Метод аппроксимации отношения правдоподобия . . . . .	57
Глава 5. Технология прогноза максимальных энергий ожидаемых землетрясений . . . . .	70
5.1. Введение . . . . .	70
5.2. Основные подходы к прогнозированию максимальных магнитуд ожидаемых землетрясений . . . . .	71
5.3. Построение прогнозной карты максимальных магнитуд землетрясений Кавказского региона . . . . .	75
Глава 6. Построение прогнозных карт нефтегазоносности . . . . .	91
6.1. Введение . . . . .	91
6.2. Надрегиональный отдельный прогноз нефтегазоносности в пределах Западно-Сибирской платформы . . . . .	93
6.3. Прогноз нефтегазовых месторождений по комплексу электроразведочных и сейсмических данных (Паннонская депрессия, Венгрия) . . . . .	99
6.4. Прогноз концентраций углеводородов в пределах месторождения в Восточном Китае . . . . .	105

Глава 7. Построение карт минерагенического прогноза . . . . .	108
7.1. Введение . . . . .	108
7.2. Прогноз свинцово-цинковых месторождений Болгарии . . . . .	109
7.3. Региональный прогноз медных месторождений Болгарии . . . . .	116
Глава 8. Пространственный прогноз на основе экспертных знаний . . . . .	130
8.1. Введение . . . . .	130
8.2. Локальный прогноз нефтегазоносности . . . . .	131
8.3. Оценка устойчивости городской территории к динамическим воздействиям. . . . .	138
8.4. Оценка ущерба от землетрясений. . . . .	146
8.5. Прогнозирование при локальном взаимодействии компонент пространственной структуры . . . . .	151

### III. Пространственно-временной прогноз

Глава 9. Пространственно-временной анализ процессов подготовки геологических катастроф . . . . .	158
9.1. Введение . . . . .	158
9.2. Предметная область . . . . .	159
9.3. Технология поиска пространственно-временных предвестников геологических катастроф. . . . .	160
9.4. Базовые функции системы ГеоТайм . . . . .	163
9.5. Методы обнаружения нестационарностей в динамических полях . . . . .	166
9.6. Выделение пространственно-временных аномалий в динамических полях. . . . .	170
Глава 10. Исследование пространственно-временных процессов подготовки землетрясения . . . . .	175
10.1. Введение . . . . .	175
10.2. Исследование процесса подготовки землетрясения по каталогам . . . . .	176
10.3. Исследование процесса подготовки землетрясения по данным гео-мониторинга . . . . .	187

### IV. Прогнозирование и анализ в сетевых ГИС

Глава 11. Сетевые аналитические геоинформационные технологии и системы . . . . .	204
11.1. Введение . . . . .	204
11.2. Потребители сетевых аналитических ГИС . . . . .	205
11.3. Схемы сетевых ГИС. . . . .	206
11.4. Вывод и объяснение при сетевом анализе ГИС . . . . .	209
11.5. Методы правдоподобного вывода ГИС ГеоПроцессор и КОМПАС. . . . .	216
11.6. Сетевые ГИС ГеоПроцессор и КОМПАС . . . . .	218
Глава 12. Комплексный анализ в сетевых ГИС . . . . .	228
12.1. Введение . . . . .	228
12.2. Оценка возможного ущерба от землетрясений . . . . .	228
12.3. Оценка сейсмической опасности географического объекта по каталогу землетрясений . . . . .	229
12.4. Выделение возможных зон возникновения землетрясений. . . . .	236
Список литературы . . . . .	244

## Предисловие

Геоинформатика изучает методы анализа информации о локализованных в пространстве объектах, явлениях и процессах реального мира. Научные исследования в области геоинформационного анализа приобретают всё большее признание в качестве одного из ключевых факторов создания информационного общества. В настоящее время разработаны сотни географических информационных систем (ГИС), которые обеспечивают возможность хранения, доступа, анализа и картографического представления географической информации в виде интерактивных электронных карт. Одновременно происходит экспоненциальный рост объёмов цифровой информации с географической привязкой, относящейся к природной среде, инфраструктуре и обществу. В связи с этим возникают задачи освоения этой информации, задачи её комплексной переработки, задачи использования географической информации для прогноза процессов и явлений, задачи извлечения из географической информации нетривиальных закономерностей и знаний. Этому направлению посвящена данная монография.

В книге изложены принципы геоинформационного подхода к исследованию пространственно-временных процессов и явлений. Цель книги состоит в том, чтобы снабдить специалистов руководством как для разработки аналитических ГИС, так и для создания предметно-ориентированных ГИС-приложений, предназначенных для решения задач пространственно-временного прогноза в фундаментальных исследованиях и прикладных областях. Книга объединяет в себе методологию, методы и алгоритмы решения задач геоинформационного анализа, а также примеры решения практических задач сейсмологии, геофизики, геологии и экономики.

Книга состоит из четырёх частей. Первая часть посвящена основам геоинформационного прогнозирования. В ней даны основные понятия аналитических ГИС, сформулированы типы задач и рассмотрена концепция геоинформационного подхода к прогнозу пространственно-временных объектов. Вторая часть относится к проблематике прогноза свойств среды, не зависящих от времени. Здесь рассмотрена технология пространственного прогноза, методы нахождения правил прогноза и примеры предметной ориентации технологии на задачи прогноза сейсмической опасности, прогноза нефтегазоносности и рудных полезных ископаемых, экологии и оценки ущерба от природно-техногенных воздействий. В третьей части исследуются проблемы прогноза пространственно-временных явлений и процессов. Предметная ориентация этой технологии относится к анализу сейсмологических и геофизиче-

ских предвестников при подготовке землетрясений. Последняя часть посвящена интенсивно развивающейся проблеме интеллектуальной обработки данных в сетевых аналитических ГИС. Рассмотренные в этой части методы анализа географической информации и примеры предметной ориентации отражают особенности сетевой геоинформационной технологии, связанные с изменением типа пользователя сетевых ГИС от специалистов в геоинформационных технологиях к специалистам в предметной области.

В основу книги положены исследования сектора Геоинформационных технологий и систем Института проблем передачи информации РАН. Работы по анализу пространственно-временных данных были начаты в ИППИ РАН в 1971 году по инициативе И. Т. Турбовича, В. И. Буне и Г. И. Рейснера в связи с исследованиями по проблеме составления прогнозных карт максимальных магнитуд ожидаемых землетрясений. Части книги 1, 3 и 4, главы 3, 4, и разделы 4.1, 4.2, 4.3, 8.1, 8.3, 8.4, 8.5 написаны В. Г. Гитисом, главы 5, 6, 7 и раздел 8.2 написаны В. Г. Гитисом и Б. В. Ермаковым, раздел 4.4 написан Е. Ф. Юрковым.

Авторы выражают признательность целому ряду лиц, непосредственно или косвенно содействующих подготовке книги. Мы считаем своим приятным долгом поблагодарить В. В. Зяблова и Т. Гергея за полезные дискуссии при разработке концепции геоинформационного прогноза пространственно-временных процессов, Е. Ф. Юркова, внёсшего существенный вклад в разработку главы 9 и разделов 5.2, 6.2, 8.3, Б. В. Ошера, который являлся одним из основных разработчиков ГИС ГЕО, ГеоТайм, ГеоРиск и ГеоПроцессор и участником исследований, рассмотренных в главах 5, 6 и 9, С. А. Пирогова, совместно с которым разработаны ряд положений главы 9 и разделов 6.3, 8.4, 8.5 и 10.3.7, Е. Н. Петрову, участвовавшую в разработке раздела 8.5, Ю. К. Шукину, совместно с которым исследовались задачи глав 5, 7 и раздела 6.2, В. Вычева, В. Шенка и З. Шенкову, с которыми развивались методические основы главы 5, Э. А. Лихачёву и Г. Л. Коффа, совместно с которыми разрабатывались разделы 8.3 и 8.4 соответственно, А. С. Горюнова, предложившего модель для решения задачи раздела 6.3, Г. А. Соболева, А. В. Пономарёва, совместно с которыми выполнялись исследования, представленные в главах 9 и 10, Чжан Чжаочэн, Ван Гусюан, Чин Синси и Тиадонг Тянь, участвовавших в исследованиях раздела 10.3, В. В. Казакова, участвовавшего в исследованиях раздела 10.2, А. П. Вайнштока, участвовавшего в исследованиях главы 11 и разделов 8.3, 8.4, А. В. Чехова и Н. А. Никулину, которыми написана сетевая ГИС КОМПАС, Г. Л. Андриенко и Н. В. Андриенко, внёсших существенный вклад в исследования главы 12, И. И. Стенину, совместно с которой выполнены экспериментальные исследования по разделу 4.3.5. Получение ряда представленных в книге результатов в области пространственно-временного прогноза и геоинформационных технологий было бы невозможно без финансовой поддержки РФФИ и Министерства науки и образования РФ.

# **I. ОСНОВНЫЕ ПОНЯТИЯ ГЕОИНФОРМАЦИОННОГО ПРОГНОЗА**

---

## **Глава 1**

### **ВВЕДЕНИЕ В ГЕОИНФОРМАЦИОННЫЙ АНАЛИЗ**

#### **1.1. Географическая информация**

Географическая информация (ГИ) включает в себя любую информацию, относящуюся к объектам, явлениям и процессам, локализованным в географическом пространстве. Геоинформатика изучает методы исследования ГИ: методы её хранения, представления, отображения, переработки, передачи и использования (детальный анализ понятия «геоинформатика» даётся в работе *Берлянд, 1996*). По определению *Burrough and McDonnel, 1998*, «Географическая информационная система (ГИС) представляет собой универсальную совокупность средств сбора, хранения, поиска, преобразования и отображения пространственных данных реального мира».

О важности ГИ можно судить по экспертной оценке Л. Уоррела (*Worral, 1991*), состоящей в том, что 80% данных, используемых менеджерами и лицами, принимающими решения, имеют географическую привязку. Традиционным средством организации доступа к ГИ являлось и является её представление в виде карт на бумажном носителе. С появлением ГИС широкое развитие получило представление ГИ в виде электронных карт. Они представляют собой изображения на экране дисплея, которые с помощью встроенных программных средств можно увеличивать, сдвигать и реализовать выбор атрибутивной и документальной информации по координатному запросу.

Геоинформационные технологии возникли как базы пространственных данных. Считается, что первой ГИС является Канадская Географическая Информационная система, разработанная в 1965 году по заказу Канадского министерства горной промышленности и ресурсов (*Frank, Raubal, and Vlugt, editors, 2000*). Дальнейшее развитие ГИС тесно связано с бурным прогрессом средств компьютерных технологий: быстрого действия, памяти и средств визуализации (*Hirtle and Frank eds. 1997*). В настоящее время созданы тысячи версий геоинформационных систем и инструментов, которые применяются практически во всех видах человеческой деятельности, имеющей дело с пространственной информацией. В первом приближении геоинформационные технологии можно классифицировать по трём проблемным областям: адресно-спра-

вочные, навигационные и рассматриваемые в настоящей монографии аналитические.

Аналитические геоинформационные технологии направлены на исследование пространственных и пространственно-временных данных и извлечение из них существенной географической информации и знаний (*Openshow and Openshow, 1997, Pyle, 1999, Malczewski, 1999, Miller and Han, editors, 2001, Klosgen and Zytkow, editors, 2002*). Анализ требует представления географического мира исследуемой предметной области в виде структурированной цифровой модели.

Географический мир состоит из природной среды и созданной человеком инфраструктуры. Методы анализа пространственно временной информации о географическом мире требуют введения некоторых ограничений. Учитывая это обстоятельство, описание географического мира можно построить с помощью трёх базовых понятий: *сущностей*, их *свойств* и *связей* (рис. 1.1).

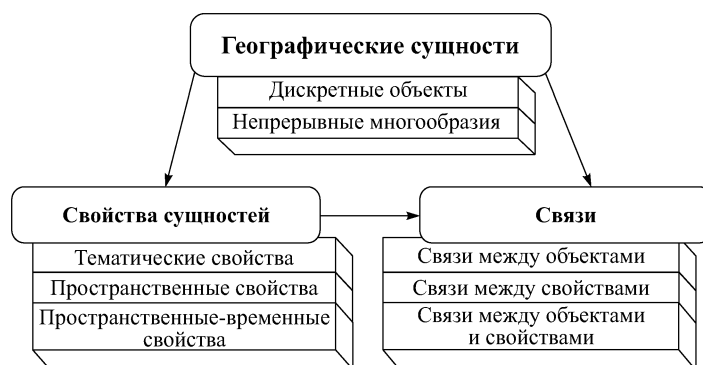


Рис. 1.1. Геоинформационное представление географического мира

Достаточно выделять два типа сущностей: *дискретные сущности* (географические объекты) и *гладкие непрерывные многообразия*. Примерами географических объектов являются единицы административного деления, месторождения, эпицентры и очаговые зоны землетрясений, озёра, реки, геологические разломы, транспортные магистрали, линии связи, природные и техногенные катастрофы, социальные взрывы, демографические процессы, природные и техногенные процессы и т. д. Непрерывные многообразия могут представлять как свойства природной среды, так и пространственно-временные свойства множеств географических объектов. Например, природную среду можно описывать с помощью геодезических, геохимических, геологических, геофизических и метеорологических полей, а пространственно-временные свойства географических объектов можно представить функциями плотности пространственно-временного распределения объектов или близости к объектам, такими, как плотность геологических разломов,



близость к водным ресурсам, плотность эпицентров землетрясений, их фрактальная размерность, плотность транспортной сети и т. д.

*Свойства* представляют собой понятия, которые характеризуют сущности. Обычно свойства классифицируются на *тематические, пространственные и пространственно-временные*. Свойства пространственной зависимости и неоднородности ГИ являются весьма важными в геоинформационном анализе. Наличие зависимости усложняет методы статистического анализа ГИ. Степень зависимости обычно оценивается функцией пространственной автокорреляции. Для различных аспектов анализа представляются существенными также ряд других способов классификации свойств географических сущностей: по способу получения информации (измеряемые свойства, экспертные решения и гипотезы, свойства, выведенные в результате обработки данных), по возможности построения логических объяснений (невербализованные и дескриптивные свойства), по типам измерительной шкалы (числовые, интервальные, порядковые, номинальные шкалы), по модели порождения (детерминистские и стохастические).

*Связи* представляют собой понятия, которые описывают отношения между географическими объектами, между свойствами и между объектами и свойствами. Связи могут представлять, например, детерминированные и стохастические причинно-следственные отношения, пространственно-временные отношения.

Пространственно-временные отношения являются весьма существенными для геоинформационного анализа и прогнозирования. Базовыми в геоинформатике являются три типа отношений пространственно-временного соседства, которые определяются топологией, расстояниями и направлениями. Из этих отношений с помощью теоретико-множественных операций могут быть получены более сложные. Топологические отношения для двух связанных объектов основаны на положении их границ, внутренних частей и дополнений. К ним относятся отношения пересечения границ, вложенности и покрытия. Топологические отношения инвариантны к непрерывным пространственно-временным преобразованиям. Отношения расстояний основаны на сравнении дистанции между объектами с константой. Отношения направлений позволяют упорядочить объекты по их расположению в пространственно-временных координатных системах. Пространственно-временные отношения между объектами обычно отображаются с помощью графов.

Цифровые модели географических сущностей состоят из географической привязки и атрибутивных данных. Для географических объектов обычно различают три типа моделей: точки, линии и полигоны. Точки задаются своими координатами, линии и полигоны представляются в виде последовательностей отрезков прямых линий с координатно привязанными точками излома. Для полигонов дополнительно указывается внутренняя область. Следует заметить, что во многих случаях один географический объект может состоять из нескольких точек, или нескольких линий, или нескольких полигонов. Например,

из нескольких полигонов может состоять государственная граница, а единая транспортная система может состоять из нескольких линий. Что касается атрибутов, то важно отметить, что они описывают свойства географического объекта, как единого целого. Атрибуты могут включать числовые значения, векторные ряды, символьные индексы, ссылки на другие объекты и источники ГИ, а также на документальную информацию: тексты, изображения, видео- и аудиозаписи. Форма, координаты и атрибутивные значения географических объектов могут изменяться во времени. В геоинформатике для объектов используется так называемые векторные форматы данных.

Цифровые модели гладких непрерывных многообразий представляют собой функции пространственных или пространственно-временных координат. Дискретизация функции может производиться с помощью триангуляции, мозаичного или сеточного представлений (*Burrough, 1992*). Наиболее распространённой формой является представление функции на прямоугольном растре с регулярной координатной сеткой. Часто географическая привязка растровой ГИ задаётся координатами двух диагонально противоположных углов сетки и шагом сетки, который зависит от точности съёма информации и степени пространственно-временной изменчивости данных. Значения в узлах сетки могут быть заданы числовыми, векторными или строковыми типами данных. В ряде случаев удобно относить эту информацию к особому виду атрибутивных данных. Этот формат данных в геоинформатике называется сеточным или растровым.

В общем случае географический мир независимо от масштаба его исследования (от детальности) представляется непрерывными и дискретными пространственно-временными сущностями. ГИС, предназначенные для изучения этого мира, поддерживают векторный и сеточный форматы ГИ и называются векторно-растровыми. В то же время существует широкий класс задач, географический мир которых может быть представлен только дискретными сущностями. Для анализа ГИ этого типа используются ГИС, называемые векторными.

Обычно цифровые модели географических сущностей группируются в информационные слои, которые состоят из однотипных семантически однородных компонентов. Например, страны, заданные государственными границами, административные регионы, озёра, геологические зоны определённого типа могут быть представлены как различные слои полигонов; дороги, реки, геологические разломы — как различные слои линий; населённые пункты, эпицентры землетрясений, места заложения промысловых нефтяных скважин — как различные слои точек. Каждое пространственное или пространственно-временное свойство удобно отождествлять с отдельным слоем. Например, земная кора приближённо может быть представлена сеточными цифровыми моделями поверхности дневного рельефа, подошвы осадочного чехла, моделями поверхностей отражающих горизонтов, моделями магнитных, гравитационных и тепловых полей, моделями скоростных неоднород-

ностей и т. д. При этом каждая сеточная модель рассматривается как отдельный информационный слой. Информационные слои могут быть жёстко заданы, а могут гибко изменяться пользователем с помощью формирования запросов к базе географических данных. Структурирование географических сущностей в информационные слои позволяет задавать выполнение групповых операций для всех компонентов слоя.

## 1.2. Типы задач

Информационные технологии пространственно-временного прогноза и аналитические ГИС используются как в научных исследованиях, так и для поиска эффективных решений в практических приложениях. Аналитические ГИС объединяют пространственно-временные цифровые модели географического мира со специализированными средствами извлечения существенной информации и знаний о географических объектах, явлениях и процессах. Это позволяет не только получить одномоментный результат, но, пополняя данные, изучать проблему в развитии, моделировать различные версии решения и выбирать из них наилучшие.

Сложность решения задач геоинформационного анализа существенно зависит от полноты имеющейся информации (*Кузнецов, 2001*).

Задачи с *полной информацией* обычно встречаются при ознакомлении с географическими понятиями по комплексу всех имеющихся данных. Анализ в таких задачах обычно состоит в выявлении качественных пространственных или пространственно-временных особенностей ГИ, в вычислении новых свойств ГИ с помощью заранее известных преобразований и в оценивании стандартных статистических характеристик векторных и сеточных данных.

Задачи с *неполной информацией* возникают при исследовании географических понятий и, в частности, при прогнозе объектов, явлений и пространственно-временных свойств геологической среды. Такие задачи являются очень сложными как с точки зрения формализации и структуризации исходной информации, так и её дальнейшей обработки и анализа. В задачах с неполной информацией знания о прогнозируемом явлении часто носят описательный характер, инструментальные измерения параметров среды, прямо связанных с математическими моделями прогнозируемых явлений, отсутствуют, а имеющиеся инструментальные косвенные измерения сильно зашумлены. Сложность задач дополнительно возрастает в связи с тем, что инструментальные измерения некоторых параметров среды либо невозможны, либо требуют дополнительной интерпретации, не всегда однозначной, и поэтому часто заменяются экспертными оценками или дополняются экспертными решениями.

Задачи с неполной информацией встречаются не только при обработке ГИ, но и в других дисциплинах, где накоплен большой набор экспериментальных фактов, а теория в основном носит описатель-

ный характер. Существенная особенность обработки ГИ обусловлена дуальностью: с одной стороны ГИ представляет тематические свойства объектов, а с другой — пространственно-временные отношения между ними, например, такие как отношение соседства, направлений и принадлежности пространственно-временной области. Аналогичные пространственно-временные отношения исследуемых объектов и их графическое представление характерны для задач обработки изображений. Отличие задач анализа ГИ состоит в многообразии типов данных. Данные, которые необходимо комплексно обрабатывать, часто включают в себя двух и трёхмерные растровые поля, векторные слои полигонов и линий, точечные маркированные слои и временные ряды с координатной привязкой.

Можно сформулировать четыре основные проблемы, для которых разрабатываются методы аналитических геоинформационных технологий.

1. Оценивание и интерпретация связей между свойствами ГИ (например, картографическое исследование взаимного расположения элементов нескольких информационных слоёв, оценивание статистических параметров свойств, в частности, оценивание прямых экономических потерь от землетрясений, оценивание запасов месторождений полезных ископаемых).

2. Оценивание и интерпретация отношений между географическими объектами (например, оценивание меры связи между природными процессами или географическими объектами, в частности, оценивание косвенных потерь от землетрясения).

3. Прогнозирование, обнаружение и распознавание целевых заранее неизвестных свойств ГИ и заранее неизвестных географических объектов (например, прогнозирование максимальных магнитуд ожидаемых землетрясений по комплексу геолого-геофизических данных, прогнозирование месторождений, обнаружение предвестников землетрясений).

4. Прогнозирование развития пространственно-временных взаимодействующих процессов (например, моделирование сценариев развития катастрофы после землетрясения).

### 1.3. Средства аналитических ГИС

Для решения аналитических задач геоинформатики в основном используются три способа анализа:

1. *Визуальное исследование (картографический и графический анализ ГИ).*
2. *Аналитические преобразования (применение заранее заданного оператора к имеющимся географическим данным для вычисления нового представления и новых свойств ГИ).*
3. *Правдоподобный вывод (нахождение оператора аналитического преобразования).*

Методы и операции визуального анализа схематически показаны на рис. 1.2. Картографические операции направлены на выявление образа пространственных и пространственно-временных свойств слоя, выявление взаимных пространственно-временных свойств нескольких географических слоёв и картографическое управление анализом ГИ.

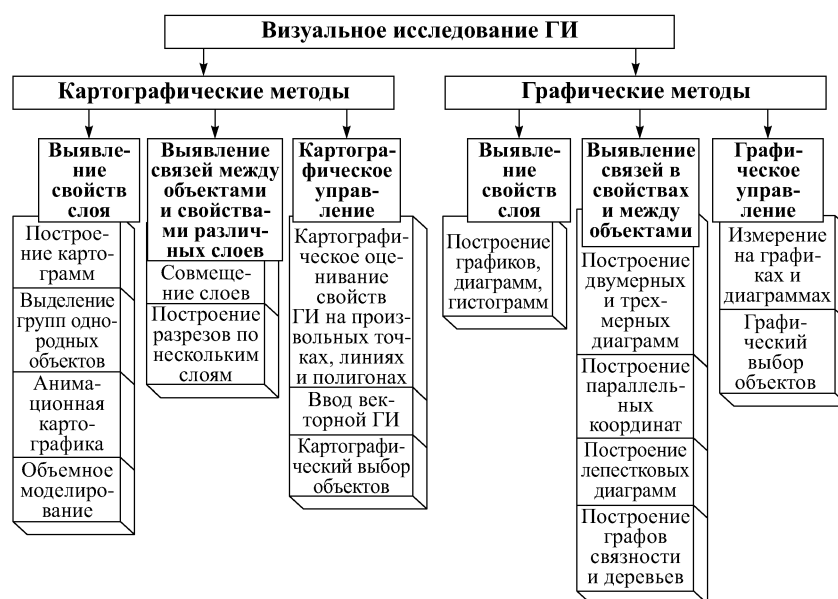


Рис. 1.2. Картографические и графические операции по извлечению существенной пространственно-временной информации из ГИ

Для выявления особенностей географического и временного расположения объектов одного слоя используются следующие типы картографических операций:

- Интерактивное картографическое представление одного или нескольких атрибутов векторной ГИ с помощью цвета, закрашки, типа и размеров пиктограмм.
- Интерактивная классификация векторной ГИ на группы однородных по одному или нескольким свойствам географических элементов.
- Интерактивное выделение зон сеточной ГИ с помощью настройки палитры на семантически обоснованные значения анализируемого свойства.
- Пространственно-временное представление динамики изменения значений растровой и векторной ГИ с помощью анимационного изменения цвета заливки, типа закрашки, типа и размеров пиктограмм.

- Построение моделей освещённости растровой ГИ и трёхмерных моделей.

Для выявления пространственно-временных связей между объектами и свойствами различных слоёв используются следующие типы операций:

- Построение карт с использованием совмещения нескольких информационных слоёв ГИ.
- Отображение разрезов нескольких слоёв ГИ по произвольному профилю.

В картографическом управлении наиболее типичными операциями являются:

- Интерактивное картографическое оценивание функционалов от свойств одного или нескольких слоёв в буферных зонах произвольных точек, линий и полигонов.
- Интерактивный ввод векторной ГИ.
- Картографический выбор объектов анализа.

Графические методы и операции поддерживают наглядное представление тематических свойств слоя и графическое управление (*Wills and Kein, 2002; Theus, 2002*).

Для представления свойств слоя и сравнения индивидуальных свойств объектов обычно используются операции представления графиков, диаграмм и гистограмм. Для графического представления многомерных свойств географических объектов и их подмножеств используются двумерные и трёхмерные диаграммы, лепестковые диаграммы и параллельные координаты (*Inselberg, 1998*). Для отображения отношений между объектами и их классификации используются направленные графы и деревья решений.

Графическое управление включает в себя графический интерфейс для измерения значений отображаемых атрибутов, для выбора способов наложения графических информационных слоёв и для изменения параметров диаграмм и гистограмм.

Для визуального исследования данных весьма важным является поддержка интерактивности картографических и графических операций (*MacEachren, 1994, MacEachren and Kraak, 1997, MacEachren and Kraak, 2001, Andrienko and Andrienko, 1999*). Интерактивность позволяет динамически управлять изменением закрашки карт и размерами пиктограмм, трёхмерным представлением ГИ, динамически изменять отображение данных на диаграммах при пересчёте значений показателей, динамически выделять на карте группы объектов в скользящем временном интервале (например, эпицентры землетрясений, координаты техногенных катастроф или правонарушений), одновременно подсвечивать группы анализируемых объектов в картографических, графических и табличных окнах. Последняя операция позволяет графически представить анализируемую информацию одновременно в географическом пространстве и в пространстве тематических свойств.

Например, при выборе группы объектов с помощью указания интервала на гистограмме распределения некоторого тематического свойства, в картографическом окне подсвечиваются объекты со значениями атрибута, принадлежащими соответствующему интервалу гистограммы, в окне представления тематических свойств с помощью параллельных координат выделяются линии этих объектов, в текстовом окне подсвечиваются соответствующие имена объектов и значения свойства и т. д.



Рис. 1.3. Аналитические методы извлечения существенной информации из ГИ с помощью операций преобразования

*Аналитические преобразования* данных позволяют по имеющимся данным вычислить новые тематические и пространственные свойства географической информации с помощью заранее заданных операторов (Andrienko, Andrienko, Gitis, 2003). При анализе векторно-растровой ГИ возможны шесть типов операций (рис. 1.3).

1. Преобразование **Сеточные слои**  $\Rightarrow$  **Сеточный слой**: например, преобразования растровой фильтрации, вычисление произвольных функций от одного или нескольких растров.

2. Преобразование **Сеточные слои**  $\Rightarrow$  **Векторный слой**: например, вычисление линий или областей по одному или нескольким растрам.

3. Преобразование **Сеточные слои и Векторный слой**  $\Rightarrow$  **Атрибуты векторного слоя**: например, вычисление статистик сеточных слоёв для буферных зон точечных, линейных и зональных географических объектов.

4. Преобразование **Векторный слой**  $\Rightarrow$  **Сеточный слой**: например, вычисление раstra (взвешенной) плотности точек или линий, пространственная интерполяция, вычисление раstra расстояний до географических объектов.

5. Преобразование **Векторные слои**  $\Rightarrow$  **Векторный слой**: например, вычисление векторного слоя с помощью теоретико-множественных операций пересечения, объединения и вычитания.

6. Преобразование **Векторные слои**  $\Rightarrow$  **Атрибуты векторного слоя**: например, вычисление метрических свойств географических объектов, вычисление произвольных функций от имеющихся атрибутов слоя, вычисление атрибутов слоя с учётом пространственных отношений и атрибутов других слоёв, в том числе вычисление атрибутов слоя по атрибутам вложенных объектов другого слоя.

Методы *правдоподобного вывода* позволяют решать задачи с неполной информацией. Эти методы направлены на нахождение аналитических преобразований, которые позволяют решать задачи прогноза, сформулированные в предыдущем разделе. Как указывалось ранее, такие задачи относятся к наиболее сложным задачам комплексного исследования ГИ. Применение методов визуального исследования и аналитических преобразований в этих задачах не позволяет получить окончательных решений. В то же время методы визуализации здесь необходимы для выдвижения и обоснования причинно-следственных моделей исследуемых данных, а методы аналитических преобразований необходимы для выявления существенных для решения задач свойств данных.

Целевые задачи правдоподобного вывода практически совпадают с проблемами, перечисленными в разделе 1.2 (рис. 1.4):

1. Оценивание зависимостей между свойствами ГИ.
2. Оценивание отношений между географическими объектами.
3. Прогнозирование заранее неизвестных свойств ГИ и обнаружение заранее неизвестных географических объектов.
4. Прогнозирование развития пространственно-временных процессов.

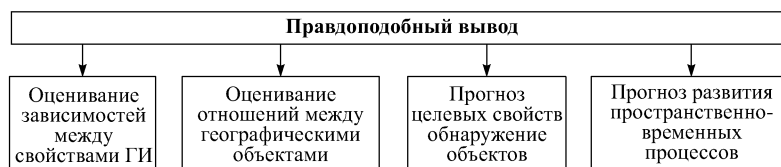


Рис. 1.4. Задачи правдоподобного вывода

Для решения этих задач используются аппарат многомерного статистического анализа, методы распознавания образов, регрессионный анализ, имитационное моделирование и методы искусственного интеллекта.



## Глава 2

# МЕТОДОЛОГИЯ ПРОСТРАНСТВЕННО- ВРЕМЕННОГО ПРОГНОЗА

### 2.1. Введение

В данной главе мы рассмотрим методологию решения задач пространственно-временного прогноза с неполной информацией. Для конкретности выберем предметную область анализа свойств геологической среды.

Мы будем исходить из того, что любой прогноз — это, прежде всего, экстраполяция, т.е. распространение выводов, полученных из наблюдений за одной частью явления на другую его часть. Носителем этих выводов являются свойства географической информации, относящиеся к исследуемым явлениям. Например, в задачах геолого-геофизического прогноза носителем выводов являются характеристики геологической среды. Сами выводы чаще всего выражаются в виде закономерности, которая связывает прогнозируемую величину со свойствами изучаемых сущностей, в частности, с геолого-геофизическими признаками. Эта зависимость обычно называется *прогнозирующей функцией*. В задачах с неполной информацией прогнозирующая функция заранее неизвестна. Она ищется по имеющимся знаниям и данным в соответствии с некоторой моделью, которая может корректироваться по мере решения задачи. Прогнозирующая функция позволяет вычислить значения прогноза для всех исследуемых явлений в пространстве и во времени. Таким образом, результат решения задачи прогноза, как правило, включает в себя три компонента: значения прогноза, прогнозирующую функцию и модель, которая должна дать возможность объяснить прогноз, и прогнозирующую функцию через базовые понятия предметной области. Далее мы рассмотрим как эти понятия можно представить в виде информационной модели исследуемого явления.

В задачах геолого-геофизического прогноза с неполной информацией исследуются две характеристики геологической среды:

- пространственное распределение не зависящих от времени (стационарных) свойств геологической среды,
- пространственно-временное распределение изменяющихся во времени (динамических) свойств геологической среды.

Комплексный анализ стационарных свойств геологической среды используется при решении таких задач, как прогноз максимальных возможных магнитуд ожидаемых землетрясений (*Буне и др., 1974*;

Рейснер, 1980; Ризниченко, 1985, Уломов, 1999), оценивание объёмов запасов месторождений (*Vainshtok et al.*, 1998), оценивание динамических свойств грунта (*Лихачёва и др.*, 1993), прогноз месторождений (*Воронин и Черемисина, 1971, Гитис и др., 1989б, Гитис и др., 1989в, Конторович и др., 1981*), прогноз возможности землетрясений с магнитудой, превышающей некоторую пороговую величину (*Schenk et al.*, 1991).

Динамические свойства геологической среды исследуются при прогнозе землетрясений, при поиске предвестников землетрясений (*Гвишиани и др.*, 1988, *Мячкин и др.* 1975, *Пономарёв и др.*, 1999, *Соболев, 1993, Соболев и Пономарёв, 2003, Zhang Guomin and Zhang Zhaocheng, 1992*), при экологическом мониторинге или при изучении сценариев развития геологических процессов (*Gitis et al.*, 1994, *Gitis et al.*, 1996, *Гитис и др.*, 2004).

Задачи прогноза стационарных и динамических свойств геологической среды имеют ярко выраженную прикладную направленность. Их решение приводит к уменьшению социального и экономического риска природных катастроф и уменьшению экономических затрат при поиске месторождений. Применение для решения рассмотренных классов задач предметно-ориентированных информационных технологий позволяет не только получить формальное решение, но и исследовать и установить связи между прогнозируемыми сущностями и свойствами геологической среды. Поэтому получаемые результаты органично связаны с развитием фундаментальных проблем сейсмотектоники, освоения недр, прогноза землетрясений, инженерной геологии и социально-экономического прогнозирования.

## **2.2. Критерии качества решения задач прогноза с неполной информацией**

Высокая неопределённость в имеющихся знаниях и данных, которые используются при решении задач пространственно-временного прогноза с неполной информацией, не позволяет оставаться в рамках чисто математических методов. В этих условиях наилучшее решение задачи может быть получено за счёт комплексного использования всей доступной информации. Эта информация включает в себя исходные данные, знания предметной области, экспертные решения и некоторое предположение о связи рассматриваемого свойства с характеристиками исследуемой среды, сформулированное в виде причинно-следственной модели. Задача состоит в том, чтобы в рамках модели найти зависимость исследуемого свойства от характеристик среды и его картографическое представление. При этом каждая версия формального решения представляет собой гипотезу, которая порождается в результате совместной обработки всех имеющихся знаний и данных.

Применение математических критериев для оценки качества версии решения требует выполнения ряда предположений, которые чаще всего

в таких задачах трудно считать обоснованными. Компромиссное решение состоит в проверке условий соответствия полученного результата с причинно-следственной моделью, предложенной экспертом в предметной области.

Будем считать, что решение не опровергает модель, если выполняются следующие три условия:

- *Совместимость*: результат не опровергается на имеющихся данных.
- *Значимость*: критерий качества решения на контрольных примерах значимо показывает на наличие причинно-следственной связи.
- *Непротиворечивость*: результат может быть объяснён и обоснован с помощью имеющихся знаний и данных.

Все три условия оценки качества версии прогноза не являются формальными, хотя для их проверки используются формальные процедуры.

Условие совместимости означает отсутствие грубых ошибок прогноза по отношению к имеющимся экспериментальным фактам. Условие значимости связано с методом нахождения прогнозирующей функции. Обычно зависимость между прогнозируемой величиной и данными ищется путём оптимизации некоторой целевой функции на выборке примеров (учебная выборка). Значимость найденной зависимости проверяется по значению целевой функции на выборке примеров, которые не использовались при обучении (контрольная или тестовая выборка). Чем ближе значение целевой функции на контрольной выборке к оптимуму, тем больше значимость найденной зависимости.

Анализ результатов на непротиворечивость по отношению к предложенной причинно-следственной модели приводит к необходимости использовать классическую методологию решения прогнозных задач экспертами. Обычно специалисты при прогнозе используют два метода построения рассуждений:

- построение логических высказываний о связи исследуемого свойства геологической среды с признаками,
- прогноз свойств среды по аналогии с прецедентом.

Применение первого метода опирается на *дескрипторные* характеристики геологической среды. К ним относятся те характеристики, для которых связь с исследуемым свойством геологической среды известна. Например, при прогнозе нефтегазовых месторождений такими характеристиками являются геометрические особенности строения перспективных структурных комплексов, которые используются в моделях оценки запасов углеводородов, а при прогнозе сейсмической опасности — физико-механические характеристики земной коры. Дескрипторные характеристики позволяют эксперту формулировать логические высказывания о связи прогнозируемого свойства с характеристиками геологической среды.

Второй метод рассуждений применяется при использовании *невербализованных* характеристик, которые связаны с прогнозируемым понятием опосредованно. Например, гравитационные аномалии силы тяжести связаны с вещественным составом и структурными неоднородностями геологической среды. Но эта информация не входит напрямую в известные модели оценки запасов углеводородов или оценки сейсмической опасности.

В некоторых случаях невербализованные характеристики удаётся преобразовать в дескрипторные признаки с помощью заранее подготовленного универсального набора операторов. Последовательные цепочки из этих операторов представляют знания предметной области и являются одним из важнейших элементов предметной ориентации технологии. Однако для подавляющего большинства невербализованных характеристик такие преобразования неизвестны. В этом случае при решении задачи такие характеристики преобразуются в косвенные признаки, которые более адекватно описывают исследуемое явление. Так, например, по векторным, точечным и сеточным данным могут генерироваться сеточные цифровые модели плотности или близости к линиям или точкам, сеточные модели, полученные в результате фильтрации растровых полей, а также в результате вычисления алгебраических и логических функций от нескольких растровых полей. Для решения задач с использованием косвенных признаков применяется метод рассуждений по аналогии с прецедентом.

### **2.3. Информационная модель как версия решения в условиях неопределённости**

Найденную версию решения будем называть *информационной моделью*. Модель даёт описание исследуемого явления или процесса средствами компьютерно-реализуемых формализмов (знаний, данных и алгоритмов) и неформальных экспертных знаний. Формальная часть модели состоит из следующих трёх элементов:

1. Цифровая модель пространственно-временного представления исследуемого явления (цифровая модель прогноза и цифровая модель оценки точности прогноза).
2. Правило нахождения цифровой модели прогноза по признакам (прогнозирующая функция).
3. Используемые для получения результатов набор исходных данных, учебная и контрольная выборки, признаки прогноза и цепочки операторов, которые преобразуют исходные данные в признаки.

В неформальную часть модели входят два понятия:

1. Причинно-следственная модель прогнозируемого явления.
2. Обоснование и объяснение прогноза.

Например, при пространственном прогнозе формальными компонентами информационной модели являются:

1. Двумерные сеточные модели прогноза и его точности.
2. Прогнозирующая функция.
3. Векторные, растровые и табличные исходные данные, выборки примеров с известными оценками прогноза, сеточные модели признаков, полученные из всех типов исходных данных, и набор операторов преобразования точечных, векторных и растровых данных в отобранные при решении признаки.

При пространственно-временном анализе геологических процессов, в частности, при поиске предвестников землетрясений компонентами информационной модели являются:

1. Трёхмерная сеточная модель динамического поля предвестников землетрясения и поля оценок значимости выявленных аномалий.
2. Алгоритмы вычисления предвестников по исходным наблюдениям и отобранные в решении алгоритмы оценивания и проверки статистических гипотез,
3. Инструментальные каталоги землетрясений, временные ряды геофизических наблюдений с координатной привязкой, векторные, растровые и табличные данные о свойствах геологической среды, не зависящих от времени, трёхмерные растровые динамические поля признаков, характеризующие процесс подготовки землетрясения, и отобранные при решении последовательности операторов преобразования всех типов данных в растровые динамические поля.

Информационная модель является средством представления знания, извлечённого в результате решения задачи. В отличие от математической модели, она не позволяет создать всеобъемлющую теорию. Однако информационная модель вместе со средствами информационного моделирования позволяет объединить всю доступную информацию об изучаемом явлении: описательное знание, экспертные гипотезы, невербальное знание эксперта, методологию обработки и анализа, разнотипных пространственно-временных данных и знание, извлечённое из них при решении задачи.

Информационные модели прогноза могут быть использованы при решении других аналогичных задач. Так, например, причинно-следственная модель прогнозируемого явления и прогнозирующая функция, найденные при прогнозе определённых свойств геологической среды, могут быть взяты как основа решения таких же задач в регионах с близкими геолого-геофизическими условиями.

При решении задачи могут изменяться все составляющие модели. Модель развивается и уточняется по мере получения дополнительных данных и появления новых гипотез о причинно-следственных связях. Предсказующая сила информационных моделей возрастает по мере того как появляются новые данные и приобретаются новые экспертные знания.

## **2.4. Элементы решения задач пространственно-временного прогноза**

Процесс построения информационной модели прогноза является итеративным. Для решения задачи прогноза пользователь, как правило, должен многократно выполнять обработку и анализ данных, проверять ряд альтернативных гипотез и оценивать результаты с точки зрения возможности построения их интерпретации в терминах исследуемой предметной области. В процессе решения могут вводиться новые наблюдения. В результате итераций версии решения модифицируются и уточняются. Процесс решения задачи содержит 6 основных этапов.

### **1. Формализация задачи:**

- Формулировка причинно-следственной модели исследуемого явления.
- Формализация модели в виде некоторой картографической схемы и/или в виде множества логических высказываний.
- Определение способа представления исследуемого объекта и/или его свойства с помощью скалярной величины и способа описания объекта набором геолого-геофизических признаков.

### **2. Формализация информационного пространства задачи:**

- Выбор, структурирование и предварительная обработка исходных данных.
- Представление пространственно-временных структур, каталогов точек, растров и временных рядов с координатной привязкой в векторном, табличном и в сеточном форматах.
- Формирование выборок прецедентов.
- Выбор методов преобразования исходных данных в признаки, более адекватно описывающие исследуемое свойство.
- Представление знаний эксперта о зависимости изучаемого свойства от признаков в виде системы ограничений на область оцениваемых параметров.

### **3. Генерирование системы признаков:**

- Создание двумерных сеточных моделей полей признаков для линейных объектов.
- Создание двух- и трёхмерных сеточных моделей признаков для точечных объектов.
- Создание двух- и трёхмерных сеточных моделей признаков с помощью линейной и нелинейной фильтрации.
- Создание двух- и трёхмерных сеточных моделей признаков с помощью алгебраических и логических преобразований над несколькими исходными растровыми полями.
- Создание дескрипторных признаков с учётом тезауруса предметной области.

**4. Порождение решения:**

- Правдоподобный вывод зависимости исследуемого свойства от сеточных моделей признаков (прогнозирующая функция).
- Отбор группы наиболее информативных растровых признаков.
- Построение карты прогноза и карты оценки точности решения.

**5. Анализ и аргументация решения:**

- Анализ ошибок на контрольной выборке прецедентов.
- Проверка непротиворечивости решения и причинно-следственной модели с помощью картографического и статистического анализа.
- Анализ и обоснование решения по аналогии с прецедентами.
- Построение логических объяснений и обоснований.

**6. Модификация версии решения:**

- Модификация причинно-следственной модели.
- Формализация дополнительных данных.
- Изменение гипотез формирования признаков.
- Изменение выборки прецедентов.

## 2.5. Компоненты среды информационного моделирования

Для исследований пространственно-временных свойств географических существностей, решения задач прогноза, объяснения и обоснования необходимо объединить всю информацию об изучаемом явлении: многодисциплинарные данные, операторы преобразования данных в дескрипторные признаки, выборки прецедентов, знание, извлечённое при решении задачи, а также невербальное знание эксперта. Для этого требуется специализированная *среда информационного моделирования* (Гитис и др., 1989а, Гитис et al., 1991, Гитис, 1995). Среда включает в себя следующие компоненты:

1. *Информационное пространство задач.*
2. *Средства контекстно-зависимой формализации, структуризации и обработки данных, нахождения решения и извлечения знаний.*
3. *Информационное содержание решения.*
4. *Средства анализа и аргументации полученных формализмов.*

Схематически *информационное пространство задачи* изображено на рис. 2.1. Оно состоит из неформализованных данных, формализованных и структурированных данных, априорных знаний и экспертных гипотез.

К исходным данным относится картографическая информация, представленная на бумажных носителях и в виде цифровых моделей, временные ряды с координатной привязкой, каталоги точек. Формализация и структуризация данных состоит в преобразовании данных во внутренние форматы среды информационного моделирования, в очист-



Рис. 2.1. Информационное пространство задачи

ке данных от шумов и их стандартизации, в выполнении преобразований данных из векторных форматов изолиний, полигонов и точек в нерегулярной сетке к их растровому представлению. Априорные знания представляются в виде набора цепочек операторов генерирования признаков и в виде системы ограничений на класс функций, в котором ищется зависимость прогноза от признаков. Экспертные гипотезы могут быть представлены в виде качественных причинно-следственных моделей, в виде картографических схем, в виде выборок прецедентов с оценками прогноза и в виде логических конструкций.

*Средства контекстно-зависимой формализации, структуризации, обработки данных и нахождения решения* схематически изображены на рис. 2.2. Они позволяют программно реализовать предпо-



Рис. 2.2. Средства контекстно-зависимой формализации, структуризации, обработки данных и нахождения решения

ложения эксперта об адекватных способах формализации исходных данных и знаний, преобразовать данные в признаки, проанализировать данные и знания, найти зависимость между изучаемым свойством и признаками прогноза, построить цифровую модель прогноза.

Ввод картографических данных производится с помощью дигитайзера и сканера. Линейные элементы карт представляются в виде последовательностей географических координат в так называемом векторном формате. Алгоритмы векторизации поддерживают оцифровку и редактирование нескольких типов картографических данных. К ним относятся линейные структурные элементы (линементы и геологи-



ческие разломы, контуры топографической основы, административные границы), полигоны, изолинии, картографические элементы с точечной привязкой, пиктограммы, надписи. Оцифровка сеточных данных состоит в преобразовании растрового изображения к географическим координатам по заданным на изображении опорным точкам.

Средства конвертирования данных позволяют выполнить ряд преобразований формата: перейти от моделей карт в изолиниях или полигонах к сеточным моделям и наоборот, вычислить сеточную модель свойства по измерениям в нерегулярных точках изучаемого географического региона, вычислить трёхмерные сеточные модели процессов по каталогам пространственно-временных событий (например, по каталогам землетрясений) и по временным рядам мониторинга среды.

Для генерирования сеточных признаков используется широкий класс операторов над точечными каталогами, линейными схемами, временными рядами, двумерными и трёхмерными растровыми данными. Эти операторы позволяют оценить пространственно-временные свойства географических объектов, такие как пространственно-временные сеточные модели взвешенной плотности эпицентров землетрясений и других параметров сейсмического режима, сеточные модели плотности геологических разломов, пространственно-временные модели временных рядов геофизического мониторинга, дву- и трёхмерные растры, полученные с помощью операций линейной и нелинейной фильтрации, а также с помощью алгебраических и логических операций над несколькими исходными растрами. В результате обработки удаётся преобразовать все типы исходных данных в более адекватные задаче признаки.

Средства конструирования логических правил и вывода по аналогии с прецедентом служат для нахождения прогнозирующей функции (зависимости изучаемого свойства от признаков). Для этого используются методы искусственного интеллекта, распознавания образов и регрессионного анализа.

Цифровые модели карт пространственного прогноза стационарных свойств геологической среды находятся путём вычисления значений прогнозирующей функции во всех точках двумерного раstra. Если правило прогноза ищется методами вывода по аналогии с прецедентом, то, помимо цифровой модели карты прогноза, может быть вычислена сеточная модель оценки его точности.

В задачах пространственно-временного прогноза результат представляет собой характеристику динамического процесса, которая отображается на трёхмерную сетку, образованную двумя пространственными и одной временной координатами. В задачах прогноза землетрясений, например, для нахождения решения используются методы проверки статистических гипотез. В этом случае значениями сеточной модели являются оценки значимостей отклонений от стационарности временных рядов в каждом элементе пространственного раstra.

В процессе решения задачи порождается *информационное содержание решения*. В него входят отобранные при решении задачи при-

знаки, прогнозирующие функции, которые определяют зависимость прогноза от признаков, поле прогноза и модель точности прогноза. Можно видеть, что этот компонент среды информационного моделирования представляет собой часть информационной модели решаемой задачи. Схематически информационное содержание решения изображено на рис. 2.3.

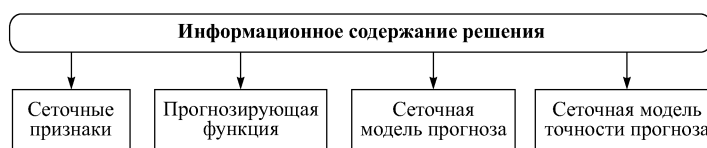


Рис. 2.3. Информационное содержание решения

*Средства анализа и аргументации* (рис. 2.4) позволяют исследовать качество решения: его совместимость, значимость и непротиворечивость. Совместимость и значимость проверяются по выборке примеров, по статистическим критериям либо по уровню ошибок на контрольной выборке примеров. Проверка условия непротиворечивости производится экспертом. Эта процедура поддерживается методами статистического анализа, картографического анализа и кластерного анализа.



Рис. 2.4. Средства анализа и аргументации

Средства статистического анализа поддерживают оценивание средних, дисперсий, гистограмм и матриц корреляции для каталогов точек и растровых данных.

Средства картографического исследования предназначены для представления и комплексного анализа данных. Они поддерживают операции по композиции карт из нескольких сеточных, векторных и точечных слоёв, чтение значений и атрибутов элементов карты, увеличение карты с интерполяцией и без интерполяции, построение по произвольному профилю разреза нескольких растровых слоёв с нанесением

распределённых по глубине точечных объектов, построение на карте произвольных областей, нанесение выборок прецедентов, анимационное представление пространственно-временных процессов, анализ по прецеденту.

Анализ по прецеденту состоит в построении карт сходства по набору признаков с несколькими анализируемыми точками. Точки пространства признаков, выбранные по сходству с анализируемыми точками, могут образовывать связные области в географических координатах. Совмещение этих областей с другими информационными слоями даёт возможность эксперту найти противоречия в полученном решении и установить причины этих противоречий.

Средства кластерного анализа позволяют выделить однородные по признакам сгустки точек, называемые кластерами или таксонами. Сравнение гистограмм растровых полей для всего региона и для зон региона, относящихся к одному кластеру, позволяет сопоставить свойства кластера по отношению к свойствам всего изучаемого региона. К этому объяснению можно добавить данные о частотах принадлежности точек выборки прецедентов к зонам региона, относящимся к одному кластеру. Далее для объяснения прогнозирующей силы кластера можно использовать отношение количества попадающих в данный кластер точек выборки одного класса к общему числу попадающих в этот же кластер точек выборки.

Взаимодействие элементов среды информационного моделирования представлено на блок-схеме рис. 2.5.

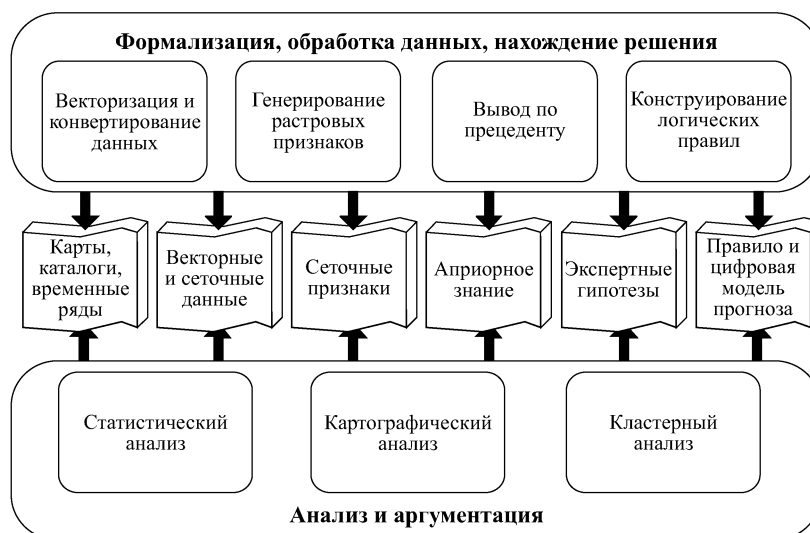


Рис. 2.5. Блок-схема взаимодействия элементов среды информационного моделирования

Решение задач с неполной информацией требует выполнения ряда итераций. Решение начинается с формализации задачи, данных, знаний и экспертных гипотез. Затем производится оцифровка, редактирование и конвертирование форматов исходных данных, генерирование системы растровых признаков и нахождение версии решения. На следующем этапе производится анализ версии решения, его объяснение и обоснование. Цель этапа состоит в том, чтобы подтвердить прогноз или представить возможные причины его неточности. При этом в соответствии с методологией решения задач с неполной информацией версия прогноза считается удовлетворительной, если результат не опровергается на имеющихся данных (*совместимость решения*), критерий качества решения на контрольных примерах значимо показывает на наличие причинно-следственной связи (*значимость решения*), результат может быть объяснён и обоснован с помощью имеющихся знаний и данных (*непротиворечивость решения*). Если эти условия не выполнены, то принимается решение о необходимости модификации экспертных гипотез, добавлении новой информации, изменении её формализации и обработки. При этом средства аргументации подсказывают эксперту возможный способ коррекции результата. После этого выполняется следующая итерация решения.

## **II. ПРОСТРАНСТВЕННЫЙ ПРОГНОЗ**

---

### **Глава 3**

### **ПРОГНОЗ СВОЙСТВ ГЕОЛОГИЧЕСКОЙ СРЕДЫ, НЕ ЗАВИСЯЩИХ ОТ ВРЕМЕНИ**

#### **3.1. Введение**

Задачи пространственного прогноза свойств геологической среды возникают при экологическом районировании территории, при анализе и оценке природной опасности и при прогнозе месторождений полезных ископаемых. В этих задачах требуется по комплексу геолого-геофизических данных оценить пространственное распределение числового или номинального свойства геологической среды, которое не зависит от времени, представить это свойство в виде прогнозной карты и найти формальное правило прогноза.

В задачах пространственного прогноза предполагается, что в пределах изучаемой области оцениваемая величина связана некоторой зависимостью с известными характеристиками геологической среды. Чаще всего эта зависимость, называемая прогнозирующей функцией, заранее неизвестна. Для её нахождения используются выборка примеров (прецедентов) и/или знания предметной области. При построении прогнозной карты прогнозирующая функция используется для вычисления цифровой модели прогнозного поля. Для этого правило прогноза применяется к известным характеристикам геологической среды во всех точках сеточного поля региона. Если признаки прогноза определены сеточными моделями с разными параметрами сетки, то значения признаков в узлах раstra прогнозного поля интерполируются. При этом в задачах пространственного прогноза обычно используется алгоритм билинейной интерполяции.

В ряде случаев прогнозная карта составляется высоко квалифицированным экспертом предметной области. Такое решение, как правило, является качественным и субъективным. При этом в некоторых зонах карты решения эксперта могут быть интуитивными и трудно интерпретируемыми. Поскольку эксперт обычно не использует формального правила прогноза, то экстраполяция его решения на другой регион с аналогичными геолого-геофизическими свойствами невозможна.

Прогнозирующая функция определяет прогнозную карту как результат выполнения последовательности операций над исходными данными. Предполагается, что исходные данные являются более объективными, чем прогноз, полученный в результате их комплексной интерпретации экспертом. Поэтому прогноз с помощью прогнозирующей функции открывает дополнительные возможности для интерпретации

полученного результата и позволяет экстраполировать правило прогноза с изученных регионов на менее изученные регионы с близким геологическим строением. В некоторых случаях экспертное решение задачи прогноза представляет собой только предварительный результат исследования. Экспертная прогнозная карта может быть использована для нахождения прогнозирующей функции и построения затем новой прогнозной карты (Gitis et al., 1998). Полученные результаты помогают эксперту более аргументировано объяснить своё решение и позволяют откорректировать окончательную версию прогнозной карты.

Прогнозирующие функции, найденные для прогноза одного и того же свойства геологической среды в различных регионах представляют самостоятельный теоретический интерес для развития представлений о связях между исследуемыми свойствами геологической среды. Такой сравнительный анализ прогнозирующих функций, найденных для различных регионов мира при решении задач прогноза максимальных магнитуд ожидаемых землетрясений, выполнен в работе Schenk et al., 1994.

В данной главе рассмотренная в гл. 2 методология развивается на проблемную область пространственного прогноза. В разделе 3.2 детально рассматриваются основные этапы построения информационных моделей прогноза пространственных свойств геологической среды, обосновывается выбор типа прогнозирующей функции в виде суммы одномерных нелинейных зависимостей, обсуждаются процедуры объяснения и обоснования результатов решения задачи. В разделе 3.3 кратко представлены основные функции аналитической ГИС ГЕО 2.5.

### 3.2. Технология построения информационной модели пространственного прогноза

Построение информационной модели пространственного прогноза выполняется итеративно. Итерации состоят из элементов, перечисленных в разделе 2.4. Технология построения модели состоит из 6 основных этапов, которые схематично изображены на рис. 3.1.

**1-й этап: формализация задачи.** На этом этапе решаются две задачи: построение и формализация *причинно-следственной модели прогнозируемого явления* для исследуемого региона и формализация понятий *объекта прогноза и признаков*, на основании которых производится прогноз.

Причинно-следственная модель *описывает* связь между прогнозируемым явлением или прогнозируемой величиной и состоянием, свойствами и историей развития геологической среды как в пределах региона, так и в его окружении. Так, например, в задачах прогноза максимальных магнитуд ожидаемых землетрясений ( $M_{\max}$ ) описывается связь сейсмического потенциала со свойствами геологической среды, с ориентацией внешних по отношению к исследуемому региону текто-

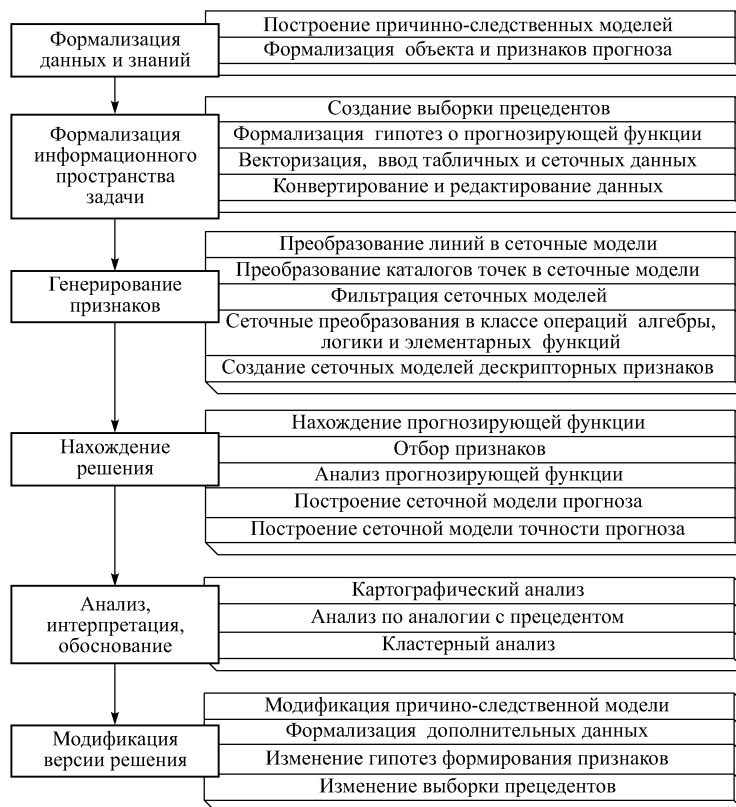


Рис. 3.1. Этапы построения информационной модели пространственного прогноза

нических сил, с расположением зон с различным типом тектонических деформаций и т. д. В задачах прогноза нефтегазовых месторождений описываются связи областей повышенной концентрации углеводородов с теми свойствами геологической среды, которые влияют на условия генерации углеводородов, их накопления, миграции и сохранения.

Причинно-следственная модель может *формализоваться* в виде некоторой схемы районирования изучаемой территории и в виде некоторого набора логических правил. Так, например, при прогнозе  $M_{\max}$  могут быть сформулированы логические высказывания, которые связывают условия накопления и разрядки тектонической энергии со свойствами геологической среды. В задачах нефтегазового прогноза логические правила могут описывать связи между степенью уверенности в наличии месторождения и расположением очагов генерации углеводородов, вероятными направлениями миграции флюидов, структурными свойствами среды и свойствами коллекторов и покрышек.

Причинно-следственная модель *используется* для выбора исходных геолого-геофизических данных, для выбора методов преобразования исходных данных в признаки, более адекватно описывающие прогнозируемую величину, для нахождения прогнозирующей функции и, наконец, для объяснения и обоснования результатов.

При *формализации объекта и признаков прогноза* требуется определить способ представления прогнозируемого объекта или его свойства с помощью некоторой величины, называемой прогнозируемой величиной, и способ описания объекта набором геолого-геофизических признаков.

Нетривиальность формализации объекта и признаков прогноза связана с тем, что обычно прогнозируемые геологические объекты могут представлять собой достаточно протяжённые зоны с заранее неизвестной формой. Математические методы обнаружения или распознавания таких зон связаны с выполнением больших переборов по всем возможным формам, и поэтому их трудно реализовать. Компромиссное решение состоит в замене объекта прогноза на некоторое подмножество точек раstra. При этом каждая точка раstra должна принимать значение прогнозируемого объекта или прогнозируемого свойства объекта, а признаки каждой точки раstra должны нести информацию, достаточную для определения этого значения. Например, при построении прогнозных карт максимальных магнитуд ожидаемых землетрясений значение  $M_{\text{max}}$  в точке раstra рассматривается как максимальная магнитуда в очаге ожидаемого землетрясения, проекция которого на поверхность земли покрывает данную точку. Геолого-геофизические признаки, относящиеся к точке раstra, должны описывать как локальные особенности геологической среды с точки зрения возможности накопления потенциальной энергии деформации в окрестности данной точки, так и, тесно связанные с размерами очагов ожидаемых землетрясений размеры, тенденции развития и динамические свойства структурных неоднородностей, к которым принадлежит указанная точка. В задачах нефтегазового прогноза для каждой точки раstra анализируется степень её принадлежности к месторождению рассматриваемого нефтегазоносного комплекса, а геолого-геофизические и геохимические признаки точки раstra формируются так, чтобы описывать как локальные особенности нефтепроизводящих, коллекторных и нефтеизолирующих свойств геологической среды в окрестности данной точки, так и геометрию структурных комплексов, к которым принадлежит эта точка.

**2-й этап: формализация информационного пространства задачи.** Важнейшую роль в задачах прогнозирования играет прогнозирующая функция. Обычно правило прогноза ищется по некоторой выборке прецедентов, в которой представлены значения признаков и оценки значений прогнозируемой величины. Можно выделить три способа *формирования выборки прецедентов*.

Способ прямого оценивания использует в качестве выборки множество точек, для которых значения прогнозирующей величины оценены



в результате инструментальных измерений, а при невозможности прямых измерений — в результате экспертного оценивания. Так, например, при оценивании  $M_{\max}$  выбирается относительно небольшое число точек региона, для которых эксперты независимо друг от друга дают оценки  $M_{\max}$  (Гитис и др., 1982). Для задач нефтегазового прогноза выборка обычно также составляется экспертами с учётом результатов бурения на нефть и газ (Гитис et al., 1994).

Способ зонального оценивания основан на использовании предварительно составленной карты прогнозируемого свойства. Эта карта может быть получена в результате прямых измерений прогнозируемой величины, выполненных для части исследуемого региона, либо может представлять собой экспертное решение задачи районирования. В этом случае координаты точек выборки порождаются по закону случайных чисел, а значения прогнозируемой величины считываются с соответствующей карты. Например, значения аномалий силы тяжести, измеренные для части региона, могут быть использованы для формирования выборки прецедентов с целью построения карты гравитационных аномалий региона по косвенным геолого-геофизическим признакам. В задаче прогноза  $M_{\max}$  такие карты районирования могут быть составлены экспертами на основе знаний о сейсмической истории региона, представлений о тектонических особенностях и их связи с  $M_{\max}$  землетрясений (Гитис et al., 1998). При нефтегазовом прогнозе могут использоваться карты районирования, относящиеся только к тем частям региона, где выполнены поисковые работы, околтурены месторождения и выделены бесперспективные зоны.

И, наконец, третий способ получения выборки предложен специально для решения задач прогноза  $M_{\max}$  землетрясений. Этот способ применяется при отсутствии высоко квалифицированных экспертов. Для формирования выборки используется каталог землетрясений. Идея алгоритма состоит в том, что выборка генерируется автоматически по каталогу землетрясений с учётом точности оценки магнитуды сильных землетрясений и координат их эпицентров, а также при использовании экспертной оценки фонового значения  $M_{\max}$  (Гитис, 1995).

При решении задачи в ряде случаев выборка прецедентов может отсутствовать, но эксперты могут сформулировать гипотезы о правиле прогноза. Иногда этих знаний достаточно для логического вывода прогнозирующей функции (Осинов, 1997). В частности, в задаче поиска и оценки значимости предвестников землетрясений правило прогноза определяется на основе математических моделей прогнозируемого явления, разработанных экспертами предметной области (Keilis-Borok and Soloviev, ed. 2003, Смирнов, 1995). В некоторых задачах комплексного экологического районирования по инженерно-геологическим данным правило прогноза удаётся сформировать и скорректировать только на основе экспертных знаний о связи инженерно-геологических условий с прогнозируемым свойством среды (Лихачёва и др., 1993).

Обычно выборка прецедентов относительно невелика. В связи с этим при оценивании прогнозирующей функции используются параметрические методы. Применение этих методов требует *формализовать экспертные гипотезы о правиле прогноза*: определить класс функций, в котором ищется аппроксимация зависимости прогнозируемой величины от признаков и ввести ограничения на область допустимых параметров. Из общих соображений зависимость прогноза от признаков предполагается нелинейной. Кроме того, в задачах оценивания значений числовой характеристики геологической среды эта зависимость предполагается и непрерывной. Действительно, если при изменении свойства не происходит качественного изменения состояния среды, то вероятнее всего малому изменению значений признаков соответствует малое изменение прогноза.

Важным элементом оценивания прогнозирующей функции является использование экспертных гипотез о характере зависимости, которые определяются как прогнозируемым свойством, так и типом изучаемого региона. Эти знания часто формулируются в виде продукций, логических конъюнкций и в виде зависимости прогнозируемой величины от каждого признака в отдельности при условии неизменности остальных. Например, при прогнозе  $M_{\max}$  предполагается, что уменьшение расстояния от точки прогноза до зоны активного тектонического разлома или увеличение в точке прогноза скорости современных вертикальных движений коры при прочих равных условиях связаны с увеличением  $M_{\max}$ . В нефтегазовом прогнозе можно предположить, что увеличение мощности коллекторного слоя, уменьшение электропроводности и приуроченность точки растра к значимой антиклинали при прочих равных условиях увеличивает вероятность наличия продуктивной залежи. Экспертные гипотезы о характере зависимости прогноза от признаков дают возможность искать решение в классе физически интерпретируемых функций. Эти знания формализуются в виде системы математических ограничений на класс прогнозирующих функций.

В предлагаемой информационной технологии прогнозирующая функция ищется в виде

$$F(\mathbf{x}, \vartheta) = \sum_{i=1}^I \varphi_i(x_i, \vartheta),$$

где  $\varphi_i(x_i, \vartheta)$  — нелинейная функция признака  $x_i$ , а  $\vartheta$  — вектор оцениваемых параметров. Чаще всего в практических приложениях достаточно, чтобы для прогнозирующей функции выполнялось условие непрерывности. Условие гладкости является менее существенным. Поэтому в практических приложениях функции  $\varphi_i(x_i, \vartheta)$  выбираются кусочно-линейными. При этом функции  $\varphi_i(x_i, \vartheta)$  задаются в виде разложения по системе кусочно-линейных базисных функций с фиксированными узловыми точками по осям  $x_i$ . Вектор оцениваемых параметров  $\vartheta$  определяет значения функций в точках излома. Такой вид параметризации прогнозирующей функции позволяет учесть нелинейный характер ап-

проксимируемой зависимости от признаков прогноза при малом числе оцениваемых параметров (Гитис, 1974, Гитис 1977, Ларичев, 1989, Юрков, 1974). Он удобен для формализации экспертных знаний, так как в задачах с неполной информацией экспертам часто известен вид зависимости прогнозируемой величины от каждого признака в отдельности. Эти знания просто ввести в виде ограничений на параметры кусочно-линейных функций. Сумма нелинейных функций одной переменной позволяет интерпретировать сеточную модель прогнозного поля как сумму нелинейно преобразованных полей сеточных признаков.

Исходные данные геологических задач пространственного прогноза обычно имеют вид карт, схем линейных структур, каталогов точек, и растровых данных. На этом этапе выбранные линейные данные вводятся в систему с помощью дигитайзера, сканера или с магнитного носителя в виде линий, границ зон, изолиний, маркированных точек, пиктограмм и полигонов. Линейные и точечные данные представляются в геоинформационных системах в так называемом векторном формате. Поэтому процесс их оцифровки обычно называют *векторизацией*. Растровые данные, табличные каталоги и точечные значения в нерегулярной сетке вводятся со сканера или магнитных носителей. При необходимости данные редактируются и конвертируются из векторного и точечного форматов в сеточный формат. Например, в сеточный формат с помощью специальных методов пространственной интерполяции преобразуются векторные данные, представляющие изолинии или полигоны, а также точечные данные, снятые в нерегулярной сетке.

**3-й этап: генерирование системы признаков.** На этом этапе производится преобразование данных в цифровые сеточные поля (признаки), адекватные причинно-следственной модели прогнозируемого явления. Например, создаются двумерные сеточные поля линейных объектов, которые характеризуют пространственное распределение плотности линий, суммарную длину всех линий в текущем окне произвольного радиуса, меру близости к линиям и т. д. Создаются двумерные растровые поля для точечных объектов, такие как, например, поле сейсмической активности, поле наклона графика повторяемости землетрясений, поля фрактальных размерностей пространственного распределения эпицентров землетрясений, поля плотности, взвешенной плотности и расстояний до точечных объектов. Для сеточных полей могут выполняться операции двумерной фильтрации, могут вычисляться элементарные функции от значений нескольких сеточных полей, вычисляться произвольные алгебраические и логические функции от нескольких исходных полей.

В некоторых приложениях для нахождения сеточных признаков используется моделирование. Так, например, для обнаружения нефтяных ловушек для ГИС ГЕО 2.5 разработана программа, которая моделирует заполнение кровли структурного горизонта углеводородами из зон их возможной генерации. Этот подход моделирования позволяет обнаруживать все типы ловушек, позволяет учесть географическое рас-

положение и оценку мощности нефтегенерирующей толщи и процесс горизонтальной миграции углеводородов. Моделирование производится итеративно. На каждой итерации оценивается заполнение кровли исследуемой структуры одним слоем углеводородов заданной толщины из зоны нефтегенерации. Выходом программы на каждой итерации является цифровая сеточная модель распределения толщи накопленных углеводородов и цифровая сеточная модель распределения объемов углеводородов в каждой отдельной ловушке.

**4-й этап: оценивание прогнозирующей функции и построение прогноза.** Этот этап является основным при решении задачи пространственного прогноза. Целью этапа является нахождение версии решения задачи и проверка версии на совместимость и значимость (раздел 2.2).

Для нахождения прогнозирующей функции в ГИС ГЕО 2.5 используются методы аппроксимации функций многих переменных, применяемые в условиях существенной априорной неопределённости: нахождение нелинейной зависимости между прогнозируемым свойством и описанием объектов по данным, содержащим экспертные оценки (*Гитис и др., 1982, Гитис, 1987*), и нахождение зависимости между логарифмом отношения правдоподобия двух гипотез и описанием объектов (*Юрков, 1983*). Эти методы позволяют использовать для нахождения прогнозирующей функции нечёткие знания нескольких независимых экспертов о значении прогнозируемой величины в точках выборки. Одновременно с нахождением прогнозирующей функции производится *отбор группы наиболее информативных признаков* и поддерживается *детальный анализ найденного прогнозного правила*. Для построения цифровой модели прогнозного поля значения прогнозирующей функции вычисляются для всех узлов растровой сетки региона. Могут быть построены два типа прогнозного поля: поле скалярного свойства физического явления и поле апостериорной вероятности некоторого события, связанного с физическим явлением (или поле логарифма отношения правдоподобия гипотез).

Недостаточное обоснование статистических моделей в практических задачах прогноза делает практически невозможным построение доверительных интервалов оценок прогнозируемых величин растрового поля. В этих условиях для оценки точности цифровой растровой модели прогноза можно исходить из того, что прогнозирующая функция находится из условий наилучшей аппроксимации точек учебной выборки. Можно предположить, что точность аппроксимации уменьшается пропорционально увеличению расстояния в пространстве признаков от точек выборки. Расстояния между всеми точками выборки и узлами сетки прогнозного поля можно измерять, например, либо в метрике Махаланобиса, либо, в частном случае, в евклидовой метрике, нормированной по среднеквадратичным отклонениям признаков прогноза. Результаты вычислений представляют собой растровое поле расстояний. Функцию близости, принимающую значение 1 на нулевых расстояниях до точек выборки и линейно убывающую до нуля при увеличении

расстояния до некоторой пороговой величины, можно интерпретировать как степень уверенности в прогнозе. Нормировка пространства признаков позволяет выбирать пороговую величину в долях среднеквадратичных отклонений.

**5-й этап: анализ, интерпретация и обоснование решения.** Цель этапа состоит в проверке решения на непротиворечивость (раздел 2.2). Для этого необходимо на основании имеющихся данных и знаний либо подтвердить прогноз, либо представить возможные причины его неточности, принять решение о необходимости изменения информационной модели и найти возможный способ её коррекции. Интерпретация и обоснование в задачах геологического пространственно-временного прогноза опирается на методологию рассуждений эксперта, основу которой составляют

- картографическая форма представления данных,
- использование принципа прецедента при анализе и обосновании прогноза,
- построение объяснения прогноза на языке логических конструкций.

Картография обеспечивает эксперта наиболее адекватным представлением входных и выходных данных. Например, процессор картографии ГИС ГЕО 2.5 работает в многооконном режиме, чтобы предоставить доступ к нескольким картам. Наряду с набором стандартной графики, процессор поддерживает основные картографические проекции, картографическое измерение значений нескольких сеточных полей и текстовых атрибутов векторных данных, совмещение нескольких информационных слоёв, состоящих из каталогов точечных объектов, условных обозначений, схем линейных структур, сеточных полей в форме изолиний и в цветовой заливке, обеспечивает представление схем разрезов по произвольному профилю для нескольких сеточных и точечных полей. Процессор картографии также поддерживает комплексный анализ по сходству с прецедентом и обладает элементами анимации для визуализации пространственно-временных процессов.

Использование при обосновании результатов блока анализа данных по принципу прецедента реализует метод экспертного рассуждения по аналогии, который зачастую используется специалистами при решении задач геолого-геофизического прогноза. При анализе по аналогии у эксперта имеются две опции: (1) сравнить по некоторому комплексу признаков анализируемую точку со всеми точками изучаемого региона и отобразить в географических координатах сеточное поле значений функции сходства, либо (2) построить вокруг анализируемой точки произвольную прямоугольную область в пространстве признаков и отобразить сеточное поле, выделив на нём группы точек, находящихся внутри и вне построенной области. В результате картографически выделяются зоны, похожие по комплексу признаков на анализируемую точку. В эти зоны могут попадать прецеденты — точки с известными оценками прогноза. В некоторых случаях сходство анализируемой

точки с прецедентом по признакам может войти в противоречие с представлениями эксперта. Этот конфликт подсказывает эксперту, какие модификации данных следует произвести, чтобы указанные точки разделялись по признакам. Иногда для разделения требуется ввести дополнительные исходные данные или по старым данным сгенерировать новые признаки.

Объяснение прогноза в терминах предметной области является одной из важных компонент информационной модели. Сложность построения объяснения состоит в следующем. В геологии и геофизике прогнозирующая функция обычно аппроксимирует некоторую физическую закономерность. В связи с этим аппроксимация ищется в классе непрерывных функций многих переменных. Содержательное объяснение такого решающего правила в терминах использованных в нём признаков может быть основано на построении текстового описания зависимости прогнозируемой величины от каждого из признаков в отдельности при фиксированных значениях остальных признаков. Однако легко видеть, что такое объяснение не является достаточно полным из-за наличия зависимости между признаками. Тем не менее, объяснение этого типа, реализованное с помощью использования текстовых шаблонов и лингвистических переменных, было опробовано в системе ГЕО 1.2 [43, 145], которая являлась одной из предыдущих версий системы ГЕО 2.5. Результаты экспериментальных исследований показали, что данный тип построения объяснения мало интересен для экспертов предметной области. В связи с этим в версии ГЕО 2.5 от данного типа объяснений пришлось отказаться.

С другой стороны, хорошо известно, что знания предметных областей, связанные с геологическим прогнозом, часто могут быть сформулированы на языке логических высказываний. Обычно эти высказывания носят качественный характер. Поэтому они не обладают достаточно высокой степенью определённости для построения удовлетворительной прогнозирующей функции. Однако этот язык удобен для объяснения прогноза: он имитирует способ рассуждений эксперта и дополняет объяснения формального решающего правила.

Идея построения объяснения прогноза в виде логических конструкций состоит в аппроксимации в пространстве признаков области с некоторым интервальным значением прогноза набором многомерных прямоугольников с рёбрами, параллельными координатным осям (*Турбович и др., 1971*). Каждый такой прямоугольник может быть представлен в виде конъюнкции от соответствующих интервалов признаков, а множество прямоугольников определяется с помощью дизъюнкции конъюнкций. Для построения такого покрытия можно воспользоваться процедурой иерархического кластерного анализа (*Gitis et al., 1992*).

Рассмотрим процедуру иерархического кластерного анализа, реализованную в ГИС ГЕО 2.5. Алгоритм кластеризации состоит в следующем:

1. Выбирается подмножество признаков. При построении объяснения оно, как правило, совпадает с признаками, используемыми в прогнозирующей функции. Признаки нормируются по среднеквадратичным отклонениям.

2. Выбирается метрика для вычисления межкластерных расстояний. Выбирается способ определения топологической связности в географических координатах: (соседними точками кластера, например, могут считаться точки, примыкающие к нему по горизонталям и вертикалям, либо примыкающие к нему по горизонталям, вертикалям и диагоналям). Кластер *A* считается связным с кластером *B*, если соседняя точка кластера *A* принадлежит кластеру *B*.

3. Вычисляются межкластерные расстояния между связными кластерами (вначале кластерами считаются все точки растра). Пара ближайших кластеров объединяется в один кластер.

4. Выход из процедуры производится, если минимальное межкластерное расстояние превосходит некоторое пороговое значение. При этом фиксируются все кластерные разбиения, которые предшествовали значимым скачкам минимального межкластерного расстояния.

В процессе выполнения кластеризации после достижения кластерами достаточно больших размеров условие топологической связности обычно снимается. Поэтому в географических координатах кластерам могут соответствовать неодносвязные области.

Для каждого кластера можно построить формальное описание, используя его представления в географических координатах и в пространстве признаков, на которых проводилась кластеризация. Расположение кластера в географических координатах используется для оценки прогнозных свойств кластеров. В качестве такой оценки может использоваться, например, отношение количества попадающих в зону кластера (в географических координатах) точек выборки прецедентов одного класса к общему количеству прецедентов в зоне кластера. Чтобы получить формальное описание свойств кластера, множество его точек в пространстве признаков покрывается минимальным прямоугольником с рёбрами, параллельными координатным осям. Легко видеть, что каждый такой прямоугольник может быть представлен в виде конъюнкции соответствующих интервалов признаков. В результате может быть построено объяснение, связывающее прогнозные свойства кластера и описание его свойств с помощью признаков. Полученное логическое выражение легко вербализовать с помощью тезауруса предметной области и удобно использовать для объяснения свойств кластеров.

**6-й этап: модификация версии решения.** На этом этапе принимается решение о выполнении следующей итерации или об окончании процесса решения задачи.

Версия прогноза считается удовлетворительной, если выполняются условия совместности, значимости и непротиворечивости, сформулированные в разделе 2.2. В противном случае средства системы аргумен-

тации могут помочь эксперту найти способ модификации версии, после чего выполняется следующая итерация. При этом возможны несколько вариантов: модификация причинно-следственной модели, ввод и формализация дополнительных данных, изменение гипотез формирования признаков и изменение выборки прецедентов.

Результатом решения задачи могут быть несколько альтернативных информационных моделей. Версии моделей могут отличаться по набору признаков, по значению средней ошибки аппроксимации оценок прогноза на контрольной выборке и по разбросу ошибок аппроксимации, обусловленному неточностью измерений признаков. Например, типичным является случай, когда при построении прогнозной карты  $M_{\text{max}}$  строится версия с использованием сейсмологических признаков и версия без использования сейсмологических признаков. Это связано с тем, что сейсмологические признаки в этой задаче являются наиболее информативными, но вместе с тем их надёжность зависит от длительности инструментальных наблюдений. Одна из карт отбирается экспертом в качестве рабочей версии. По мере обновления входной информации о геологической среде все версии могут использоваться для разработки более совершенных информационных моделей. При решении аналогичных задач прогноза в регионах со сходными геологическими условиями элементы версий информационных моделей могут быть использованы в качестве предварительного варианта решения.

### **3.3. Аналитическая ГИС ГЕО 2.5**

Первая версия аналитической ГИС ГЕО была разработана в 1987 г. (*Гитис и др., 1989а*) как средство помощи специалисту в решении широкого класса задач геолого-геофизического прогноза, таких как построение прогнозных полей параметров сейсмического режима, прогноз полезных ископаемых и их запасов, распознавание геологических объектов, геоэкологический прогноз. Версия ГИС ГЕО 2.5. содержит шесть подсистем: подсистему подготовки данных, подсистему генерирования сеточных признаков, подсистему индуктивного вывода закономерностей и построения прогнозных полей, подсистему анализа, объяснения и обоснования, подсистему управления и контекстной помощи пользователю.

Подсистема подготовки данных поддерживает векторизацию и редактирование вводимых с дигитайзера или со сканера векторных данных (линий, границ зон, изолиний, надписей, точек, пиктограмм и закрашенных областей), ввод и редактирование сеточных цифровых моделей, вычисление сеточных числовых моделей по изолиниям и по значениям, снятым в нерегулярной сетке, вычисление зональных сеточных полей с номинальными значениями по полигонам, составление подкаталогов событий, ввод знаний о графических объектах и полях, построение и анализ карт полей, преобразование форматов данных из формата ГЕО в текстовый формат и наоборот.



Подсистема генерирования сеточных признаков позволяет преобразовать исходные данные в признаки, адекватные модели прогнозируемого явления. Подсистема позволяет создавать сеточные модели, выявляющие пространственные свойства точечных и линейных объектов, такие как, например, поле сейсмической активности, поле наклона графика повторяемости землетрясений, поля плотности и взвешенной по атрибутивным значениям плотности точек, поля расстояний до точечных или линейных объектов, поля суммарной длины линий в скользящем окне произвольного радиуса, производить нелинейную фильтрацию растровых полей, вычислять произвольные функции нескольких исходных сеточных полей, конструируемых из элементарных функций с использованием алгебраических и логических операций.

В подсистеме индуктивного вывода закономерностей и построения прогнозных карт реализованы методы аппроксимации функций многих переменных, применяемые в условиях существенной априорной неопределённости: нахождение нелинейной зависимости между прогнозируемым свойством и описанием объектов либо зависимости между логарифмом отношения правдоподобия двух гипотез и описанием объектов. Подсистема также отбирает группы наиболее информативных признаков и поддерживает детальный анализ найденного решения. Для построения прогнозных полей подсистема вычисляет значения прогнозирующей функции для всех точек регулярной координатной сетки региона и строит прогнозные поля двух типов: поля скалярного свойства физического явления и поля апостериорной вероятности некоторого события, связанного с физическим явлением (или поле логарифма отношения правдоподобия гипотез).

Подсистема анализа, объяснения и обоснования включает в себя процессор картографии, модуль оценивания одномерных статистических характеристик сеточных моделей и их корреляционных матриц и блок кластерного анализа на сеточных данных. Наиболее существенную роль в анализе результата играет картографическое исследование. Процессор картографии работает в многооконном режиме. Наряду с набором стандартной графики он поддерживает отображение в различных картографических проекциях любых сочетаний каталогов точечных объектов, условных обозначений, схем линейных структур, полей в форме изолиний и цветовых изображений, представление схем разрезов по одному или нескольким полям, выделение зон, однородных по заданному набору свойств, визуализацию пространственно-временных процессов. Блок кластерного анализа позволяет разбить любую зону региона на однородные области и построить формальное описание каждой области по комплексу геолого-геофизических свойств.

Подсистемы управления и контекстной помощи поддерживают диалог пользователя с системой при помощи позиционного меню, дают возможность получить справочную информацию об управлении системой и рекомендации по методике решения задач проблемной области.

## ВЫВОД ЭМПИРИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ В УСЛОВИЯХ НЕОПРЕДЕЛЁННОСТИ

### 4.1. Введение

Карта пространственного прогноза строится на основании прогнозирующей функции. Прогнозирующая функция аппроксимирует зависимость прогнозируемой величины от признаков. Обычно она находится путём оптимизации некоторого критерия качества, который определяется математической моделью прогноза. Качество прогноза в практических приложениях в большой степени зависит от полноты описания прогнозируемой величины с помощью признаков, от представительности экспериментальной выборки примеров прогноза, от уровня шумов в данных и от того, насколько целевая функция оптимизации и класс, в котором ищется прогнозирующая функция, согласованы с решаемой задачей.

Для поиска прогнозирующей функции по выборке примеров используются методы многомерной статистики (*Андерсон Т., 1963.*), распознавания образов (*Вапник и Червоненкис, 1974, Загоруйко, 1999, Турбович и др., 1971, Fukunaga, 1972, Zhuravlev, 1998*) и регрессионного анализа (*Дрейпер, Смит, 1973.*). В задачах распознавания часто класс, к которому принадлежит точка выборки, оценивается экспертом. В задачах восстановления регрессионных зависимостей обычно значения прогнозируемой величины на выборке примеров измерены инструментально. В то же время в ряде задач восстановления непрерывных функциональных зависимостей по эмпирическим данным значения прогнозируемой величины в точках выборки не могут быть инструментально измерены и оцениваются экспертами на основании их знаний и опыта. В этом случае возникают вопросы о способах формализации экспертного знания, о моделях экспертного оценивания и методах аппроксимации функциональной зависимости по данным, содержащим экспертные оценки (*Осипов, 1997, Bradley and Terry, 1952, Hayes-Roth F., Waterman D.A., and Lenat, 1983*).

Часто решение эксперта о значении функции представляется в виде числа и для нахождения функциональной зависимости используется метод наименьших квадратов. Во многих случаях адекватность такого подхода вызывает сомнения, поскольку с одной стороны точечная оценка не всегда позволяет эксперту выразить свои знания об оцениваемой величине, а, с другой стороны, оценивание параметров

функциональной зависимости по методу наименьших квадратов не всегда соответствует модели рассуждений эксперта. По существу вид экспертной оценки определяет язык, на котором эксперт выражает своё знание о значении прогнозируемой величины. Очевидно, что такой язык должен быть простым и удобным для однозначного понимания его всеми участниками экспертного опроса. Это уменьшает шум экспертного оценивания. С другой стороны, вид оценки должен дать возможность эксперту достаточно полно формализовать своё знание о значении прогноза.

В данной главе рассматриваются три метода вывода эмпирических зависимостей в условиях неопределённости: в разделе 4.2 — метод интервальных экспертных оценок (Гитис и др., 1982), в разделе 4.3 — метод балльных экспертных оценок (Гитис, 1987), и в разделе 4.4 — метод аппроксимации отношения правдоподобия (Юрков, 1983). Первые два метода разработаны для нахождения восстановления непрерывных функциональных зависимостей. Метод аппроксимации отношения правдоподобия решает задачу распознавания образов. С его помощью находится прогнозирующая функция для разделения объектов на два класса.

## 4.2. Метод интервальных экспертных оценок

**4.2.1. Интервальные экспертные оценки.** Методика оценивания значений функции экспертами с помощью интервалов была предложена для формализации знаний эксперта при решении задач построения прогнозных карт максимальных магнитуд ( $M_{\max}$ ) ожидаемых землетрясений (Гитис и др., 1982).

Значения  $M_{\max}$  нельзя измерить инструментально. Для оценивания  $M_{\max}$  эксперты выбирают наиболее изученные точки региона с разной сейсмичностью и геологическими условиями. Давая оценки  $M_{\max}$ , они пользуются историческими данными о сейсмичности, результатами геолого-геофизических исследований изучаемого региона и своими представлениями о связи сейсмичности с состоянием и свойствами геологической среды. Некоторые допущения экспертов можно сформулировать следующим образом (Ризниченко, 1985, Садовский и Писаренко, 1991, Бунэ и Горшков, ред., 1980):

- Значение  $M_{\max}$  в изучаемом регионе находится в интервале от фонового значения магнитуды, которое достижимо в любом месте региона, до некоторого максимального значения, которое может быть оценено статистическими методами.
- Отсутствие исторических данных о больших разрушительных землетрясениях в течение 1000–2000 лет в центрах, где документировались исторические события, свидетельствует о пониженных значениях  $M_{\max}$ .
- Оценкой снизу значения  $M_{\max}$  ожидаемого землетрясения в некоторой точке является значение  $m_{\max}$  максимальной магнитуды

землетрясения, которое ранее было зарегистрировано в окрестности данной точки. Степень уверенности в выполнении неравенства  $M_{\max} \geq m_{\max}$  зависит от точности оценок магнитуды  $m_{\max}$  и координат эпицентра.

- Оценкой сверху значения  $M_{\max}$  в некоторой точке является значение  $m_{\max}$  максимальной магнитуды землетрясения, зарегистрированного ранее в пределах выделенной экспертом локальной однородной геологической зоны, которой принадлежит данная точка. Степень уверенности в выполнении неравенства  $M_{\max} \leq m_{\max}$  зависит от точности оценок магнитуды  $m_{\max}$ , от близости величины  $m_{\max}$  к оценке максимального землетрясения региона, от степени уверенности эксперта в представительности сейсмостатистики выделенной им зоны и от уверенности в сеймотектонической однородности этой зоны.

Исходя из этих соображений, были предложены интервальные экспертные оценки вида:

$$Q_{qn} = \left( m_{qn}^{(1)}, m_{qn}^{(2)}, w_{qn}^{(1)}, w_{qn}^{(2)} \right), \quad (4.1)$$

где  $m_{qn}^{(1)}, m_{qn}^{(2)}$  — границы интервала, внутри которого все значения прогнозируемой величины в точке  $n$  являются, по мнению  $q$ -го эксперта, наиболее вероятными и равновероятными,  $m_{qn}^{(1)} \leq m_{qn}^{(2)}$ ;  $w_{qn}^{(1)} > 0$ ,  $w_{qn}^{(2)} > 0$  — весовые коэффициенты, с помощью которых  $q$ -й эксперт указывает степень своей уверенности в возможности того, что значение прогнозируемой величины в точке  $n$  может быть меньше или больше соответствующей границы интервала  $m_{qn}^{(1)}$  или  $m_{qn}^{(2)}$ .

**4.2.2. Выборка прецедентов.** Рассмотрим некоторые способы получения выборки оценок прогнозируемой величины:

1. *Метод независимого экспертного оценивания.*
2. *Метод получения выборки по карте.*
3. *Метод получения выборки по каталогу наблюдений.*

В первом случае вначале экспертами составляется выборка прецедентов, а затем для каждой точки выборки эксперты независимо друг от друга назначают интервальные оценки прогнозируемой величины. При этом экспертам гарантируется анонимность результатов опроса.

Во втором случае для получения выборки используется карта прогнозируемой величины, составленная либо для всего изучаемого региона либо для его фрагментов. Эта карта может представлять собой экспертное решение или содержать результаты измерений прогнозируемой величины. Для получения выборки генерируется последовательность случайных точек, а оценки прогнозируемой величины считаются с исходной карты (*Gitis et al., 1998*).

На третьем методе мы остановимся более подробно и рассмотрим его на примере получения выборки для построения прогнозной карты  $M_{\max}$  ожидаемых землетрясений по каталогу землетрясений (*Gitis et al., 1992*).

Очевидно, что зарегистрированные в каталоге землетрясения являются оценками  $M_{\max}$  снизу. В каталоге содержатся ошибки в определении магнитуды и координат землетрясений. При автоматическом генерировании выборки наблюдений ошибка в определении магнитуды землетрясения влияет на оценку нижней границы  $M_{\max}$ . Ошибки в определении координат эпицентров нелинейно влияют на значения вектора признаков. Идея метода состоит в генерировании выборки по каталогу событий с моделированием шумов, которые содержатся в инструментальных данных.

Будем считать, что каждому событию каталога с номером  $n$ , с магнитудой  $M_n$  и координатами  $(\lambda_n, \varphi_n)$  соответствует подвыборка инструментальных наблюдений

$$\{m_k(M_n), \boldsymbol{\Psi}_k(\lambda_n, \varphi_n)\}, \quad k = 1, \dots, K, \quad (4.2)$$

где случайная величина  $m_k$  и двумерная случайная величина  $\boldsymbol{\Psi}_k$  имеют нормальные распределения

$$m_k \sim N(M_n, \sigma_{1n}), \quad (4.3)$$

$$\boldsymbol{\Psi}_k \sim N\left((\lambda_n, \varphi_n), \begin{vmatrix} \sigma_{2n} & 0 \\ 0 & \sigma_{3n} \end{vmatrix}\right), \quad \sigma_{2n} = \sigma_{3n}. \quad (4.4)$$

Для генерирования выборки применяется следующий алгоритм.

1. Назначение для изучаемого региона значения  $M_0$ , равного фоновому значению  $M_{\max}$ .
2. Выбор из каталога всех событий с магнитудой  $m \geq M_0$ .
3. Генерирование для каждого выбранного  $n$ -го события по  $K$  наблюдений  $(Q_{nk}, \mathbf{x}_{nk})$  в соответствии с распределениями (4.3) и (4.4). Интервальная оценка  $Q_{nk}$  определяется выражением

$$m_{nk}^{(1)} = M_n, \quad m_{nk}^{(2)} = M_n + \sigma_{1n}, \quad w_{nk}^{(1)} = \frac{1}{\sigma_{1n}}, \quad w_{nk}^{(2)} = \frac{w_{nk}^{(1)}}{c}, \quad (4.5)$$

где  $c$  — коэффициент, который показывает во сколько раз штраф за завышение  $F(\mathbf{x}_{nk}, \boldsymbol{\theta}) > m_{nk}^{(2)}$  меньше, чем штраф за занижение  $F(\mathbf{x}_{nk}, \boldsymbol{\theta}) < m_{nk}^{(1)}$ .

Значения вектора признаков  $\mathbf{x}_{nk}$  снимаются в точках с координатами

$$(\lambda_{nk}, \varphi_{nk}) = (\alpha_n + \zeta_{nk}, \beta_n + \xi_{nk}), \quad (4.6)$$

где  $\zeta_{nk}, \xi_{nk}$  — независимые случайные величины с нормальным распределением  $\sim N(0, \sigma_{2n})$ .

4. Исключение из региона окрестности точек,  $(\alpha_n, \beta_n)$ , выбранных из каталога в пункте 2. Внутри оставшейся области генерирование случайной выборки наблюдений с фоновыми значениями  $M_{\max}$

$$m_t^{(1)} = M_0 - \Delta, \quad m_t^{(2)} = M_0, \quad w_t^{(1)} = w_0^{(1)}, \quad w_t^{(2)} = w_0^{(2)}, \quad (4.7)$$

где  $\Delta$  — ширина интервала оценок,  $w_t^{(1)}$  и  $w_t^{(2)}$  — выбираются с учётом средних значений аналогичных весовых коэффициентов для выборки  $m > M_0$ .

**4.2.3. Алгоритм аппроксимации.** Можно считать, что экспертная оценка  $Q$  соответствует некоторой функции субъективной плотности вероятности  $f(Y, Q)$ , которая отображает знания эксперта о возможном значении  $Y$  в данной точке выборки. Из условий экспертного оценивания очевидно, что эта функция должна принимать постоянное значение внутри интервала  $[m^{(1)}, m^{(2)}]$  и убывать с весами  $w^{(1)}$  и  $w^{(2)}$  соответственно влево и вправо от границ интервалов.

Этим условиям удовлетворяет, в частности, следующая функция

$$f(Y, Q) = C \exp \left\{ - \left( w^{(1)} \frac{|m^{(1)} - Y| + m^{(1)} - Y}{2} + w^{(2)} \frac{|m^{(2)} - Y| - m^{(2)} + Y}{2} \right)^p \right\}, \quad (4.8)$$

где:  $m^{(1)} \leq m^{(2)}$ ,  $w^{(1)} \geq 0$ ,  $w^{(2)} \geq 0$ ,  $p \geq 1$ , а константа  $C$  определяется условием

$$\int_{-\infty}^{\infty} f(y, Q) dy = 1.$$

Пусть имеется выборка наблюдений  $\{Q_{qn}, \mathbf{x}_n\}$ , где  $q$  — номер эксперта,  $n$  — номер точки выборки. Требуется аппроксимировать функцию  $Y(\mathbf{x})$  в некотором классе функций  $\{F(\mathbf{x}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ , где  $\Theta$  — область допустимых значений вектора параметров  $\boldsymbol{\theta}$ .

Заменим в (4.8)  $Y$  значением прогнозирующей функции  $F(\mathbf{x}, \boldsymbol{\theta})$  и рассмотрим функцию

$$r(\mathbf{x}, \boldsymbol{\theta}) = -\ln f(F, Q) + \ln C = \left( w^{(1)} \frac{|m^{(1)} - F| + m^{(1)} - F}{2} + w^{(2)} \frac{|m^{(2)} - F| - m^{(2)} + F}{2} \right)^p. \quad (4.9)$$

Функция  $r(\mathbf{x}, \boldsymbol{\theta})$  определяет величину штрафа за неточность аппроксимации экспертной оценки  $Q$  значением  $F$  прогнозирующей функции. Если  $F \in [m^{(1)}, m^{(2)}]$ , то прогноз не противоречит экспертной оценке и штраф  $r = 0$ . Когда значение прогноза  $F$  выходит из границ интервала, то штраф возрастает как  $[w^{(1)}(m^{(1)} - F)]^p$  при  $F < m^{(1)}$  и как  $[w^{(2)}(F - m^{(2)})]^p$  при  $F > m^{(2)}$ . Для нахождения оценки  $\boldsymbol{\theta}$  минимизируется средний штраф на выборке.

Оценка вектора параметров  $\hat{\boldsymbol{\theta}}$  имеет вид

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_n \sum_q r(F(\mathbf{x}_n, \boldsymbol{\theta}), Q_{qn}). \quad (4.10)$$

Легко проверить, что если прогнозирующая функция  $F(\mathbf{x}, \boldsymbol{\theta})$  является линейной по параметрам, то минимизируемый функционал (4.10) является выпуклым. Если при этом область  $\Theta$  допустимых значений вектора также является выпуклой, то при оценивании можно воспользоваться итерационными алгоритмами градиентного типа.

Из (4.9) и (4.10) нетрудно видеть, что в частном случае, если используются экспертные оценки, в которых всегда  $m^{(1)} = m^{(2)}$ , и  $w^{(1)} = w^{(2)}$ , то при  $p = 1$  алгоритм оценивания (4.10) совпадает с методом наименьших модулей, а при  $p = 2$  — с методом наименьших квадратов. В следующем разделе показано, что при некоторых предположениях оценка (4.10) является оценкой максимального правдоподобия.

**4.2.4. Статистическая модель.** Рассмотрим вероятностную модель назначения экспертных оценок и покажем, что при некоторых предположениях оценка (4.10) является оценкой максимального правдоподобия.

Пусть рассматриваемая выборка наблюдений  $\{Q_{qn}, \mathbf{x}_n\}$  получена случайным выбором в соответствии с некоторой плотностью распределения  $f(\mathbf{x}, Q; \boldsymbol{\theta})$ .

Обозначим

$$m = \frac{m^{(1)} + m^{(2)}}{2}, \quad d = \frac{m^{(2)} - m^{(1)}}{2}. \quad (4.11)$$

В новых обозначениях величину  $m$  можно рассматривать как решение эксперта о значении  $Y$  в данной точке выборки, а величины  $d$ ,  $w^{(1)}$  и  $w^{(2)}$  — как оценку экспертом точности своего решения.

Модель состоит из следующих основных предположений.

- Зависимость  $Y(x)$  описывается функцией  $F(\mathbf{x}, \boldsymbol{\theta})$  при некотором заранее неизвестном значении  $\boldsymbol{\theta}_0$  вектора параметров  $\boldsymbol{\theta} \in \Theta$ , т. е.

$$Y(\mathbf{x}) = F(\mathbf{x}, \boldsymbol{\theta}). \quad (4.12)$$

- Маргинальное распределение случайных величин  $d$ ,  $w^{(1)}$  и  $w^{(2)}$  не зависит от вектора параметров  $\boldsymbol{\theta}$ , т. е.

$$\begin{aligned} f(\mathbf{x}, Q; \boldsymbol{\theta}) &= f(m, d, w^{(1)}, w^{(2)}, \mathbf{x}; \boldsymbol{\theta}) = \\ &= f_2(d, w^{(1)}, w^{(2)}, \mathbf{x}) \cdot f_1(m|d, w^{(1)}, w^{(2)}; \boldsymbol{\theta}). \end{aligned} \quad (4.13)$$

- Плотность условного распределения  $f_1$  имеет вид:

$$\begin{aligned} f_1(m|d, w^{(1)}, w^{(2)}; \boldsymbol{\theta}) &= \\ &= C_1 \cdot \exp \left\{ - \left( w^{(1)} \frac{|m - F(\mathbf{x}, \boldsymbol{\theta}) - d| + m - F(\mathbf{x}, \boldsymbol{\theta}) - d}{2} + \right. \right. \\ &\quad \left. \left. + w^{(2)} \frac{|m - F(\mathbf{x}, \boldsymbol{\theta}) + d| - m + F(\mathbf{x}, \boldsymbol{\theta}) - d}{2} \right)^p \right\}. \end{aligned} \quad (4.14)$$

Нетрудно видеть, что если  $d = 0$  и  $w^{(1)} = w^{(2)}$ , то (4.14) при  $p = 1$  является плотностью распределения Лапласа, а при  $p = 2$  — плотностью нормального распределения.

- Наблюдения, составляющие выборку  $\{Q_{qn}, \mathbf{x}_n\}$ , статистически независимы в совокупности.

Таким образом, вектор  $\boldsymbol{\theta}$  является неизвестным параметром плотности распределения  $f(\mathbf{x}, Q; \boldsymbol{\theta})$ . Задача состоит в нахождении оценки  $\hat{\boldsymbol{\theta}}$  по выборке  $\{Q_{qn}, \mathbf{x}_n\}$ .

Взяв в качестве оценки  $\hat{\boldsymbol{\theta}}$  оценку максимального правдоподобия, получим

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \prod_n \prod_q f(m_{qn}, d_{qn}, w_{qn}^{(1)}, w_{qn}^{(2)}, \mathbf{x}_n; \boldsymbol{\theta}). \quad (4.15)$$

Переходя к логарифму функции правдоподобия и учитывая (4.14), получаем

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_n \sum_q \left( -\ln f_1(m_{qn} | d_{qn}, w_{qn}^{(1)}, w_{qn}^{(2)}, \mathbf{x}_n; \boldsymbol{\theta}) \right) + c, \quad (4.16)$$

где

$$c = \sum_n \sum_q \left( -\ln f_2(d_{qn}, w_{qn}^{(1)}, w_{qn}^{(2)}, \mathbf{x}_n) \right).$$

С учётом того, что  $c$  не зависит от  $\boldsymbol{\theta}$ , а  $x$  и  $\vartheta$  связаны зависимостью  $F(x, \vartheta)$  это эквивалентно

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_n \sum_q \left( -\ln f_1(m_{qn} | d_{qn}, w_{qn}^{(1)}, w_{qn}^{(2)}, F(\mathbf{x}_n, \boldsymbol{\theta})) \right). \quad (4.17)$$

Подставляя в (4.17)  $m^{(1)}$  и  $m^{(2)}$  вместо  $m$  и  $d$ , согласно (4.11), учитывая (4.14) и отбрасывая константу  $\ln C_1$ , снова получаем критерий (4.10)

$$\begin{aligned} \hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} & \left\{ -\ln C_1 + \right. \\ & + \sum_n \sum_q \left( w_{qn}^{(1)} \frac{|m_{qn} - F(\mathbf{x}_n, \boldsymbol{\theta}) - d_{qn}| + m_{qn} - F(\mathbf{x}_n, \boldsymbol{\theta}) - d_{qn}}{2} + \right. \\ & \left. \left. + w_{qn}^{(2)} \frac{|m_{qn} - F(\mathbf{x}_n, \boldsymbol{\theta}) + d_{qn}| - m_{qn} + F(\mathbf{x}_n, \boldsymbol{\theta}) - d_{qn}}{2} \right)^p \right\} = \\ & = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_n \sum_q \left( w_{qn}^{(1)} \frac{|m_{qn}^{(1)} - F(\mathbf{x}_n, \boldsymbol{\theta})| + m_{qn}^{(1)} - F(\mathbf{x}_n, \boldsymbol{\theta})}{2} + \right. \\ & \left. + w_{qn}^{(2)} \frac{|m_{qn}^{(2)} - F(\mathbf{x}_n, \boldsymbol{\theta})| - m_{qn}^{(2)} + F(\mathbf{x}_n, \boldsymbol{\theta})}{2} \right)^p. \quad (4.18) \end{aligned}$$

**4.2.5. Решение с несколькими экспертами.** Алгоритм аппроксимации интервальных экспертных оценок (4.10) позволяет использовать при оценивании функции экспертные оценки, полученные путём



независимого опроса большого числа экспертов. Предполагается, что эксперты назначают свои интервальные оценки для точек выборки независимо друг от друга. При этом в экспертных оценках неизбежно встречаются противоречия.

Для получения более робастной оценки вектора  $\theta$  рекомендуется исключать выбросы в экспертных оценках. В связи с этим разработанный для ГИС ГЕО 2.5 алгоритм аппроксимации для каждой точки выборки исключает две экспертные оценки: оценку с минимальным значением  $m^{(2)}$  и оценку с максимальным значением  $m^{(1)}$ .

Однако исключение двух крайних экспертных оценок не позволяет контролировать ситуацию, когда в экспертном опросе участвует много специалистов с различной степенью квалификации. В этом случае оценки не вполне компетентных экспертов могут существенно влиять на оценки наиболее компетентных экспертов.

В основу отбора группы наиболее компетентных экспертов можно положить предположение о том, что решения наиболее компетентных экспертов позволяют точнее определить истинное пространственное распределение прогнозируемой величины, чем решения менее компетентных экспертов. Поэтому прогнозные карты, построенные по оценкам наиболее компетентных экспертов, близки к истинной карте и близки между собой. Прогнозные карты некомпетентных экспертов далеки от истинной карты и при независимом экспертном опросе скорее всего далеки друг от друга.

Процедура выбора экспертов состоит в следующем. По оценкам каждого эксперта строится растровая прогнозная карта. По взаимным расстояниям между растрами выделяется кластер наиболее близких в определённом смысле карт. Соответствующие эксперты выбираются для оценки прогнозирующей функции как наиболее компетентные.

### 4.3. Метод балльных экспертных оценок

**4.3.1. Постановка задачи.** Балльные экспертные оценки традиционно используются в задачах, в которых прогнозируемые величины измеряются в порядковых шкалах. Например, в сейсмологии такие оценки используются в качестве меры интенсивности сотрясения (балл землетрясения) (*Бунэ и Горшков ред., 1980*), в природопользовании в качестве меры выраженности патогенного поражения биосистем, в медицине как мера степени проявления заболевания или мера тяжести состояния больного (*Гитис и др. 1974, Турбович и др., 1977*).

Во многих работах при нахождении функциональной зависимости по данным, в которых значения функции в точках выборки представлены балльными экспертными оценками, используется критерий наименьших квадратов. Однако целевая функция метода наименьших квадратов никак не учитывает методики выбора балла экспертом. Метод балльных экспертных оценок развивает подход, который основывается на некоторой вероятностной модели назначения баллов экспертом.

Модель предполагает, что оцениваемая величина принимает значения из непрерывной шкалы, разбитой на интервалы. Эксперт для каждого наблюдения указывает балл: номер того интервала, в который, по его мнению, попадает оцениваемая величина. При этом вероятность назначения экспертом некоторого балла монотонно уменьшается по мере удаления истинного значения оцениваемой величины от центра соответствующего интервала.

В настоящем разделе метод аппроксимации функциональной зависимости по балльным экспертным оценкам формулируется как задача параметрического оценивания. Для метода предложена статистическая модель, разработан алгоритм, исследованы асимптотические свойства оценок и на примере решения практической задачи прогноза из области медицины изложена технология применения.

**4.3.2. Статистическая модель.** Пусть скалярное свойство  $Y$  физического явления связано с вектором свойств  $\xi = (\xi_1, \xi_2, \dots, \xi_J)$  функцией  $Y = F(\xi, \theta)$  из параметрического семейства  $\{F(\xi, \theta) : \theta \in \Theta\}$ , где  $\Theta$  — подмножество  $R^m$  допустимых значений вектора параметров  $\theta$ . Свойство  $Y$  оценивается экспертом в балльной шкале: каждому наблюдению  $\xi(n)$  эксперт ставит в соответствие оценку  $\eta(n)$  величины  $Y(n)$  в баллах. Задача состоит в том, чтобы по выборке  $n = 1, 2, \dots, N$  независимых наблюдений  $\{\xi(n), \eta(n)\}$  оценить вектор параметров  $\theta$ .

Пусть  $(\xi, \eta)$  — случайный вектор со значениями в измеримом пространстве  $\{\nu, \mathfrak{S}, P_{\theta}, \theta \in \Theta\}$ , где  $\nu$  есть прямое произведение  $\{\mathbf{x} = (x_1, x_2, \dots, x_J) : \mathbf{x}^T \mathbf{x} < C\} \times \{k = 1, 2, \dots, K\}$ ;  $\mathfrak{S} = \mathfrak{R}^J \times \mathfrak{R}_0$ ;  $\mathfrak{R}^J$  —  $\sigma$ -алгебра борелевских множеств в  $R^J$ ;  $\mathfrak{R}_0$  — совокупность всех подмножеств множества целых чисел  $1, 2, \dots, K$ ;  $P_{\theta}$  — параметризованная вероятностная мера, абсолютно непрерывная относительно меры  $\mu$  на  $\mathfrak{S}$  ( $\mu$  — мера Лебега по  $\mathbf{x}$  и считающая мера по  $k$ , относительно которой существует плотность распределения  $f(\mathbf{x}, k; \theta) = dP_{\theta}/d\mu(\mathbf{x}, k)$ );  $\theta$  — вектор параметров, принимающий значения из  $\Theta = \{\theta : \theta^T \theta < C\}$ .

Пусть  $\{\xi(n), \eta(n)\}$ ,  $n = 1, 2, \dots, N$  — независимые наблюдения вектора  $\{\xi, \eta\}$  и пусть выполняются следующие предположения.

1. Маргинальное распределение случайного вектора  $\xi$  не зависит от вектора параметров  $\theta$ , т. е.

$$f(\mathbf{x}, k; \theta) = f_1(\mathbf{x})f_2(k | \mathbf{x}; \theta). \quad (4.19)$$

2. Матрица

$$M\{\xi\xi^T\} = \int_{\nu} \mathbf{x}\mathbf{x}^T f(\mathbf{x}, k; \theta) d\mu(\mathbf{x}, k) \quad (4.20)$$

— невырожденная.

3. Зависимость  $Y$  от  $\xi$  является линейной по параметрам, т. е.

$$Y = F(\xi, \theta) = \theta^T A(\xi), \quad (4.21)$$

где  $A$  — заранее заданная непрерывная функция из  $R^J$  в  $R^m$ . Без ограничения общности далее везде будем считать, что  $A(\xi) = \xi$ , т. е.

$$Y = F(\xi, \theta) = \theta^T \xi. \quad (4.22)$$

4. Условное распределение  $f_2(k | \mathbf{x}; \vartheta)$  определяется следующей вероятностной моделью назначения экспертных оценок.

У эксперта есть представление о непрерывной шкале, в которой измеряется свойство  $Y$ . Шкала разбита на  $K > 2$  интервалов  $\{(I_k^{(1)}, I_k^{(2)})\}$ , где  $I_1^{(1)} = -\infty$ ,  $I_{k+1}^{(1)} = I_k^{(2)} > I_k^{(1)}$  для  $k = 1, 2, \dots, K-1$ ,  $I_K^{(2)} = \infty$ . Эксперт воспринимает свойство  $Y$ , «размытое» аддитивным шумом  $\zeta$ , распределение которого не зависит от номера наблюдения и имеет нулевое среднее. Следуя концепции *Бредли-Терри* (Bradley and Terry, 1952), будем считать, что шум  $\vartheta \in \Theta$  имеет плотность

$$g(t) = \frac{\alpha}{2\text{ch}^2 \alpha t}, \quad \alpha > 0, \quad (4.23)$$

и вероятность принятия экспертом решения о том, что физическое явление с вектором свойств  $\xi = \mathbf{x}$  имеет балл  $f^{1/2}(\mathbf{x}, k; \vartheta)$ , равна вероятности попадания величины  $F(\xi, \theta) + \zeta$  в соответствующий интервал  $(I_k^{(1)}, I_k^{(2)})$ , т. е.

$$\begin{aligned} P(\eta = k | \xi = \mathbf{x}; \theta) &= \int_{I_k^{(1)}}^{I_k^{(2)}} g(t - F(\mathbf{x}, \theta)) dt = \\ &= \frac{1}{2} \left[ \text{th} \alpha \left( I_k^{(2)} - F(\mathbf{x}, \theta) \right) - \text{th} \alpha \left( I_k^{(1)} - F(\mathbf{x}, \theta) \right) \right] = f_2(k | \mathbf{x}; \theta). \end{aligned} \quad (4.24)$$

**4.3.3. Алгоритм оценивания.** С учётом (4.19)–(4.24) функция правдоподобия имеет вид

$$\begin{aligned} L(\xi(n), \eta(n); \theta) &= \prod_{n=1}^N f(\xi(n), \eta(n); \theta) = \\ &= \frac{1}{2} \prod_{n=1}^N f_1(\xi(n)) \left[ \text{th} \alpha \left( I_{\eta(n)}^{(2)} - F(\xi(n), \theta) \right) - \text{th} \alpha \left( I_{\eta(n)}^{(1)} - F(\xi(n), \theta) \right) \right]. \end{aligned}$$

В следующем разделе показано, что для данной модели оценка максимального правдоподобия

$$\hat{\theta} = \arg \sup_{\theta \in \Theta} \left\{ \prod_{n=1}^N f(\xi(n), \eta(n); \theta) \right\}. \quad (4.25)$$

равномерно по  $\theta$  асимптотически нормальна со средним нуль и ковариационной матрицей  $(1/N) \mathbf{I}^{-1}(\theta)$ , где  $\mathbf{I}(\theta)$  — матрица Фишера, которая с учётом (4.24) имеет вид

$$\begin{aligned} \mathbf{I}(\theta) &= \int_{\nu} \frac{1}{f(\mathbf{x}, k; \theta)} \frac{\partial f(\mathbf{x}, k; \theta)}{\partial \theta} \left( \frac{\partial f(\mathbf{x}, k; \theta)}{\partial \theta} \right)^{\top} d\mu(\mathbf{x}, k) = \\ &= \alpha^2 M\{\xi\xi^{\top} \Phi^2(\eta | \xi; \vartheta)\} = \alpha^2 M\{\xi\xi^{\top}\} M\{\Phi^2(\eta | \xi; \theta)\}, \end{aligned} \quad (4.26)$$

$$\Phi(k | \mathbf{x}; \theta) = \text{th } \alpha(I_k^{(1)} - \theta^{\top} \mathbf{x}) + \text{th } \alpha(I_k^{(2)} - \theta^{\top} \mathbf{x}). \quad (4.27)$$

Легко показать, что при невырожденной выборочной оценке матрицы Фишера функционал (4.25) является выпуклым. Если при этом область  $\Theta$  допустимых значений вектора  $\theta$  также является выпуклой, то при оценивании можно воспользоваться итерационными алгоритмами градиентного типа.

**4.3.4. Статистические свойства оценки.** Покажем, что при предположениях (4.19)–(4.24) оценка (4.25) равномерно по  $\theta \in \Theta$  асимптотически нормальна со средним нуль и ковариационной матрицей  $(1/N) \mathbf{I}^{-1}(\theta)$ , где  $\mathbf{I}(\theta)$  — матрица Фишера (4.26), (4.27). Для этого надо доказать выполнение условий (Ибрагимов и Хасьминский, 1978) для  $f(\mathbf{x}, k; \theta)$ .

1. Для почти всех  $(\mathbf{x}, k) \in \nu$  относительно меры  $\mu$  функция  $P(\mathbf{x}, k; \theta)$  непрерывная функция на  $\Theta$ .

Из (4.19) и (4.24) очевидно, что это условие выполняется для всех  $(\mathbf{x}, k) \in \nu$ .

2. Функция  $f^{1/2}(\mathbf{x}, k; \theta)$  дифференцируема в каждой точке  $\theta \in \Theta$  в  $L_2(\mu)$  (в среднем квадратичном).

Действительно, из (4.19), (4.24) следует, что для любого  $\theta \in \Theta$  существует векторная функция

$$\psi(\mathbf{x}, k; \theta) = \frac{\partial f^{1/2}(\mathbf{x}, k; \theta)}{\partial \theta} = \alpha \mathbf{x} \Phi(k | \mathbf{x}; \theta) f^{1/2}(\mathbf{x}, k; \theta). \quad (4.28)$$

Принимая во внимание, что

$$\sup_{\{\mathbf{x}: \mathbf{x}^{\top} \mathbf{x} < C\}} \mathbf{x}^{\top} \mathbf{x} < C \quad \text{и} \quad \sup_{\{\mathbf{x}: \mathbf{x}^{\top} \mathbf{x} < C\}} \sup_{\theta \in \Theta} \Phi^2(k | \mathbf{x}; \theta) < 4,$$

получаем, что функция  $\psi(\mathbf{x}, k; \theta)$  ограничена по норме  $L_2(\mu)$ :

$$\begin{aligned} \int_{\nu} |\psi(\mathbf{x}, k; \theta)|^2 d\mu(\mathbf{x}, k) &= \alpha^2 \int_{\nu} \mathbf{x} \mathbf{x}^{\top} \Phi^2(k | \mathbf{x}; \theta) f(\mathbf{x}, k; \theta) d\mu(\mathbf{x}, k) = \\ &= \alpha^2 M\{\xi^{\top} \xi \Phi^2(\eta | \xi; \theta)\} = \alpha^2 M\{\xi^{\top} \xi\} M\{\Phi^2(\eta | \xi; \theta)\} < \alpha^2 \cdot 4C. \end{aligned} \quad (4.29)$$

3. Функция  $\psi(\mathbf{x}, k; \theta)$  непрерывна в пространстве  $L_2(\mu)$ .

Это свойство вытекает из теоремы Лебега (Колмогоров и Фомин, 1972), так как функция  $\psi(\mathbf{x}, k; \boldsymbol{\theta})$ , согласно (4.19), (4.24), (4.27) и (4.28), непрерывна по  $\boldsymbol{\theta}$  при всех  $\mathbf{x}$  и  $k$  и, согласно (4.29), ограничена по норме пространства  $L_2(\mu)$ .

4. Для любого  $\vartheta \in \Theta$  матрица Фишера (4.26) ограничена по норме и её минимальное собственное значение положительно.

Легко видеть, что

$$\begin{aligned} & \inf_{\boldsymbol{\theta} \in \Theta} \inf_{\{\mathbf{x}: \mathbf{x}^T \mathbf{x} < C\}} M\{\Phi^2(\eta | \boldsymbol{\xi}; \boldsymbol{\theta})\} > \\ & > \inf_{\boldsymbol{\theta} \in \Theta} \inf_{\{\mathbf{x}: \mathbf{x}^T \mathbf{x} < C\}} \left\{ \left[ -1 + \text{th} \alpha(I_1^{(2)} - \boldsymbol{\theta}^T \mathbf{x}) \right]^2 f_2(1 | \mathbf{x}; \boldsymbol{\theta}) + \right. \\ & + \left. \left[ 1 + \text{th} \alpha(I_K^{(1)} - \boldsymbol{\theta}^T \mathbf{x}) \right]^2 f_2(K | \mathbf{x}; \boldsymbol{\theta}) \right\} \geq \left[ 1 + \text{th} \alpha(I_K^{(1)} - I_1^{(2)}) \right]^2 \varepsilon = c > 0, \end{aligned} \quad (4.30)$$

где

$$\varepsilon = \min \{f_2(1 | \mathbf{x}; \boldsymbol{\theta}), f_2(K | \mathbf{x}; \boldsymbol{\theta})\}.$$

Отсюда с учётом (4.13) получаем

$$\begin{aligned} 0 < \alpha^2 c |M\{\boldsymbol{\xi} \boldsymbol{\xi}^T\}| < \inf_{\boldsymbol{\theta} \in \Theta} \inf_{\|\lambda\|=1} \lambda^T \mathbf{I}(\boldsymbol{\theta}) \lambda \leq \sup_{\boldsymbol{\theta} \in \Theta} |\mathbf{I}(\boldsymbol{\theta})| < \\ < \alpha^2 4 |M\{\boldsymbol{\xi} \boldsymbol{\xi}^T\}| < \alpha^2 4C. \end{aligned}$$

5. Для любого компакта  $\Theta_0 \subset \Theta$  и любого  $\delta > 0$

$$z(h) = \inf_{\boldsymbol{\theta} \in \Theta} \inf_{h \in H} \int_{\nu} \left| f^{1/2}(\mathbf{x}, k; \boldsymbol{\theta} + h) - f^{1/2}(\mathbf{x}, k; \boldsymbol{\theta}) \right|^2 d\mu(\mathbf{x}, k) > 0,$$

где

$$H = \{\mathbf{h}: \boldsymbol{\theta} + \mathbf{h} \in \Theta_0, \|\mathbf{h}\| > \delta\}.$$

Действительно, применяя теорему о конечном приращении для  $\mathbf{u} = \boldsymbol{\theta} + a\mathbf{h}$ ,  $u_i = \vartheta_i + a_i h_i$ ,  $0 < a_i < 1$  и учитывая неравенство (4.30), получим

$$\begin{aligned} z(\mathbf{h}) &= \inf_{\mathbf{u} \in \Theta_0} \inf_{h \in H} \int_{\nu} \left[ \frac{\partial}{\partial \mathbf{u}} f^{1/2}(\mathbf{x}, k; \boldsymbol{\theta})^T \mathbf{h} \right] d\mu(\mathbf{x}, k) = \\ &= \inf_{\mathbf{u} \in \Theta_0} \inf_{h \in H} \int_{\nu} (\mathbf{x}^T \mathbf{h})^2 \Phi^2(k | \mathbf{x}; \mathbf{u}) f(\mathbf{x}, k; \mathbf{u}) d\mu(\mathbf{x}, k) = \\ &= \inf_{\mathbf{u} \in \Theta_0} \inf_{h \in H} M\{(\boldsymbol{\xi}^T \mathbf{h})^2 \Phi^2(\eta | \boldsymbol{\xi}; \mathbf{u})\} > \inf_{h \in H} cM\{(\boldsymbol{\xi}^T \mathbf{h})^2\} > 0. \end{aligned}$$

**4.3.5. Технология применения метода.** В предыдущем разделе свойства оценки вектора  $\boldsymbol{\theta}$  исследованы в предположении, что остальные параметры модели (3.19)–(4.24) — границы интервалов и коэффициент  $\alpha$  — заданы заранее. В ряде практических задач аппроксимации функции по балльным экспертным оценкам это предположение не вы-

полняется. В этих случаях вместо точных значений границ интервалов и коэффициента  $\alpha$  могут быть использованы их оценки, полученные на основе дополнительной информации. При этом, как правило, абсолютные значения границ и масштаб шкалы, в которую отображается исследуемое свойство  $Y$ , могут быть выбраны достаточно произвольно. Существенными же являются относительные расстояния: отношения длин интервалов, соответствующих баллам экспертных оценок, отношения расстояний между парами объектов и т. д.

В таких задачах при выборе значений для границ интервалов требуются только сведения об относительной длине интервалов  $(I_k^{(1)}, I_k^{(2)})$ ,  $k = 2, 3, \dots, K - 1$ . В частности, для очень типичных задач такого рода с трёхбалльными экспертными оценками вопрос о выборе интервалов вообще снимается. Здесь границами интервалов являются  $-\infty, I_2^{(1)}, I_2^{(2)}, \infty$  где  $I_k^{(1)} < I_k^{(2)}$  — любые действительные числа.

При заданных границах интервалов можно получить оценку коэффициента  $\alpha$ , определяющего для модели (4.19)–(4.24) дисперсию  $D = \pi/(12\alpha^2)$  шумовой компоненты. Для этого привлекается дополнительная информация об уровне ошибок назначения баллов при проведении повторной экспертизы.

Методику применения метода балльных экспертных оценок (4.25) для решения задач, в которых заранее неизвестны границы интервалов и коэффициент  $\alpha$ , удобнее изложить на примере решения практической задачи. В качестве такого примера рассмотрим задачу аппроксимации зависимости между результатами почечной биопсии и значениям почечных функций при заболевании хроническим гломерулонефритом.

Биопсия почки при хроническом гломерулонефрите проводится с целью определения формы заболевания, с которой связаны сроки развития хронической почечной недостаточности и выбор тактики лечения. Биопсия почки является обременительной для пациента формой хирургического вмешательства и в ряде случаев её проведение нежелательно или даже невозможно. Поэтому оценка результатов биопсии по функциональным и клиническим показателям больного представляет значительный интерес.

Одним из морфологических проявлений тяжести заболевания гломерулонефритом является тубулоинтерстициальный компонент (ТИК), с помощью которого оценивается степень повреждений канальцев и стромы почек (Ратнер и др., 1984). Для решения задачи требуется аппроксимировать зависимость между ТИК и значениями почечных функций больного.

Для получения учебной и контрольной выборок примеров значения ТИК оценивались экспертом-патоморфологом по результатам однократного гистологического исследования в следующих терминах: 1 — отсутствие повреждений канальцев и стромы почек; 2 — начальная стадия повреждений; 3 — умеренные повреждения; 4 — выраженные повреждения. Шум экспертного оценивания обусловлен случайностью

локализации гистологической пробы в почке больного и качественными критериями, которыми пользуется эксперт для определения ТИК.

По объяснению эксперта степень выраженности ТИК устанавливается им в соответствии с долей площади, занятой поражёнными элементами среза биопата почки. При этом увеличению градаций ТИК соответствует линейное увеличение площади поражения. Из этих соображений было принято, что градациям 2 и 3 на шкале ТИК соответствуют равные интервалы. Абсолютные значения границ интервалов на шкале ТИК были приняты следующими:  $-\infty, 1, 2, 3, \infty$ .

Задавшись границами интервалов, можно перейти к оцениванию параметра  $\alpha$ . Для этого были привлечены данные о повторных экспертизах, проведённых для части гистологических проб одним и тем же экспертом. Было установлено, что в 25% случаев результаты экспертизы не совпадают. Отсюда можно принять, что вероятность назначения экспертом ошибочного балла

$$P(e) = \sum_{k=1}^K \pi_k \left[ 1 - \int_{I_k^{(1)}}^{I_k^{(2)}} P(y|k) f_2(k|y) dy \right] = 0,25,$$

где  $\pi_k$  — вероятность того, что  $I_k^{(1)} \leq Y \leq I_k^{(2)}$ ;  $P(y|k)$  — условное распределение  $Y$  в  $k$ -м интервале;  $f_2(k|y)$  — вероятность (2.24) принятия экспертом решения о том, что  $I_k^{(1)} \leq Y \leq I_k^{(2)}$ . Положим также, что

$$\pi_k \equiv 0,25, \quad I_3^{(1)} - I_2^{(1)} = I_4^{(1)} - I_3^{(1)} = \Delta = 1,$$

$$P(y|1) = 1/\Delta \quad \text{при условии } I_1^{(2)} - \Delta \leq y < I_1^{(2)},$$

$$P(y|1) = 0 \quad \text{при невыполнении условия,}$$

$$P(y|1) = 1/\Delta \quad \text{при условии } I_k^{(1)} \leq y < I_k^{(1)} + \Delta,$$

$$P(y|1) = 1/\Delta \quad \text{при невыполнении условия,}$$

$$k = 2, 3, 4.$$

Тогда

$$P(e) = 0,75 \left( 1 - \frac{1}{\alpha} \ln \operatorname{ch} \alpha \right).$$

Подставляя в полученное выражение  $P(e) = 0,25$ , получим  $\alpha \cong 2$ .

Зависимость между ТИК и значениями почечных функций искалась в виде суммы кусочно-линейных функций одной переменной, т. е. (4.21) было выбрано в виде

$$F(\mathbf{x}, \vartheta) = \sum_{i=1}^J \varphi_i(x_i, \vartheta) + \vartheta_0, \quad \text{где } \varphi_i(x_i, \vartheta) = \sum_{t=T_{i-1}+1}^{T_i} \vartheta_t H_t(x_i),$$

$$H_t(x_i) = \begin{cases} 1 & \text{при } x_i \geq g_i^{(t+1)}, \\ \frac{x_i - g_i^{(t)}}{g_i^{(t+1)} - g_i^{(t)}} & \text{при } g_i^{(t+1)} > x_i \geq g_i^{(t)}, \\ 0 & \text{при } x_i < g_i^{(t)}, \end{cases}$$

$\vartheta_0$  — постоянная составляющая;  $H_t(x_i)$  —  $t$ -я базисная функция для признака  $x_i$ ;  $g_i$  —  $t$ -я узловая точка для признака  $x_i$ ;  $T_i - T_{i-1}$  — число базисных функций для признака  $x_i$ ; ( $T_0 = 1$ ).

Задача решалась в 2 этапа.

На первом этапе была проанализирована выборка из 155 больных, для которых были известны значения для всех 10-ти участвовавших в рассмотрении функциональных признаков. Основываясь на значениях функции правдоподобия (4.25) в точках  $\hat{\theta}$  для различных моделей связи, были отобраны признаки и проверена устойчивость полученного решения для нескольких вариантов разбиения выборки на учебную и контрольную. Наилучший результат оказался для решения по 2-м признакам  $x_1$  — клиренс креатинина и  $x_2$  — максимальное осмотическое концентрирование. Этому решению соответствует 6 оцениваемых параметров, т. е. размерность вектора  $\hat{\theta}$  в (2.25) равна 6.

На втором этапе для нахождения зависимости между ТИК и отобранными признаками выборка была увеличена до 390 больных. Найденная зависимость имеет вид:

$$F(\mathbf{x}, \hat{\theta}) = \hat{\vartheta}_0 + \varphi_1(x_1, \hat{\theta}_1) + \varphi_2(x_2, \hat{\theta}_2),$$

где

$$\hat{\vartheta}_0 = 3,5,$$

$$\varphi_1(x_1, \hat{\theta}_1) = \begin{cases} 0 & \text{при } x_1 \leq 80, \\ -\frac{7(x_1 - 80)}{200} & \text{при } 80 < x_1 \leq 100, \\ -0,7 - \frac{2(x_1 - 100)}{800} & \text{при } 100 < x_1 \leq 180, \\ -0,9 & \text{при } 180 < x_1. \end{cases}$$

$$\varphi_2(x_2, \hat{\theta}_2) = \begin{cases} 0 & \text{при } x_2 \leq 350, \\ -\frac{63(x_2 - 350)}{25\,000} & \text{при } 350 < x_2 \leq 600, \\ -0,63 - \frac{112(x_2 - 600)}{20\,000} & \text{при } 600 < x_2 \leq 800, \\ -1,75 - \frac{55(x_2 - 800)}{55\,000} & \text{при } 800 < x_2 \leq 1350, \\ -2,3 & \text{при } 1350 < x_2. \end{cases}$$



Балл ТИК, находимый по этой зависимости, для 50% больных совпал с баллом, назначаемым экспертом, для 36% больных отличался на 1, для 10% больных — на 2 и для 4% больных — на 3.

#### 4.4. Метод аппроксимации отношения правдоподобия

**4.4.1. Отношение правдоподобия как прогнозирующая функция.** Многие задачи геологии и геофизики приходится решать, не имея достаточно полной информации об изучаемом объекте или явлении. Таковы, в частности, задачи прогноза землетрясений и задачи поиска месторождений полезных ископаемых. Один из способов действия в условиях неполной информации состоит в использовании вероятностного подхода. Такой подход положен в основу излагаемого здесь метода решения класса задач прогнозирования.

Будем считать, что задача состоит в прогнозе наступления некоторого события  $C$  по комплексу данных  $\mathbf{x}$ . Событие  $C$  можно связать, например, с наличием месторождения или с землетрясением, интенсивность которого превосходит заданный порог. Под комплексом данных здесь понимается  $I$ -мерный вектор признаков  $\mathbf{x} = (x_1, \dots, x_I)$ , компоненты которого являются геологическими или геофизическими полями (картами), построенными для изучаемой территории.

Считая, что событие  $C$  и вектор  $\mathbf{x}$  имеют вероятностную природу, введём для вектора  $\mathbf{x}$  два условных (многомерных) распределения:  $P(\mathbf{x}/C)$  и  $P(\mathbf{x}/\bar{C})$ , где  $\bar{C}$  — событие, противоположное событию  $C$ . С математической точки зрения дать прогноз о событии  $C$  на основании вектора  $\mathbf{x}$  означает указать функцию от  $\mathbf{x}$ , которую можно рассматривать как прогнозирующую. С информационной точки зрения наилучшей прогнозирующей функцией будет та, которая содержит наибольшую информацию о событии  $C$ .

Функции от  $\mathbf{x}$ , несущие о событии  $C$  всю содержащуюся в  $\mathbf{x}$  информацию, называются достаточными статистиками. Хорошо известным примером достаточной статистики является функция отношения правдоподобия, представляющая в нашем случае отношение двух введённых условных распределений:  $P(\mathbf{x}/C)$  и  $P(\mathbf{x}/\bar{C})$ . Прогностический смысл функции отношения правдоподобия состоит в том, что большие её значения указывают на большую возможность наступления события  $C$ . Функцию отношения правдоподобия можно было бы считать искомой прогнозирующей функцией, если бы введённые условные распределения были известны. Однако эти распределения в рассматриваемых задачах неизвестны. Один из способов обойти указанную трудность состоит в приближённом представлении функции отношения правдоподобия и использовании экспериментальных данных (обучающей выборки) для оценки параметров полученного приближения.

Описываемый метод состоит в том, что функция отношения правдоподобия аппроксимируется многочленом. Для оценки качества приближения вводится критерий аппроксимации, позволяющий найти как

порядок, так и коэффициенты аппроксимирующего многочлена. Этот критерий используется также для отбора информативных компонент вектора  $\mathbf{x}$  — прогностически наиболее важных признаков.

**4.4.2. Аппроксимация отношения правдоподобия и модель апостериорной вероятности.** Обозначив логарифм функции отношения правдоподобия как

$$F(\mathbf{x}) = \ln \frac{P(\mathbf{x}/C)}{P(\mathbf{x}/\bar{C})}, \quad (4.31)$$

представим функцию  $F(\mathbf{x})$  в виде ряда:

$$F(\mathbf{x}) = \sum_{i=1}^{\infty} \vartheta_i U_i(\mathbf{x}), \quad (4.32)$$

где  $\vartheta_i$  — коэффициенты,  $U_i(\mathbf{x})$  — базисные функции ряда.

Ограничиваясь  $k$  первыми членами ряда, получим аппроксимацию  $k$ -го порядка для функции  $F(\mathbf{x})$ :

$$F_{\vartheta}^k(\mathbf{x}) = \sum_{i=1}^k \vartheta_i U_i(\mathbf{x}), \quad (4.33)$$

где  $\boldsymbol{\theta} = (\vartheta_1, \dots, \vartheta_k)$  —  $k$ -мерный вектор коэффициентов (параметров) ряда. Задача состоит в нахождении коэффициентов  $\vartheta_i$  усечённого ряда (4.33) и числа членов ряда  $k$ . Для решения этой задачи необходимо ввести критерий качества аппроксимации, позволяющий судить, насколько выражения (4.32) и (4.33) отличаются друг от друга. Критерий аппроксимации, рассматриваемый в следующем разделе, основан на преобразовании выражений (4.32) и (4.33) с помощью так называемой логистической функции (Cox, 1970):

$$Z(F) = \left( 1 + \frac{P(\bar{C})}{P(C)} e^{-F} \right)^{-1}, \quad (4.34)$$

где  $P(C)$  и  $P(\bar{C}) = 1 - P(C)$  будем рассматривать как априорные вероятности событий  $C$  и  $\bar{C}$ . Подставляя в выражение (4.34) в качестве аргумента функцию  $F(\mathbf{x})$  из выражения (4.31), получим

$$Z(F(\mathbf{x})) = \left( 1 + \frac{P(\bar{C})}{P(C)} \exp \left( -\ln \frac{P(\mathbf{x}/C)}{P(\mathbf{x}/\bar{C})} \right) \right)^{-1} = P(C/\mathbf{x}). \quad (4.35)$$

Легко видеть, что два последних члена, соединённые в выражении (4.35) знаком равенства, просто представляют собой (при прочтении справа налево) хорошо известную формулу Байеса, в которой  $P(C/\mathbf{x})$  называют апостериорной вероятностью события  $C$ . Соотношение (4.35) показывает, что апостериорная вероятность и логарифмическая функция правдоподобия связаны логистической зависимостью (4.34). Так как эта зависимость монотонна, то апостериорная вероятность, как и отношение правдоподобия, является достаточной

статистикой и поэтому статистики  $P(C/\mathbf{x})$  и  $F(\mathbf{x})$  информационно эквивалентны.

Заменяя в выражении (4.34) функцию  $F(\mathbf{x})$  её аппроксимацией (4.33), и обозначая это выражение как  $P_{\vartheta}^k(C/\mathbf{x})$ , получим

$$P_{\vartheta}^k(C/\mathbf{x}) = Z(F_{\vartheta}^k(\mathbf{x})) = \left( 1 + \frac{P(\bar{C})}{P(C)} \exp \left( \sum_{i=1}^k \vartheta_i U_i(\mathbf{x}) \right) \right)^{-1}. \quad (4.36)$$

Выражение (4.36) назовём моделью апостериорной вероятности  $k$ -го порядка, соответствующей аппроксимации (4.33).

**4.4.3. Критерий аппроксимации.** В математической статистике для сравнения двух вероятностных распределений  $P_1(\mathbf{x})$  и  $P_2(\mathbf{x})$  вводится информационная мера — расстояние Кульбака–Лейблера (*Kullback and Leibler, 1951*):

$$\rho(P_1, P_2) = \int \ln \frac{P_1(\mathbf{x})}{P_2(\mathbf{x})} dP_1(\mathbf{x}) = M \ln \frac{P_1(\mathbf{x})}{P_2(\mathbf{x})} = M \ln P_1(\mathbf{x}) - M \ln P_2(\mathbf{x}). \quad (4.37)$$

Эта мера обладает как свойствами информации, так и свойствами расстояния: она неотрицательна и равна нулю тогда и только тогда, когда оба сравниваемых распределения совпадают.

В нашем случае роль распределений  $P_1(\mathbf{x})$  и  $P_2(\mathbf{x})$  будут играть условные распределения  $P(v/\mathbf{x})$  и  $P_{\vartheta}^k(v/\mathbf{x})$ , соответствующие апостериорной вероятности  $P(C/\mathbf{X})$  и её модели  $P_{\vartheta}^k(C/\mathbf{x})$ :

$$P(v/\mathbf{x}) = [P(C/\mathbf{x})]^v [1 - P(C/\mathbf{x})]^{1-v}, \quad (4.38)$$

$$P_{\vartheta}^k(v/\mathbf{x}) = [P_{\vartheta}^k(C/\mathbf{x})]^v [1 - P_{\vartheta}^k(C/\mathbf{x})]^{1-v}, \quad (4.39)$$

где  $v$  — случайная величина, принимающая два значения:  $v = 1$ , если произошло событие  $C$  и  $v = 0$ , если произошло событие  $\bar{C}$ . В соответствии с (4.37) определим расстояние между введёнными условными распределениями (4.38) и (4.39) как

$$\rho(P, P_{\vartheta}^k) = \int \ln \frac{P(v/\mathbf{x})}{P_{\vartheta}^k(v/\mathbf{x})} dP(v, \mathbf{x}) = M \ln P(v/\mathbf{x}) - M \ln P_{\vartheta}^k(v/\mathbf{x}). \quad (4.40)$$

Так как распределение (4.38) зависит от  $F(\mathbf{x})$ , что следует из (4.35), а распределение (4.39) — от  $F_{\vartheta}^k(\mathbf{x})$ , что следует из (4.36), то определённая выражением (4.40) величина оказывается мерой близости между функцией отношения правдоподобия  $F(\mathbf{X})$  и её аппроксимацией  $F_{\vartheta}^k(\mathbf{x})$ .

**4.4.4. Оценка коэффициентов аппроксимации.** Пусть  $\{v_n, \mathbf{x}_n\}_{n=1}^N$  — обучающая выборка объёма  $N$  для случайных величин  $v$  и  $X$ . Оценивая в выражении (4.37) математическое

ожидание выборочным средним, получим в качестве оценки расстояния Кульбака–Лейблера (*Kullback and Leibler, 1951*)

$$\hat{\rho}(P, P_{\vartheta}^k) = M \ln P(v/\mathbf{x}) - \frac{1}{N} \sum_{n=1}^N \ln P_{\vartheta}^k(v_n/\mathbf{x}_n). \quad (4.41)$$

Оценка расстояния Кульбака–Лейблера (4.41) зависит от функций  $F(\mathbf{x})$  и  $F_{\vartheta}^k(\mathbf{x})$ , из которых первая (отношение правдоподобия) неизвестна, а вторая (аппроксимация) — известна с точностью до векторного параметра  $\theta$ , при условии, что базисные функции  $U_i(\mathbf{x})$  и порядок аппроксимации  $k$  заданы. Естественно подчинить выбор вектора параметров  $\theta$  условию минимума величины критерия (4.41). Оценка вектора параметров  $\theta$ , найденная из этого условия, должна иметь вид

$$\hat{\theta} = \arg \min_{\theta} \hat{\rho}(P, P_{\vartheta}^k) = \arg \max_{\theta} \frac{1}{N} \sum_{n=1}^N \ln P_{\vartheta}^k(v_n/\mathbf{x}_n). \quad (4.42)$$

Легко заметить, что оценка (4.42) совпадает с оценкой, доставляющей максимум условной функции правдоподобия  $\prod_{n=1}^N P_{\vartheta}^k(v_n/\mathbf{x}_n)$ .

В работе (*Юрков, 1983*) показано, что при определённых условиях оценка  $\hat{\theta}$  обладает свойством состоятельности относительно  $\theta$ .

Подстановка  $\theta = \hat{\theta}$  в выражение (4.33) даёт искомую аппроксимацию  $k$ -го порядка для функции логарифма отношения правдоподобия

$$F_{\hat{\vartheta}}^k(\mathbf{x}) = \sum_{i=1}^k \hat{\theta}_i U_i(\mathbf{x}). \quad (4.43)$$

Та же самая подстановка в выражение (4.41) даёт оценку критерия качества аппроксимации в виде

$$\hat{\rho}(P, P_{\hat{\vartheta}}^k) = \text{const} - \frac{1}{N} \sum_{n=1}^N \ln P_{\hat{\vartheta}}^k(v_n/\mathbf{x}_n). \quad (4.44)$$

Первое слагаемое в правой части выражения (4.44) обозначено как  $\text{const}$ , чтобы подчеркнуть, что оно не зависит ни от  $k$ , ни от  $\theta$ .

**4.4.5. Оценка порядка аппроксимации.** В предыдущем разделе порядок аппроксимации  $k$  считался заданным. Критерий аппроксимации (4.44) позволяет обосновать выбор величины  $k$ . Очевидно, что с ростом  $k$  величина

$$\hat{\rho}_k = \hat{\rho}(P, P_{\hat{\vartheta}}^k)$$

монотонно уменьшается, однако скорость её уменьшения, как показывает опыт, быстро падает. Целесообразно остановиться на таком значении  $k$ , дальнейшее увеличение которого не приводит к существенному уменьшению величины критерия  $\hat{\rho}_k$ . Этот подход к выбору  $k$  является качественным, так как он не учитывает, какое уменьшение величи-

ны критерия считать существенным. Возможен также количественный подход к выбору значения  $k$ , излагаемый ниже.

Предположим, что существуют такие значения  $k$  и  $\theta$ , что логарифм отношения правдоподобия (4.32) может быть представлен в виде конечного ряда (4.33), т. е. справедливо равенство

$$F(\mathbf{x}) = F_{\vartheta}^k(\mathbf{x}). \quad (4.45)$$

Это предположение означает, что распределения  $P(\mathbf{x}/C)$  и  $P(\mathbf{x}/\bar{C})$  принадлежат к так называемому экспоненциальному семейству распределений, примерами которого могут служить семейства нормальных, биномиальных, пуассоновских и ряд других известных распределений (Закс, 1975). В рамках допущения (4.45) возможна проверка гипотезы  $H_0$  о принадлежности распределения (4.38) к параметрическому классу распределений (4.39) для значений  $i$ , меньших, чем  $k$ :

$$H_0 : P(v/\mathbf{x}) = P_{\vartheta}^i(v/\mathbf{x}), \quad i < k. \quad (4.46)$$

Гипотеза  $H_0$  описывает подмножество распределений для модели (4.45) и означает, что, на самом деле, порядок модели ниже, чем предполагалось первоначально.

Обозначая оценку вектора параметров, найденную для модели  $i$ -го порядка как  $\theta$ , рассмотрим статистику:

$$\begin{aligned} \xi &= 2N(\hat{\rho}_1 - \hat{\rho}_k) = 2\left(\sum_n \ln P_{\vartheta}^k(v_n/\mathbf{x}_n) - \sum_n \ln P_{\vartheta}^i(v_n/\mathbf{x}_n)\right) = \\ &= 2 \ln \left( \prod_n P_{\vartheta}^k(v_n/\mathbf{x}_n) / \prod_n P_{\vartheta}^i(v_n/\mathbf{x}_n) \right), \quad (4.47) \end{aligned}$$

которая является удвоенным логарифмом отношения максимальных правдоподобий. Из результатов теории проверки гипотез известно, что когда гипотеза  $H_0$  верна, правая часть равенства (4.47) имеет асимптотическое хи-квадрат распределение с числом степеней свободы  $r = k - i$  (Кендалл и Стьюарт, 1973). Этот результат даёт возможность проверить гипотезу  $H_0$ , сравнив статистику  $\xi$  с порогом  $t_r(\alpha)$ , зависящим от уровня значимости проверяемой гипотезы  $\alpha$  и числа степеней свободы  $r$ . В качестве уровня значимости  $\alpha$  выбирается достаточно малая величина, лежащая обычно в диапазоне 0,1–0,01. Если  $\xi < t_r(\alpha)$ , то считается, что выборочные данные не противоречат гипотезе  $H_0$  и она принимается. В противном случае гипотеза  $H_0$  отвергается.

В соответствии с указанным подходом для определения порядка аппроксимации следует, задавшись достаточно большим значением  $k$ , произвести проверку неравенства  $\xi < t_{k-i}(\alpha)$ , полагая  $i = 1, \dots, k - 1$ . Первое значение  $i$ , для которого это неравенство окажется выполненным, определит искомый порядок аппроксимации.

**4.4.6. О выборе базисных функций.** Будем рассматривать увеличение порядка модели как её усложнение. Со статистической точки

зрения более сложная модель характеризуется большим числом статистических связей. С этой точки зрения наиболее простой является вероятностная модель, в которой компоненты вектора  $\mathbf{x} = (x_1, \dots, x_I)$  в двух условных распределениях  $P(\mathbf{x}/C)$  и  $P(\mathbf{x}/\bar{C})$ , образующих функцию отношения правдоподобия (4.31), статистически независимы. Более сложной является модель, в которой статистические зависимости распространяются не более, чем на пары компонент. Следующая по сложности модель должна охватывать тройки компонент и т. д.

Легко проверить, что логарифм отношения правдоподобия для случая независимых компонент должен иметь вид  $\sum_i \varphi_i(x_i)$ , для случая парных зависимостей — вид  $\sum_i \sum_j \varphi_{ij}(x_i, x_j)$ , для случая зависимостей троек — вид  $\sum_i \sum_j \sum_k \varphi_{ijk}(x_i, x_j, x_k)$  и т. д. Достаточно неопределённый характер функций  $\varphi_{ij}(x_i, x_j)$ ,  $\varphi_{ijk}(x_i, x_j, x_k)$  и т. д. можно уточнить, положив  $\varphi_{ij}(x_i, x_j) = \varphi_{ij}(x_i x_j)$ ,  $\varphi_{ijk}(x_i, x_j, x_k) = \varphi_{ijk}(x_i x_j x_k)$  и т. д., полагая тем самым функции  $\varphi_{ij}(\cdot)$ ,  $\varphi_{ijk}(\cdot)$  и т. д. одномерными. В этом частном случае аппроксимирующий многочлен запишется в виде:

$$F(\mathbf{x}) = \sum_i \varphi_i(x_i) + \sum_i \sum_j \varphi_{ij}(x_i x_j) + \sum_i \sum_j \sum_k \varphi_{ijk}(x_i x_j x_k) + \dots \quad (4.48)$$

Характер одномерных функций в выражении (4.48) можно далее уточнить, считая, что каждая из них является линейно-ломаной непрерывной функцией. Значения аргументов, в которых одномерная функция терпит излом, будем считать заданными. Вид каждой такой функции определяется её значениями в точках излома, сами эти значения можно рассматривать как неопределённые параметры функции, подлежащие нахождению. Нетрудно проверить, что при этом выражение (4.48) сводится к виду (4.32), в котором неопределённые коэффициенты обозначены как  $\vartheta_i$ . Другим очевидным частным случаем выражения (4.48) является степенной полиномиальный ряд

$$F(\mathbf{x}) = \sum_i \alpha_i x_i + \sum_i \sum_j \beta_{ij} x_i x_j + \sum_i \sum_j \sum_k \gamma_{ijk} x_i x_j x_k + \dots, \quad (4.49)$$

полученный из (4.48) в предположении, что каждая одномерная функция линейна по своему аргументу.

Известно, что степенной полином позволяет, как угодно точно, аппроксимировать непрерывную функцию (теорема Вейерштрасса о равномерном приближении непрерывных функций). Тем более это относится к ряду вида (4.48), который имеет то преимущество перед степенным полиномом (4.49), что, как показывает опыт, во многих практических случаях вполне удовлетворительная аппроксимация достигается уже при использовании только членов вида  $\sum_i \varphi_i(x_i)$ . Отметим также, что представление прогнозирующей функции  $F(\mathbf{x})$  в виде

суммы одномерных функций позволяет содержательно интерпретировать результаты решения многих практических задач.

**4.4.7. Информативность признаков.** Введённый критерий качества аппроксимации (4.44) может быть использован для выбора информативных признаков. Для простоты изложения рассмотрим случай, когда аппроксимирующее выражение для логарифма отношения правдоподобия имеет вид  $\sum_i \varphi_i(x_i)$ . Величина критерия (4.44), которую будем обозначать здесь как  $\hat{\rho}$ , может быть вычислена для любого набора (подмножества) признаков. Это даёт возможность сравнения различных наборов признаков: чем меньше величина  $\hat{\rho}$ , т. е. чем лучше аппроксимация, тем больше информативность данного набора.

Чтобы найти наиболее информативный набор, содержащий  $q$  признаков, необходимо сравнить все наборы из  $q$  признаков и выбрать тот набор, для которого величина  $\hat{\rho}$  окажется наименьшей. Отметим, что убывающая последовательность значений  $\hat{\rho}$  при  $q = 1, 2, \dots, I$  определяет наборы наиболее информативных признаков различной (увеличивающейся) размерности. Возникает вопрос, какова разумная величина  $q$ , определяющая число информативных признаков?

Поставленную задачу можно связать с задачей проверки статистической гипотезы. Для этого разобьем множество признаков  $\mathbf{X}$  на два непересекающихся подмножества:  $\mathbf{X}^{(1)}$  и  $\mathbf{x}^{(2)} = \mathbf{X} \setminus \mathbf{X}^{(1)}$  — дополнение подмножества  $\mathbf{X}^{(1)}$  до множества  $\mathbf{X}$ . Рассмотрим гипотезу вида

$$\varphi_i(x_i, \theta) = 0, \quad x_i \in \mathbf{X}^{(2)}. \quad (4.50)$$

Гипотеза (4.50) утверждает, что каждое слагаемое прогнозирующей функции  $F(\mathbf{x}) = \sum_i \varphi_i(x_i)$ , связанное с подмножеством  $\mathbf{X}^{(2)}$ , равно нулю. Эту гипотезу можно проверить, сравнив два значения величины критерия:  $\hat{\rho}^{(1)}$  и  $\hat{\rho}$ , где значение  $\hat{\rho}^{(1)}$  вычисляется для подмножества  $\mathbf{X}^{(1)}$ , а  $\hat{\rho}$  — для всего множества признаков  $\mathbf{X}$ . Незначительное отличие величин  $\hat{\rho}^{(1)}$  и  $\hat{\rho}$  свидетельствует в пользу гипотезы (4.50).

Для проверки того, значимо ли отличается разность  $\hat{\rho}^{(1)} - \hat{\rho}$  от нуля, воспользуемся статистикой отношения максимальных правдоподобий (4.47), которая в данном случае имеет вид

$$\xi = 2N(\hat{\rho}^{(1)} - \hat{\rho}). \quad (4.51)$$

Когда гипотеза (4.50) верна, статистика (4.51) имеет асимптотическое хи-квадрат распределение с числом степеней свободы, равным суммарному числу параметров  $r$ , участвующих в гипотезе (4.50). Другими словами, в случае гипотезы (4.50) справедливо (с точностью до асимптотики) уравнение

$$P(\xi > t_r(\alpha)) = \alpha, \quad (4.52)$$

где  $\alpha$  — уровень значимости критерия проверки гипотезы (4.50),  $t_r(\alpha)$  — табличное значение хи-квадрат распределения, соответствующее

шее уровню  $\alpha$  и числу степеней свободы  $r$ . Гипотеза (4.50) отклоняется (с уровнем значимости  $\alpha$ ), если  $\xi > t_r(\alpha)$ , в противном случае принимается.

Другой, эквивалентный, способ проверки гипотезы (4.50) состоит в следующем. Ввиду того, что функция  $t_r(\alpha)$  является по  $\alpha$  монотонно убывающей, введём обратную к ней функцию  $t_r^{-1}(\cdot)$ . В этом случае уравнение (4.52) эквивалентно уравнению

$$P(t_r^{-1}(\xi) < \alpha) = \alpha. \quad (4.53)$$

Другими словами, гипотеза (4.50) отклоняется, если

$$\varepsilon = t_r^{-1}(2N(\hat{\rho}^{(1)} - \hat{\rho})) < \alpha, \quad (4.54)$$

и принимается в противном случае. Введённая выражением (4.54) статистика  $\varepsilon$  называется наблюдаемым уровнем значимости (*Кокс и Хинкли, 1978*). Таким образом, если наблюдаемый уровень значимости  $\varepsilon$  велик (например, больше, чем значение  $\alpha = 0,1-0,01$ ), то это свидетельствует в пользу гипотезы (4.50), т. е. в пользу неинформативности подмножества  $\mathbf{X}^{(2)}$ , что является основанием для исключения его из участия в прогнозе.

**4.4.8. Алгоритм выбора информативных признаков.** Изложенный в предыдущем разделе способ нахождения наиболее информативных наборов признаков, основанный на полном переборе подмножеств признаков, вряд ли можно считать практичным, так как даже при относительно небольшой размерности вектора признаков  $\mathbf{x}$  требуется перебор весьма большого числа вариантов. Опишем квазиоптимальный алгоритм выбора информативных признаков, требующий значительно меньших вычислительных затрат. Описываемый алгоритм аналогичен алгоритму, известному в регрессионном анализе как метод включения (*Дрейпер и Смит, 1983*). Согласно этому алгоритму выбор информативного подмножества признаков производится поэтапно.

На первом этапе алгоритма временно считается, что прогнозирующая функция содержит только один признак  $x_i$  и имеет вид:

$$F(\mathbf{x}, \boldsymbol{\theta}) = \vartheta_0 + \varphi_i(x_i, \boldsymbol{\theta}), \quad i = 1, \dots, I,$$

где  $\vartheta_0$  — подбираемая алгоритмом константа (постоянная составляющая). Признак  $x_i$  выбирается из условия минимума значения критерия  $\hat{\rho}$  по всем  $i$ . Пусть, например, оказалось, что наименьшее значение критерия  $\hat{\rho}$  достигается при  $i = 1$ . Тогда рассматривается прогнозирующая функция вида

$$F(\mathbf{x}, \boldsymbol{\theta}) = \vartheta_0 + \varphi_1(x_1, \boldsymbol{\theta})$$

и проверяется гипотеза вида  $\varphi_1(x_1, \boldsymbol{\theta}) = 0$  так, как описано в предыдущем разделе. Если эта гипотеза принимается, то алгоритм выбора признаков завершается и считается, что информативные признаки отсутствуют. Если же указанная гипотеза отклоняется, то считается,



что прогнозирующая функция содержит, по крайней мере, один признак ( $x_1$ ) и переходят ко второму этапу алгоритма.

На втором этапе алгоритма временно считается, что прогнозирующая функция содержит два признака, один из которых ( $x_1$ ) выбран на первом этапе:

$$F(\mathbf{x}, \boldsymbol{\theta}) = \vartheta_0 + \varphi_1(x_1, \boldsymbol{\theta}) + \varphi_i(x_i, \boldsymbol{\theta}), \quad i = 2, \dots, I.$$

Признак  $x_i$  выбирается из условия минимума критерия  $\hat{\rho}$  по указанным значениям  $i$ . Пусть, например, оказалось, что наименьшее значение критерия  $\hat{\rho}$  достигается при  $i = 2$ . Тогда рассматривается прогнозирующая функция вида

$$F(\mathbf{x}, \boldsymbol{\theta}) = \vartheta_0 + \varphi_1(x_1, \boldsymbol{\theta}) + \varphi_2(x_2, \boldsymbol{\theta})$$

и проверяется гипотеза вида  $\varphi_2(x_2, \boldsymbol{\theta}) = 0$ . Если эта гипотеза принимается, то алгоритм выбора признаков завершается и считается, что прогнозирующая функция содержит единственный признак  $x_1$ , найденный на первом этапе алгоритма. Если же указанная гипотеза отклоняется, то считается, что прогнозирующая функция содержит, по крайней мере, два информативных признака ( $x_1$  и  $x_2$ ) и переходят к третьему этапу алгоритма, на котором выбирают третий информативный признак и т. д.

Хотя описанный алгоритм выбора информативных признаков не является оптимальным (в том смысле, что число выбранных признаков может оказаться завышенным), опыт показывает, что с его помощью исходный набор признаков удаётся значительно сократить.

**4.4.9. Доверительные интервалы для логарифма отношения правдоподобия.** Точность оценок логарифма отношения правдоподобия (4.31), а также апостериорной вероятности (4.36), можно охарактеризовать доверительными интервалами. Построение доверительных интервалов основано на допущении (4.45) о том, что распределения  $P(\mathbf{x}/C)$  и  $P(\mathbf{x}/\bar{C})$  принадлежат экспоненциальному семейству распределений. В этом случае, как отмечалось в разделе 4, оценка  $\hat{\boldsymbol{\theta}}$ , найденная в результате минимизации оценки Кульбака–Лейблера в соответствии с выражением (4.42), является одновременно оценкой максимального правдоподобия.

Как известно, распределение оценки максимального правдоподобия сходится к нормальному распределению, при этом характер сходимости имеет вид:

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N_k(0, \mathbf{I}_{\vartheta}^{-1}), \quad (4.55)$$

где  $\hat{\boldsymbol{\theta}}$  — оценка для  $k$ -мерного вектора параметров  $\boldsymbol{\theta}$ , полученного по выборке объёма  $N$ ,  $\mathbf{I}_{\vartheta}$  — информационная матрица Фишера в точке  $\boldsymbol{\theta}$ . Знак  $\sim$  означает асимптотическую сходимость: случайная величина, стоящая слева от него, с ростом  $N$  сходится по распределению к  $k$ -мерному нормальному распределению  $N_k$  с нулевым вектором математи-

ческого ожидания и ковариационной матрицей  $\mathbf{I}_{\vartheta}^{-1}$ . Из предельного соотношения (4.55) следует, что для всякого ненулевого вектора  $\lambda$

$$\lambda^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})/\widehat{\sigma} \sim N_1(0, 1), \quad (4.56)$$

где

$$\widehat{\sigma} = \sqrt{\frac{1}{N} \lambda^T \mathbf{I}_{\widehat{\vartheta}}^{-1} \lambda},$$

матрица

$$\mathbf{I}_{\widehat{\boldsymbol{\theta}}} = \frac{1}{N} \sum_{n=1}^N \frac{\partial}{\partial \boldsymbol{\theta}} \ln P_{\widehat{\boldsymbol{\theta}}}(v_n/\mathbf{x}_n) \left( \frac{\partial}{\partial \boldsymbol{\theta}} \ln P_{\widehat{\boldsymbol{\theta}}}(v_n/\mathbf{x}_n) \right)^T$$

является оценкой информационной матрицы Фишера в точке  $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ . Выражение (4.56) означает, что при достаточно большом  $N$  имеет место приближённое соотношение:

$$\begin{aligned} 1 - \alpha &= \frac{1}{\sqrt{2\pi}} \int_{-k_{\alpha/2}}^{k_{\alpha/2}} e^{-t^2/2} dt \approx P\{|\lambda^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})/\widehat{\sigma}| < k_{\alpha/2}\} = \\ &= P\{\lambda^T \widehat{\boldsymbol{\theta}} - k_{\alpha/2} \widehat{\sigma} < \lambda^T \boldsymbol{\theta} < \lambda^T \widehat{\boldsymbol{\theta}} + k_{\alpha/2} \widehat{\sigma}\} = \\ &= P\{G(\lambda^T \widehat{\boldsymbol{\theta}} - k_{\alpha/2} \widehat{\sigma}) < G(\lambda^T \boldsymbol{\theta}) < G(\lambda^T \widehat{\boldsymbol{\theta}} + k_{\alpha/2} \widehat{\sigma})\}, \end{aligned} \quad (4.57)$$

где  $k_{\alpha/2}$  — квантиль нормального распределения, отвечающего вероятности  $1 - \alpha/2$ ,  $G(\cdot)$  — произвольная монотонно возрастающая функция.

Логарифм отношения правдоподобия с учётом допущения (4.45) можно записать в виде

$$F(\mathbf{x}) = F_{\vartheta}^k(\mathbf{x}) = \sum_{i=1}^k \vartheta_i U_i(\mathbf{x}) = \mathbf{U}^T(\mathbf{x}) \boldsymbol{\theta},$$

где  $\mathbf{U}^T(\mathbf{x}) = (U_1(\mathbf{x}), \dots, U_k(\mathbf{x}))$  — векторная функция от  $\mathbf{x}$ . Полагая  $G(\cdot) = Z(F)$  (см. выражение (4.34)),  $\lambda = \mathbf{U}(\mathbf{x})$ , а также учитывая, что

$$\lambda^T = \mathbf{U}^T(\mathbf{x}) \boldsymbol{\theta} = F(\mathbf{x}), \quad \lambda^T \widehat{\boldsymbol{\theta}} = F_{\widehat{\vartheta}}^k(\mathbf{x}),$$

и подставляя значения этих величин в (4.57), получим, что с доверительной вероятностью  $1 - \alpha$  асимптотический доверительный интервал для логарифма отношения правдоподобия  $F(\mathbf{x})$  равен

$$\left[ F_{\widehat{\vartheta}}^k(\mathbf{x}) - k_{\alpha/2} \widehat{\sigma}, F_{\widehat{\vartheta}}^k(\mathbf{x}) + k_{\alpha/2} \widehat{\sigma} \right], \quad (4.58)$$

а для апостериорной вероятности  $P(C/\mathbf{x})=Z(F(\mathbf{x}))$  — равен

$$\left[ Z(F_{\widehat{\vartheta}}^k(\mathbf{x}) - k_{\alpha/2} \widehat{\sigma}), Z(F_{\widehat{\vartheta}}^k(\mathbf{x}) + k_{\alpha/2} \widehat{\sigma}) \right]. \quad (4.59)$$

**4.4.10. Реализация метода аппроксимации отношения правдоподобия в системе ГЕО.** Описанный выше метод реализован в системе ГЕО как один из методов статистического вывода в задачах геолого-геофизического прогнозирования. Алгоритм аппроксимации отношения правдоподобия рассчитан на диалоговый режим работы с пользовате-

лем. В режиме диалога осуществляется выбор информативных геолого-геофизических признаков и узлов линейно-ломанных одномерных функций. Значения одномерных функций в узлах, т.е. значения вектора параметров  $\vartheta$ , находятся в соответствии с выражением (4.42) с помощью оптимизационного алгоритма Дэвида–Флетчера–Пауэлла (Банди, 1988). Этот алгоритм позволяет эффективно искать экстремум строго выпуклой функции многих переменных. В работе (Юрков, 1983) показано, что функция, имеющая вид (4.41), в невырожденном случае является строго выпуклой функцией  $\vartheta$ . Так как строго выпуклая функция на ограниченном множестве имеет единственный экстремум, то используемый алгоритм оптимизации гарантирует единственность найденного вектора параметров  $\hat{\theta}$  в невырожденном случае.

В редких (вырожденных) случаях функция (4.41) может оказаться не строго выпуклой по  $\vartheta$ . Однако эти случаи (проявляющиеся обычно при малых обучающих выборках) связаны либо с чрезмерно большим числом признаков, либо с неудачным выбором узлов одномерных функций. Первый случай легко предотвратить включением режима поиска информативных признаков. Второй случай обычно легко обнаруживается по ненормально большим значениям некоторых одномерных функций. Простой способ избежать указанной неприятности состоит в уменьшении числа узлов для этих функций.

Для выбора числа информативных признаков в системе ГЕО используются две величины: оценка расстояния Кульбака–Лейблера  $\hat{\rho}$  и наблюдаемый уровень значимости  $\varepsilon$  (см. разделы 4.4.7 и 4.4.8). Первая из них служит мерой информативности выбираемого признака в совокупности с набором уже выбранных признаков. Вторая является статистикой проверки гипотезы вида  $\varphi_i(x_i, \theta) = 0$  и помогает пользователю принять решение о включении признака  $x_i$  в прогнозирующую функцию. Малые значения статистики  $\varepsilon$  (например, значения  $\varepsilon < 0,01 - 0,1$ ) свидетельствуют против этой гипотезы и указывают на желательность включения признака  $x_i$  в прогнозирующую функцию. Большие значения статистики  $\varepsilon$  (значения порядка 0,5 и выше) свидетельствуют в пользу этой гипотезы и указывают на бесполезность включения признака  $x_i$  в прогнозирующую функцию.

Общая рекомендация состоит в выборе признака, которому соответствует минимальная величина  $\hat{\rho}$ , при условии, что величина  $\varepsilon$  достаточно мала. Впрочем, пользователь может выбрать и любой другой признак, если, по его мнению, он более уместен по содержательным соображениям.

Реализация метода аппроксимации отношения правдоподобия в системе ГЕО предусматривает также оценку качества решения задачи в виде вероятностей ошибок принятия решения (ошибок распознавания) 1-го и 2-го рода. Ошибки решения определяются с помощью обучающей (или контрольной) выборки, для чего обучающая (контрольная) выборка разбивается на две подвыборки в зависимости от значения величины  $v$  (см. раздел 4.4.3). Ошибки 1-го рода определя-

ются по подвыборке, для членов которой  $v = 1$  (наличие события  $C$ ). Считается, что ошибка 1-го рода произошла, если для  $n$ -го члена этой подвыборки выполняется неравенство  $P_{\hat{\vartheta}}^k(C/\mathbf{x}_n) < t$ , где  $t$  — порог, значение которого принадлежит интервалу  $[0, 1]$ . Вероятность ошибки 1-го рода оценивается как доля ошибок 1-го рода в данной подвыборке. Вероятность ошибки 1-го рода возрастает с увеличением  $t$ .

Ошибки 2-го рода определяются по подвыборке, для членов которой  $v = 0$  (отсутствие события  $C$ ). Считается, что ошибка 2-го рода произошла, если для  $n$ -го члена этой подвыборки выполняется неравенство  $P_{\hat{\vartheta}}^k(C/\mathbf{x}_n) \geq t$ . Вероятность ошибки 2-го рода оценивается как доля ошибок 2-го рода в данной подвыборке. Вероятность ошибки 2-го рода с увеличением  $t$  убывает.

Особый интерес представляет то значение порога  $\hat{t}$ , для которого вероятности ошибок 1-го и 2-го рода равны. Величины этих значений вероятностей ошибок можно рассматривать как критерий успешного решения задачи. Малые вероятности ошибок означают, что классы (образы)  $C$  и  $\bar{C}$  почти безошибочно разделяются сравнением величины  $P_{\hat{\vartheta}}^k(C/\mathbf{x}_n)$  с порогом  $\hat{t}$ .

**4.4.11. Интерпретация результатов аппроксимации.** В системе ГЕО метод аппроксимации отношения правдоподобия ориентирован, в основном, на решение двух типов геолого-геофизических задач. Первый тип задач — вероятностный прогноз мест залегания полезных ископаемых. В задачах этого типа событие  $C$  интерпретируется как наличие месторождения определённого вида, которое может быть обнаружено в пределах заданной площадки, а событие  $\bar{C}$  — как отсутствие месторождения. Площадки образуются в результате разбиения территории градусной сеткой на ячейки стандартного размера, каждая из которых характеризуется вектором признаков  $\mathbf{x}$ . Найденную аппроксимацию функции отношения правдоподобия можно интерпретировать как отношение шансов за то, что в площадке, характеризующейся вектором признаков  $\mathbf{x}$ , будет найдено месторождение, к шансам против этого события.

Наряду с аппроксимацией отношения правдоподобия предусматривается аппроксимация апостериорной вероятности в соответствии с выражением (4.36). В это выражение входит априорная вероятность  $P(C)$ , которая оценивается как отношение числа площадок, в которых месторождение данного вида было обнаружено, к общему числу площадок рассматриваемой территории. Апостериорную вероятность можно интерпретировать как отношение числа площадок, в которых месторождение будет обнаружено, к общему числу площадок, при условии что все эти площадки характеризуются одним и тем же значением вектора признаков  $\mathbf{x}$ .

Второй тип задач, решаемых рассматриваемым методом, — прогноз мест сильных землетрясений. В задачах этого типа событием  $C$  (сильным землетрясением) считается землетрясение, интенсивность которо-

го превосходит заданный экспертом порог, событием  $\bar{C}$  — землетрясение, интенсивность которого меньше указанного порога.

При интерпретации результатов решения таких задач необходимо указать временной интервал, для которого рассчитан прогноз сильных землетрясений. Так, апостериорная вероятность, которую обозначим здесь как  $P_T(C/\mathbf{x})$ , интерпретируется как отношение числа площадок, характеризующихся значением вектора признаков  $\mathbf{x}$ , в которых за время наблюдения  $T$  произойдет хотя бы одно сильное землетрясение, к общему числу площадок с тем же самым значением вектора признаков  $\mathbf{x}$ . Априорная вероятность  $P_T(C)$  представляет собой долю площадок, в которых за время  $T$  наблюдалось хотя бы одно сильное землетрясение. В качестве  $T$  фигурирует тот временной интервал, на котором наблюдались сильные землетрясения, вошедшие в обучающую выборку. Землетрясения с интенсивностью, менее заданной, включаемые в обучающую выборку, выбираются случайным образом в соответствии с равномерным вероятностным распределением на изучаемой территории.

Вероятностный прогноз сильных землетрясений может быть рассчитан и для произвольного интервала  $T_1$ , большего или меньшего  $T$ . Если величина  $P_T(C/\mathbf{x})$  известна, то апостериорная вероятность наступления сильного землетрясения за время  $T_1$  рассчитывается по следующей формуле:

$$P_{T_1}(C/\mathbf{x}) = 1 - (1 - P_T(C/\mathbf{x}))^{T_1/T}.$$

Значения функции апостериорной вероятности (отношения правдоподобия) изображаются в системе ГЕО в виде карт на изучаемой территории. Места с большими величинами апостериорной вероятности (отношения правдоподобия) указывают на высокую возможность появления события — обнаружения полезного ископаемого в задачах первого типа либо наступления сильного землетрясения в задачах второго типа.

## Глава 5

# ТЕХНОЛОГИЯ ПРОГНОЗА МАКСИМАЛЬНЫХ ЭНЕРГИЙ ОЖИДАЕМЫХ ЗЕМЛЕТРЯСЕНИЙ

### 5.1. Введение

Одно из наиболее важных прикладных направлений геоинформационного пространственного прогноза относится к построению прогнозных карт потенциальных природных опасностей по комплексу геолого-геофизических данных. Эти карты необходимы в первую очередь для разработки нормативных документов, регламентирующих нормы строительства, и для выбора площадок строительства специальных сооружений и размещения особо опасных в экологическом отношении объектов.

Построение прогнозных карт природных опасностей состоит в оценивании энергетических и вероятностных характеристик ожидаемых источников природных опасностей и в вычислении меры их воздействий на наземные сооружения. Уровень воздействия обычно оценивается по корреляционным зависимостям, учитывающим параметры источника природной опасности, расстояние до него и эмпирическую зависимость затухания энергии. Для оценивания вероятностных характеристик источников используется статистика зарегистрированных ранее наблюдений за природными катастрофами. Наиболее сложной и важной является задача выделения возможных источников природных катастроф и оценки максимальной возможной энергии ожидаемой катастрофы. Сложность задачи заключается в необходимости оценки редкого события, для которого не имеется достаточно представительной выборки наблюдений, а важность связана с социальными и экономическими последствиями, обусловленными занижением или завышением этой характеристики.

В данной главе рассматривается применение технологии построения информационной модели пространственного прогноза к прогнозу максимальных магнитуд  $M_{\max}$  ожидаемых землетрясений. В разделе 5.2 сопоставляются основные методики, применяемые для составления прогнозных карт  $M_{\max}$  землетрясений. В разделах 5.3 и 5.4 излагается метод построения прогнозных карт  $M_{\max}$  землетрясений на примере Кавказского региона (*Gitis et al., 1992*).

## 5.2. Основные подходы к прогнозированию максимальных магнитуд ожидаемых землетрясений

В основе методов построения прогнозной карты максимальных магнитуд  $M_{\max}$  ожидаемых землетрясений лежат хорошо известные факты о повторяемости землетрясений. В связи с этим все методы построения прогнозной карты  $M_{\max}$  базируются на принципе прецедента. Но история сейсмических наблюдений является очень короткой по отношению к скорости тектонических процессов, а землетрясения с магнитудами, близкими к максимальным, происходят относительно редко. Поэтому иногда сильные землетрясения наблюдаются в тех местах, где надёжные исторические свидетельства о предшествующих сильных землетрясениях отсутствуют. Сравнительно недавними примерами таких землетрясений на территории СССР являются Газлийское, Спитакское и Рачинское землетрясения. Чтобы компенсировать этот эффект недостаточной представительности выборки зарегистрированных сильных землетрясений, предпринимаются попытки экстраполировать значения максимальных магнитуд землетрясений. Такая экстраполяция производится на основе предположения о том, что максимальная энергия землетрясения зависит от сеймотектонических свойств геологической среды, которые очень медленно изменяются во времени (Габриелов и др. 1986, Рейснер, 1980, Ризниченко, 1985).

Известен ряд подходов к составлению прогнозных карт  $M_{\max}$  землетрясений. Приведём основные из них.

Традиционным подходом является *неформальное экспертное районирование* региона. Данный подход основан на описательном знании сеймотектоники, на неформальном опыте эксперта, знаниях эксперта об исторической сейсмичности региона с учётом степени его исторической изученности (Арефьев и Татевосян, 1991, Бунэ и Горшков ред. 1980). Эксперт при решении производит качественный анализ карт эпицентров землетрясений и очаговых зон, карт разломной тектоники и других геологических и геофизических материалов. Недостатками метода являются субъективность прогнозной карты  $M_{\max}$  землетрясений и сложности в её обосновании. Однако, принимая во внимание отсутствие приемлемой физико-математической модели оценивания  $M_{\max}$ , недостаточно представительный каталог землетрясений, составленный, как правило, за относительно короткий период времени и содержащий поэтому ряд пропусков сильнейших землетрясений, учитывая существенную неопределённость в знании неотектонических процессов, легко понять, почему неформальное экспертное районирование всё ещё широко используется на практике и почему экспертное решение является одним из основных среди подходов при оценке сейсмической опасности.

Предпринималось множество попыток модифицировать подход неформального экспертного районирования и придать ему большую

объективность. Этому, в частности, посвящены работы (Бончев и др., 1983, Bonchev et al., 1982). В них предложена методика построения карты сейсмического районирования на основе коллегиально-экспертного решения. Методика состоит из трёх основных этапов. На первом этапе оценивается степень влияния на значение  $M_{\max}$  геологических, геофизических и сейсмологических факторов и экспертами оценивается степень важности влияния каждого из факторов (вес). На втором этапе учитывается интегральное влияние факторов на значение  $M_{\max}$  путём подсчёта взвешенных сумм факторов. В результате авторы получают два формальных правила, с помощью которых они ранжируют всю изучаемую территорию по уровню так называемого сейсмического потенциала и по грациям комплексных линеаментов. На третьем этапе карты сейсмического потенциала и линеаментов совмещаются. Для каждого элемента новой карты вычисляется суммарный ранг и строится карта зон одного ранга. После этого по имеющимся наблюдениям максимальных магнитуд землетрясений, зарегистрированных в зонах одного ранга, устанавливается соответствие между рангом и  $M_{\max}$ . На последнем этапе это соответствие используется для пересчёта рангов выделенных зон в значения  $M_{\max}$ . Привлекательным в данной методике является использование авторами в рассуждениях двух основных факторов, объясняющих физическую сущность  $M_{\max}$  в терминах плотности потенциальной энергии и максимального ранга геологической структурной неоднородности, от которой может зависеть размер очага максимального землетрясения. Но всё же степень формализации в данном подходе слишком слаба, чтобы преодолеть субъективность неформального экспертного районирования, так как прогноз основан на зависимости, которая оценивается экспертами путём голосования.

*Сейсмостатистический подход* состоит в оценивании  $M_{\max}$  землетрясений по сейсмологическим наблюдениям, попадающим в скользящее пространственное окно (Полякова и др., 1993, Burton, 1979). Для оценивания обычно используются порядковые статистики и, в частности, аппарат статистики экстремальных значений (Боровков, 1984, Писаренко, 1989). Главная проблема подхода состоит в необходимости иметь достаточно большую выборку сейсмических наблюдений для получения приемлемых асимптотических оценок. Для этого приходится выбирать размеры скользящего пространственного окна достаточно большими. Так, в работе (Burton, 1979) и в других работах этого направления размер окна выбирается равным  $4^\circ \times 4^\circ$ . Однако при таких больших размерах в окно попадают зоны, которые могут существенно отличаться по своим сейсмо-тектоническим свойствам, что делает выборку сейсмических наблюдений неоднородной. Кроме того, статистические методы оценивания экстремальных значений на реальных каталогах землетрясений очень чувствительны к ошибкам в определении магнитуд и даже малые ошибки могут привести к се-



рѐзным неточностям в оценивании  $M_{\max}$  (Osher, 1996). В заключение можно добавить, что данный подход не позволяет экстраполировать оценки  $M_{\max}$  на зоны с неподставительными сейсмологическими данными.

Следующая группа методов состоит в *экспертном разбиении региона на односвязанные однородные зоны и статистическом оценивании  $M_{\max}$* . Этот подход позволяет объединить неформальный опыт сейсмогеолога и аппарат статистического оценивания экстремальной статистики (Молчан, 1970, Бунэ и Горшков ред., 1980). Однако полученная по этой методике карта  $M_{\max}$  практически определяется указанными экспертом геологическими зонами и максимальными, зарегистрированными в этих зонах, землетрясениями. Методика не предусматривает экстраполяцию значений  $M_{\max}$  с зон, для которых имеются многолетние данные о землетрясениях, на зоны с аналогичными сеймотектоническими свойствами, но с неподставительным сейсмологическим материалом.

Следующий подход, основан на применении *техники кластерного анализа*. Он использует цифровые модели сеточного представления данных о геологической среде (например, Рейснер и Иогансон, 1993). Каждому элементу регулярной сетки ставится в соответствие некоторый заранее выбранный экспертом набор геологических и геофизических признаков. Подход основан на выделении методами кластерного анализа в пространстве заданных геолого-геофизических признаков однородных в соответствии с критерием кластеризации узлов координатной сетки (кластеров). Это процедура позволяет разбить все точки региона на кластеры, каждый из которых отображается на географической карте в виде одной или нескольких зон, состоящих из точек с близкими значениями признаков. Затем максимальные магнитуды землетрясений, зарегистрированные в пределах каждой зоны, присваиваются всем ячейкам соответствующих кластеров. На этом процедура построения карты  $M_{\max}$  заканчивается. Недостатки подхода обусловлены не вполне адекватным применением к задаче прогноза  $M_{\max}$  метода кластерного анализа, который не предназначен для совместной обработки признаков и значений прогнозируемых величин. Дело в том, что разбиение множества точек на однородные группы методами многомерного кластерного анализа  $M_{\max}$  в значительной степени определяется набором признаков, способом измерения сходства между кластерами, типом алгоритма кластеризации и, наконец, критерием остановки процесса кластеризации. К сожалению, в работах, развивающих данный подход, эти параметры по существу выбираются произвольно. С этим можно было бы мириться, если бы информация о значениях прогнозируемой величины полностью отсутствовала. Но в данном случае авторы предлагаемого подхода используют в качестве оценок  $M_{\max}$  зарегистрированные ранее значения сильных землетрясений. Для случая, когда существует выборка, для которой известны и признаки и значения прогноза, разработаны методы прогноза, в которых набор признаков,

мера сходства и сложность решающего правила могут быть выбраны в процессе оптимизации некоторого критерия качества (Ванник и Червоуенкис, 1974, Загоруйко, 1999, Турбович и др., 1971, Fukunaga, 1972, Zhuravlev, 1998).

Итак, мы видим, что в основе большинства рассмотренных методов решения задачи лежат две гипотезы:

1. гипотеза о повторяемости землетрясений,
2. гипотеза о том, что  $M_{\max}$  землетрясения зависит от свойств геологической среды.

Использование только гипотезы 1 приводит к ошибкам из-за существенных пропусков зарегистрированных землетрясений, обусловленных относительно коротким периодом наблюдений. Использование только гипотезы 2 приводит к ошибкам, так как зависимость максимальной энергии землетрясения от измеряемых свойств геологической среды известна лишь качественно.

Подводя итог приведённому сравнительному анализу методов, мы приходим к заключению о том, что степень неполноты информации в задачах построения прогнозных карт  $M_{\max}$  землетрясений слишком велика для применения чисто математических методов или чисто экспертных методов. Очевидно, что технология решения данной задачи должна полноценно использовать обе гипотезы. На первом шаге необходимо комплексно использовать всю имеющуюся информацию (каталог землетрясений, все типы данных о геологической среде, знания предметной области, а также формализованное знание, опыт и интуицию эксперта) для нахождения по выборке прецедентов зависимости  $M_{\max}$  от свойств среды по выборке прецедентов. На втором шаге необходимо экстраполировать найденную зависимость на весь регион. Желательно при этом найти объяснение зависимости и прогноза на языке предметной области.

Подобная технология разрабатывалась в 70-х – 90-х годах в ИППИ РАН. В этой технологии вначале эксперт формулирует качественно свои представления о причинах наиболее сильных землетрясений изучаемого региона. Затем по выборке прецедентов и на основании имеющихся геолого-геофизических данных в классе функций, которые могут быть интерпретированы экспертом, ищется прогнозирующая зависимость и по ней строится версия прогнозной карты. После этого экспертом производится анализ решения на соответствие со сформулированной гипотезой о природе сильнейших землетрясений региона и принимается решение о необходимости модификации решения. Такой результат решения задачи прогноза удобен для обсуждения со специалистами предметной области: в своей основе он содержит сейсмологическую гипотезу. Функция, используемая для прогноза, и сеточная модель прогноза получены на основе имеющихся наблюдений, имеются количественные и качественные данные в защиту выдвинутой экспертом гипотезы. При обсуждении может быть выдвинута альтернативная

гипотеза. Для неё, возможно, потребуется сгенерировать по имеющимся исходным данным иные признаки. После чего может быть найдено альтернативное решение, которое также можно обсуждать в терминах предметной области.

Эта технология совершенствовалась в течение ряда лет на большом числе практических задач оценки сейсмической опасности. Её основы были сформулированы в работах по высоко сейсмичным регионам Кавказа, Крыма, Западной Туркмении (*Буне и др., 1974, Буне и др., 1975, Буне и др., 1984*) и Болгарии (*Гитис и др., 1982*). Дальнейшее развитие данная технология получила при построении прогнозных карт  $M_{\max}$  землетрясений Карибского региона (*Gitis et al., 1989, Gitis et al., 1993*), Чешского массива (*Schenk et al., 1990, Schenk et al., 1991*) и Коста Рики (*Colombo et al., 1997*). Технология применялась совместно с ИФЗ РАН при выполнении проектов по оцениванию сейсмической опасности площадок строительства АЭС. Завершение разработки методики связано с работами по Кавказскому региону (*Gitis, 1995*) и Ставрополью (*Гитис и др., 1997*). На основании цикла работ по построению прогнозных карт максимальных возможных магнитуд землетрясений была предпринята попытка интерпретировать и обобщить найденные зависимости между  $M_{\max}$  и геолого-геофизическими признаками с учётом различных типов земной коры (*Schenk et al., 1994*).

### **5.3. Построение прогнозной карты максимальных магнитуд землетрясений Кавказского региона**

**5.3.1. Причинно-следственная модель региона.** Как следует из раздела 3.2, причинно-следственная модель в задачах построения прогнозных карт  $M_{\max}$  землетрясений должна представлять собой экспертную гипотезу о геодинамических особенностях региона и его окружения, включающую ориентировку внешних горизонтальных сил, расположение зон различных типов деформаций и связь характеристик среды с сейсмичностью.

Согласно широко распространённой точке зрения (*Зоненшайн и др., 1979, Konn, 1982, Philip et al., 1989*) динамические процессы в регионе определяются конвергенцией Аравийской и Евразийской плит. Регион находится в условиях горизонтального сжатия, ориентированного в меридиональном направлении. Границы коллидирующих плит определяются неоднозначно, особенно это относится к северному краю Аравийской плиты.

На основе анализа современного рельефа и геологических данных можно предположить, что Малый Кавказ представляет собой раздроблённый и деформированный северный выступ Аравийской плиты или самостоятельную микроплиту, в которой хребты, ориентированные в широтном направлении, отражают зоны максимальных деформаций в условиях сжатия. В этих зонах развиваются складки, надвиги

и взбросы. Косоориентированные по отношению к оси максимальных сжатий хребты должны испытывать сдвиговые деформации с образованием сдвигов и присдвиговых складок. В пределах изометрических плато и меридионально ориентированных хребтов вероятны проявления растяжений с формированием сбросов. Небольшая ширина хребтов, соответствующих зонам деформаций, позволяет предполагать, что дислокациям подвергается верхний слой консолидированной земной коры, который залегает на относительно более пластичном среднем слое.

Рассматриваемая часть Большого Кавказа состоит из трёх крупнейших морфологических структур: Куринской низменности, Большого Кавказского хребта и района предгорий. Эти структуры выражены в особенностях геолого-геофизических признаков, таких как рельеф поверхности Мохоровичича, гравиметрические аномалии в различных редукциях, амплитуды вертикальных неотектонических движений. Можно предполагать, что, в отличие от Малого Кавказа, деформации Большого Кавказа охватывают всю земную кору, а возможно, и верхнюю мантию. Общее запад-северо-западное простирание крупных морфоструктур является косоориентированным по отношению к направлению преобладающих сжатий. Поднятие Большого Кавказского хребта, связанное со сжатиями и нагнетанием материала, сопровождается его «разваливанием» и надвиганием в ограничивающие впадины. Косая ориентировка Кавказских структур к направлению максимального сжатия определяет существенную сдвиговую составляющую вдоль надвигов. Чистые надвиговые смещения преобладают в дугообразных сегментах хребтов.

В пределах Большого Кавказа сбросовые деформации в условиях растяжений должны преобладать вблизи осевой зоны Большого Кавказского хребта и, возможно, в центральной части Куринской низменности. В бортовых зонах хребта предполагаются условия сжатия и развитие сдвиго-надвиговых деформаций, при этом дугообразно изогнутые части прибортовых хребтов будут зонами развития чистых надвигов.

Описанные представления показаны на геодинамической схеме, где выделены зоны развития надвигов, сдвигов и сбросов. В работе *Sibson, 1974* на моделях показано, что напряжения, способные вызвать смещения по надвигам, сдвигам и сбросам, соотносятся приблизительно как 15 : 5 : 1.

Геодинамическая схема ранжирована в соответствии с этим отношением. При этом зонам развития надвигов приданы значения 20 и 30, зонам с преобладанием сдвигов — значения 9, 10, 12, 15, а остальной части региона придано значение 2 (рис. 5.1). Данная схема представляет причинно-следственную модель максимальных землетрясений изучаемого региона. Её можно интерпретировать как региональную модель распределения напряжений.

Анализ предложенной геодинамической схемы показал её высокую информативность при использовании в качестве признака прогноза

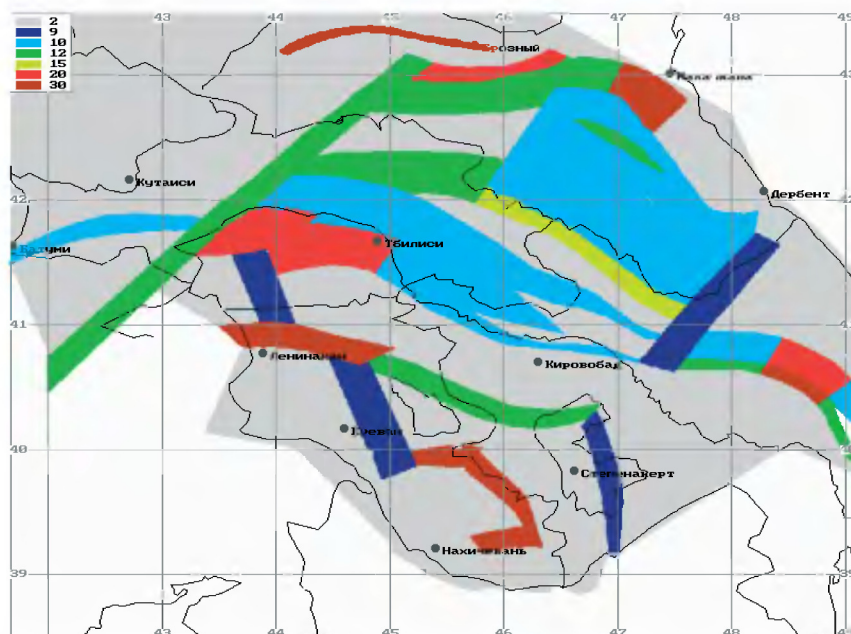


Рис. 5.1. Геодинамическая схема региона. Зонам развития надвигов приданы значения — 20 и 30, зонам с преобладанием сдвигов — значения 9, 10, 12, 15, остальной части региона придано значение 2.

$M_{\max}$ . Однако схема является результатом геологической интерпретации, её прямое использование привело бы к получению необъективных результатов. Зато гипотеза, лежащая в основе схемы, и процедура её построения достаточно просты и понятны для специалистов. Можно надеяться, что геолого-геофизические признаки, которые вносят наибольший вклад в аппроксимацию обсуждаемой схемы, должны быть достаточно просто интерпретированы в рамках модели и вместе с тем должны вносить достаточно высокий вклад в аппроксимацию оценок  $M_{\max}$ . Поэтому схема использовалась для отбора признаков прогноза. В результате анализа был сгенерирован ряд морфологических признаков рельефа и ряд признаков, характеризующих особенности разломной тектоники.

**5.3.2. Исходные данные.** При решении прогнозной задачи использовались следующие исходные данные.

**А) Карты числовых полей.**

- Гипсометрическая карта СССР масштаба 1 : 200 000.
- Карта градиентов скоростей вертикальных тектонических движений за постсарматское время ( $T = 17$  млн лет) масштаб 1 : 1 000 000, по данным В. И. Куликова, 1973 г.

- Карта строения поверхности консолидированной коры Кавказа масштаба 1 : 1 000 000, по данным Л. И. Андреева, 1988 г.
- Карта рельефа поверхности Мохоровичича юга СССР, масштаба 1 : 2 500 000, по данным Г. В. Краснопевцевой, 1984 г.
- Карта новейшей тектоники юга СССР масштаба 1 : 1 000 000, Полканова ред., 1972.
- Карта временных аномалий первых коровых волн юга СССР масштаба 1 : 2 500 000, по данным Г. В. Краснопевцевой, 1990 г.
- Карта скоростных горизонтальных вариаций верхней мантии Кавказа масштаба 1 : 1 500 000, по данным А. Х. Баграмяна, 1982 г.
- Карта теплового потока масштаба 1 : 1 000 000, по данным Л. И. Андреева и др., 1988 г.
- Карта аномалий силы тяжести в редукции Буге для территории СССР масштаба 1 : 200 000, 1975 г.
- Карта изостатических аномалий силы тяжести территории СССР масштаба 1 : 2 500 000, *Артемьев, 1976.*
- Карта глубинных аномалий силы тяжести территории СССР, *Рябой, 1979.*
- Карта аномального магнитного поля (дельта Та) СССР масштаба 1 : 200 000, 1978 г.

#### **Б) Зональные карты.**

- Тектоническая карта юга СССР масштаба 1 : 10 000 000, *Дикенштейн ред., 1975.*

#### **В) Линейные структурные элементы.**

- Карта разломов территории СССР и сопредельных стран масштаба 1 : 2 500 000, *Сидоренко, ред., 1980.*
- Карта космических линеаментов, по данным В. И. Макарова, 1978 г.

#### **Г) Сейсмические данные.**

- Новый каталог сильных землетрясений на территории СССР с древнейших времён до 1975 г., *Кондорская и Шебалин, 1977.*
- *Землетрясения в СССР, 1978–1990.*

**5.3.3. Признаки.** По исходным картам (раздел 5.3.2) были оцифрованы следующие группы первичных данных: величины залегания поверхностей (дневного рельефа, консолидированного фундамента, Мохоровичича), амплитуды неотектонических движений, градиент скорости постсорматских вертикальных движений, отклонение времён пробега первых коровых волн, отклонение времён пробега продольных волн в верхней мантии, значения теплового потока, значения аномалий силы тяжести (в редукции Буге, изостатические и глубинные), аномалий магнитного поля, прочностные свойства пород на поверхности, разломы, ранжированные по возрасту заложения, по возрасту

последней активизации, по типу движений и по тектонической значимости.

Большинство сеточных полей первичных признаков было представлено в сетке  $5' \times 5'$ .

**5.3.4. Выборка.** Выборка прецедентов генерировалась автоматически по каталогу землетрясений с учётом точности оценок магнитуды и координат эпицентров (раздел 4.2.2). Автоматическое генерирование выборки даёт возможность проверить устойчивость версий прогноза путём анализа решений, полученных для многих выборок, и позволяет в определённой степени устранить субъективизм при генерировании выборки. Тем не менее, привлечение экспертных знаний при оценивании  $M_{\max}$  всегда полезно, поскольку решения эксперта основываются не только на данных каталога, но и на макросейсмических, исторических данных, а также на его интуиции. Желательно, поэтому, наряду с автоматическим генерированием выборки использовать экспертные решения.

**5.3.5. Информационная модель прогноза.** В процессе решения исследовались различные версии прогноза, различающиеся по составу признаков и по способу генерирования выборок. Так, например, решались задачи на выборках, сгенерированных по землетрясениям, произошедшим на Малом Кавказе, Большом Кавказе и во всём регионе, использовались различные методы ранжирования разломов (по геологической значимости, по возрасту заложения, по возрасту последней активизации, по кинематике), исследовался широкий набор преобразований над точечными, линейными и растровыми данными, такие как сглаживание путём вычисления скользящего среднего, медианное сглаживание, низкочастотная и высокочастотная фильтрация; вычисление градиента, лапласиана, производных по направлению и дисперсии в скользящем окне; вычисление растровых полей суммарной длины разломов, полей расстояний и близости к разломам; вычисление растровых полей сейсмической активности, плотности эпицентров; вычисление алгебраических функций от нескольких растровых полей и т. д. В общей сложности проанализировано более 200 растровых признаков. При анализе версий отдавалось предпочтение тем, в которых достигается наименьший уровень ошибки, выбранные признаки и прогнозирующая функция наиболее просто интерпретируются в рамках сформулированной геодинамической модели.

Решение, наиболее адекватное причинно-следственной модели раздела 5.5.1, получено для версии прогноза, в которой обучающаяся выборка генерировалась только по данным Малого Кавказа. Эта версия рассматривается ниже.

Для нахождения прогнозирующей функции использовалась выборка из 475 точек, сгенерированная по каталогу землетрясений Малого Кавказа. Каталог содержит 59 событий с  $M > 5,0$ . Фоновое значение максимальных магнитуд землетрясений было выбрано  $M_0 = 5,0$ . При

этом  $\sigma_1$  изменялось в диапазоне от 0,1 до 1,0 долей магнитуды, а  $\sigma_2$  изменялось в диапазоне от 0,05 до 1,0 градуса. Выборка содержит 270 точек с оценками  $M_{\max}$  в диапазоне магнитуд от 5,0 до 7,1 и 205 точек со значениями  $M_{\max}$  от 4,5 до 5,0.

Наиболее информативными оказались следующие признаки:

$x_1$  — полусумма растровых полей близости к надвиговому ( $y_1$ ) и сдвиговому ( $y_2$ ) разломам, активным в кайнозое:

$$x_{1t} = \frac{y_{1t} + y_{2t}}{2},$$

где

$$y_{st} = 100 \left( 1 - \frac{\rho_{st}}{R} \right),$$

$\rho_{st}$  — расстояние от точки растра  $t$  до ближайшего надвига при  $s = 1$  или сдвига при  $s = 2$ , радиус  $R = 50$  км;

$x_2$  — модуль градиента скорости вертикальных движений за постсарматское время;

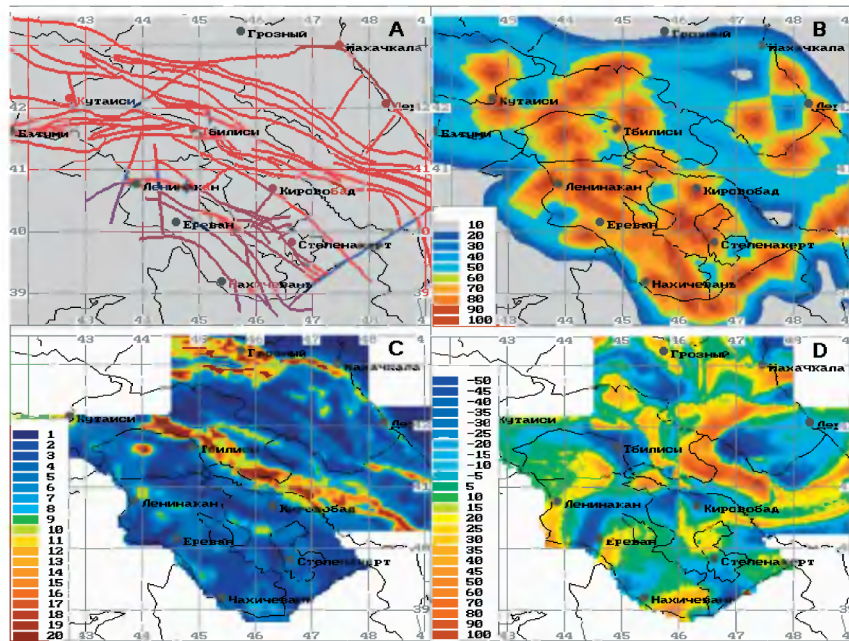


Рис. 5.2. Наиболее информативные геолого-геофизические данные: «А» — Надвиговые (красный цвет) и сдвиговые (синий цвет) разломы, активные в кайнозое; «В» — Признак  $x_1$ , равный полусумме растровых полей близости к надвиговому ( $y_1$ ) и сдвиговому ( $y_2$ ) разломам, активным в кайнозое; «С» — Признак  $x_2$ , равный модулю градиента скорости вертикальных движений за постсарматское время; «D» — Признак  $x_3$ , равный аномалии отклонений времён пробега продольных волн в верхней мантии



$x_3$  — аномалии отклонений времён пробега продольных волн в верхней мантии

$$x_{3t}(\alpha, \beta) = y_{3t}(\alpha, \beta) - \int_{u^2+v^2=R^2} y_t(\alpha - v, \beta - u) du dv,$$

где  $R = 30$  км,  $y_{3t}$  — отклонения времён пробега продольных волн в верхней мантии;

$x_4$  — аномалии силы тяжести в редукции Буге.

Карты наиболее информативных данных представлены на рис. 5.2. Поле признака  $x_1$  хорошо согласуется с геодинамической причинно-следственной моделью, согласно которой максимальные напряжения возникают в зонах надвигов и сдвигов. С этой моделью также согласуется поле признака  $x_2$ : хребты, ориентированные ортогонально или под углом к направлению сближения Аравийской и Евразийской плит, имеют максимальные значения градиентов скоростей вертикальных движений. Согласно модели, они испытывают наибольшие сдвиговые и надвиговые деформации. Поле признака  $x_3$  труднее интерпретировать в рамках геодинамической модели. Но легко видеть, что этот признак характеризует особенности геологической среды, влияющие на сей-

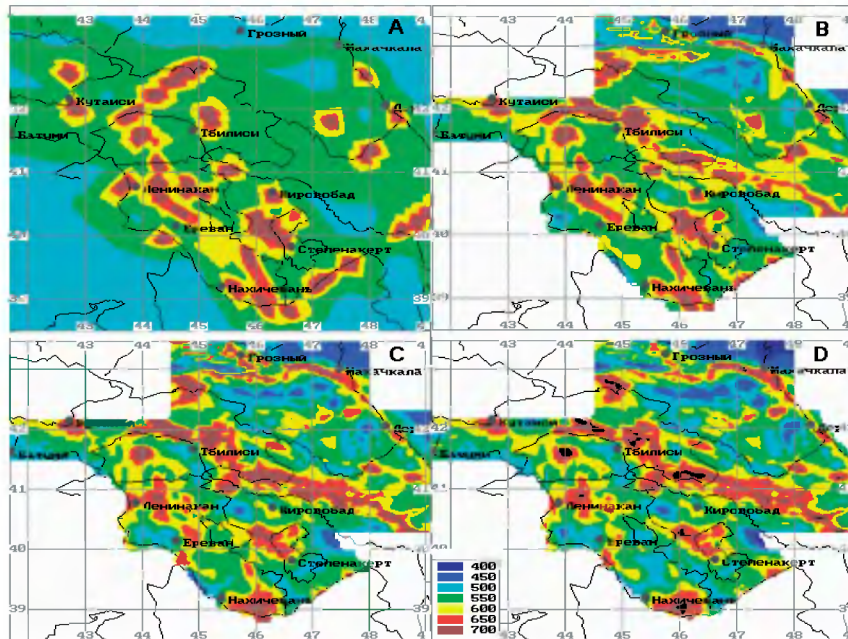


Рис. 5.3. Изменение прогнозного поля  $M_{\max}$  в зависимости от числа признаков. «А»: признак  $x_1$ ; «В»: признаки  $x_1$  и  $x_2$ ; «С»: признаки  $x_1$ ,  $x_2$  и  $x_3$ ; «D»: признаки  $x_1$ ,  $x_2$ ,  $x_3$  и  $x_4$

смичность: положительные аномалии времён пробега в верхней мантии отвечают наиболее жёсткому веществу, влияющему на накопление напряжений в вышележащей земной коре.

Изменение прогнозного поля  $M_{\max}$  в зависимости от числа признаков показано на рис. 5.3. Наиболее информативным признаком является поле  $x_1$ . На карте прогнозного поля  $M_{\max}$ , построенного по этому признаку (рис. 5.3, «А»), видно, что зоны возможных значений  $M_{\max} > 6,0$  приурочены к областям сближения надвигов и сдвигов.

На карте прогнозного поля  $M_{\max}$ , построенного по двум признакам  $x_1$  и  $x_2$  (рис. 5.3, «В») наметились зоны с высокими значениями  $M_{\max}$  вдоль южного и северного склонов Большого Кавказа. На Малом Кавказе выделяется широкая зона северо-западного простираения с большими значениями  $M_{\max}$ . Прогнозные поля, полученные по трём признакам  $x_1$ ,  $x_2$  и  $x_3$  (рис. 5.3, «С»), и по четырём признакам  $x_1$ ,  $x_2$ ,  $x_3$  и  $x_4$  (рис. 5.3, «D») практически не отличаются друг от друга (в качестве признака  $x_4$  программой аппроксимации было отобрано поле гравитационных аномалий в редукции Буге). Существенное сходство рис. 5.3, «С» и рис. 5.3, «D» свидетельствует в пользу того, что рационально остановиться на более простой версии прогноза, полученной по трём признакам  $x_1$ ,  $x_2$  и  $x_3$ .

Найденная прогнозирующая функция имеет вид:

$$F(x) = 4,1 + \varphi_1(x_1) + \varphi_2(x_2) + \varphi_3(x_3),$$

где

$$\varphi_1(x_1) = \begin{cases} 0 & 0 < x_1 \leq 25, \\ (x_1 - 25) \cdot 0,027 & 25 < x_1 \leq 51, \\ 0,7 + (x_1 - 51) \cdot 0,004 & 51 < x_1 \leq 77, \\ 0,8 + (x_1 - 77) \cdot 0,056 & 77 < x_1 \leq 93, \\ 1,7 & 93 < x_1 \leq 100; \end{cases}$$

$$\varphi_2(x_2) = \begin{cases} 0 & 0 < x_2 \leq 2,3, \\ (x_2 - 2,3) \cdot 0,15 & 2,3 < x_2 \leq 6,3, \\ 0,6 & 6,3 < x_2 \leq 10,6, \\ 0,6 + (x_2 - 10,6) \cdot 0,156 & 10,6 < x_2 \leq 21,3, \\ 2,3 & 21,3 < x_2 \leq 30; \end{cases}$$

$$\varphi_3(x_3) = \begin{cases} 0 & -10 < x_3 \leq -3,6, \\ (x_3 - 3,6) \cdot 0,082 & -3,6 < x_3 \leq 1,3, \\ 0,4 + (x_3 - 1,3) \cdot 0,29 & 1,3 < x_3 \leq 3,0, \\ 0,9 & 3,0 < x_3 \leq 10. \end{cases}$$

Функции  $\varphi_i(x_i)$  показаны на рис. 5.4. Все функции  $\varphi_i(x_i)$  являются неубывающими. Геолого-геофизическая интерпретация этого состоит в том, что увеличение значений каждого признака при прочих равных условиях приводит к увеличению (точнее не уменьшает) значений максимальной возможной магнитуды землетрясения.

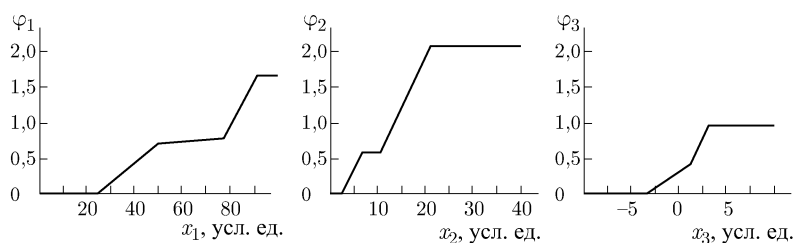


Рис. 5.4. Функции  $\varphi_i(x_i)$

Средняя ошибка аппроксимации оценок  $M_{\max}$  на контрольной выборке равна 0,3. С учётом точности оценок  $M_{\max}$  и диапазона аппроксимируемых значений  $M_{\max}$  эта величина свидетельствует о достаточно высоком качестве аппроксимации.

На рис. 5.5, «А» показано пространственное распределение ошибок прогноза  $M_{\max}$  на учебной и контрольной выборках: серые ромбики указывают на то, что ошибка аппроксимации оценок невелика, красные ромбики указывают на ошибки завышения значений прогноза, а синие — на ошибки занижения прогноза, размер ромбиков увеличивается при увеличении ошибок. Поля вкладов признаков в прогноз — поля  $\varphi_i(x_i)$  — показаны на рис. 5.5, «В», «С» и «D». Карта вклада признака  $x_1$  в прогнозную карту (рис. 5.5, «В») выделяет зоны сближения надвигов и сдвигов. Карта вклада признака  $x_2$  (рис. 5.5, «С») содержит три наиболее значимые полосы, протягивающиеся в восток-северо-восточном направлении вдоль южного и северного склонов Большого Кавказа, а также в центральной части Малого Кавказа. Вклад этого признака существенно приближает прогнозную карту к геодинамической модели. Карта вклада признака  $x_3$  (рис. 5.5, «D») имеет сложный рисунок сочетающий дугообразные, изометрические и линейные зоны. Вклад этого признака характеризует влияние жёсткой упругой мантии на возникновение больших землетрясений.

На рис. 5.6 представлены результаты прогноза по признакам  $x_1$ ,  $x_2$ ,  $x_3$ . Полученное прогнозное поле (рис. 5.6, «А»), согласуется с геоди-

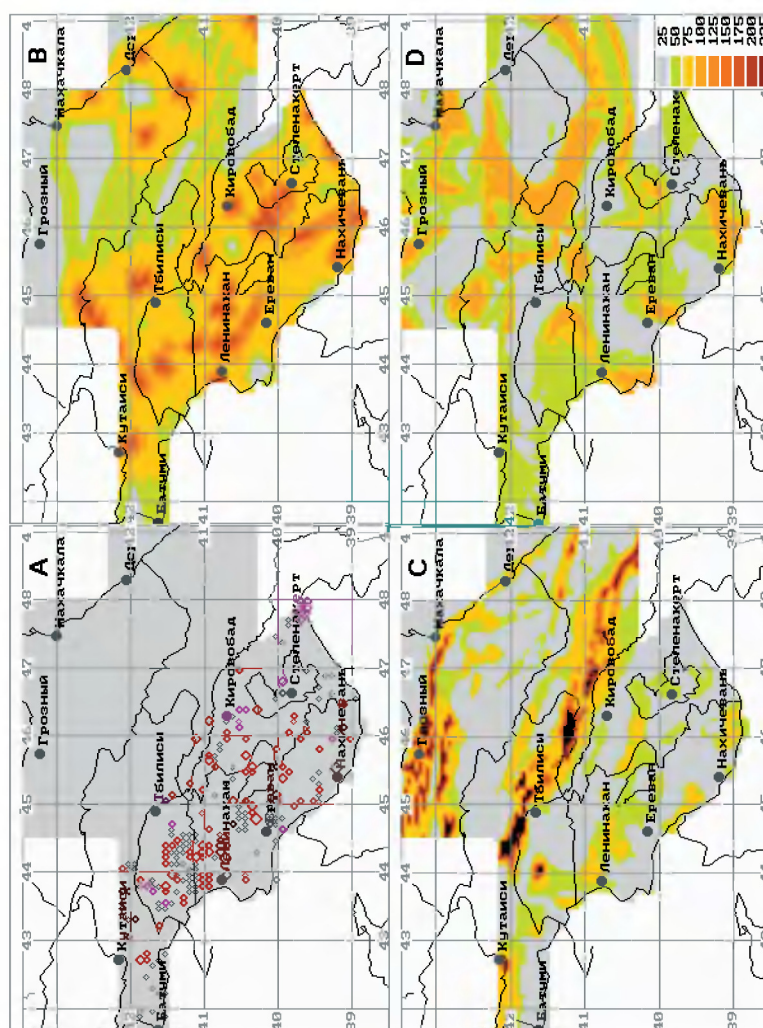


Рис. 5.5. Пространственное распределение ошибок прогноза на выборке и вклады признаков в прогноз (значения вкладов умножены на 100). «А»: серые ромбики указывают на то, что ошибка аппроксимации оценок невелика, красные ромбики указывают на ошибки завышения значений прогноза, а синие — на ошибки занижения прогноза, размер ромбиков увеличивается при увеличении ошибок. «В», «С» и «D»: поля вкладов признаков  $x_1$ ,  $x_2$ ,  $x_3$  в прогноз — поля  $\varphi_i(x_i)$

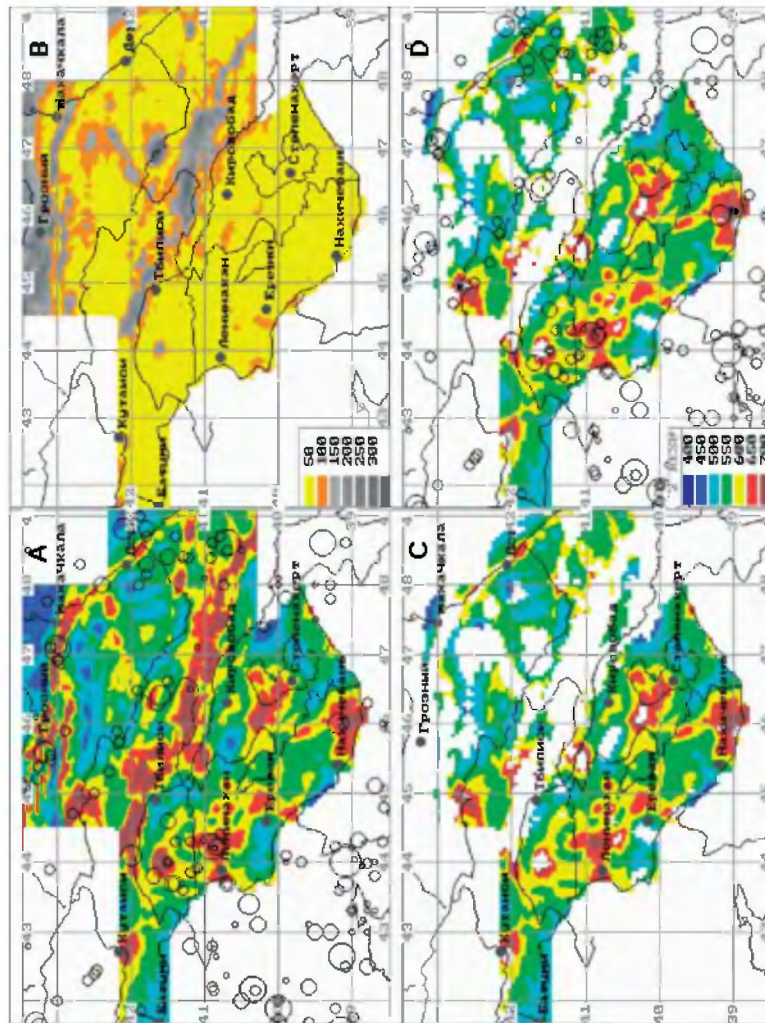


Рис. 5.6. Результаты прогноза по признакам  $x_1$ ,  $x_2$ ,  $x_3$  (значения магнитуд и расстояний умножены на 100). «А»: прогнозные поле и эпицентры землетрясений с магнитудой более 5; «В»: растровое поле евклидовых расстояний (в нормированном по дисперсиям трёхмерном пространстве признаков) между множеством точек выборки и остальными точками растра; «С» и «D»: область прогнозной карты, состоящая из точек растра, для которых расстояние до точек выборки не превышает 0,5

намической моделью и предполагаемым распределением напряжений, показанным на рис. 5.1.

Часть прогнозного поля  $M_{\max}$ , относящаяся к Большому Кавказу, показывает достаточно хорошую экстраполяцию найденной на Малом Кавказе прогнозирующей функции. Эпицентры почти всех крупных землетрясений восточной части Большого Кавказа находятся внутри или вблизи прогнозных зон с соответствующими значениями  $M_{\max}$ . На Малом Кавказе выявляются четыре зоны высоких значений  $M_{\max}$ . Одна из них имеет широтное простирание и располагается на северо-западе Малого Кавказа. Вторая зона проходит южнее и параллельно первой. К ней приурочено Спитакское землетрясение декабря 1988 г.

Третья зона также имеет субширотное простирание и локализуется в самой южной части Малого Кавказа. В её пределах располагается Зангезурское землетрясение 27 апреля 1931 г ( $M = 6,3$ ). Четвёртая зона приурочена к северо-западной части Нагорного Карабаха и также вытянута в широтном направлении.

На рис. 5.6, «В» показана сеточная модель поля измеренных в трёхмерном пространстве нормированных по дисперсиям признаков евклидовых расстояний между множеством точек выборки и остальными точками раstra (расстояния умножены на 100). Можно предполагать, что точность прогноза (экстраполяции) падает с увеличением расстояния от точек выборки, на которых оценивалась прогнозирующая функция. На рис. 5.6, «С» и 5.6, «D» показаны только точки раstra прогнозной карты, для которых расстояние от точек выборки не превышает 0,5.

После решения производился анализ прогнозного поля  $M_{\max}$  по прецеденту для ряда точек раstra. В частности, производился анализ точек из зоны с высоким прогнозными значениями  $M_{\max}$ , расположенной в северо-западной части Нагорного Карабаха. В этой зоне в XX столетии не наблюдалось сильных землетрясений. Анализ показал, что в регионе имеются только две области со значениями признаков, близкими к значениям в анализируемой точке. Одна из них находится в окрестности анализируемой точки, а другая приурочена к эпицентральной зоне Спитакского землетрясения. Сходство этих зон по признакам подтверждает полученный прогноз.

Рассмотрим в качестве примера анализ для расположенной к северо-востоку от Степанакерта точки с прогнозным значением  $M_{\max}$  около 7 и с координатами  $\lambda = 46^\circ$ ,  $\varphi = 40,25^\circ$ . В пространстве признаков был построен прямоугольник с центром в анализируемой точке и с координатами  $x_1 = 91 \pm 5$ ,  $x_2 = 9 \pm 3$  [10/год],  $x_3 = 1,8 \pm \pm 0,3$  [сек]. Анализ показал, что в пределах региона имеются только две точки раstra, векторы признаков которых попадают в выбранную прямоугольную область. Одна из точек находится в окрестности анализируемой точки, а другая — в пределах очага Спитакского землетрясения. Близость анализируемой точки в пространстве найденных при решении задачи признаков геологической среды

к точкам очаговой зоны Спитакского землетрясения, где землетрясение с подобной магнитудой произошло, является подтверждением высокого значения  $M_{\max}$  в анализируемой точке.

Рассмотрим теперь использование рассмотренной в разделе 3.2 техники кластерного анализа на примере объяснения прогноза значений  $M_{\max} \geq 6,0$ .

На рис. 5.7 показано разбиение прогнозной зоны со значениями  $M_{\max} \geq 6,0$  на 8 кластеров, на которое нанесены надвиговые и сдвиговые кайнозойские разломы.

На рис. 5.8 показаны гистограммы значений признаков  $x_1$ ,  $x_2$ ,  $x_3$  и значений прогноза  $F$ . Синим цветом показаны гистограммы для всего исследуемого региона, а жёлтым — гистограммы этих же характеристик для прогнозной зоны со значениями  $F \geq 6,0$ . Можно видеть, что значения прогноза  $F \geq 6,0$  имеют место в основном для зон с тектоническими нарушениями, однако значения признаков  $x_3$  и  $x_2$  принимают широкий диапазон значений и поэтому интерпретация решения затруднена. Кластеризация позволяет сформулировать более точные высказывания о свойствах этой прогнозной зоны. Приведём некоторые примеры.

На рис. 5.9 показаны гистограммы кластера с наибольшей площадью с номером 2. По гистограммам можно сделать следующий вывод (для данного региона):

- если* значения аномалий вариаций времён пробега Р-волн признаков (признак  $x_3$ ) — положительные
- и* значения модуля градиента скорости постсарматских вертикальных движений (признак  $x_2$ ) — невелики
- и* значения признака  $x_1$  — большие (точнее: точки, находятся на расстоянии до 30 км от узлов пересечения надвиговых и сдвиговых разломов),
- то* значение  $M_{\max} \geq 6,0$ .

На рис. 5.10 показаны гистограммы кластера с номером 4. Из рисунка следует, что

- если* значения признаков  $x_3$  и  $x_2$  — малы
- и* значения признака  $x_1$  — очень велики (точки, находящиеся на расстоянии до 10 км от узлов пересечения надвиговых и сдвиговых разломов),
- то* значение  $M_{\max} \geq 6,0$ .

На рис. 5.11 показаны гистограммы кластера с номером 5. Из рисунка следует, что

- если* значения признака  $x_3$  — положительные
- и* значения признака  $x_2$  — большие
- и* значения признака  $x_1$  — средние (точки, находящиеся на расстоянии около 10 км от надвиговых и сдвиговых разломов и далеко отстоящие от узлов их пересечения),
- то* значение  $M_{\max} \geq 6,0$ .

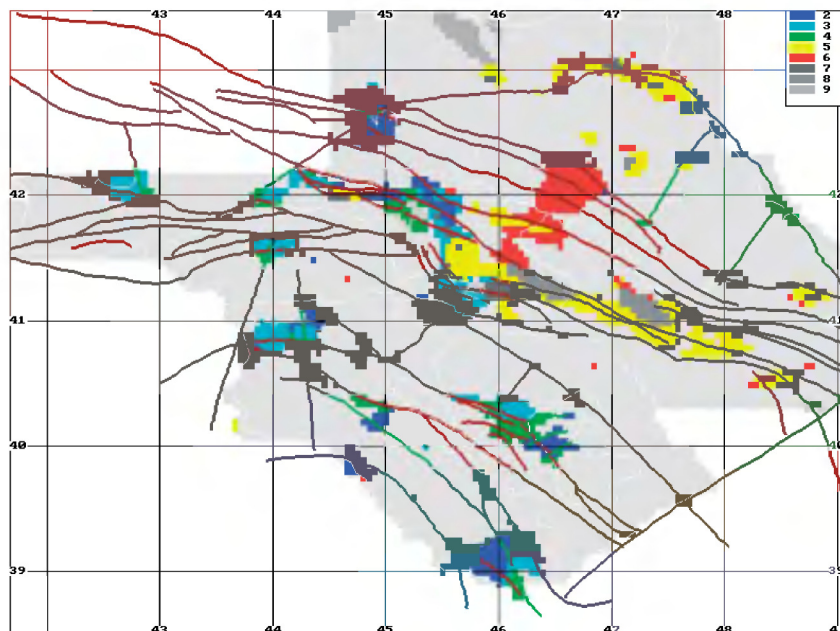


Рис. 5.7. Карта кластеров для прогнозной зоны со значениями  $M_{\max} \geq 6,0$

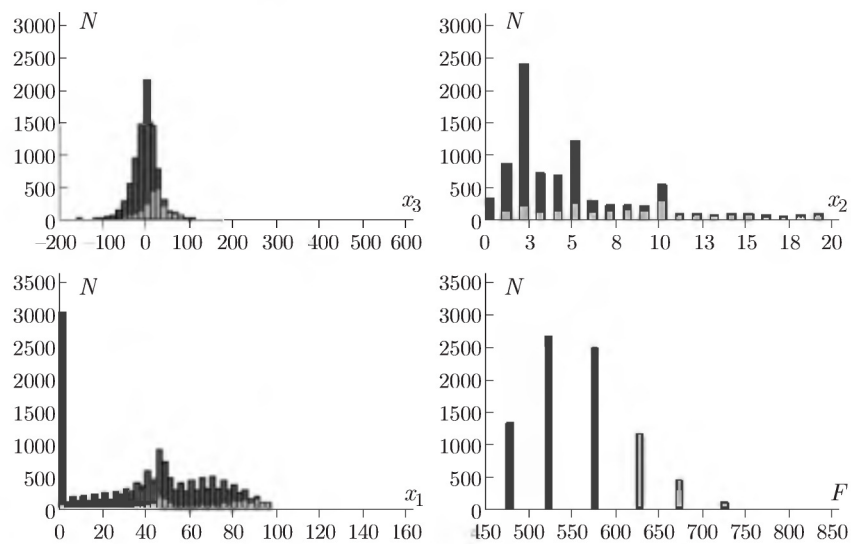


Рис. 5.8. Гистограммы значений признаков  $x_1$ ,  $x_2$ ,  $x_3$  и прогноза  $F \geq 6,0$  для всего исследуемого региона (чёрный цвет) и для прогнозной зоны со значениями  $M_{\max} \geq 6,0$  (серый цвет)



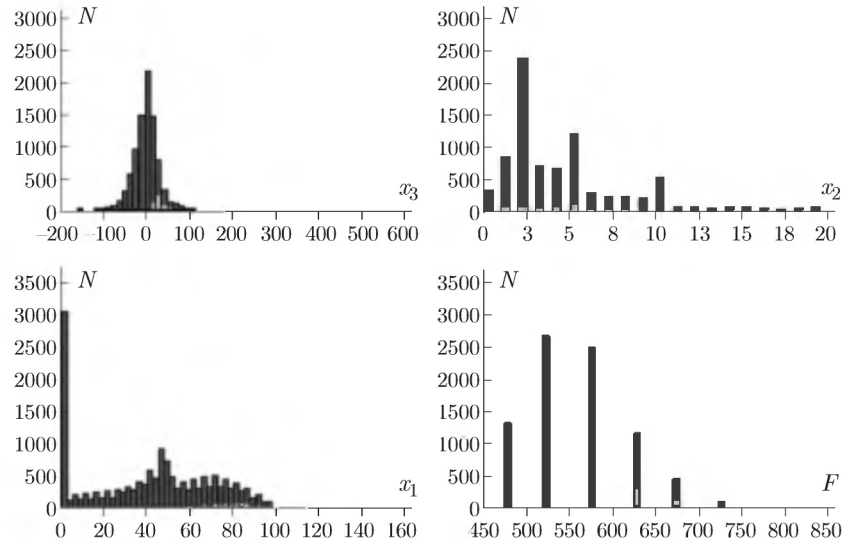


Рис. 5.9. Гистограммы кластера с номером 2. Чёрным цветом показаны гистограммы для всего исследуемого региона, а серым — гистограммы этих же характеристик для прогнозной зоны со значениями  $F \geq 6,0$

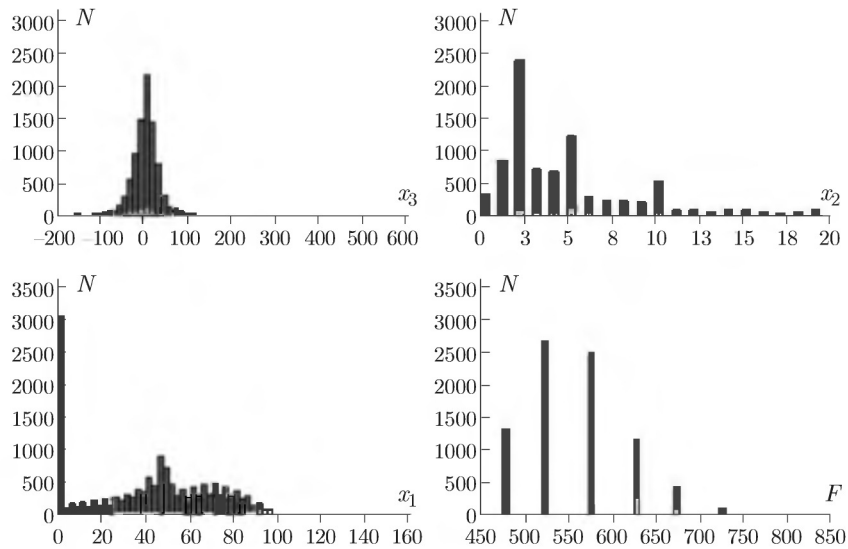


Рис. 5.10. Гистограммы кластера с номером 4. Чёрным цветом показаны гистограммы для всего исследуемого региона, а серым — гистограммы этих же характеристик для прогнозной зоны со значениями  $F \geq 6,0$

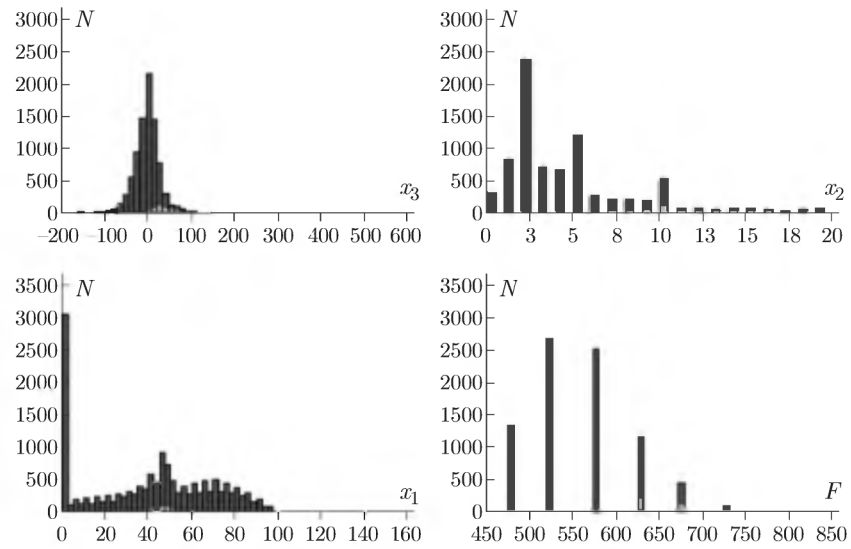


Рис. 5.11. Гистограммы кластера с номером 5. Чёрным цветом показаны гистограммы для всего исследуемого региона, а серым — гистограммы этих же характеристик для прогнозной зоны со значениями  $F \geq 6,0$

## Глава 6

# ПОСТРОЕНИЕ ПРОГНОЗНЫХ КАРТ НЕФТЕГАЗОНОСНОСТИ

### 6.1. Введение

Прогнозные карты нефтегазоносности необходимы для планирования геологоразведочных работ, оптимального размещения поисковых и оценочных скважин, проведения лицензионной политики. В существующих подходах широко используется как методология экспертного оценивания, так и технологии, основанные на методах распознавания образов. Ориентация рассмотренной в гл. 3 технологии пространственного прогноза на предметную область нефтегазового прогноза касается только первых двух этапов решения: этапа формализации задачи и этапа формализации информационного пространства задачи.

Формирование причинно-следственной модели в задачах нефтегазового прогноза имеет целью качественное описание условий генерации углеводородов на исследуемой территории, условий их миграции, аккумуляции и консервации. Модель формализуется в виде логических высказываний, устанавливающих связь между наличием месторождения и качеством нефтематеринских толщ, возможными путями миграции, наличием структурных и неструктурных ловушек, ёмкостными свойствами коллекторов и надёжностью покрышек. Модель используется для выбора исходных геолого-геофизических и геохимических данных, для выбора методов преобразования исходных данных в признаки, при нахождении прогнозирующей функции и для объяснения и обоснования результатов прогноза.

Формализация объектов прогноза (провинций, месторождений, мест заложения скважин) и признаков прогноза использует принципы, которые были сформулированы в гл. 5 при предметной ориентации общей технологии решения задач пространственного прогноза на задачи построения прогнозных карт максимальных магнитуд землетрясений. Объект прогноза заменяется прогнозируемой величиной, которая обычно рассматривается как оценка степени принадлежности точки карты к прогнозируемому объекту (в исследуемом нефтегазоносном комплексе) или соответствует продуктивности скважины. Формализация описания с помощью признаков состоит в выборе таких геолого-геофизических и геохимических признаков, которые должны описывать как локальные особенности нефтепроизводящих, коллекторских и флюидоизолирующих свойств геологической среды в окрестности данной точки, так и геометрию и другие особенности ловушек, к которым принадлежит эта точка.

Формализация гипотез о правиле прогноза основана на общих соображениях о нелинейности зависимости наличия месторождения от признаков. Знания предметной области относятся к характеру связи возможности наличия месторождения с каждым признаком в отдельности и формализуются в виде системы математических ограничений на класс прогнозирующих функций.

В задачах нефтегазового прогноза используются два метода формирования выборки: учёт прямых измерений по результатам бурения на нефть и газ и использование информации об исследованных участках территории. Во втором случае на исследованных участках порождается случайная выборка точек, для которых определяется наличие или отсутствие месторождения.

Предметная ориентация информационной технологии пространственного прогноза на нефтегазовую предметную область отрабатывалась на нескольких разномасштабных задачах:

- Надрегиональный прогноз нефтегазоносности СССР по признакам глубинного строения <sup>1)</sup>, таким как строение консолидированной части земной коры, раздел Мохо, глубинные разломы, толщины крупных комплексов осадочного чехла и т. д. Полученная по этим признакам карта хорошо совпадает с картой нефтегазоносности, выполненной под руководством В. В. Семеновича на основе традиционных геологических данных.

- Надрегиональный отдельный прогноз на нефть и газ в пределах Западной Сибири по данным о глубинном строении, таким как раздел Мохо, структурные карты кровли, мощности и скоростных неоднородностей нижнего, промежуточного и верхнего слоёв консолидированной коры, структурные карты подошвы верхнемеловых и среднеюрских отложений, магнитные аномалии и разломы, ограничивающие и смещающие грабены (*Gitis et al., 1994*). Ошибка прогноза на тестовой выборке составила около 3% для прогноза газовых месторождений и около 17% для прогноза нефтяных месторождений.

- Региональный прогноз нефтегазоносности структур, установленных по сейсмическим данным на юге Оренбургской области в масштабе 1 : 200 000 и 1 : 50 000 и прогноз нефтегазоносности участка Паннонской впадины по комплексу электроразведочных и сейсмических данных в масштабе 1 : 50 000 <sup>2)</sup>.

- Локальный прогноз в масштабе 1 : 10 000 с выявлением высокодебитных зон в пределах одного из месторождений северо-восточного Китая по данным бурения 56 скважин и прогноз мест заложения продуктивных поисковых скважин для небольшой поисковой площади в Удмуртии <sup>3)</sup>.

<sup>1)</sup> Работа выполнена совместно с ВНИИ Геофизика МПР РФ.

<sup>2)</sup> Работа выполнена совместно с ЗАО «НПЦ «Геонефтегаз».

<sup>3)</sup> Обе работы выполнены совместно с ЗАО «НПЦ «Геонефтегаз».

В данной главе разработанная технология исследуется на трёх примерах построения прогнозных карт нефтегазоносности: надрегionalный раздельный прогноз областей нефте- и газоносности в пределах Западно-Сибирской платформы — раздел 6.2., прогноз нефтегазовых месторождений по комплексу электроразведочных и сейсмических данных (район Паннонской депрессии, Венгрия) — раздел 6.3. В разделе 6.4 рассматривается технология локального прогноза концентраций углеводородов на месторождении в северо-восточном Китае.

## **6.2. Надрегionalный раздельный прогноз нефтегазоносности в пределах Западно-Сибирской платформы**

**6.2.1. Причинно-следственная модель.** Основная особенность данного примера состоит в использовании нетрадиционного для задач нефтегазового прогноза набора признаков. В основном это геофизические признаки, которые описывают глубинные свойства земной коры, полученные в центре ГЕОН в результате интерпретации данных по глубинному зондированию Земли с помощью мощных взрывов.

При прогнозе нефтегазовых провинций Западно-Сибирской платформы была предпринята попытка раздельного прогноза газоносности и нефтеносности. Известно, что нефтегазоносные провинции или их части различаются по преобладающему фазовому состоянию скоплений углеводородов. Прогноз фазового состояния может внести коррективы в стратегию поисковых работ и оценку предполагаемых финансовых затрат.

Различия в фазовом состоянии углеводородов согласно современным представлениям (Конторович и др. 1981, Никитин, 1986) определяются составом органического вещества и уровнем его катагенетической преобразованности, который главным образом зависит от величины термального воздействия. При относительно низких (до 50 °С) и высоких температурах (более 150 °С) из органического вещества преимущественно образуется газ, при средних температурах образуется нефть. Некоторые исследователи считают дегазацию мантии одним из факторов, способствующих формированию газа.

В данной задаче принята модель, согласно которой фазовое состояние углеводородов определяется историей тектонического развития, выраженной в особенностях строения земной коры и верхней мантии.

Для газовых месторождений характерно ухудшенное качество исходного органического вещества, обусловленное увеличенной скоростью осадконакопления, привнесом большого количества терригенного материала и формированием мощных (свыше 10 км) осадочных толщ. Мощное осадконакопление вызывает уплотнение пород фундамента, что приводит к увеличению их теплопроводности и увеличивает разогрев нижней части осадочной толщи. Увеличение плотности пород

фундамента проявляется в увеличении скорости продольных сейсмических волн на поверхности фундамента. Высокая скорость прогибания сопровождается утонением земной коры за счёт уменьшения толщины верхнего слоя её консолидированной части. Утонённая земная кора в меньшей степени препятствует дегазации мантии.

Для нефтяных месторождений показательно высокое качество органического вещества, накапливающегося при средних скоростях осадконакопления, формирующих осадочную толщу сравнительно небольшой (до 5 км) мощности. Наличие относительно толстой консолидированной коры с хорошо развитым гранитным слоем способствует умеренному разогреву нижней части осадочного чехла и уменьшает влияние дегазации мантии. Бесперспективными относительно нефтяных и газовых месторождений являются части провинций с маломощным (до 1,5 км) осадочным чехлом.

Предложенная модель применена для прогноза нефтегазоносности Западно-Сибирской провинции. В её пределах известны 11 нефтегазоносных областей (Конторович и др., 1981). В 5 из них, расположенных в северной части провинции, преобладают газовые и газоконденсатные месторождения. В 6 остальных, охватывающих центральную часть провинции, преимущественным развитием пользуются нефтяные месторождения. Эта информация привлекалась только при анализе качества решения.

**6.2.2. Исходные данные.** При решении задачи использовались данные центра ГЕОН, ЗапСибНИГНИ, а также опубликованные материалы, представленные в виде карт в масштабе 1 : 2 500 000. Рассматривался следующий набор признаков:

- $x_1$  — рельеф поверхности Мохоровичича,
- $x_2$  — рельеф поверхности нижнего слоя консолидированной коры,
- $x_3$  — рельеф поверхности промежуточного слоя консолидированной коры,
- $x_4$  — рельеф поверхности кристаллического фундамента,
- $x_5$  — рельеф поверхности складчатого фундамента,
- $x_6$  — толщина нижнего слоя консолидированной коры,
- $x_7$  — толщина промежуточного слоя консолидированной коры,
- $x_8$  — толщина верхнего слоя консолидированной коры,
- $x_9$  — толщина консолидированной части коры,
- $x_{10}$  — скорость продольных сейсмических волн на поверхности Мохоровичича,
- $x_{11}$  — скорость продольных сейсмических волн на поверхности кристаллического фундамента,
- $x_{12}$  — скорость продольных сейсмических волн на поверхности складчатого фундамента,
- $x_{13}$  — глубина залегания кровли верхнемеловых отложений,
- $x_{14}$  — глубина залегания кровли сеноманских отложений,
- $x_{15}$  — глубина залегания кровли среднеюрских отложений,

$x_{16}$  — глубина залегания подошвы плитного комплекса

$x_{17}$  — аномалии магнитного поля ( $\Delta T_a$ )

$x_{18}$  — разломы, ограничивающие грабен-рифтовые структуры.

Сеточные модели построены в сетке 30' по долготе и 20' по широте.

**6.2.3. Выборка.** Выборка формировалась по данным бурения на 1985 г. В неё вошли 115 пунктов с месторождениями газа, 167 пунктов с месторождениями нефти и 206 пунктов, в которых месторождения при бурении не были обнаружены.

**6.2.4. Прогнозирующая функция.** Для прогноза используется метод аппроксимации отношения правдоподобия, рассмотренный в разделе 4.4. Регион разбивается на непересекающиеся ячейки. Каждой ячейке соответствует вектор признаков  $\mathbf{x} = (x_1, x_2, \dots, x_I)$  компоненты которого являются значениями геологических или геофизических полей. Считается, что произошло событие  $C$  или  $\bar{C}$  в зависимости от того, обнаружено или не обнаружено в ячейке месторождение рассматриваемого вида. Задача обнаружения месторождения рассматривается как статистическая задача прогноза наступления события  $C$  по комплексу данных  $\mathbf{x}$ . Предполагается, что событие  $C$  и вектор  $\mathbf{x}$  имеют вероятностную природу. Логарифм отношения правдоподобия условных распределений  $P(\mathbf{x}/C)$  и  $P(\mathbf{x}/\bar{C})$

$$y(\mathbf{x}) = \ln \frac{P(\mathbf{x}/C)}{P(\mathbf{x}/\bar{C})}$$

аппроксимируется выражением вида

$$F(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^I \varphi_i(x_i, \boldsymbol{\theta})$$

где  $\varphi_i(x_i, \boldsymbol{\theta})$  — кусочно-линейная функция признака  $x_i$ ,  $\boldsymbol{\theta}$  — вектор оцениваемых параметров,  $I$  — число информативных признаков.

Вектор параметров  $\boldsymbol{\theta}$  оценивается методом максимального правдоподобия по выборке  $N$  независимых наблюдений  $\{\nu_n, \mathbf{x}_n\}_1^N$  случайных величин  $\nu$  и  $\mathbf{x}$ , где  $\nu = 1$ , когда имеет место событие  $C$ , и  $\nu = 0$  в противном случае:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \sum_{n=1}^N \ln P_{\vartheta}(\nu_n / \mathbf{x}_n),$$

где

$$P_{\vartheta}(\nu / \mathbf{x}) = [P_{\vartheta}(C / \mathbf{x})]^{\nu} [1 - P_{\vartheta}(C / \mathbf{x})]^{1-\nu},$$

$$P_{\vartheta}(C / \mathbf{x}) = \left(1 + \frac{1 - \pi(C)}{\pi(C)} \exp(-F(\mathbf{x}, \boldsymbol{\theta}))\right)^{-1},$$

$\pi(C)$  — априорная вероятность события  $C$ , которая оценивается как доля событий  $C$  в выборке наблюдений, либо задаётся на основании экспертной оценки доли территории, где наблюдалось событие  $C$ .

Подмножество наиболее информативных признаков находится неполным перебором, аналогично методу шаговой регрессии.

**6.2.5. Информационная модель прогноза газовых месторождений.** Для прогноза газовых месторождений привлекалась выборка прецедентов из двух классов: к первому классу были отнесены 115 пунктов с газовыми месторождениями, а ко второму — все остальные 373 пункта, из них 167 пунктов относились к нефтяным месторождениям, а 206 пунктов — к местам, где ни газ и ни нефть не были обнаружены.

Вся выборка примеров была разбита случайным выбором на учебную и тестовую в соотношении 2 к 1.

Кроме перечисленных выше 18 признаков в решении участвовали признаки, полученные средствами ГИС ГЕО 2.5, такие как модули градиентов геофизических растровых полей, сглаженные поля, геологически оправданные логические и алгебраические функции от нескольких растровых полей и другие. Прогнозная функция использует два наиболее информативных признака:

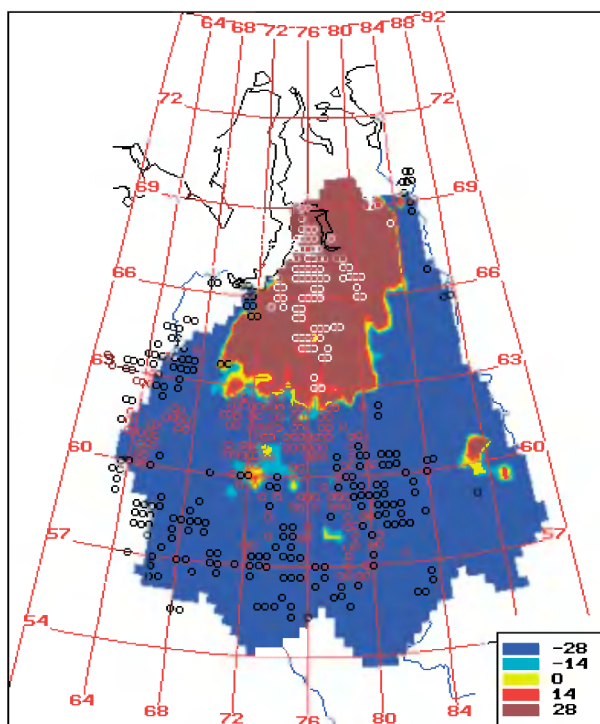


Рис. 6.1. Прогнозная карта газовых месторождений. Жёлтые и коричневые тона соответствуют газовым месторождениям, синие — отсутствию газовых месторождений. Кружками показаны пункты выборки прецедентов: белый цвет — газ, красный — нефть, чёрный — отсутствие углеводородов



- глубину залегания кровли среднеюрских отложений,  $x_{15}$ ,
- полусумму скоростей продольных сейсмических волн на поверхности кристаллического ( $x_{11}$ ) и складчатого ( $x_{12}$ ) фундаментов,  $x_{19}=(x_{11} + x_{12})/2$ .

Прогнозирующая функция имеет вид:

$$F^{(1)}(x_{15}, x_{19}) = \varphi_{15}^{(1)}(x_{15}) + \varphi_{19}^{(1)}(x_{19}),$$

где:

$$\varphi_{15}^{(1)}(x_{15}) = \begin{cases} 30 & \text{при } x_{15} < -4,3, \\ 30 - 28(x_{15} + 4,3) & \text{при } -4,3 \leq x_{15} < -2,8, \\ -12 - 1,0(x_{15} + 2,8) & \text{при } -2,8 \leq x_{15} < -0,8, \\ -10 & \text{при } -0,8 \leq x_{15} < 0; \end{cases}$$

$$\varphi_{19}^{(1)}(x_{19}) = \begin{cases} -9 & \text{при } x_{19} < 5,6, \\ -9 + 15(x_{19} - 5,6) & \text{при } 5,6 \leq x_{19} < 6,2, \\ 30(x_{19} - 6,2) & \text{при } 6,2 \leq x_{19} < 6,6, \\ 12 & \text{при } 6,6 \leq x_{19} < 7. \end{cases}$$

Зависимости  $\varphi_{15}^{(1)}(x_{15})$  и  $\varphi_{19}^{(1)}(x_{19})$  не противоречат причинно-следственной модели. Они свидетельствуют о том, что увеличение глубины залегания кровли среднеюрских отложений ( $x_{15}$ ), отвечающее возрастанию общей мощности плитного комплекса, и увеличение скорости на поверхности складчатого и кристаллического фундамента ( $x_{19}$ ) повышает вероятность обнаружения газовых месторождений. Точность прогноза газовых месторождений на тестовой выборке составила 3%.

Прогнозное поле показано на рис. 6.1. Из рисунка видно, что высокие значения прогноза образуют компактный район, расположенный в северной части Западно-Сибирской нефтегазоносной провинции. Этот район совпадает с известными преимущественно газоносными областями, отвечает Пур-Тазовской зоне интенсивного прогибания земной коры, характеризуется максимальной мощностью осадочного чехла (до 14 км), сокращённой мощностью верхнего слоя консолидированной коры и большими значениями продольных сейсмических волн на поверхности фундамента, что резко отличает его от остальной территории рассматриваемой провинции.

**6.2.6. Информационная модель прогноза нефтяных месторождений.** Для построения прогнозной карты нефтяных месторождений выборка была разбита на 2 класса следующим образом: к первому классу были отнесены 167 пунктов с нефтяными месторождениями, а ко второму — 321 пункт её оставшейся части, из них 115 точек

относились к газовым месторождениям, а 206 пунктов — к местам, где ни газ и ни нефть не были обнаружены.

Вся выборка была разбита случайным выбором на учебную и тестовую в соотношении 2 к 1.

Наиболее информативными признаками оказались:

- глубина залегания кровли среднеюрских отложений ( $x_{15}$ ),
- глубина залегания кровли верхнемеловых отложений ( $x_{13}$ ),
- толщина верхнего слоя консолидированной коры ( $x_8$ ).

Прогнозирующая функция для прогноза нефтяных месторождений выглядит следующим образом:

$$F^{(1)}(x_{15}, x_{13}, x_8) = \varphi_{15}^{(2)}(x_{15}) + \varphi_{13}^{(2)}(x_{13}) + \varphi_8^{(2)}(x_8),$$

где:

$$\varphi_{15}^{(2)}(x_{15}) = \begin{cases} -8 & \text{при } x_{15} < -4,2, \\ -8 + 8,33(x_{15} + 4,2) & \text{при } -4,2 \leq x_{15} < -3, \\ 2 & \text{при } -3 \leq x_{15}; \end{cases}$$

$$\varphi_{13}^{(2)}(x_{13}) = \begin{cases} 1 & \text{при } x_{13} < -0,8, \\ 1 - 23,33(x_{13} + 0,8) & \text{при } -0,8 \leq x_{13} < -0,5, \\ -6 & \text{при } -3 \leq x_{13}; \end{cases}$$

$$\varphi_8^{(2)}(x_8) = \begin{cases} 1 & \text{при } x_8 < 5, \\ 1 + 2,14(x_8 - 5) & \text{при } 5 \leq x_8 < 12, \\ 2,5 + 0,375(x_8 - 12) & \text{при } 12 \leq x_8 < 16, \\ 4 & \text{при } 16 \leq x_8. \end{cases}$$

Заметим, что поведение функции  $\varphi_{15}^{(2)}(x_{15})$  существенно отличается от поведения функции этого же признака в решении предыдущей задачи. В данном случае эта функция выделяет благоприятную для нефтяных месторождений область, в пределах которой кровля среднеюрских отложений залегает на глубине около 2,5 км. Вид функций двух других признаков показывает, что вероятность обнаружения нефтяных месторождений возрастает при увеличении глубины залегания верхнего мела и при увеличении мощности верхнего слоя консолидированной коры.

Прогнозное поле показано на рис. 6.2, из которого видно, что высокие значения отношения правдоподобия приурочены к среднему Приобью, где сосредоточены крупнейшие нефтяные месторождения. Этот район характеризуется мощностью чехла около 4,5 км, дифференцированными значениями скоростей продольных сейсмических волн в кровле фундамента и хорошо развитым верхним слоем консолидированной коры.

### 6.3. Прогноз нефтегазовых месторождений по комплексу электроразведочных и сейсмических данных (Паннонская депрессия, Венгрия)

**6.3.1. Исходные данные.** Для решения задачи использовались следующие исходные данные:

- структурная карта поверхности кайнозойского фундамента для восточной части изучаемой территории, полученная по данным сейсморазведки,
- структурная карта поверхности паннонского яруса для восточной части изучаемой территории, полученная по данным сейсморазведки,
- растровое поле гравитационных аномалий в редукции Буге в сетке  $100 \times 100$  м,
- амплитуды и фазы результатов электромагнитных измерений полученных с помощью станции VEGA: измерения проводились в 134 пунктах, излучатели имели 8 конфигураций, работы проводились на 7 частотах, каждый раз регистрировалось 2 амплитуды

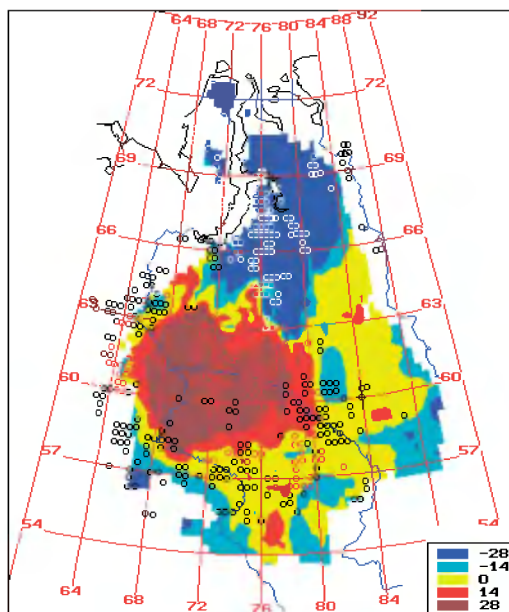


Рис. 6.2. Прогнозная карта нефтяных месторождений. Жёлтые и коричневые тона соответствуют газовым месторождениям, синие — отсутствию газовых месторождений. Кружками показаны пункты выборки прецедентов: белый цвет — газ, красный — нефть, чёрный — отсутствие углеводородов

ды, две фазы напряжённости электрического поля и 3 компоненты (вертикальная и две горизонтальных) магнитного поля,

- данные бурения о наличии или отсутствии нефтегазовых месторождений в 34 точках региона (19 продуктивных и 15 непродуктивных скважин).

**6.3.2. Информационная модель прогноза.** С помощью ГИС ГЕО 2.5 исходные данные были векторизованы, преобразованы в растровый формат и по ним были вычислены растровые признаки, такие как мощность слоя между кровлей кайнозойского фундамента и кровлей паннона, локальные аномалии (антиклинального и синклинального типа) поверхности кровли фундамента и паннона, модули градиентов структурных поверхностей и аномалий Буге. Объём электромагнитных исходных измерений в данной задаче составил 392 поля. Для сокращения перебора электромагнитные измерения были проанализированы и обработаны до вычисления растровых полей. В соответствии с предложенной экспертом причинно-следственной моделью по электромагнитным данным были вычислены растровые поля усреднённой амплитуды и девиации фазы напряжённости электрического поля по двум ортогональным направлениям для двух типов излучателей и трёх нижних и трёх верхних частот зондирования. Компоненты магнитного поля в решении не анализировались.

Для прогноза территория была разбита на восточную и западную области в соответствии с наличием структурных данных о геологическом строении. Для каждой области была найдена своя прогнозирующая функция. При этом для восточной части использовались как гравитационные аномалии и электромагнитные измерения, так и геологические структурные карты, а для западной части — только гравитационные и электромагнитные данные. Выборка прецедентов также оказалась разбитой на две части: 10 продуктивных и 3 непродуктивных скважин оказалось в восточной части карты и 9 продуктивных и 12 непродуктивных скважин — в западной части.

Для восточной области наиболее информативными оказались 2 признака: девиация фазы напряжённости электромагнитного поля в широтном направлении для излучателя типа диполь ( $x_1$ ) и локальные поднятия в кровле кристаллического фундамента ( $x_2$ ). Прогнозирующая функция имеет вид:

$$F^{(1)}(\mathbf{x}) = -290,05 + \varphi_1^{(1)}(x_1) + \varphi_2^{(1)}(x_2),$$

где графики функций  $\varphi_1^{(1)}(x_1)$  и  $\varphi_2^{(1)}(x_2)$  показаны на рис. 6.3.

Для западной области наиболее информативными оказались 3 признака:  $y_1$  — девиация фазы напряжённости электромагнитного поля в направлении долготы для излучателя типа пара разнесённых диполей,  $y_2$  — усреднённая по трём нижним частотам амплитуда для излучателя того же типа и  $y_3$  — девиация фазы напряжённости элек-

ромагнитного поля в широтном направлении для излучателя типа диполь. Прогнозирующая функция имеет вид:

$$F^{(2)}(\mathbf{y}) = -283,35 + \varphi_1^{(2)}(y_1) + \varphi_2^{(2)}(y_2) + \varphi_3^{(2)}(y_3),$$

где графики функций  $\varphi_1^{(2)}(y_1)$ ,  $\varphi_2^{(2)}(y_2)$  и  $\varphi_3^{(2)}(y_3)$  показаны на рис. 6.4. Решающее правило, составленное из функций  $F^{(1)}(\mathbf{x})$  для восточной области и  $F^{(2)}(\mathbf{y})$  для западной области, даёт безошибочный прогноз на имеющейся выборке.

Прогнозная карта нефтегазоносности приведена на рис. 6.5, А. Красные тона соответствуют продуктивным областям, голубые — непродуктивным, степень уверенности в прогнозе возрастает по мере увеличения насыщенности цветов. На рисунке серым цветом показана граница между западной и восточной областями. Кружками показаны пункты выборки: большие кружки — продуктивные скважины, маленькие — непродуктивные. Из-за малого объёма выборки точность прогноза неодинакова для разных точек исследуемой территории. Можно полагать, что более точный прогноз даётся для точек территории, которые наиболее близки по комплексу признаков к точкам учебной выборки. В таком случае точность прогноза можно оценивать с помощью нормированного евклидова расстояния от точек растра до множества точек выборки (нормировка производится по среднеквадратичным отклонениям признаков). На рис. 6.5, Б приведена прогнозная карта для тех точек, у которых указанное расстояние меньше 0,5.

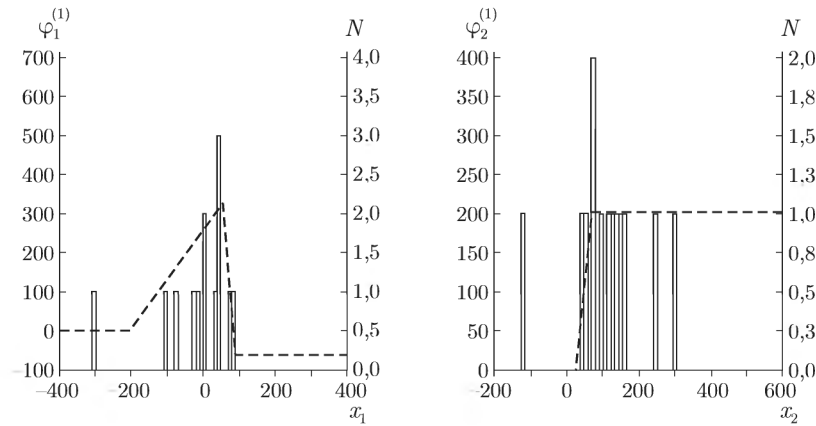


Рис. 6.3. Графики прогнозирующей функции для  $(\varphi_1^{(1)}, \varphi_2^{(1)})$  восточной области и гистограммы точек ( $N$ ) учебной выборки.  $x_1$  — признак девиация фазы напряжённости электромагнитного поля в широтном направлении для излучателя типа диполь,  $x_2$  — признак наличия локального поднятия в кристаллическом фундаменте

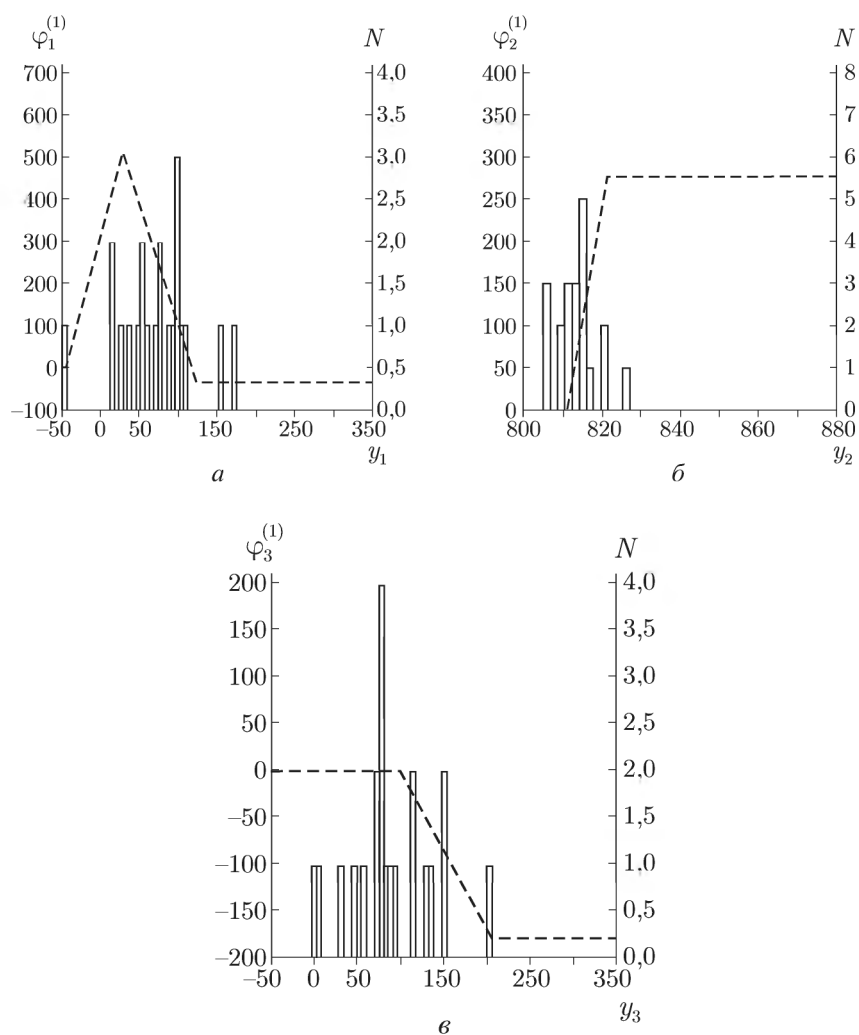


Рис. 6.4. Графики прогнозирующей функции ( $\varphi_1^{(2)}, \varphi_2^{(2)}, \varphi_3^{(2)}$ ) для западной области и гистограммы точек ( $N$ ) учебной выборки.  $y_1$  — признак девиация фазы напряжённости электромагнитного поля в направлении долготы для излучателя типа пара разнесённых диполей,  $y_2$  — признак усреднённая по трём нижним частотам амплитуда для излучателя того же типа и  $y_3$  — признак девиация фазы напряжённости электромагнитного поля в широтном направлении для излучателя типа диполь

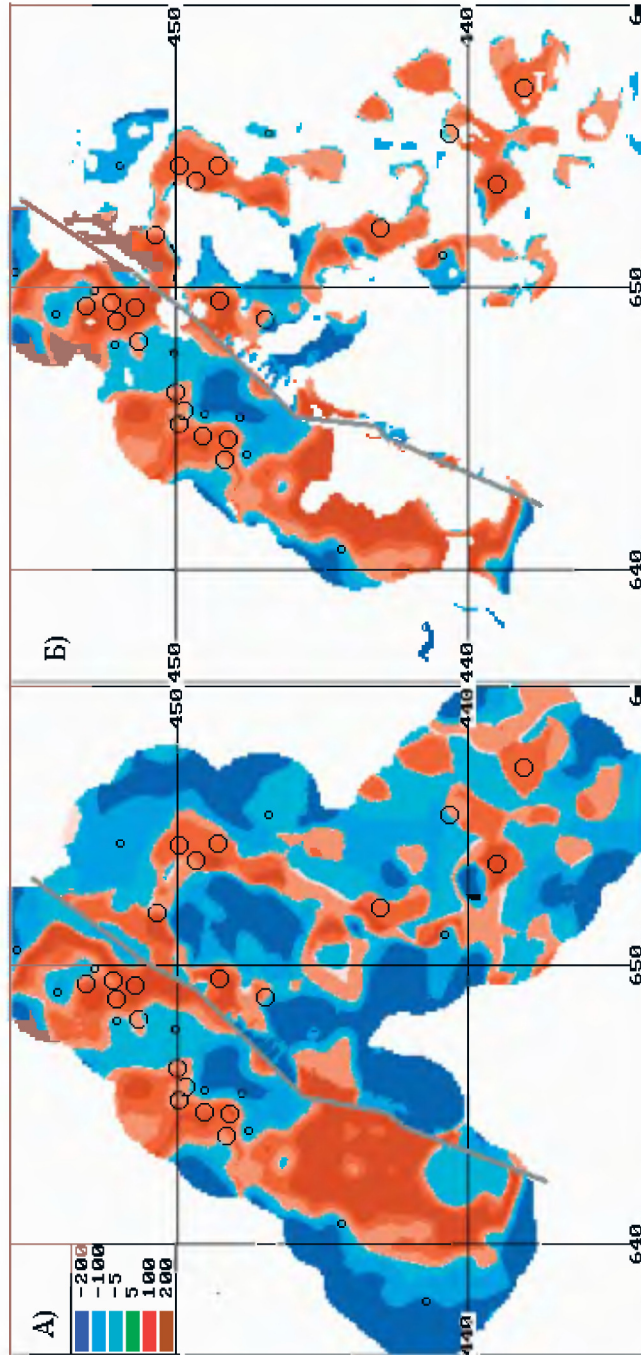


Рис. 6.5. Прогноз нефтегазовых месторождений. А — полная карта прогноза, Б — часть исследуемой области с наиболее точными значениями прогноза. Обозначения: синие тона — неперспективные области, красные тона — перспективные области, насыщенность цвета соответствует степени уверенности прогноза. Координатная сетка — в километрах, размер элемента раstra —  $100 \times 100$  м, большие кружки — продуктивные скважины, малые — сухие

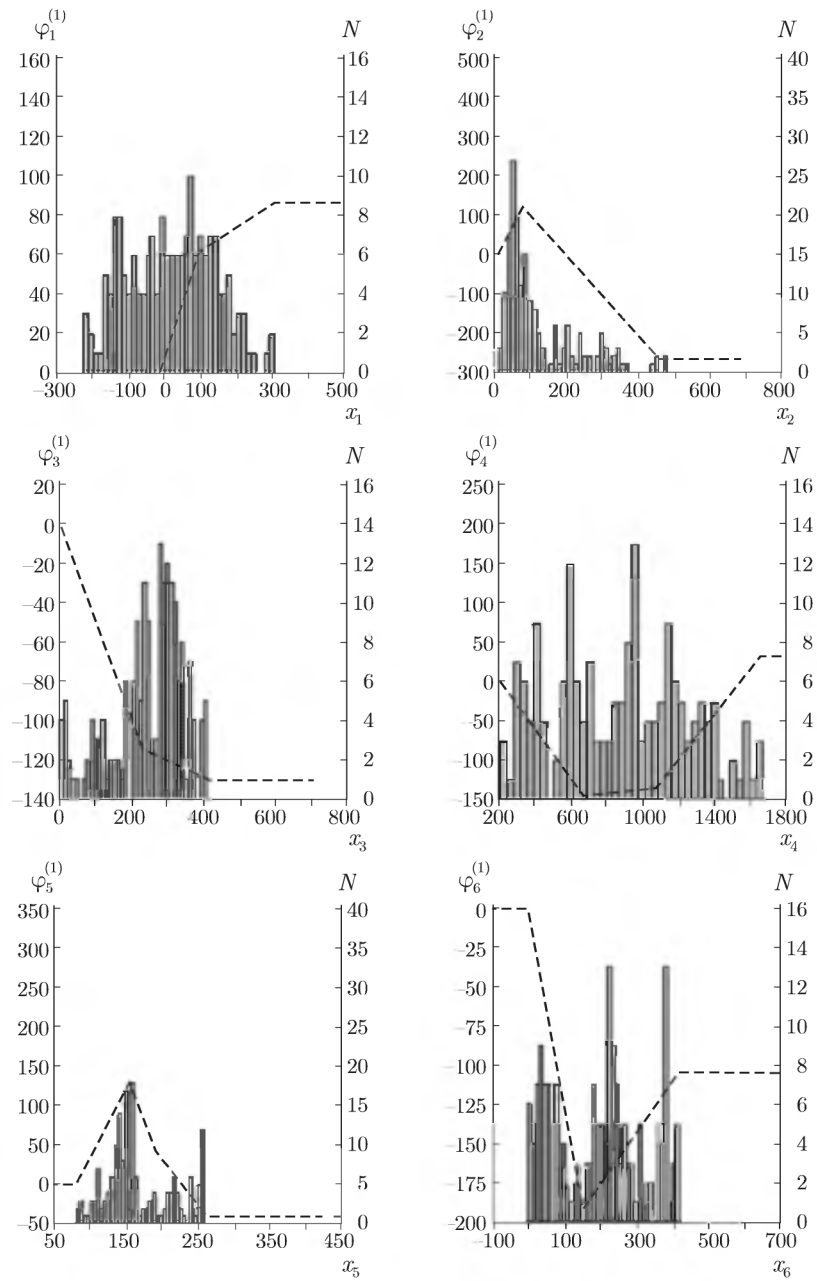


Рис. 6.6. Графики прогнозной функции ( $\varphi_i$ ) и гистограммы точек ( $N$ ) выборки по отобранным признакам



## 6.4. Прогноз концентраций углеводородов в пределах месторождения в Восточном Китае

**6.4.1. Причинно-следственная модель.** Исследуемое месторождение углеводородов Восточного Китая характеризуется сложной тектонической структурой, широким развитием надвигов и крутопадающих разломов. Форма залежи не установлена. Скорее всего единой залежи не существует, а углеводороды сосредоточены в зонах трещиноватости или в отдельных структурных ловушках. Задача прогноза состоит в выявлении наиболее перспективных участков для постановки бурения.

**6.4.2. Исходные данные.** Одна из особенностей задачи состоит в том, что исходные данные представлены каталогом результатов бурения 56 скважин, распределённых неравномерно на участке  $7,5 \times 9,5$  км.

В каталоге содержатся данные о глубинах залегания кровли 9 горизонтов, обозначенных  $L_1-L_9$ :

- $L_1$  — кровля 3-го слоя среднего палеогена,
- $L_2$  — кровля 4-го слоя среднего палеогена,
- $L_3$  — кровля нижнего мела,
- $L_4$  — кровля первого слоя средне-верхнеюрской толщи,
- $L_5$  — кровля второго слоя средне-верхнеюрской толщи,
- $L_6$  — кровля третьего слоя средне-верхнеюрской толщи,
- $L_7$  — кровля четвёртого слоя средне-верхнеюрской толщи,
- $L_8$  — кровля нижнего палеозоя,
- $L_9$  — кровля архея,

а также данные о количестве эффективных нефтегазоносных слоёв, общая мощность эффективных слоёв, возраст пород, выходящих на поверхность нижнего палеозоя, проявление карста, продуктивность скважин, положение залежей относительно поверхности нижнего палеозоя. Два последних параметра при решении задачи были закодированы следующим образом: сухие скважины — 0, малопродуктивные внутри палеозоя — 1, малопродуктивные на поверхности палеозоя — 2, высокопродуктивные внутри палеозоя — 3, высокопродуктивные на поверхности палеозоя — 4.

Данные каталога были интерполированы в растровые поля с шагом  $100 \times 100$  м для горизонтов  $L_1-L_9$  (структурные карты), а также поля распределения эффективных слоёв и их суммарной мощности. По этим полям найдены производные параметры: поле мощности 8-ми горизонтов ( $L_1-L_2$ ,  $L_2-L_3$  и т. д.), слагающих разрез месторождения, градиентов структурных карт и лапласиана горизонта  $L_8$  (кровля основного нефтесодержащего горизонта) и др. Большая часть скважин сконцентрирована на площадке  $4 \times 8$  км. Эта площадка выбрана для прогноза, чтобы уменьшить краевые эффекты интерполяции растровых полей.

**6.4.3. Информационная модель прогноза.** Для решения прогнозной задачи использовались данные о продуктивности пород внутри нижнепалеозойской толщи. Низкопродуктивные и высокопродуктивные 15 скважин составили один класс, а 25 сухих скважин были включены в альтернативный класс. Предполагалось, что продуктивность пород одинакова в радиусе 250 м от скважин. Исходя из этого, по данным бурения было построено растровое поле, на котором выделены продуктивные и непродуктивные области. По этому полю сгенерирована выборка из 400 точек, которые были отобраны из соответствующих областей случайным образом. Продуктивным скважинам присвоено значение 100, сухим — значение -100. Для нахождения прогнозирую-

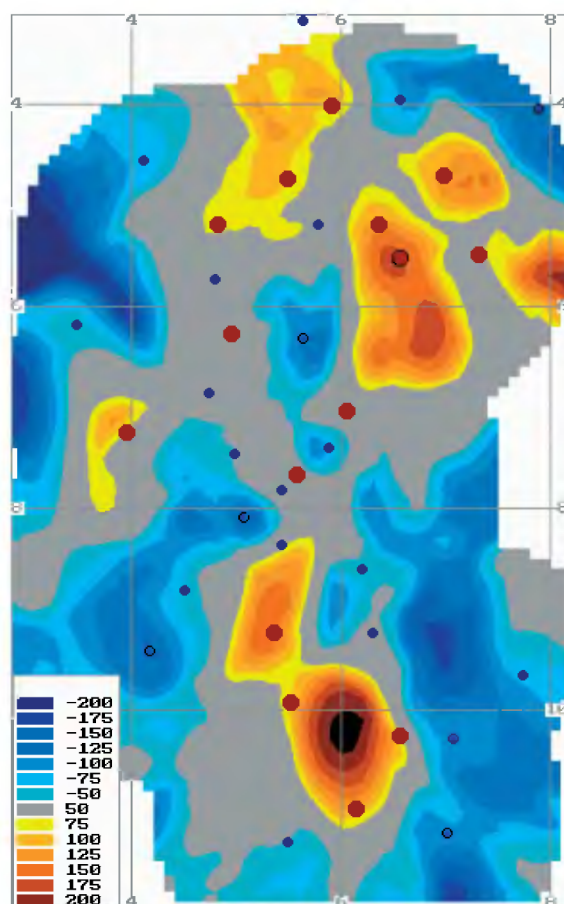


Рис. 6.7. Карта прогнозного поля. Большие кружки — продуктивные скважины, малые кружки — непродуктивные

шей функции использовался предложенный в гл. 3 метод интервальных экспертных оценок. Прогнозирующая функция имеет вид:

$$F(\mathbf{x}) = 184 + \sum_{i=1}^6 \varphi_i(x_i),$$

где  $x_1$  — лапласиан 8-го горизонта ( $L_8$ ),  $x_2$  — мощность 2-го горизонта (разность высот горизонтов  $L_2$  и  $L_3$ ),  $x_3$  — мощность 5-го горизонта (разность высот горизонтов  $L_5$  и  $L_6$ ),  $x_4$  — мощность 8-го горизонта (разность высот горизонтов  $L_8$  и  $L_9$ ),  $x_5$  — суммарная мощность эффективных слоёв,  $x_6$  — мощность 6-го горизонта (разность высот горизонтов  $L_6$  и  $L_7$ ).

Прогнозные функции и гистограммы выборки по признакам изображены на рис. 6.6.

Прогнозное поле приведено на рис. 6.7. Уверенность прогноза перспективных (неперспективных) областей увеличивается с увеличением (уменьшением) значения прогнозного поля. Средняя ошибка аппроксимации по выборке составляет +50, из этого следует, что на прогнозной карте значения от -50 до +50 интерпретируются как область неопределённости. Кружками большого радиуса обозначено расположение продуктивных скважин, малого радиуса — сухих.

## Глава 7

# ПОСТРОЕНИЕ КАРТ МИНЕРАГЕНИЧЕСКОГО ПРОГНОЗА

### 7.1. Введение

Карты минерагенического прогноза необходимы при лицензировании участков недр и при выборе направлений геологоразведочных работ. Формализация задачи минерагенического прогноза производится в соответствии с принципами, сформулированными в гл. 3. Причинно-следственная модель прогноза, как и в предметной области нефтегазового прогноза, описывает условия генерации высоких концентраций полезного компонента на исследуемой территории. Объекты прогноза (районы, зоны, месторождения) заменяются прогнозируемой величиной, рассматриваемой как степень принадлежности точки карты к прогнозируемому объекту. Признаки точки выбираются так, чтобы характеризовать как локальные особенности вещественного состава, геодинамики и структуры геологической среды в окрестности данной точки, так и особенности региональной структуры, к которой принадлежит данная точка.

Зависимость степени уверенности в принадлежности точки региона к объекту прогноза от признаков прогноза предполагается нелинейной. Знание предметной области позволяет указать характер связи прогнозируемой величины с каждым признаком в отдельности. Это знание формализуется в виде системы математических ограничений на класс прогнозирующих функций.

В задачах минерагенического прогноза, как и в задачах прогноза нефтегазоносности, выборка примеров для обучения и контроля формируется на основе прямых свидетельств наличия месторождения и по информации об исследованных участках территории.

Предметная ориентация информационной технологии пространственного прогноза на предметные области минерагенического прогноза отрабатывалась на решении задач свинцово-цинковых и медных месторождений Болгарии и Чехословакии. В разделе 7.2 рассмотрена задача регионального прогноза свинцово-цинковых месторождений Болгарии по геолого-геофизическим признакам, отражающим вещественные особенности, возраст, историю геологического развития, структуру земной коры, разломы и геодинамику (*Гитис и др., 1989б*). В разделе 7.3 рассмотрена задача регионального прогноза медных месторождений Болгарии по тем же геолого-геофизическим признакам (*Гитис и др., 1989в*).

## 7.2. Прогноз свинцово-цинковых месторождений Болгарии

**7.2.1. Причинно-следственная модель.** Большая часть свинцово-цинковых месторождений Болгарии принадлежит к гидротермальному среднетемпературному типу. Наиболее крупными месторождениями являются Мадан, Звездел, Рудозем, Устрем, Руен и др. Модель образования месторождений этого типа основана на работах *Вольфсон и Дружинин, 1975, Шалдун и Добровольская, 1976, Богданов Б. и др. 1978, Косыгин и Кулиш, ред., 1984.*

Модель состоит в следующем. Консолидированная кора толщиной 35–40 км подвергается кайнозойской магматической активизации. Гипабиссальные интрузии среднего состава, представленные серией даек и штоков, обеспечивают мобилизацию гидротермальной деятельности и формирование рудных жильных тел на глубинах 1000–1500 м. Протяжённость оруденения по глубине составляет около 1000 м. Для того чтобы оруденение стало доступным для эксплуатации, район должен испытать тектоническое поднятие и эрозию. Протяжённость вскрытых рудных тел составляет первые километры при ширине до нескольких десятков метров.

Причинно-следственная модель определяется тремя группами признаков. Первая из них охватывает вещественные параметры месторождений и вмещающих пород, а также геофизические признаки, связанные с веществом. Она включает характеристики магнитного поля, состава магматических образований, аномалий силы тяжести в редукции Буге. Вторая — объединяет структурные признаки вмещающей среды: градиенты потенциальных полей, плотность разломов, гетерогенность, глубину поверхности Мохоровичича, мощность гранитно-метаморфического слоя и др. Третью группу составляют геодинамические признаки, определяющие вскрытие месторождений. К ним относятся неотектонические движения, характеристики рельефа, изостатические аномалии силы тяжести, тепловой поток и др.

**7.2.2. Исходные данные.** Региональная информация, связанная с месторождениями полезных ископаемых Болгарии, сосредоточена на картах масштаба 1 : 500 000, 1 : 1 000 000 и 1 : 1 500 000. Информационная нагрузка карт определила выбор шага координатной сетки 12' по широте и 6' по долготе (это приблизительно составляет 11,5 и 16,5 км соответственно). Значения признаков снимались для узлов сетки.

При решении использовались следующие 24 геолого-геофизических признака.

### Группа вещественных признаков среды.

$x_1$  — интенсивность тектономагматической проработки. Для каждого узла сетки определялось количество проявлений разновозрастного магматизма по геологической карте из работы *Атлас, 1983*. Использовалась следующая кодировка: 0 — отсутствие проявлений магматиз-

ма, значения от 1 до 4 — количество тектономагматических циклов. Связь значений признака с прогнозом имеет следующий характер. Окрестности точек с нулевым значением признака слабо перспективны в отношении свинцово-цинковых месторождений. Значения 2–3 высоко перспективны в связи с повторной мобилизацией и возможной концентрацией рудного вещества. Значения 4 менее перспективны, поскольку многократная активизация вещества может приводить к его рассеиванию;

$x_2$  — возраст последних проявлений магматизма по карте *Димитрова, ред., 1978*. Использовалась следующая кодировка: 0 — отсутствие проявлений магматизма, 1 — последние проявления магматизма в докембрии, 2 — в карбоне, 3 — в перми, 4 — в триасе, 5 — в юре, 6 — в мелу, 7 — в палеоцене, 8 — в эоцен-олигоцене, 9 — в миоцене. Перспективность точки региона считается тем выше, чем моложе последние проявления магматизма;

$x_3$  — состав последних проявлений магматизма. Данные снимались с карты *Димитрова, ред., 1978*. Использовалась следующая кодировка: 0 — ультраосновной состав, 1 — отсутствие проявлений, 2 — основной, 3 — кислый, 4 — средний составы. Перспективность точки возрастает при увеличении значения признака;

$x_4$  — региональное магнитное поле в гаммах (*Григоров Б., 1961*). Наиболее перспективны области проявления среднего и кислого магматизма, значение магнитного поля в которых должны быть близким к 0;

$x_5$  — вертикальная составляющая локального магнитного поля в гаммах (*Григоров Б., 1961*). Учитывая состав магматических образований, контролирующих оруденение, наиболее перспективными считаются точки, в окрестности которых имеются небольшие положительные значения данного параметра;

$x_6$  — гравитационное поле в редукции Буге в миллигалах (*Добрев и Щукин, 1974*). Состав рудовмещающих пород позволяет считать наиболее перспективными ячейки со средними значениями поля.

#### **Группа структурных признаков.**

$x_7$  — суммарная длина разломов в км в ячейке  $11,5 \times 16,5 \text{ км}^2$  с центром в узле сетки по геологической карте (*Атлас, 1983*). Наиболее точки сетки со значениями признака 3–20 км. При больших значениях среда обладает повышенной проницаемостью, и рудоносные растворы рассеиваются. Среда с меньшими значениями плотности разломов представляет собой экран для этих растворов;

$x_8$  — гетерогениость. Определялась по геологической карте (*Атлас, 1983*) путём подсчёта количества геологических границ в ячейке  $11,5 \times 16,5 \text{ км}^2$  с центром в узле сетки. Зоны глубинных разломов проявляются в повышенной гетерогенности. Поэтому, чем больше значение признака, тем перспективнее данная точка сетки;

$x_9$  — толщина гранито-метаморфического слоя в км по соответствующей карте (*Добрев и Щукин, 1974*). Наиболее перспективные точки вероятно расположены в областях со средними значениями толщины

слоя, приблизительно равными 10–30 км. При значениях, больших 30 км, мощность слоя затрудняет энергетический обмен с нижними частями земной коры и мантией, необходимый для мобилизации рудного вещества и формирования интрузий среднего — кислого состава. При значениях мощности слоя менее 10 км сокращается объём активизируемого вещества;

$x_{10}$  — глубина залегания поверхности Мохоровичича в км по карте *Димитрова, ред., 1978*. Глубины залегания этой поверхности, меньшие 35 км, могут быть связаны с процессами базальтификации коры и активизации основного магматизма. Значения 35–40 км наиболее благоприятны для месторождений свинца и цинка. При залегании поверхности Мохоровичича глубже 40 км затрудняется вещественно-термальный обмен с мантией;

$x_{11}$  — модуль градиента регионального магнитного поля;

$x_{12}$  — модуль градиента вертикальной составляющей локального магнитного поля;

$x_{13}$  — модуль градиента гравитационного поля в редукции Буге;

$x_{14}$  — модуль градиента поля изостатических аномалий;

$x_{15}$  — модуль градиента толщины гранито-метаморфического слоя;

$x_{16}$  — модуль градиента неотектонических движений;

$x_{17}$  — модуль градиента глубины залегания поверхности Мохоровичича.

Модули градиента  $x_{11}$ – $x_{17}$  вычислены средствами ГИС ГЕО 2.5. Увеличение значений модуля характеризует увеличение перспективности точки карты, поскольку можно предполагать, что градиентные зоны коррелированы с зонами геологических разломов.

#### **Группа геодинамических признаков.**

$x_{18}$  — амплитуда новейших тектонических движений в метрах (*Атлас, 1983*). Области опусканий и области поднятий с амплитудой до 200 м не перспективны. Поднятия с амплитудой 200–500 м — слабо перспективны, 500–700 м — высоко перспективны. Поднятия с амплитудами более 1700 м слабо перспективны в связи с разрушением эрозией значительной части месторождений;

$x_{19}$  — гравитационные изостатические аномалии в миллигалах (*Добрев и Щукин, 1974*). Поле отражает степень компенсации уплотнённых и разуплотнённых объёмов в земной коре и верхней мантии. Области разуплотнённых объёмов могут быть связаны с проницаемыми, возможно раздроблёнными зонами. Точки региона, принадлежащие таким зонам, более перспективны;

$x_{20}$  — тепловой поток в единицах теплового потока по *Смирнову и др. 1986 г.* Области неотектонической активизации с проявлениями магматизма среднего — кислого состава отличаются повышенными значениями теплового потока. Эти же области содержат большую часть свинцово-цинковых месторождений. Отсюда, перспективность точки тем больше, чем выше значение теплового потока в её окрестности.

**7.2.3. Выборка.** Для нахождения прогнозирующей функции была сформирована выборка из 133 узлов сетки. В неё вошли 40 точек, соответствующих известным месторождениям Болгарии (Йовчев, 1960). Большая часть точек выборки, относящихся к зонам, не содержащим месторождений, выбрана экспертом в амагматических областях, с преимущественным развитием слабо деформированных мезо-кайнозойских отложений, испытывающих в нетектонический этап погружения. Кроме того, выбирались ячейки, в которых выявлены месторождения других типов полезных ископаемых, слабо совместимых со свинцово-цинковыми, например, хромитовые и железо-марганцевые.

Оценки давались одним экспертом. Для нахождения прогнозирующей функции использовался метод аппроксимации интервальных экспертных оценок (раздел 4.2). Интервальные оценки назначались следующим образом: для точек, относящихся к месторождениям, границы интервалов принимались  $m^{(1)} = 600$ ,  $m^{(2)} = 700$ , для точек, относящихся к зонам, не содержащим месторождений, границы интервалов принимались  $m^{(1)} = 400$ ,  $m^{(2)} = 500$ . Для всех точек, кроме тех, которые соответствовали крупным и средним месторождениям, весовые коэффициенты брались  $w^{(1)} = w^{(2)} = 1$ . Для точек, относящихся к крупным и средним месторождениям, выбрано соответственно  $w^{(1)} = 10$  и  $w^{(2)} = 5$ .

**7.2.4. Прогнозирующая функция.** В процессе решения задачи было выполнено несколько итераций построения версий регионального прогноза свинцово-цинковых месторождений. В них исследовалось влияние на прогноз различных групп геолого-геофизических признаков и различных способов организации учебной выборки.

Сравнение версий прогноза по выбранным для прогноза группам признаков способствовало уточнению моделей образования месторождений. Этот анализ подсказал методы формализации некоторых признаков, таких как, например,  $x_2$  и  $x_3$ , характеризующих возраст и состав последних магматических проявлений.

В окончательной версии прогнозирующая функция найдена по всей выборке (рис. 7.1). Она содержит восемь оцениваемых параметров и имеет вид (порядок признаков в прогнозирующей функции соответствует порядку их выбора в процедуре шаговой регрессии):

$$F(\mathbf{x}, \mathbf{a}) = a_0 + \varphi_{18}(x_{18}) + \varphi_{19}(x_{19}) + \varphi_2(x_2) + \varphi_8(x_8) + \varphi_{12}(x_{12}),$$

где  $\mathbf{x}$  — вектор признаков,  $\mathbf{a} = (a_1, a_2, \dots, a_7)$  — вектор параметров,  $a_0 = 304$  — константа.

Из графиков (рис. 7.1) видно, что свинцово-цинковым месторождениями благоприятствуют неотектонические поднятия ( $x_{18} > 500$  м); не скомпенсированные вещественные неоднородности или зоны разуплотнения или переуплотнения в земной коре ( $x_{19} > 0$  мГл); проявления палеозойско-мезозойского ( $x_2 \geq 2$  условных единиц) и, особенно, позднекайнозойского магматизма ( $x_2 \geq 8$  условных единиц); проявле-



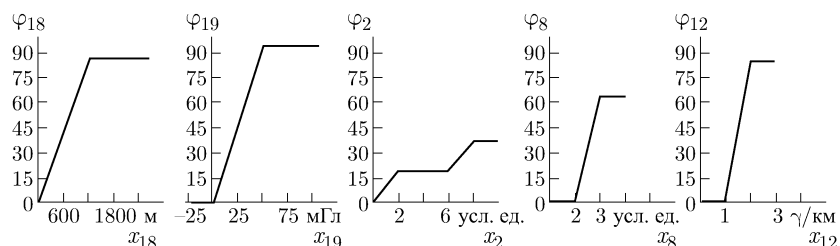


Рис. 7.1. Графики функций  $\varphi_i(x_i)$ ;  $x_{18}$  — амплитуда неотектонических движений,  $x_{19}$  — гравитационные изостатические аномалии в миллигалах,  $x_2$  — возраст последних проявлений магматизма,  $x_8$  — гетерогенность,  $x_{12}$  — модуль градиента вертикальной составляющей локального магнитного поля

ния интенсивной гетерогенности ( $x_8 \geq 3$  условных единиц) и высокие значения модуля градиента вертикальной составляющей локального магнитного поля ( $x_{12} \geq 2$  гамм/км) над зонами скрытых разломов. Это соответствует модели формирования свинцово-цинковых месторождений.

**7.2.5. Карта прогноза.** Сеточная модель регионального прогноза свинцово-цинковых месторождений показана на рис. 7.2. Внутри контура, ограниченного изолинией со значением прогнозной функции 570, попало 37 из 40 месторождений. Указанная изолиния очерчивает 9 перспективных зон. Наиболее крупная из них (I) охватывает Родопский массив. В восточной части зоны сосредоточены крупнейшие свинцово-цинковые месторождения Болгарии (Мадан, Рудозем, Звездел и др.). В западной части известно несколько месторождений с невыясненными перспективами (Йовчев И.С., 1960). Небольшая перспективная зона (IX) охватывает верховья рек Марица и Искыр. Она располагается в северной части Родопского массива и по геологическим и геодинамическим характеристикам близка к предыдущей зоне.

Две перспективные зоны локализованы в тектонической системе Крайштид. Одна из них (VII) включает месторождение Руен и несколько мелких месторождений и рудопроявлений свинца и цинка. Другая зона (VI), расположенная севернее, не содержит известных месторождений.

Две перспективные зоны размещаются в пределах тектонической системы Среднегорья, представляющей собой переработанную часть Родопского массива. Зона (IV) охватывает восточную часть системы и включает несколько месторождений, из которых крупнейшим является Устрем. Небольшая по размеру зона (VIII) расположена южнее Софии. Её перспективность подтверждается локализацией в её пределах свинцово-цинкового рудопроявления и месторождения с неясными перспективами, которые не использовались при составлении выборки примеров.

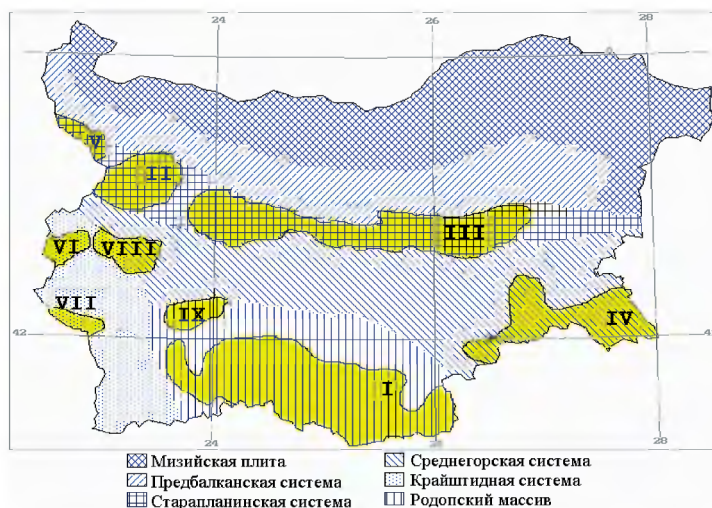


Рис. 7.2. Прогнозные зоны I–IX свинцово-цинковых концентраций и схема тектонического районирования

Три зоны приурочены к Старопланинской тектонической системе. Наиболее интересная из них протягивается от долины р. Искыр до западной границы (II). Она включает несколько полиметаллических месторождений, крупнейшим из которых является Седмочисленици. Все месторождения приурочены к восточной части зоны. В пределах линейной перспективной зоны (III) месторождения пока известны только в западной части. Перспективная зона (V) включает свинцово-цинково-железородное месторождение Чипровци, не входившее в выборку.

Прогнозное поле можно интерпретировать как сумму полей нелинейно преобразованных признаков. Вклады признаков в прогнозное поле легко определяются по функциям  $\varphi_i(x_i)$ . Наиболее дифференцированное выделение зон с наличием месторождений наблюдается по признакам: амплитуда неотектонических поднятий ( $x_{18}$ ), возраст последних проявлений магматизма ( $x_2$ ) и гетерогенность ( $x_8$ ). Гравитационные изостатические аномалии ( $x_{19}$ ) дают максимальные вклады для Старопланинской системы и Родопского массива, где действительно известны месторождения свинца и цинка, а также для восточной части Мизийской плиты и района Бургаса, малоперспективных в отношении концентраций свинца и цинка. Меньшая дифференциация наблюдается для признака модуль градиента вертикальной составляющей локального магнитного поля ( $x_{12}$ ).

Выделенные перспективные зоны отличаются соотношением вкладов признаков в прогноз. Семь из девяти этих зон располагаются в пределах трёх крупных тектонических структур: Старопланинской, Среднегорской систем и Родопского массива, которые различаются по

глубинному строению и современному геодинамическому состоянию (Добрев и Шукин, 1974). В табл. 7.1 показаны средние значения прогноза и вкладов признаков  $\varphi_i$  в прогноз для ячеек выборки с месторождениями, находящимися в указанных зонах;  $N$  — число месторождений в пределах тектонических структур.

Согласно табл. 7.1 месторождения Стараплатинской и Родопской структур отличаются от Среднегорских относительно большим вкладом амплитуды неотектонических движений ( $x_{18}$ ). Для Родопского массива вклады возраста последнего магматизма ( $x_2$ ) в два раза выше по сравнению с двумя другими структурами. Признаки  $x_8$  и  $x_{12}$  практически не различают месторождения различных тектонических структур. Выявленные различия вкладов признаков хорошо увязываются с представлениями о геодинамической обстановке этих структур. В Стараплатинской и Среднегорской системах интенсивно проявлен мезозойский магматизм, связанный с последней фазой тектоно-магматической активизации. Восточная часть Среднегорской системы, где расположена большая часть известных в ней месторождений, примыкает к Чёрному морю и отличается меньшей мощностью земной коры и изостатической неуравновешенностью.

Таблица 7.1

Тектонические структуры	$F(\mathbf{x}, \hat{\mathbf{a}})$	$\varphi_{18}$	$\varphi_{19}$	$\varphi_2$	$\varphi_8$	$\varphi_{12}$	$N$
Стараплатинская система (зоны II, III, V)	620	85	66	15	65	85	8
Родопский массив (зоны I, IX)	633	76	75	35	65	78	24
Среднегорская система (зоны IV, VIII)	585	27	91	15	65	85	4

Полученной прогнозной карте (см. рис. 7.2) можно дать вероятностную интерпретацию.

Разобьём исследуемую территорию на непересекающиеся ячейки  $11,5 \times 16,5$  км с центрами в узлах координатной сетки. Будем считать, что каждая ячейка характеризуется двумя случайными величинами: значением прогнозирующей функции  $F(\mathbf{x}, \hat{\mathbf{a}})$  и величиной  $Y$ , принимающей значение 0 при отсутствии в ячейке оруденения и значение 1 при наличии его. Оценкой апостериорной вероятности выбора ячейки с оруденением при условии, что прогнозирующая функция  $F > f$ , является величина

$$P(Y = 1/F \geq f) = \frac{n_1(F \geq f)}{n(F \geq f)},$$

где  $n_1(F \geq f)$  — число ячеек, в которых имеется свинцово-цинковое оруденение при условии, что в них значение прогноза  $F \geq f$ ;  $n(F \geq f)$  — число ячеек в регионе, где значение прогноза  $F \geq f$ . Вообще, эта оценка несколько занижена, так как полное число ячеек с оруденением для любого  $i$  больше, чем число  $n_1(F \geq f)$  известных оруденений.

Оценки  $P$  апостериорной вероятности выбора ячейки с оруденением при условии, что прогнозирующая функция  $F \geq f$ , даны в табл. 7.2.

Таблица 7.2

$f$	$n_1(F \geq f)$	$n(F \geq f)$	$P$
570	37	155	0,24
585	33	115	0,29
600	26	82	0,32
615	21	63	0,33

Из табл. 7.2 видно, что при увеличении значения прогнозирующей функции увеличивается оценка апостериорной вероятности наличия в ячейке оруденения. Это свидетельствует о том, что предсказуемая сила прогноза возрастает по мере увеличения прогнозирующей функции.

Полученные результаты позволяют сделать следующие выводы.

Прогнозная карта не противоречит металлогеническому районированию. Две зоны оконтуривают районы, в которых расположено свинцово-цинковое оруденение, не включённое в выборку.

Результаты прогноза могут быть уточнены при дополнении исходных данных геохимическими признаками, использовании более детальных геологических материалов и привлечении более совершенных моделей формирования свинцово-цинковых месторождений.

Отобранные при нахождении прогнозирующей функции признаки позволяют уточнить модель образования свинцово-цинковых месторождений Болгарии. Предполагается, что существенная роль принадлежит признакам, связанным с современным геодинамическим состоянием, таким как амплитуда неотектонических движений и изостатические аномалии, обычно не используемым при прогнозировании месторождений.

Соотношение вкладов признаков различается для крупных тектонических структур. При решении задачи более детального прогноза на зональном и локальном уровне необходимо отдельно исследовать перспективность тектонических структур, различающихся историей развития и современной геодинамической обстановкой.

### 7.3. Региональный прогноз медных месторождений Болгарии

**7.3.1. Причинно-следственная модель.** Для Болгарии характерно многообразие медных месторождений, так, *Й. Йовчев, 1960*, выделяет 12 их типов. Количество обнаруженных месторождений и рудопроявлений в каждом типе изменяется от 1 до 50. При решении прогнозной задачи рассматривались только шесть типов месторождений, представленных более чем 15 находками. В части этих типов медьсодержащие минералы не являются главными. Основные параметры месторождений всех шести типов приведены в табл. 7.3.

В соответствии с табл. 7.3 для каждого типа месторождений (в дальнейшем они именуется по их порядковым номерам) может быть предложена своя геолого-геофизическая модель — система представлений, отвечающая структурным, возрастным, вещественным и геодинамическим особенностям формирования и современного состояния типового месторождения, выраженная через значения геолого-геофизических признаков.

**Тип 1.** Месторождения этого типа свойственно наличие ультраосновных, частично серпетинизированных пород, которые могут создавать положительные аномалии регионального и локального магнитного поля, а также положительные аномалии гравитационного поля в редукции Буге. Интрузии ультраосновных пород относятся к ранним этапам геологического развития Болгарии. Они приурочены к зонам глубинных разломов с повышенной степенью раздробленности земной коры и высокой геологической гетерогенностью.

**Тип 2.** Месторождения связаны с проявлениями герцинского магматизма. Интрузии локализованы в зонах разломов, часть которых была активизирована в неотектонический этап.

**Тип 3.** Месторождения связываются с раннемеловыми интрузиями, локализация которых контролируется разломами.

**Тип 4.** Месторождения сосредоточены в эффузивах позднемелового возраста среднего состава и локализованы, как правило, в зонах их дробления.

**Тип 5.** Месторождения связаны с проявлениями мел-палеоценового магматизма среднего — кислого состава и приурочены к тектоническим контактам, дайкам и другим малым интрузиям.

**Тип 6.** Свинцово-цинковые месторождения с медьсодержащими минералами обнаруженными, главным образом, в пределах Родопского массива. Они связаны с интрузиями среднего кислого состава последнего из многократных периодов тектонической активизации. Магматические тела локализованы в зонах разломов, в районах, испытавших неотектонические поднятия.

**7.3.2. Исходные данные.** Так же как и при решении предыдущей задачи (региональный прогноз свинцово-цинковых месторождений Болгарии), использовались региональные признаки, снятые с карт масштаба 1 : 500 000, 1 : 1 000 000 и 1 : 1 500 000. Информационная нагрузка карт определила выбор шага координатной сетки 12' по широте и 6' по долготе. Это приблизительно соответствует шагу 11,5 и 16,5 км. Значения признаков снимались для узлов сетки.

При решении использовались следующие геолого-геофизические признаки.

**Группа вещественных признаков среды**

$x_1$  — состав последних проявлений магматизма. Данные снимались с карты *Димитрова, ред., 1978*: 1 — отсутствие проявлений, 2 — основной, 3 — кислый, 4 — средний составы интрузий;

Таблица 7.3. Основные параметры медных месторождений

№	Тектонический и магматический этап, стадия	Магматическая формация	Генезис	Главные рудоконтролирующие факторы	Полезные компоненты			Типовые месторождения
					главные	второстепенные	сопутствующие	
1	Докембрийский тектонический этап	Древние серпентинизированные ультраосновные породы	Раннемагматический	Дунитовые и серпентинитовые тела часто с разломными контактами	Cr	Cu, Co, Ni	V, W	Небольшие месторождения
2	Раннегерцинская стадия	Среднекарбонная габродiorитовая с гранодиоритами, гранитоидами, щелочными породами	Высоко-среднетемпературный гидротермальный и контактово-вольтасоматический	Гранитоидные плутоны на контактах с раннепалеозойскими мраморами	Fe, Mn, Pb, Au, Cu	Zn, W, Mo	Ag, Hg	Чировци
3	Раннемеловая стадия альпийского цикла	Раннемеловые малые интрузии	Телетермальный	Разломы, пористые песчаники и карбонаты пермского-юрского возраста	Fe, Mn, Ba, Pb	Cu, Zn	Ag, Co, Ni	Небольшие месторождения
4	Позднемеловая стадия альпийского этапа	Эффузивы среднего состава (андезиты, дациты, латиты) и субвулканические интрузии	Эксталационно-осадочные и гидротермальные	Разломы, гидротермальные зоны в сенонских осадках, туфы, туффиты, эффузивы	Mn, Cu	Pb, Au, Mo		Небольшие месторождения
5	Палеоценовая стадия альпийского этапа	Гранитоидные и монцитовые интрузии, субвулканические дацитовые дайки	Магматические, высоко- и среднетемпературные гидротермальные	Разломы, интрузии, дайки	Cu, Fe	Mo, W, Pb, Zn, V	Co, Ni, Ag, Zr, Cd	Росен, Граматиково, Медет, Малко-Тырново
6	Раннемиоценовая стадия альпийского этапа	Малые интрузии в Родопском кристаллическом массиве	Гидротермальные и контактово-метасоматические	Молодые (Эгейские) разломы	Pb, Zn	Cu, Ag, Cd	Mo	Руен, Устрем, Маджарово

$x_2$  — состав преобладающих магматических пород в точке (Димитрова, ред., 1978). Кодировка значений данного номинального признака выбрана так, чтобы увеличение признака соответствовало при прочих равных условиях более благоприятному для медных концентраций составу интрузий. Отсутствие магматизма — 1, кислый состав — 2, средний состав — 3, ультраосновной состав — 4, основной состав — 5;

$x_3$  — региональное магнитное поле в гаммах по работе (Григоров, 1961);

$x_4$  — вертикальная составляющая локального магнитного поля в гаммах по работе (Григоров, 1961);

$x_5$  — гравитационное поле в редукции Буге в миллигалах (Добрев и Щукин, 1974).

#### **Группа временных признаков среды**

$x_6$  — количество тектоно-магматических циклов в точке определялось по геологической карте из работы (Атлас, 1983). 0 — отсутствие проявлений магматизма в окрестности данной точки, 1, 2, 3, 4 — количество тектономагматических циклов;

$x_7$  — возраст последних проявлений магматизма. Данные снимались с карты Димитрова Е., ред., 1978: 0 — отсутствие проявлений магматизма, 1 — последние проявления магматизма в докембрии, 2 — в карбоне, 3 — в перми, 4 — в триасе, 5 — в юре, 6 — в мелу, 7 — в палеоцене, 8 — в эоцен-олигоцене, 9 — в миоцене;

$x_8$  — возраст магматических тел по карте Димитрова, ред., 1978, преобладающих в ячейке  $11,5 \times 16,5$  км с центром в точке сетки. Кодировка значений данного номинального признака выбрана так, чтобы увеличение признака соответствовало при прочих равных условиях более благоприятному для медных месторождений возрасту преобладающих магматических образований: 0 — нет магматизма, 1 — неоген-четвертичный, 2 — палеозойский и эоцен-олигоценовый, 3 — пермо-триасовый, 4 — юрский, палеоценовый, 5 — меловой.

#### **Группа структурных признаков среды**

$x_9$  — плотность разломов. Снималась суммарная длина разломов в километрах в ячейке по геологической карте (Атлас, 1983);

$x_{10}$  — гетерогенность, определялась по геологической карте (Атлас, 1983) путём подсчёта количества геологических границ в ячейке. Зоны глубинных разломов проявляются в повышенной гетерогенности;

$x_{11}$  — толщина гранито-метаморфического слоя в километрах по карте (Добрев и Щукин, 1974);

$x_{12}$  — глубина залегания поверхности Мохоровичича в километрах по карте Димитрова, ред., 1978;

$x_{13}$  — модуль градиента регионального магнитного поля;

$x_{14}$  — модуль градиента вертикальной составляющей локального магнитного поля;

$x_{15}$  — модуль градиента гравитационного поля в редукции Буге;

$x_{16}$  — модуль градиента поля изостатических аномалий;

$x_{17}$  — модуль градиента толщины гранито-метаморфического слоя;

$x_{18}$  — модуль градиента неотектонических движений.

$x_{19}$  — модуль градиента глубины залегания поверхности Мохоровичича.

Признаки  $x_{11}$ – $x_{17}$  рассчитаны средствами системы ГЕО. Можно предполагать, что увеличение значений модуля градиента геолого-геофизических полей скорее всего связано с геологическими разломами.

#### Группа геодинамических признаков

$x_{20}$  — амплитуда новейших движений в метрах (*Атлас, 1983*). Приняты следующие градации: области опусканий и области поднятий с амплитудой до 200 м; поднятия с амплитудой 200–500 м, 500–1700 м и более 1700 м;

$x_{21}$  — гравитационные изостатические аномалии в миллигалах (*Добрев и Щукин, 1974*). Поле отражает степень компенсации уплотнённых и разуплотнённых объёмов в земной коре и верхней мантии;

$x_{22}$  — тепловой поток в единицах теплового потока по *Я. Б. Смирнову и др., 1986 г.*

Возможная связь перечисленных исходных данных с моделями рассмотренных в разделе 7.3.1. шести типов медных месторождений показана в табл. 7.4.

Таблица 7.4. Признаки, которые могут участвовать в прогнозе рассматриваемых шести типов медных месторождений.

Тип месторождения	Номера признаков										
	1	2	3	4	5	6	7	8	9	10	11
1		+	+	+	+			+	+	+	
2		+						+	+	+	
3								+	+	+	
4		+						+	+		
5		+						+	+	+	
6	+					+	+	+	+	+	

Тип месторождения	Номера признаков										
	12	13	14	15	16	17	18	19	20	21	22
1		+	+	+	+	+	+	+			
2		+	+	+	+	+	+	+			
3		+	+	+	+	+	+	+			
4											
5		+	+	+	+	+	+	+			+
6		+	+	+	+	+	+	+	+	+	+

**7.3.3. Выборка.** Для нахождения прогнозирующей функции была сформирована выборка из 203 точек. В неё вошли 107 точек, которые выбраны в пределах всех шести типов медных месторождений Болга-



рии (Йовчев, 1960). Большая часть точек, относящихся к зонам без месторождений, выбрана экспертом в амагматических областях, с преимущественным развитием полого залегающих мезо-кайнозойских отложений, испытывающих нетектонический этап погружения в пределах Мизийской плиты, а также в районе выходов на поверхность древних гранитов и гранитизированных пород, практически не переработанных последующими процессами тектоно-магматической активизации в пределах Родопского массива и Среднегорской зоны. При решении задач прогноза для каждого типа медных месторождений участвовала выборка, в которую входили одни и те же точки, выбранные экспертом как бесперспективные для поиска медных месторождений, и точки, относящиеся к месторождениям и рудопроявлениям меди только данного типа.

Оценки давались одним экспертом. Для нахождения прогнозирующей функции использовался метод аппроксимации интервальных экспертных оценок (раздел 4.2). Интервальные оценки назначались следующим образом: для точек, относящихся к месторождениям, границы интервалов принимались  $m^{(1)} = 600$ ,  $m^{(2)} = 700$ , для точек, относящихся к зонам, не содержащим месторождений, границы интервалов принимались  $m^{(1)} = 400$ ,  $m^{(2)} = 500$ . Для всех точек, кроме тех, которые соответствовали крупным и средним месторождениям, весовые коэффициенты брались  $w^{(1)} = w^{(2)} = 1$ . Для точек, относящихся к крупным и средним месторождениям, выбрано соответственно  $w^{(1)} = 10$  и  $w^{(2)} = 5$ .

**7.3.4. Прогнозирующая функция. Тип 1.** При нахождении прогнозирующей функции  $f^{(1)}(\mathbf{x})$  использовалась учебная выборка, включавшая 9 ячеек с месторождениями и 95 ячеек без месторождений. Прогнозирующая функция имеет следующий вид (порядок признаков в прогнозирующей функции соответствует порядку их выбора в процедуре шаговой регрессии):

$$f^{(1)}(\mathbf{x}) = 400 + \varphi_2^{(1)}(x_2) + \varphi_1^{(1)}(x_1).$$

Прогнозирующая функция даёт безошибочное разделение на учебной выборке. Из графика на рис. 7.3 видно, что определяющим при прогнозе месторождений данного типа является преобладание основных и ультраосновных пород (значения 4 и 5 признака  $x_2$ ) и наличие молодых кислых и средних интрузий (значения 3 и 4 признака  $x_1$ ). Это соответствует модели месторожде-

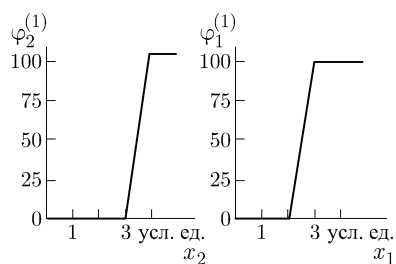


Рис. 7.3. Графики функций  $\varphi_i^{(1)}(x_i)$ , вошедших в прогнозирующую функцию медных месторождений 1 типа:  $x_2$  — состав преобладающих магматических тел в ячейке;  $x_1$  — состав последних проявлений магматизма

ний 1 типа, приведённой выше. Участие в прогнозирующей функции признака  $x_1$ , показывающего, что для месторождений этого типа важна повторная мобилизация вещества за счёт внедрения интрузий среднего состава, уточняет геологическую модель генезиса месторождений 1 типа, предложенную *И. Йовчевым, 1960*.

**Тип 2.** При нахождении прогнозирующей функции  $f^{(2)}(\mathbf{x})$  использовалась учебная выборка, включавшая 10 ячеек с месторождениями и 95 ячеек без месторождений. Прогнозирующая функция имеет следующий вид:

$$f^{(2)}(\mathbf{x}) = 406,25 + \varphi_2^{(2)}(x_2) + \varphi_1^{(2)}(x_1) + \varphi_6^{(2)}(x_6).$$

Прогнозирующая функция даёт безошибочное разделение на учебной выборке. Из графиков на рис. 7.4 видно, что, для месторождений 2 типа, определяющим является преобладание основных пород

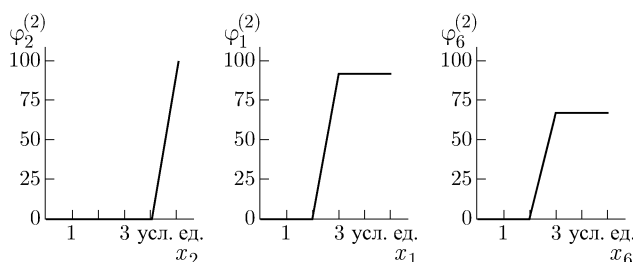


Рис. 7.4. Графики функций  $\varphi_i^{(2)}(x_i)$ , вошедших в прогнозирующую функцию медных месторождений 2 типа:  $x_2$  — состав преобладающих магматических тел в ячейке;  $x_1$  — состав последних проявлений магматизма;  $x_6$  — количество тектоно-магматических циклов

(значение 5 признака  $x_2$ ). Это соответствует модели месторождений 2 типа, приведённой выше. В прогнозирующей функции участвуют признаки  $x_1$  и  $x_6$ , показывающие, что для месторождений этого типа важна неоднократная тектоно-магматическая мобилизация (значения признака  $x_6 \geq 3$ ), причём, последняя из них должна характеризоваться внедрением интрузий среднего и кислого состава (значения признака  $x_6 \geq 3$ ). Модель образования месторождений второго типа, приведённая *И. Йовчевым, 1961*, может быть уточнена за счёт придания ведущей роли в формировании месторождений этого типа проявлениям основного магматизма, активизированным за счёт внедрения в более поздние эпохи интрузий кислого — основного состава.

**Тип 3.** При нахождении прогнозирующей функции  $f^{(3)}(\mathbf{x})$  использовалась учебная выборка, включавшая 11 ячеек с месторождениями и 95 ячеек без месторождений. Прогнозирующая функция имеет следующий вид:

$$f^{(3)}(\mathbf{x}) = 388,73 + \varphi_{10}^{(3)}(x_{10}) + \varphi_9^{(3)}(x_9) + \varphi_2^{(3)}(x_2) + \varphi_{16}^{(3)}(x_{16}).$$

Прогнозирующая функция даёт безошибочную классификацию на учебной выборке. Геологическая интерпретация графиков на рис. 7.5 позволяет уточнить модель образования медных месторождений 3 типа.

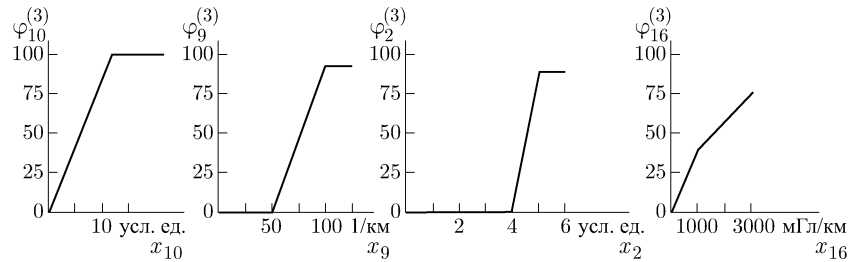


Рис. 7.5. Графики функций  $\varphi_i^{(3)}(x_i)$ , вошедших в прогнозирующую функцию медных месторождений 3 типа:  $x_{10}$  — гетерогенность;  $x_9$  — плотность разломов;  $x_2$  — состав преобладающих магматических тел в ячейке;  $x_{16}$  — модуль градиента поля изостатических аномалий

Последние приурочены к изостатически нескомпенсированным зонам дробления, содержащим интрузии преимущественно основного состава.

**Тип 4.** При нахождении прогнозирующей функции  $f^{(4)}(\mathbf{x})$  для месторождений 4 типа использовалась учебная выборка, включавшая 10 ячеек с месторождениями и 95 ячеек без месторождений.

$$f^{(4)}(\mathbf{x}) = 500 + \varphi_8^{(4)}(x_8).$$

Согласно модели, на прогноз месторождений данного типа влияют возраст и состав магматических продуктов, а также раздроблённость пород. Однако возраст и состав магматических продуктов коррелированы. Поэтому добавление других признаков не улучшает качества прогноза. Прогнозирующая функция даёт безошибочное разделение на учебной выборке.

**Тип 5.** Прогнозирующая функция месторождений 5 типа найдена по учебной выборке, включавшей 24 ячейки с месторождениями и 95 ячеек без месторождений. Прогнозирующая функция, несмотря на разные учебные выборки, получилась такой же, как и для месторождений 4 типа,  $f^{(5)}(\mathbf{x}) = f^{(4)}(\mathbf{x})$ , и дала безошибочное разделение классов на учебной выборке.

**Тип 6.** При нахождении прогнозирующей функции  $f^{(6)}(\mathbf{x})$  использовалась учебная выборка, включавшая 30 ячеек с месторождени-

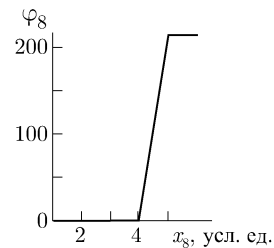


Рис. 7.6. График функции  $\varphi_8^{(4)}(x_8)$ , вошедшей в прогнозирующую функцию медных месторождений 4 и 5 типов:  $x_8$  — возраст преобладающих в ячейке магматических тел

ями и 95 ячеек без месторождений. Прогнозирующая функция имеет следующий вид:

$$f^{(6)}(\mathbf{x}) = 388,02 + \varphi_{20}^{(6)}(x_{20}) + \varphi_8^{(6)}(x_8) + \varphi_{14}^{(6)}(x_{14}) + \varphi_{10}^{(6)}(x_{10}) + \varphi_6^{(6)}(x_6).$$

Прогнозирующая функция даёт безошибочную классификацию на учебной выборке. Как показывают графики функций (рис. 7.7), в ячейках, перспективных на обнаружение медьсодержащих месторождений

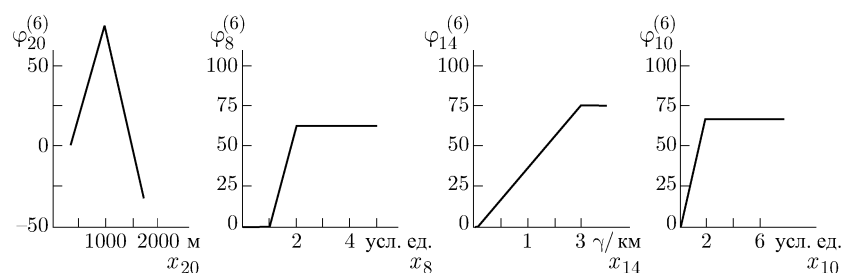


Рис. 7.7. Графики функций  $f^{(6)}(\mathbf{x})$ , вошедших в прогнозирующую функцию медных месторождений 6 типа:  $x_{20}$  — амплитуда новейших движений;  $x_8$  — возраст преобладающих в ячейке магматических тел;  $x_{14}$  — модуль градиента вертикальной составляющей локального магнитного поля;  $x_{10}$  — гетерогенность

6 типа, амплитуда новейших поднятий должна составлять 500–1700 м. Если поднятия ниже этих величин, месторождение не будет вскрыто на поверхности, если — выше, то оно будет в значительной степени эродировано. Возраст преобладающих в такой ячейке магматических пород может быть любым, кроме неоген-четвертичного. Перспективная ячейка располагается над зоной скрытого глубинного разлома или зоной дробления и интенсивных деформаций. Информативные прогнозные признаки месторождений 6 типа практически полностью совпадают с признаками, отвечающими модели Й. Йовчева (1961).

Информативные признаки для всех типов медных месторождений показаны в табл. 7.5. При сравнении с табл. 7.4, отражающей возможные геолого-геофизические признаки, согласующиеся с генетическими моделями различных типов медных месторождений, можно отметить следующее. Для 4, 5, 6 типов информативные признаки полностью выбраны из модельных признаков, что определяет соответствие генетических моделей прогнозным. Для первых трёх типов медных месторождений информативные признаки позволяют уточнить генетические модели Й. Йовчева (1961).

**7.3.5. Карты прогноза.** Рассмотренные в предыдущем разделе прогнозирующие функции позволили построить карты регионального прогноза медных месторождений всех шести типов.

Таблица 7.5. Информативные признаки, выбранные при прогнозе шести типов медных месторождений

Тип место- рождения	Номера признаков										
	1	2	3	4	5	6	7	8	9	10	11
1	+	+									
2	+	+				+					
3		+							+	+	
4								+			
5								+			
6								+		+	

Тип место- рождения	Номера признаков											
	12	13	14	15	16	17	18	19	20	21	22	
1												
2												
3					+							
4												
5												
6			+						+			

**Карта прогноза медных месторождений типа 1 (рис. 7.8).** Точки с высокими значениями прогнозной функции, превышающими 600 условных единиц, можно объединить в три крупные зоны. Наиболее чётко выраженной является зона I северо-западного простирания. Эта зона пересекает несколько крупных тектонических структур. Известные рудные проявления и месторождения приурочены к юго-восточной части зоны, пересекающей Родопский массив.

Зона II имеет более сложную форму, но в целом она параллельна первой зоне. Открытые месторождения и проявления меди также расположены в той части второй зоны, которая пересекает Родопский массив.

Зона III имеет широтное простирание, располагается в юго-восточной части Болгарии, южнее г. Бургаса. В её пределах пока не обнаружены месторождения данного типа.

Поскольку вклад в прогнозную функцию вносят только интрузивы различного состава и возраста, то выделенные на карте прогноза зоны скорее всего локализованы над глубинными проницаемыми структурами. Их глубинный характер подчёркивается основным и ультраосновным составом интрузий. Эти структуры, по-видимому, имели длительную историю развития и служили подводящими каналами для более молодых магматитов среднего и кислого состава. Все известные месторождения и рудопроявления меди I типа приурочены к Родопскому массиву. Очевидно, особенности строения коры этой структуры

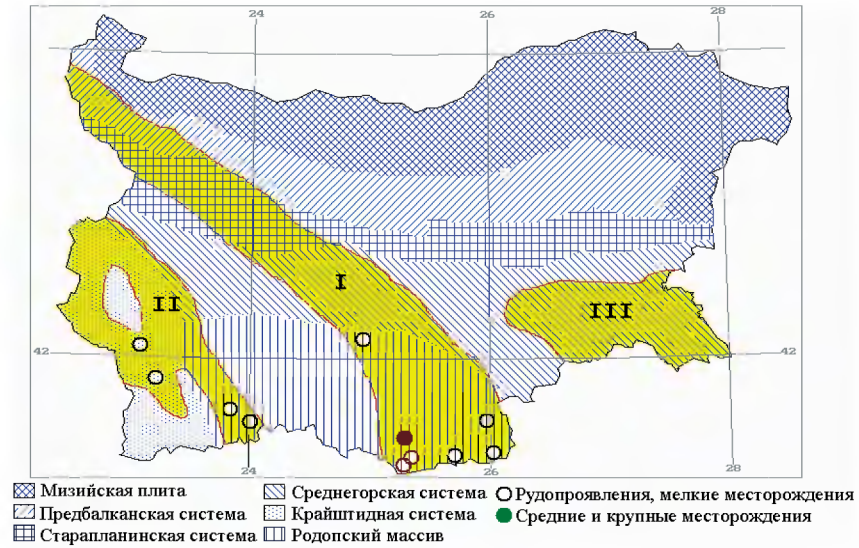


Рис. 7.8. Карта прогнозных зон I–III медных месторождений 1 типа

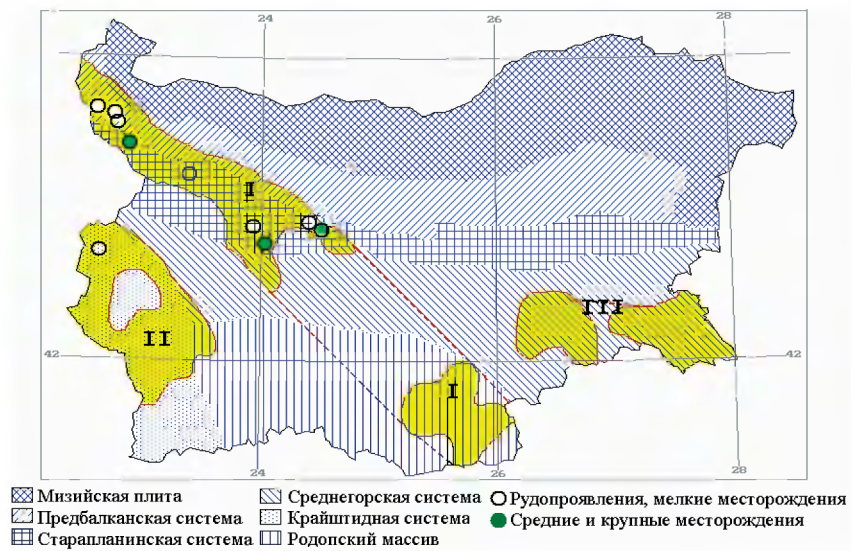


Рис. 7.9. Карта прогнозных зон I–III медных месторождений 2 типа

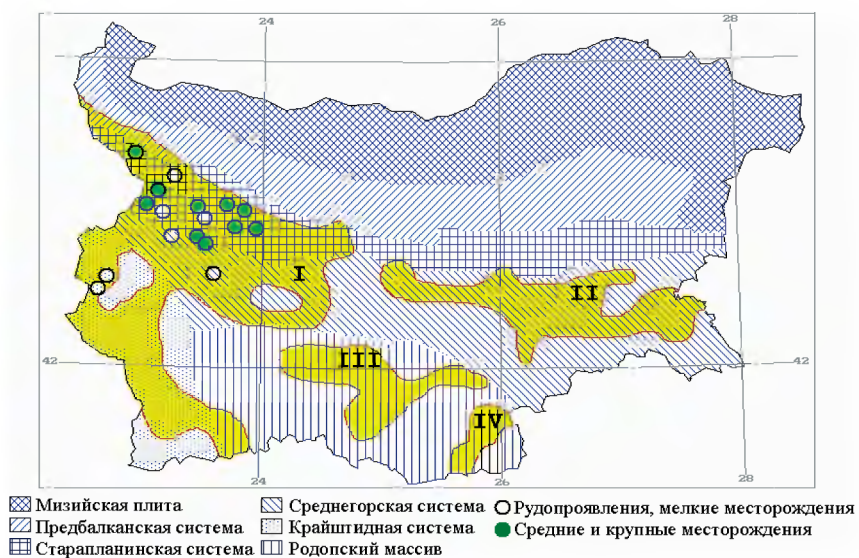


Рис. 7.10. Карта прогнозных зон I–IV медных месторождений 3 типа

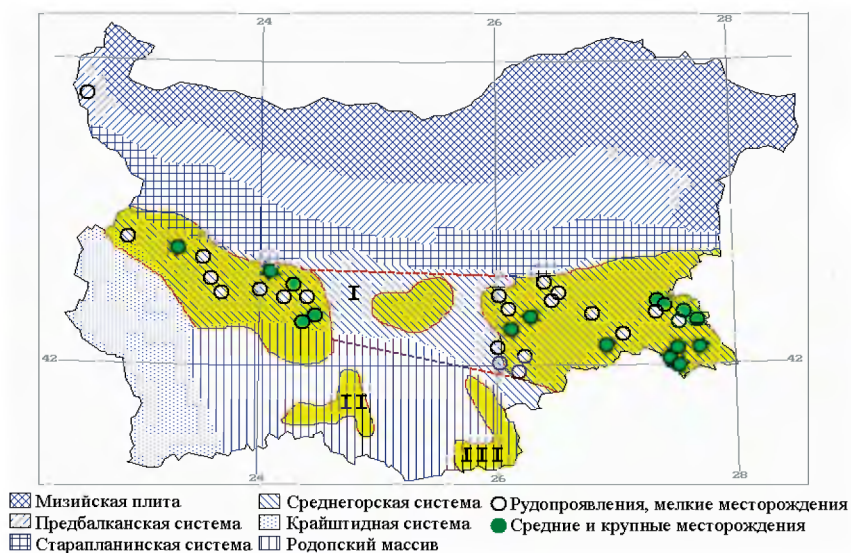


Рис. 7.11. Карта прогнозных зон I–III медных месторождений 4, 5 типов

и истории её эволюции определяли специфику проявления здесь зон глубинных разломов, благоприятно отразившуюся на концентрациях полезных компонентов. В связи с этим наиболее перспективными следует считать части прогнозных зон, расположенных в пределах массива.

**Карта прогноза медных месторождений типа 2 (рис. 7.9).** На этой карте точки со значениями прогнозной функции, превышающими 600 условных единиц, объединены в три зоны. Конфигурация зон напоминает схему районирования, показанную на рис. 7.8. Однако, имеется и ряд отличий. Так, линейная зона I юго-восточного простираения чётко разбивается на два района, из которых в северо-западном сосредоточены рудопроявления рассматриваемого типа, в то время как в юго-восточном пока нет ни месторождений, ни рудопоявлений. Зона II не продолжается в пределы Родопского массива. Зона III идентична соответствующей прогнозной зоне месторождений 1 типа. Как было показано выше, прогнозные признаки месторождений 2 типа охватывают различные аспекты магматической эволюции Болгарии, что позволяет предполагать связь прогнозных зон с глубинными разломами, которые были активны и при формировании медных месторождений 1 типа. Наиболее благоприятными для поисков медных месторождений 2 типа можно считать те части прогнозных зон, которые располагаются вне Родопского массива.

**Карта прогноза медных месторождений типа 3 (рис. 7.10).** На карте выделено 4 зоны. Первая из них, зона I, является наиболее крупной и охватывает западную часть Болгарии. Она имеет сложную конфигурацию с хорошо проявленными северо-западными и широтными элементами. К северо-западной части этой зоны приурочены все выявленные медные месторождения 3 типа. Зона II отличается широтным простираением и располагается в восточной части Старопланинской системы. Широтное простираение характерно и для зоны III, локализованной в пределах Родопского массива. Там же расположена и зона IV. Она невелика по размерам и не имеет чёткой ориентировки. Набор информативных прогнозных признаков медных месторождений 3 типа позволяет предполагать, что прогнозные зоны связаны с глубинными разломами, среди которых преобладают субширотные. Разломы северо-западного простираения также сохранились, но их роль по сравнению с месторождениями 1 и 2 типов значительно уменьшилась.

**Карта прогноза медных месторождений типов 4, 5 (рис. 7.11).** Точки сетки, в которых значения прогнозной функции превышают 600 условных единиц, могут быть объединены в три неравнозначных зоны. Наиболее крупная из них вытянута в широтном направлении и совпадает со Среднегорской тектонической системой. В её пределах расположены все крупные медные месторождения Болгарии. Две других зоны имеют ограниченные размеры и локализованы в пределах Родопского массива. В них медные месторождения пока не обнаружены. Зона I распадается на три рудных района: западный (а), центральный (б)



и восточный (в), каждый из которых по своим размерам превосходит зоны II и III.

Прогнозная карта медных месторождений 4 и 5 типов сопоставлялась с материалами Б. Богданова, опубликованными в 1986 и 1988 гг., т.е. через 25 лет после работы *И.Иовчева, 1961*. В этих материалах показаны контуры пяти крупных меднорудных районов: Софийского (1), Панагюрско-Етропольского (2), Прохоровского (3) и Малко-Тырновского (4). Эти контуры попадают в западный (а) и восточный (б) районы зоны I. Центральный район «б» прогнозной зоны I пока не содержит медных месторождений и может считаться первоочерёдным для детальных поисковых работ.

**Карта прогноза медных месторождений 6 типа (рис. 7.12).** Карта зон медьсодержащих месторождений 6 типа напоминает прогнозные зоны свинцово-цинковых месторождений (раздел 7.3). Можно отметить, что на карте 7.12 прогнозные области подчёркиваются линейны-

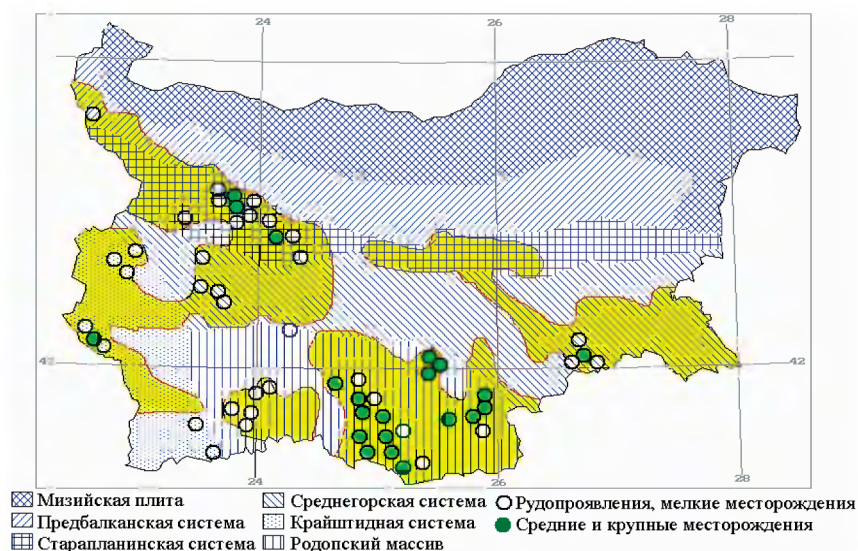


Рис. 7.12. Карта прогнозных зон медных месторождений 6 типа

ми элементами широтного и северо-западного простирания. Согласно набору информативных признаков для месторождений 6 типа эти элементы, скорее всего, отвечают глубинным разломам, испытавшим активизацию в неотектонический этап. Наибольший интерес представляет широтная линейная структура в центральной части карты, в пределах которой медьсодержащие месторождения 6 типа пока не обнаружены.

## Глава 8

# ПРОСТРАНСТВЕННЫЙ ПРОГНОЗ НА ОСНОВЕ ЭКСПЕРТНЫХ ЗНАНИЙ

### 8.1. Введение

Рассмотренная в главах 5, 6 и 7 технология пространственного прогноза базировалась на прогнозирующих функциях, которые индуктивно выводились по выборкам прецедентов с привлечением знаний предметной области. В ряде задач геоинформационного прогноза подобные выборки примеров прогноза отсутствуют. Решение этих задач возможно только на основе знаний предметной области. В данной главе обсуждаются несколько примеров решения подобных задач пространственного прогноза.

Первые три примера, рассмотренные в разделах 8.2, 8.3 и 8.4, относятся к проблеме прогноза некоторого целевого свойства пространственной среды или географических объектов по другим свойствам. При решении этих задач используются знания о связях между свойствами географических сущностей. В задаче локального прогноза концентраций углеводородов (раздел 8.2) используются знания о моделях необходимого комплекса условий нефтегазоносности; в задаче районирования урбанизированной территории по устойчивости к динамическим воздействиям (раздел 8.3) используются экспертные знания о связях между геоморфологическими и литологическими характеристиками геологической среды и степенью её устойчивости к различного рода сотрясениям; в задаче прогноза ущерба от сильных землетрясений и сейсмического риска (раздел 8.4) использованы общепринятые модели сейсмичности и знания инженерной сейсмологии.

Последние два примера, рассмотренные в разделе 8.5, относятся к проблеме прогноза стационарного состояния процесса, которое достигается в результате взаимодействия элементов некоторой структуры. При решении этих задач используются знания о связях между географическими сущностями: объектами или пространственными элементами среды. Примерами здесь являются задача анализа цепочек вторичных природных и природно-техногенных катастроф, возникающих как последствия спонтанных воздействий, таких как землетрясения и задача распространения загрязняющих веществ с поверхностным стоком.

## 8.2. Локальный прогноз нефтегазоносности

**8.2.1. Постановка задачи.** Задача локального нефтегазового прогноза состоит в выявлении ловушек любого генетического типа, с наибольшей вероятностью заполненных углеводородами. Проверка прогноза возможна только бурением поисковой скважины. Поэтому более конкретная постановка задачи локального прогноза, результат решения которой можно проверять прямыми методами, состоит в выборе места заложения продуктовой поисковой скважины. Локализация поисковой скважины опирается на следующие три фактора.

- Структурный: как правило, скважина выбирается в наиболее гипсометрически высокой части ловушки.
- Ёмкостной: скважина выбирается в области с наиболее высокими ёмкостными свойствами коллектора, обладающего крышкой (флюидоупором).
- Продуктивный: скважина выбирается в контуре, в котором прогнозируется нефтенасыщенность коллектора.

При решении задачи может быть известна следующая исходная информация:

1. Геолого-генетическая модель части нефтегазоносной провинции, включающая в себя исследуемый участок. В модели описывается:

- Наиболее распространённый тип ловушек.
- Нефтегазоносные или прогнозные нефтегазоносные комплексы.
- Литологическая выдержанность коллекторов и крышек по простиранию.
- Наличие несогласий и соотношение их с нефтегазоносными комплексами и отражающими горизонтами. Характер унаследованности структур.

2. Результаты бурения в пределах исследуемого участка или смежных районов с данными геофизического исследования скважин и сведениями о возможном или установленном положении водонефтяного контакта (ВНК), газоводяного контакта (ГВК) и газонефтяного контакта (ГНК).

3. Результаты интерпретации объёмных (3D) и профильных (2D) геофизических измерений.

- Структурные признаки
  - Сейсмические профили с выделенными отражающими горизонтами.
  - Структурные карты отражающих горизонтов (в случае 2D — точность около 10 м, при 3D — точность первые метры).
- Динамические признаки, отражающие ёмкостные свойства нефтегазоносных и возможно нефтегазоносных горизонтов в пределах продуктивных и возможно продуктивных комплексов, пачек горизонтов.
  - Профили и карты акустических жёсткостей.

- Карты амплитуд отражений и мгновенных частот отдельных горизонтов.
- Карты распределения ёмкостных свойств (пористости, эффективных толщин коллекторов, удельной ёмкости).
- Карты распределения толщин и проницаемости покрышек.
- Карты точности измерений.
- Признаки нефтенасыщенности
  - Карты и профили распределения электрического сопротивления.
  - Карты и профили аномалий вызванной поляризации с указаниями точности привязки аномалий сопротивления и поляризации.

Следует заметить, что вычисление 2D признаков сейсморазведки и электромагнитных признаков производится на профилях, заложенных при проведении полевых работ. Переход к сеточной цифровой модели производится с помощью интерполяции. Сеточная модель использует информацию по всем профилям, что позволяет в каждой точке карты оценить пространственные свойства геологической среды и получить представление о размерах прогнозируемых объектов. Вместе с тем, очевидно, что более точные значения признаков относятся к профилям.

**8.2.2. Модель локального прогноза.** Наиболее общая модель залежи углеводородов (УВ) основана на наличии обязательного комплекса условий нефтегазоносности:

- Наличие ловушки любого генетического типа.
- Наличие коллектора с достаточно хорошей пористостью и проницаемостью. Критическая пористость (т. е. пористость при которой извлечение УВ рентабельно) определяется плотностью и вязкостью нефти, а также литологией коллектора.
- Наличие надёжного флюидоупора (надёжность связана с проницаемостью покрышки и её толщиной, достаточной для того, чтобы удерживать УВ, часто находящихся в условиях аномально высоких давлений).
- Наличие нефтегазонасыщения.

Обязательные условия нефтегазоносности — коллектор, покрышка, ловушка и УВ насыщение — определяются по следующим признакам.

Коллектор определяется распределением ёмкостных свойств, связанных с:

- акустической скоростью сейсмических волн в предполагаемом продуктивном горизонте, соответствующей значениям повышенной пористости,
- толщиной пласта-коллектора,
- эффективной толщиной пласта (т. е. суммарной ёмкостью пласта-коллектора).

Покрышка определяется признаками:

- скорость сейсмических волн во флюидоупоре, соответствующая малой проницаемости,
- толщина флюидоупора.

Ловушка определяется структурными признаками:

- мера принадлежности точки к структурному элементу (поднятию, рифу, литологическому или стратиграфическому выклиниванию),
- объём ловушки, вычисленный с учётом условного ВНК.

Нефтегазонасыщенность определяется электроразведочными признаками:

- пониженная электрическая проводимость (высокое сопротивление),
- аномалии вызванной поляризации на профиле.

Все перечисленные признаки являются неоднозначными, так как при их вычислении возможны ошибки интерпретации исходных измерений и ошибки интерполяции от профилей к произвольным точкам карты. Поэтому для оценки нефтегазоносности введём частные функции уверенности признаков в выполнении необходимых условий нефтегазоносности,

$$f(\text{ловушка}), f(\text{коллектор}), f(\text{покрышка}), f(\text{продукт}),$$

принимающие значения от 0 до 1. Для оценки точности интерполяции введём функцию точности интерполяции, также принимающую значения от 0 до 1.

В пределах исследуемого участка количество примеров (скважин) с наличием и отсутствием продукта, как правило, невелико. Поэтому оценить частные функции уверенности признаков в наличии геологических сушностей на основе индуктивного обучения проблематично. В связи с этим частные функции уверенности должны задаваться экспертами на основе их опыта и с учётом геолого-физической модели месторождения. Чтобы учесть степень уверенности в выполнении всего комплекса условий нефтегазоносности, необходимо определить функцию

$$F = \min \{f(\text{ловушка}), f(\text{коллектор}), f(\text{покрышка}), f(\text{продукт})\}.$$

Функция  $F$  является мерой совпадения проявления всех необходимых признаков нефтегазоносности и представляет собой конъюнкцию нечёткой логики от частных функций уверенности  $f(\cdot)$ .

Функцию точности интерполяции  $A$  определим следующим образом:

$$A = 0,5(S_s + S_e),$$

где:

$$S_s = \begin{cases} 1 - \frac{\rho}{R_s} & \text{при } \rho < R_s, \\ 0 & \text{при } \rho \geq R_s, \end{cases} \quad S_e = \begin{cases} 1 - \frac{\rho}{R_e} & \text{при } \rho < R_e, \\ 0 & \text{при } \rho \geq R_e, \end{cases}$$

$S_s$  и  $S_e$  — функции близости точек карты к профилям сейсмо- и электроразведки соответственно,  $\rho$  — расстояние от профиля до произвольной точки сетки,  $R_s$  и  $R_e$  — расстояния от профилей, при которых точность интерполяции сейсморазведочных и электромагнитных признаков считается неудовлетворительной. Нетрудно видеть, что  $A$  изменяется от 0 до 1.

Теперь можно определить обобщённую функцию степени уверенности в наличии нефтегазоносности, которая учитывает как все частные степени уверенности в наличии каждого необходимого признака нефтегазоносности, так и точность интерполяции:

$$\varphi = F - \alpha(1 - A),$$

где:  $\alpha$  — задаваемый экспертом весовой коэффициент, который учитывает точность интерполяции. Легко видеть, что при  $\alpha = 0$  функция уверенности не уменьшается от ошибок интерполяции, при  $\alpha = 1$  на профилях сейсморазведки и электромагнитной разведки функция  $\varphi = 0,5F$ , на пересечении профилей функция  $\varphi = F$ , на расстояниях, больших  $R_s$  и  $R_e$  от профилей, функция  $\varphi = 0$ .

Очевидно, что по функции  $\varphi$  можно построить цифровую сеточную модель карты уверенности в наличии нефтегазоносности и выбрать места заложения скважины в зонах с максимальными значениями уверенности.

Приведённые рассуждения относятся к ситуации, когда нефть или газ находятся в одном горизонте. Но, как известно, с одной точкой карты можно связать несколько нефтегазоносных горизонтов. Задача состоит в выборе места заложения скважины, которая вскрывала бы несколько горизонтов

Для решения этой задачи требуется определить функцию уверенности в наличии нефтегазоносности на нескольких горизонтах. Выбор функции уверенности зависит от требований, которым должно удовлетворять выбранное место бурения. Оказывается, что формулировка этих требований вызывает существенные трудности у экспертов. Поэтому мы рассмотрим здесь несколько примеров выбора функции уверенности.

Допустим, критерий состоит в том, чтобы свойством нефтегазоносности непременно обладали все  $M$  горизонтов. Тогда функцию уверенности можно определить, как

$$\Psi^{(1)} = \min(\varphi_1, \varphi_2, \dots, \varphi_M).$$

В некоторых случаях можно предположить, что функции уверенности  $\varphi_i$  имеют вероятностный смысл и интерпретируются как мера возможности наличия нефтегазоносности в  $i$ -м горизонте. Тогда по аналогии с формулой вероятности нескольких независимых событий можно определить функцию уверенности в наличии нефтегазоносности

в  $M$  горизонтах, как

$$\Psi^{(2)} = \prod_{i=1}^M \varphi_i.$$

При выборе частных функций уверенности  $\varphi_i$  эксперт может принимать во внимание возможный объём запасов вскрываемых горизонтов. В этом случае можно определить функцию уверенности в наличии нефтегазоносности в  $M$  горизонтах как полный объём запасов в вскрываемых залежах

$$\Psi^{(3)} = \sum_{i=1}^M \varphi_i.$$

Наконец, можно определить функцию уверенности в наличии нефтегазоносности в  $M$  горизонтах, как линейную комбинацию предыдущих критериев:

$$\Psi = \frac{w_1 \Psi^{(1)} + w_2 (\Psi^{(3)})^{1/M} + \frac{w_3}{M} \Psi^{(3)}}{w_1 + w_2 + w_3}.$$

Таким образом, при решении задач локального прогноза определён набор значимых признаков. Эти признаки могут быть получены только при проведении комплексных геолого-геофизических работ. В ряде случаев задачи локального прогноза приходится решать при ограниченном наборе признаков. Локальный прогноз иллюстрируется примером выбора мест продуктивного бурения на одной из разведочных площадей в Удмуртии.

### 8.2.3. Прогнозирование мест заложения продуктивных поисковых скважин на одной из разведочных площадей Удмуртии.

**8.2.3.1. Причинно-следственная модель.** Предполагалось, что геологическое строение разведочной площади аналогично строению геологической среды, относящейся к более изученным территориям соседних месторождений. В соседних месторождениях установлено наличие пластов-коллекторов в трёх нефтегазоносных комплексах: среднекаменноугольном, нижнекаменноугольном и верхнедевонско-нижнекаменноугольном. Эти комплексы разделены региональными, надёжными флюидоупорами. Пласты-коллекторы в пределах комплексов, как правило, перекрыты слабо проницаемыми пластами небольшой толщины, что определяет возможность гидродинамической связи между коллекторами. В среднекаменноугольном комплексе выявлено четыре карбонатных пласта-коллектора, в нижнекаменноугольном три терригенных пласта-коллектора и верхнедевонско-нижнекаменноугольном три карбонатных пласта-коллектора.

Разведочная площадь изучена сейсморазведкой 2D и высокоразрешающей электроразведкой с измерением параметра вызванной поляризации в скважинно-наземной и наземной модификациях. По данным сейсморазведки, проведённой специалистами Удмуртгеофизики, выделены три антиклинальные структуры. В НПЦ «Геонефтегаз» под руководством Г. Е. Руденко проведён динамический анализ сейсмических

данных и получены оценки сеточных моделей акустических жёсткостей, пористости, эффективной толщины 10 пластов-коллекторов во всех трёх нефтегазоносных комплексах.

Высокоразрешающая электроразведка, выполненная в НПЦ «Геонефтегаз» под руководством Е. С. Киселёва позволила выявить электрическое сопротивление и аномалии вызванной поляризации в средних и нижнекаменноугольных нефтеносных комплексах. По этим данным были построены сеточные модели нефенасыщенности коллекторов.

#### *8.2.3.2. Исходные данные.*

1. Структурная карта кровли карбонатных отложений верейского горизонта, московского яруса среднего карбона, отражающий горизонт ОГ-1, масштаб 1 : 25 000.
2. Структурная карта кровли терригенных отложений яснополянского надгоризонта визейского яруса нижнего карбона, отражающий горизонт ОГ-2, масштаб 1 : 25 000.
3. Структурная карта подошвы карбонатных отложений верхнего девона, отражающий горизонт ОГ-3, масштаб 1 : 25 000.
4. Электрические сопротивления в каменноугольных отложениях по профилям.
5. Аномалии вызванной поляризации в каменноугольных отложениях.
6. Карты нефтенасыщенности каменноугольных пластов-коллекторов, масштаб 1 : 25 000.
7. Карты акустических жёсткостей, пористости, эффективной толщины 10 пластов-коллекторов в трёх нефтегазоносных комплексах и удельной ёмкости (произведение эффективной толщины коллектора на пористость).

*8.2.3.3. Информационная модель прогноза.* Прогноз мест заложения поисковых скважин основан на совпадении трёх необходимых условий наличия месторождений: существования ловушки, коллектора и признаков нефтеносности.

Существование ловушки оценивается по всем имеющимся структурным картам отражающих горизонтов. Предполагается, что самые глубокие замкнутые изолинии, оконтуривающие поднятия на структурных картах отражающих горизонтов, совпадают с водонефтяными контактами (ВНК). Визуальный анализ зон антиклинальных поднятий, находящихся выше соответствующих уровней ВНК, позволил выделить три антиклинальных поднятия, которые в некотором приближении совпадают пространственно для средних и нижних каменноугольных отложений всех трёх нефтеносных комплексов. Эти поднятия представлены на рис. 8.1 наиболее глубокими замкнутыми изолиниями горизонта ОГ-2.

Признаки нефтеносности оценивались по данным высокоразрешающей электроразведки с измерением параметров вызванной поляризации (ВРЭ-ВП). Для прогноза мест заложения поисковых скважин выделены только те зоны с признаками нефтегазоносности, которые



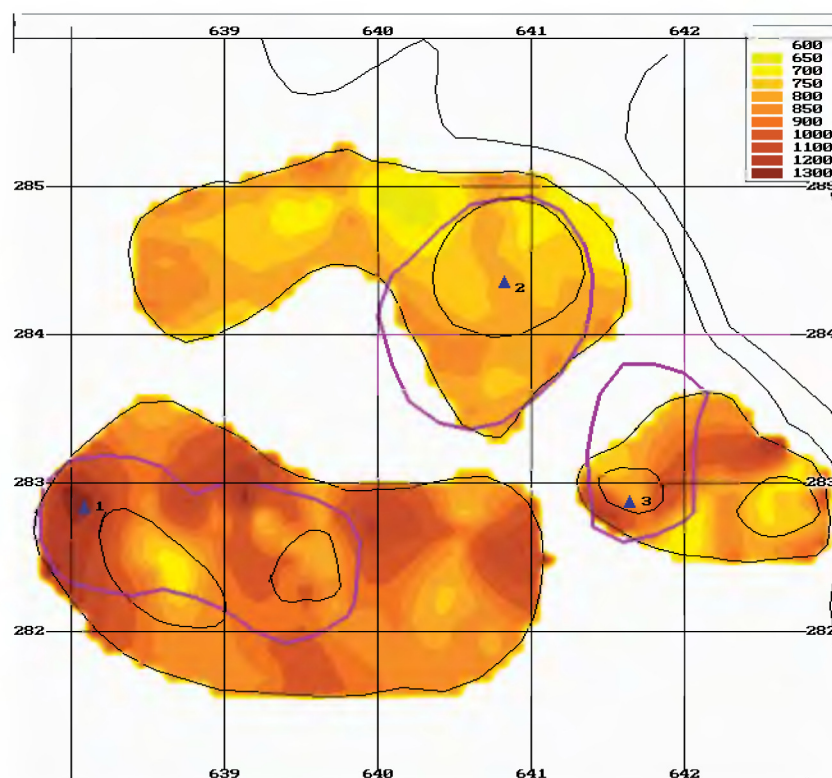


Рис. 8.1. Прогноз мест заложения поисковых скважин на одном из разведочных участков Удмуртии. На карте показаны: сеточная модель суммарной удельной ёмкости резервуаров среднего и нижнего карбона в см в пределах антиклинальных поднятий отражающего горизонта ОГ-2, тонкими линиями показаны часть изолиний глубины залегания ОГ-2, жирными линиями показаны прогнозные контуры нефтеперспективных объектов среднего и нижнего карбона по данным высокоразрешающей электроразведки с измерением параметров вызванной поляризации (ВРЭ-ВП), треугольниками с номерами 1, 2 и 3 указаны проектные скважины

пересекаются с указанными выше поднятиями. Контуры этих зон показаны на рис. 8.1 жирными красными линиями. Можно предполагать, что зоны пересечения контуров возможного наличия нефтеносности по данным ВРЭ-ВП и наиболее глубоких замкнутых изолиний горизонта ОГ-2 соответствуют границам нефтяных залежей в каменноугольных отложениях.

Для оценивания свойств коллекторов использовались средства аналитических преобразований ГИС ГЕО 2.5. С их помощью были вычислены сеточные модели удельных ёмкостей коллекторов трёх нефтегазоносных комплексов, равные произведению толщины пласта коллектора

на его пористость. Затем была найдена суммарная удельная мощность всех 10 каменноугольных пластов коллекторов. Сеточная модель суммарной удельной ёмкости в пределах антиклинальных поднятий показана на рис. 8.1. Выбор мест заложения поисковых скважин производился по максимальному значению суммарной удельной ёмкости в пределах пересечения контуров возможного наличия нефтеносности по данным ВРЭ-ВП и наиболее глубоких замкнутых изолиний горизонта ОГ-2. На рис. 8.1 синими треугольниками показаны выбранные места заложения трёх поисковых скважин.

Очевидно, что учёт площади поднятий и распределения значений суммарной удельной ёмкости позволяет оценить возможные нефтенасыщенные объёмы антиклинальных ловушек как произведение их площади на удельную ёмкость. Этот параметр также может быть использован как при выборе мест бурения, так и для установления последовательности заложения поисковых скважин.

К настоящему времени результаты данной работы частично подтверждены геологоразведочным предприятием Удмуртии. Было проведено бурение поисковой скважины в точке № 2. Бурение подтвердило наличие залежи нефти в каменноугольных отложениях.

### **8.3. Оценка устойчивости городской территории к динамическим воздействиям <sup>1)</sup>**

**8.3.1. Экспертные знания.** Эффект динамических воздействий зависит от интенсивности и спектрального состава воздействия, от удаления и взаимного расположения источника и объекта воздействия, волновых свойств геологической среды, а также от технических характеристик инженерных сооружений.

Устойчивость территории к динамическим воздействиям определяется тремя группами свойств геологической среды: 1) способностью компонентов геологической среды передавать динамические воздействия на инженерные сооружения; 2) резонансными свойствами грунтов; 3) изменением геологической среды под воздействием динамических нагрузок — степенью изменчивости прочностных характеристик грунтов, морфометрических и морфологических характеристик рельефа.

Из исследований по сейсмическому микрорайонированию (*Медведев ред., 1977*) и на основе экспертных знаний влияние каждого отдельного фактора на устойчивость территории к динамическим воздействиям можно оценить в баллах шкалы МСК. Однако зависимости между компонентами геологической среды носят сложный и неоднозначный характер. Поэтому комплексная оценка устойчивости терри-

<sup>1)</sup> Работа выполнена по данным, подготовленным в Институте Геоэкологии РАН и оцифрованным в ИППИ РАН.

тории даётся в условной шкале, полученной на основе суммирования баллов от отдельных факторов.

Для данной работы были выбраны следующие характеристики геологической среды (*Лихачёва и др., 1993*):

- Рельеф: морфометрия (уклоны и глубина расчленения), генезис.
- Грунт: литология, тип геологического разреза.
- Грунтовые воды: глубина залегания (минимальная).
- Техногенные изменения рельефа и грунтовых условий: техногенные отложения (мощность).

В качестве экспериментальной территории взят Хорошевский район Москвы, расположенный на северо-западной части города в долине реки Москвы на её левом берегу. Здесь отмечаются все три террасы реки и значительно развита пойма. Долина Москвы-реки на данном участке города имеет унаследованный характер и по геологическим данным здесь установлена так называемая доюрская долина. На большей части территории четвертичные существенно песчаные аллювиальные и лювигляциальные отложения залегают на меловых и юрских глинах, являющихся главным водоупорным горизонтом в геологическом разрезе столицы. Исключением является участок палеодолины, где юрские глины размыты, и четвертичные отложения залегают на трещиноватых и частично закарстованных карбоновых известняках. Рельеф Хорошевского района достаточно спокойный — с незначительными уклонами и расчленённостью, но вдоль реки отмечаются крутые склоны, среди которых есть и оползневые. В процессе градостроительства рельеф района претерпел некоторые изменения — прежде всего, засыпаны небольшие овраги, проведены подсыпки и срезки неровностей рельефа.

Из литературы (*Медведев ред., 1977, Сергеев 1978*) известно, что обводненные грунты менее устойчивы к динамическим воздействиям, чем необводненные. В пределах верхней 10-метровой толщи повышение грунтовых вод влечёт постоянное приращение балльности. При залегании грунтовых вод около 4 м от поверхности в песчаных, супесчаных и глинистых грунтах балльность территории повышается примерно на 0,5 балла; если грунтовые воды залегают близко от поверхности — на 1 балл. При залегании грунтовых вод глубже 10 м их влияние на приращение балльности не сказывается.

Четвертичные отложения — пески, суглинки и супеси — характеризуются приблизительно одинаковой устойчивостью к динамическим воздействиям. Отличаются лишь мелкие и пылеватые пески, увеличивающие неустойчивость приблизительно на 0,5 балла и, особенно, рыхлые суглинки и супеси (к которым могут быть отнесены техногенные отложения), повышающие неустойчивость на 1–2 балла.

Отмечаются факты значительного усиления разрушающего динамического воздействия на сооружения, расположенные на крутых склонах холмов, оврагов и на обрывистых берегах. Для склонов, сложенных супесчано-суглинистыми и песчаными несцементированными грунта-

ми, при крутизне поверхности склона более 10 градусов возможно увеличение интенсивности динамического воздействия на 1 балл.

### 8.3.2. Построение комплексной функции неустойчивости.

Для оценивания модели устойчивости территории к динамическим нагрузкам введём функцию, которая монотонно возрастает при усилении факторов, уменьшающих степень устойчивости. Будем называть эту зависимость функцией неустойчивости. Будем конструировать функцию неустойчивости в виде суммы нелинейных функций от отдельных факторов:

$$F(\mathbf{x}) = \sum_{i=1}^5 \varphi_i(x_i),$$

где  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_I)$  — вектор факторов, определяющих неустойчивость,  $\varphi_i(x_i)$  — функция от  $i$ -го фактора, определяющая его вклад в неустойчивость. Функция неустойчивости показывает, как каждый из факторов в отдельности и их совокупность влияет на неустойчивость территории. Факторы  $x_i$  и функции вкладов  $\varphi_i(x_i)$  подбирались экспертом на основе экспертных представлений о значимости каждого фактора и характере его влияния на устойчивость территории и с использованием неформального знания об устойчивости участков изучаемой территории. Для этого применялась следующая процедура поиска версии решения:

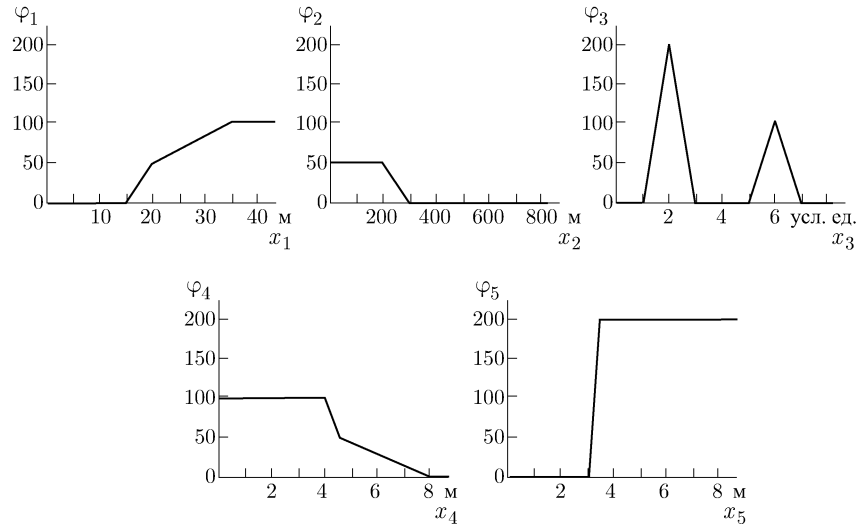


Рис. 8.2. Функции  $\varphi_i(x_i)$ :  $x_1$  — показатель вертикальной расчленённости рельефа;  $x_2$  — высота рельефа над уровнем реки;  $x_3$  — тип литологического строения по карте инженерно-геологического строения;  $x_4$  — уровень грунтовых вод;  $x_5$  — мощность техногенных отложений

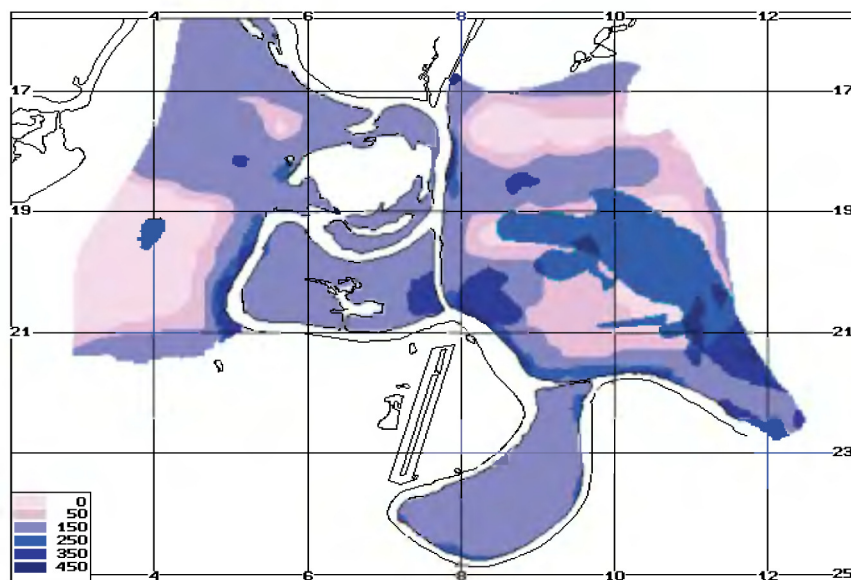
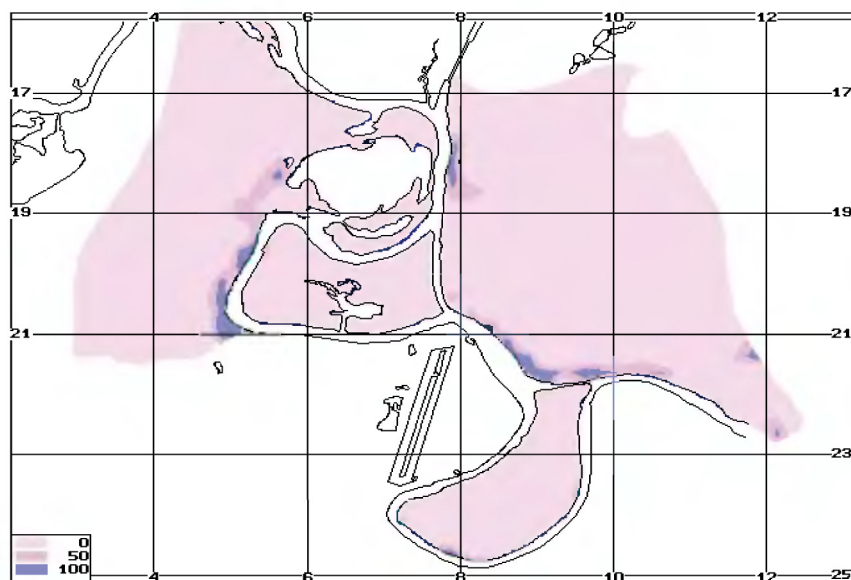


Рис. 8.3. Модель неустойчивости территории

Рис. 8.4. Модель вклада признака расчленённость рельефа  $\varphi_1(x_1)$

1. Формализация экспертных знаний в виде функции неустойчивости.
2. Построение версии цифровой сеточной модели неустойчивости для изучаемой территории и моделей вкладов в неустойчивость геологических факторов.
3. Анализ версии решения экспертом. При обнаружении участков территории, степень неустойчивости которых противоречит экспертному представлению, перейти к п. 4. В противном случае принять полученную версию прогноза.
4. Модификация функций  $\varphi_i(x_i)$  и переход к п. 2.

Таким образом, получаемое решение основано только на экспертных знаниях и на данных о геологической среде. Задача эксперта состоит в том, чтобы структурировать знания об устойчивости территории к динамическим воздействиям в виде формального правила и построить цифровую сеточную модель районирования. Применение ГИС позволяет эксперту проанализировать ряд альтернативных прогнозирующих функций, выбрать наиболее адекватную модель районирования и откорректировать правило районирования, используя свои неформальные знания.

В окончательной версии решения для описания свойств грунтовых условий, влияющих на устойчивость территории к динамическим воздействиям, выбраны следующие пять факторов.

1.  $x_1$  — показатель вертикальной расчленённости территории в метрах, полученный как разность максимальной и минимальной отметок высоты на квадратной площадке со стороной 250 м.
2.  $x_2$  — высота территории над уровнем реки в метрах.
3.  $x_3$  — тип литологического строения по геологическому разрезу. Эти данные были взяты с карты инженерно-геологического строения территории района.
4.  $x_4$  — уровень грунтовых вод в метрах.
5.  $x_5$  — мощность техногенных отложений в метрах.

Факторы  $x_1, x_2, x_3, x_4, x_5$  представлены в виде растровых полей в координатной сетке с шагом 25 на 25 м. В табл. 8.1 приведена матрица корреляции полей  $x_1, x_2, x_4, x_5$ . Включение в матрицу корреляции фактора  $x_3$  не имеет смысла, так как он представлен в номинальной шкале.

Таблица 8.1. Матрица корреляции полей факторов  $x_1, x_2, x_4, x_5$

	$x_1$	$x_2$	$x_4$	$x_5$
$x_1$	1	-0,27	-0,21	0,09
$x_2$	-0,27	1	0,63	
$x_4$	-0,21	0,63	1	0,04
$x_5$	-0,09	0,05	0,04	1

Из матрицы видно, что корреляция между факторами практически отсутствует за исключением пары факторов  $x_2, x_4$ . Слабая корреляция является некоторым обоснованием того, что функция неустойчивости от комплекса факторов может быть сконструирована в виде суммы функций неустойчивости отдельных факторов. Коэффициент корреляции факторов  $x_2, x_4$  равен 0,63. Тем не менее, в комплексную оценку включены оба фактора. Добавление к фактору  $x_4$  фактора  $x_2$  позволяет подчеркнуть увеличение неустойчивости территории в районе поймы, где неустойчивость растёт как за счёт уровня грунтовых вод, так и за счёт наличия лёссовых отложений.

Функции  $\varphi_i(x_i)$ , заданные экспертом в отобранной версии районирования, показаны на рис. 8.2. Для функции  $\varphi_1(x_1)$  предполагалось, что, начиная с 15 м при увеличении  $x_1$ , функция вклада возрастает от 0 до 100 баллов. Функция  $\varphi_2(x_2)$  формализует предположение о том, что территория, расположенная выше уровня реки до отметки 1 м, является поймой, а территория, расположенная выше уреза реки на 3 м, уже не является поймой. Влияние пойменных грунтов на устойчивость территории оценено экспертом в 50 баллов. При построении функции  $\varphi_3(x_3)$  вклад участков с различным типом геологического строения в неустойчивость оценивался следующим образом: 100 баллов для оползневого склона ( $x_3 = 6$ ), 200 баллов для палеодолины ( $x_3 = 2$ ) и 0 для прочих типов литологического строения. Для учёта влияния уровня грунтовых вод использовалась аппроксимация формулы Медведева (*Медведев ред., 1977*), согласно которой  $\varphi_4(x_4) = B \exp(-0,04x_4)$ , где  $B$  — коэффициент, зависящий от типа грунтов. Для данного примера принято  $B = 1$ . Предполагалось, что влияние  $\varphi_5(x_5)$  отложений свыше 3,5 м увеличивает динамическое воздействие на 200 баллов, отложения до 3 м не дают увеличения неустойчивости, а от 3 до 3,5 м воздействие плавно (линейно) увеличивается до 200 баллов.

**8.3.3. Цифровая модель неустойчивости территории.** Цифровая сеточная модель, вычисленная по функции неустойчивости  $F(x_1, x_2, x_3, x_4, x_5)$  для всех точек координатной сетки Хорошевского района, изображена на рис. 8.3.

На карте представлены 4 группы зон со значениями функции неустойчивости: 0–49, 50–149, 150–299 и от 300-х и более условных единиц (у.е.). Выделенные территории различаются по сложности инженерно-геологических условий: чем выше балл, тем большее число характеристик геологической среды вносит вклад в динамическую неустойчивость территории и тем дороже обходятся инженерно-строительные мероприятия при её освоении.

К устойчивым отнесены участки, оценённые в 0–49 у.е. Они составляют 30 % площади исследуемой территории. Относительно устойчивыми могут считаться участки, получившие в сумме 50–149 у.е. (эти значения могут быть получены за счёт вклада одного-двух показателей геологической среды), составляющие 29 % площади. Участки, получившие 150–299 у.е., как правило, неустойчивы за счёт комплекса

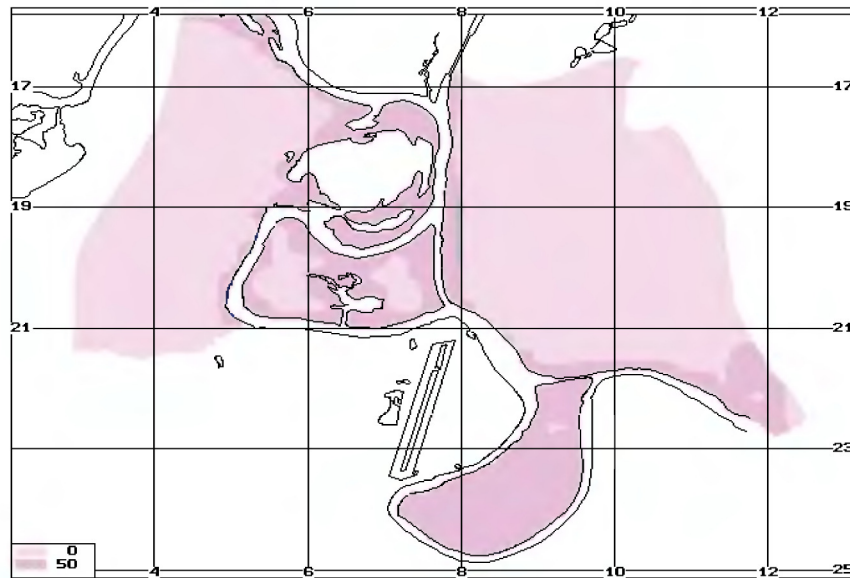


Рис. 8.5. Модель вклада признака высота территории над уровнем реки  $\varphi_2(x_2)$

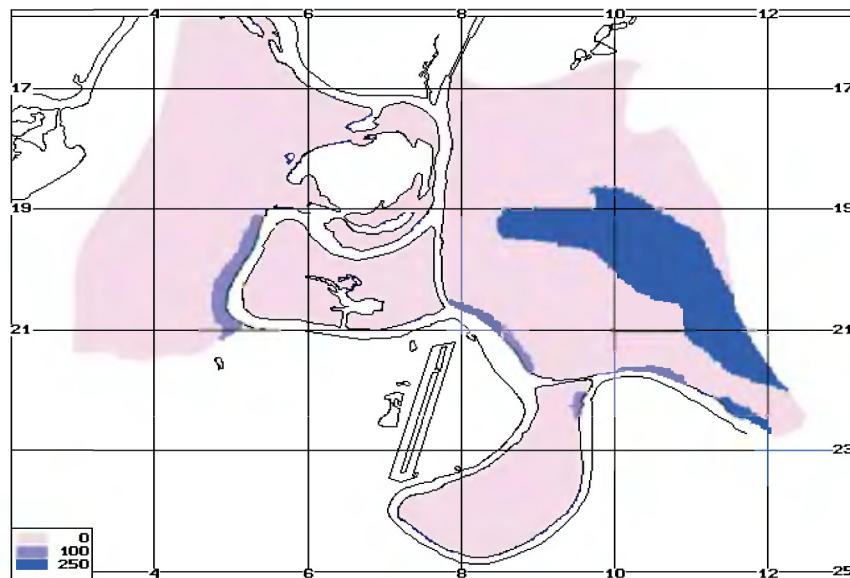
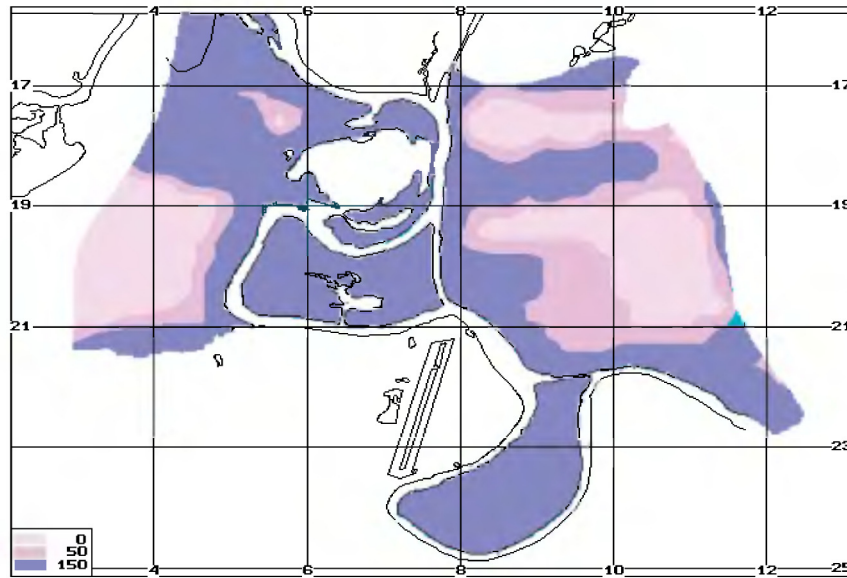
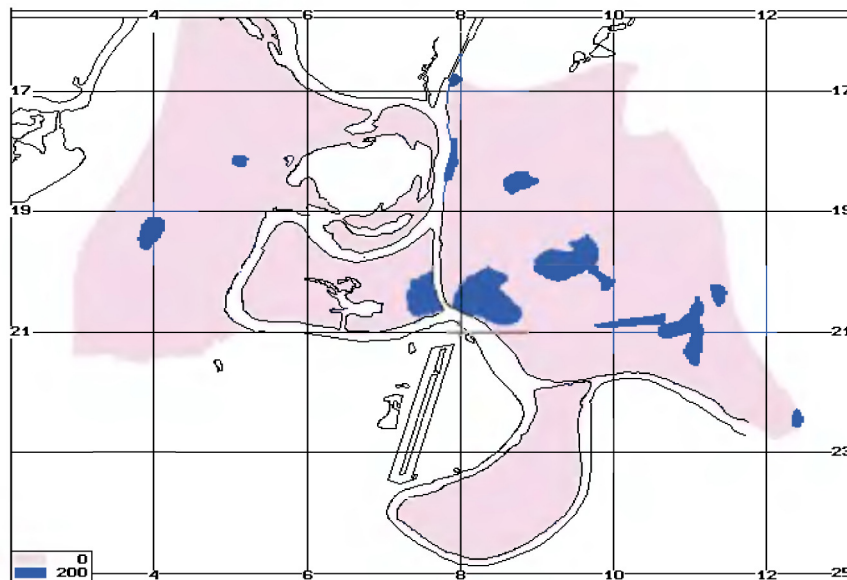


Рис. 8.6. Модель вклада признака тип литологического строения  $\varphi_3(x_3)$



Рис. 8.7. Модель вклада признака уровень грунтовых вод  $\varphi_4(x_4)$ Рис. 8.8. Модель вклада признака мощность техногенных отложений  $\varphi_5(x_5)$

факторов: оползневые склоны, подтопленная пойма, карстовосуффозионный участок палеодолины р. Москвы. Эти участки составляют 0,35 % площади. Особо неустойчивыми являются участки, получившие при комплексной оценке от 300 у.е. и выше. Они составляют 6 % площади.

На рис. 8.4–8.8 показаны сеточные модели вкладов  $\varphi_i(x_i)$  геологических факторов в динамическую неустойчивость территории  $F(\mathbf{x})$ . На основании этих моделей можно проанализировать компоненты грунтовых условий, определяющие степень неустойчивости для каждого участка территории.

В заключение следует заметить, что данный пример в существенной степени является иллюстративным. Комплексная оценка устойчивости территории в примере дана лишь в первом приближении. Так, например, не учитывались инженерно-строительные мероприятия по увеличению устойчивости, такие как дренаж или укрепление склонов. Не рассматривался также временной аспект устойчивости, вызванный необходимостью учитывать изменения во времени характеристик элементов геологической среды. В то же время полученная модель устойчивости геологической среды к динамическим воздействиям может быть использована для построения карты экологического дискомфорта от динамических техногенных источников.

## 8.4. Оценка ущерба от землетрясений

**8.4.1. Основные понятия.** В данном разделе рассматриваются элементы технологии оценки ущерба от землетрясений и пример построения карт экономического ущерба от землетрясений и сейсмического риска для г. Иркутска, в котором использованы данные, любезно предоставленные ИЗК СО РАН, ВостсибГИСИЗ и ДИГЕСТА. Для решения задач оценки сейсмического риска в ИППИ РАН разработана ГИС ГеоРиск (*Gitis, 1999*). Система применялась (*Кофф и др., 1995, Гитис и др. 1997, Amendola et al., 2000*) для оценки сейсмического риска и ущерба от землетрясений для территорий городов Иркутска, Улан-Удэ и населённых пунктов на островах Курильской гряды.

Экономический ущерб и людские потери при землетрясении определяются с одной стороны интенсивностью, спектром, длительностью и моментом времени сейсмического воздействия, а с другой стороны — устойчивостью сооружений к воздействию, структурными связями между элементами инфраструктуры, назначением, размерами и удельной стоимостью сооружений.

Устойчивость сооружений к сейсмическому воздействию задаётся уязвимостью. Уязвимость зависит от типа сооружений и от параметров сейсмических колебаний. Она равна среднему отношению затрат на восстановление сооружения, разрушенного при определённом воздействии, к первоначальной стоимости сооружения.

При анализе экономических потерь обычно используются два основных понятия: экономический ущерб и сейсмический риск.

Причиной ущерба при землетрясении являются следующие три фактора:

- Прямое сейсмическое воздействие (прямой ущерб).
- Косвенные последствия разрушения от землетрясения (косвенный ущерб).
- Вторичные природные и техногенные воздействия, индуцированные землетрясением (вторичный ущерб).

Под сейсмическим риском понимается вероятность того, что за определённое время на фиксированной территории ущерб от землетрясений будет превышать некоторую заданную величину. Обычно в расчётах используется средний риск, определяемый как математическое ожидание ущерба на фиксированной территории за определённое время.

Прямой экономический ущерб определяется стоимостной оценкой потерь из-за негативного изменения или разрушения сооружений. Косвенный ущерб определяется влиянием разрушения одних объектов инфраструктуры на функционирование других, в том числе непосредственно неповреждённых объектов.

**8.4.2. Алгоритм расчёта сейсмического риска прямого ущерба.** Рассматриваемый алгоритм оценки сейсмического риска прямого ущерба применим для точечных объектов (т.е. объектов, размеры которых много меньше размеров неоднородностей поля сейсмических воздействий от всех возможных источников) и для таких неточечных объектов, ущерб которых может быть представлен как сумма ущербов их элементарных частей.

Предполагается, что сейсмические воздействия подчиняются закону Пуассона. Тогда вероятность того, что за время  $t$  в точке  $x$  произойдёт хотя бы одно землетрясение с интенсивностью в интервале от  $I_{\min}(x)$  до  $I_{\max}(x)$

$$p(x, t) = 1 - \exp \{-\lambda(x)t\}, \quad (8.1)$$

а вероятность того, что это сейсмическое воздействие имеет интенсивность  $i$  равна

$$p(i/x, t) = \frac{\lambda_i(x)}{\lambda(x)}, \quad (8.2)$$

где:  $\lambda(x)$  — интенсивность суммарного потока сейсмических воздействий класса от  $I_{\min}(x)$  до  $I_{\max}(x)$ ;  $\lambda_i(x)$  — интенсивность потока сейсмических воздействий класса  $i$ ;

Обозначим ущерб в точке  $x$  через  $D(x)$ . Ущерб от события класса  $i$

$$D_i(x) = \sum_{j=1}^J d_{ij} S_j(x), \quad (8.3)$$

где:  $d_{ij} = v_{ij} C_j$  — удельный ущерб при сейсмическом воздействии  $i$  для сооружения типа  $j$ ,  $v_{ij}$  — уязвимость сооружения типа  $j$  при воздействии  $i$ ,  $C_j$  — стоимость единицы площади сооружения типа  $j$ ;  $S_j(x)$  — плотность застройки типа  $j$  в точке  $x$ .

Средний ущерб от события в классе от  $I_{\min}(x)$  до  $I_{\max}(x)$  равен

$$\bar{D}(x) = M\{D(x)\} = \sum_{i=I_{\min}}^{I_{\max}} D_i(x)p(i/x, t) = \sum_{i=I_{\min}}^{I_{\max}} D_i(x) \frac{\lambda_i(x)}{\lambda(x)}. \quad (8.4)$$

Теперь можно записать формулы для оценки риска. Средний удельный риск в точке  $x$  за время  $t$  определим как математическое ожидание потерь (ущерба) на единицу площади застройки в точке  $x$ .

1. Средний суммарный риск от всех событий за время  $t$

$$r_s(x, t) = \bar{N}(x)\bar{D}(x) = \lambda(x)t\bar{D}(x). \quad (8.5)$$

2. Средний риск от первого события за время  $t$

$$r_1(x, t) = p(x, t)\bar{D}(x). \quad (8.6)$$

3. Средний риск от максимального события за время  $t$

$$\begin{aligned} r_{\max}(x, t) &= \\ &= \exp \left\{ -\lambda(x)t \sum_{i=I_{\min}}^{I_{\max}} D_i(x) \left( \exp \left\{ \sum_{k=I_{\min}}^i \lambda_k(x)t \right\} - \exp \left\{ \sum_{k=I_{\min}}^{i-1} \lambda_k(x)t \right\} \right) \right\}. \end{aligned} \quad (8.7)$$

Заметим, что при подсчёте риска в (8.7) можно добавить также ущерб от функции экономических затрат на антисейсмические мероприятия, которые приводят к уменьшению прямых потерь. В этом случае возможна оптимизация суммарного риска по величине затрат на повышение устойчивости сооружений к сейсмическим воздействиям.

Риск по произвольной области  $S$  определяется как математическое ожидание потерь по этой области

$$R(S, t) = M_t \int_{x \in S} ds D(x) = \int_{x \in S} ds M_t D(x) = \int_{x \in S} ds r(x, t). \quad (8.8)$$

**8.4.3. Система ГеоРиск.** Система ГеоРиск (*Gitis, 1999*) предназначена для оценивания прямого и косвенного ущерба и риска от природных катастроф. Система состоит из 6 подсистем.

*Подсистема преобразования формата* картографических, текстовых и табличных данных из ASCII формата во внутренний формат ГЕО и наоборот. Эта подсистема обеспечивает интерфейс системы ГеоРиск с другими геоинформационными системами и базами данных.

*Подсистема векторизации и редактирования картографических данных*, вводимых с дигитайзера или сканера. Подсистема обеспечивает оцифровку и редактирование картографической информации (контуры, пиктограммы, тексты, области, изолинии, границы зон, пункты, линейные объекты) в векторном и растровом форматах.

*Подсистема оценивания и построения карт прямого ущерба и среднего риска.* Подсистема поддерживает обработку данных, связанную с построением исходной топологии, построением пересечения карт сейсмического районирования и карт застройки территории. Подсистема строит карты суммарного риска от всех событий за время  $t$

(8.5), риска от первого события за время  $t$  (8.6), риска от максимального события за время  $t$  (8.7) и оценивает риск по произвольной области  $S$  (8.8) для площадных, линейных и точечных объектов.

*Подсистема оценивания косвенного ущерба и риска.* Подсистема содержит средства для расчёта косвенного риска и возможного максимального ущерба (экономического и социального), расчёта условного риска или ущерба, сравнения фактического ущерба с расчётным, средства для построения электронных карт ущерба/риска, сетевой диаграммы связей объектов для косвенного ущерба/риска и текстового отчёта.

*Подсистема оценивания прямого и косвенного ущерба от различных типов катастроф (землетрясений, пожаров, наводнения, цунами) для отдельных типов объектов, таких как сельскохозяйственные угодья, жилая застройка, промышленные объекты и элементы инфраструктуры.*

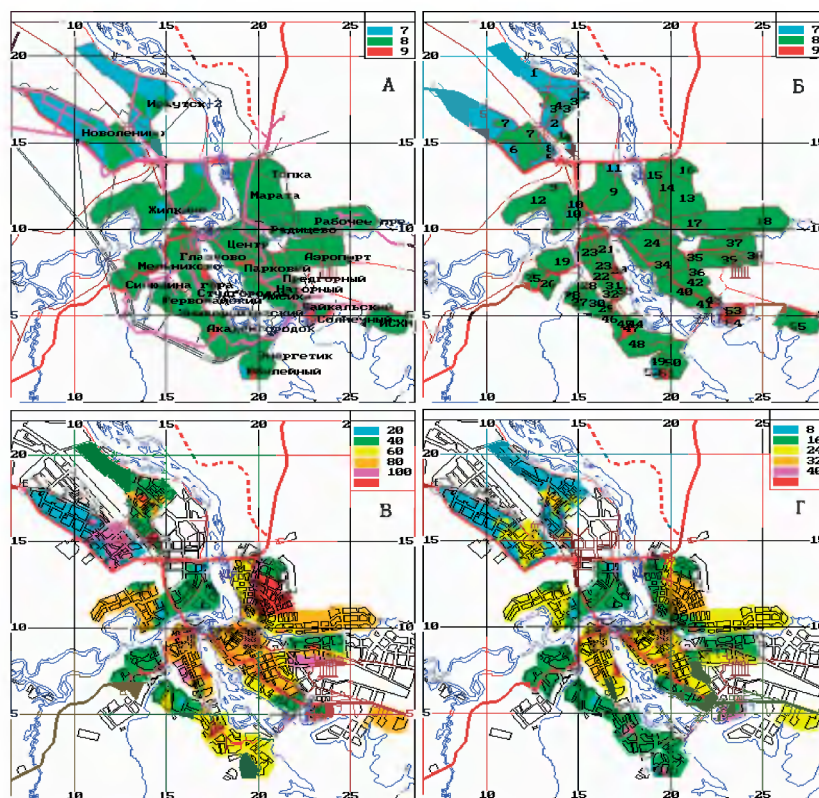


Рис. 8.9. Результаты применения ГИС ГеоРиск для анализа возможных последствий землетрясений в г. Иркутске

Подсистема научной картографии, обеспечивающая визуализацию и комплексный анализ картографической информации.

**8.4.4. Оценка прямого ущерба и риска для территории г. Иркутска.** Жилая застройка г. Иркутска состоит из 9 типов сооружений. По карте сейсмического микрорайонирования (СМР) (*Карта Сейсмического микрорайонирования г. Иркутска, 1989*) в городе выделяются зоны, в которых возможны максимальные сотрясения в 7, 8 и 9 баллов.

В табл. 8.2 дана уязвимость каждого типа сооружений при сотрясениях в 7, 8, и 9 баллов и цена одного квадратного метра застройки в условных единицах.

Таблица 8.2. Уязвимость сооружений от землетрясений

№	Тип застройки	Код	Цена 1 м <sup>2</sup> в условных единицах	Интенсивность сотрясения в баллах		
				7	8	9
1	Крупнопанельная	7КП	516,5	0,035	0,119	0,37
2	Крупнопанельная	7-8КП	516,5	0,035	0,119	0,37
3	Крупнопанельная	8КП	516,5	0,024	0,066	0,221
4	Кирпичная	БКЧ	470,4	0,103	0,37	0,757
5	Кирпичная	7КЧ	516,5	0,058	0,246	0,583
6	Кирпичная	8КЧ	516,5	0,04	0,152	0,439
7	Каркасная	7КС	516,5	0,04	0,117	0,471
8	Каркасная	8КС	516,5	0,029	0,103	0,336
9	Деревянная	ВД	470,4	0,035	0,134	0,404

В табл. 8.3 даны вероятности хотя бы одного землетрясения интенсивности  $I$  баллов в зонах с сейсмичностью  $j$  баллов за 1 год.

На рис. 8.9-А показано разбиение территории г. Иркутска на районы и на зоны с одинаковой бальностью по карте сейсмического микрорайонирования, условными знаками показаны элементы инфраструктуры (дороги, железнодорожные и трамвайные пути, линии электропередачи и теплоцентрали, мосты, вокзалы, порты и пр.).

Таблица 8.3. Вероятность сотрясений за 1 год

Балл по СМР	Вероятность хотя бы 1 землетрясения с интенсивностью		
	7	8	9
7	0,005	—	—
8	0,008	0,005	—
9	0,02	0,008	0,005

На рис. 8.9-Б показано разбиение территории г. Иркутска на зоны с однотипной застройкой и на зоны с одинаковой балльностью по карте СМР. На рис. 8.9-В приведена карта прямого экономического удельного ущерба от максимального возможного землетрясения. Удельный ущерб для каждой зоны усреднён по типам застройки. На рис. 8.9-Г приведена карта среднего удельного сейсмического риска от первого события (8.6) за 50 лет.

## 8.5. Прогнозирование при локальном взаимодействии компонент пространственной структуры

**8.5.1. Математическая модель.** Наиболее сложные задачи пространственно-временного геоинформационного анализа относятся к исследованию внутренних и взаимных связей в географических данных. Изучение связей между свойствами географических сущностей позволяет по комплексу известных пространственных и пространственно-временных свойств прогнозировать заранее неизвестные свойства или обнаруживать и распознавать заранее неизвестные географические объекты и явления. Изучение связей между географическими объектами и явлениями позволяет исследовать пространственно-временные процессы их взаимодействия и прогнозировать их стационарные свойства (*Gitis, et al., 1994, 1996*).

В данном разделе рассматривается модель процесса взаимодействия компонент сложной пространственной структуры, которая может состоять как из географических объектов, так и из условно выделенных фрагментов региона. Основное предположение модели состоит в том, что компоненты структуры взаимодействуют только локально, то есть только при наличии прямых пространственных или функциональных связей. В качестве примеров рассматриваются распространение загрязняющих веществ с поверхностным стоком, а также цепочки вторичных природных и природно-техногенных катастроф, возникающих как последствия спонтанных катастроф, таких как землетрясения (*Гитис, Петрова, Пирогов, 2004*).

Пространственная структура описывается конечным ориентированным графом, вершины которого соответствуют компонентам рассматриваемой структуры, а направленные рёбра — возможным прямым воздействиям одной компоненты пространственной структуры на другую. Количественно структура описывается вектором  $\mathbf{m} = (m_1, m_2, \dots, m_n)$ , где  $m_i$  — некоторая количественная характеристика  $i$ -й компоненты, смысл которой устанавливается из контекста конкретной задачи. Вектор  $\mathbf{m}$  мы будем называть вектором потенциалов структуры, а величину  $m_i$  — потенциалом  $i$ -й компоненты структуры. Кроме того, считаются заданными величины воздействий извне на каждую компоненту структуры, описываемые вектором  $f = (f_1, \dots, f_n)$ , а также степени прямого воздействия компонент пространственной структуры друг на друга, распространяющиеся по рёбрам направленного графа: ребру  $(i, j)$  приписывается величина  $p_{ij}$  воздействия компоненты  $i$  на

компоненту  $j$ . Таким образом, имеется матрица внутренних воздействий  $\mathbf{P} = (p_{ij})$ . В данной модели предельный (стационарный) вектор потенциалов структуры удовлетворяет уравнению

$$\mathbf{m} = \mathbf{f} + \mathbf{mP} \quad (8.9)$$

и, следовательно,

$$\mathbf{m} = \mathbf{f}(\mathbf{E} - \mathbf{P})^{-1} = \mathbf{f}(\mathbf{E} + \mathbf{P} + \mathbf{P}^2 + \dots + \mathbf{P}^K + \dots) \quad (8.10)$$

Перейдём к рассмотрению задач, которые могут быть формализованы при помощи описанной модели.

**8.5.2. Цепочки взаимозависимых природных и природно-технологических катастроф.** Рассмотрим структуру, компонентами которой являются географические объекты, например, ледники, плотины, заводы, водохранилища, а также элементы среды некоторой пространственной области. Предположим, что любой объект может находиться в одном из двух состояний: нормальном и аномальном. При этом, исходя из конкретной ситуации, пользователю предоставляется самому провести разграничение между нормальным и аномальным состояниями объекта. Если объект  $i$  находится в аномальном состоянии, мы будем говорить, что он подвергся катастрофе. Считается, что объекты могут переходить в аномальное состояние в результате прямых внешних воздействий либо воздействий друг на друга. Прямое воздействие возможно, если объект, индуцирующий аномальное состояние у других объектов, либо находится в непосредственной близости от них, либо связан с ними функционально.

Перенумеруем все подверженные воздействиям объекты, относящиеся к исследуемой области, числами  $i = 1, 2, \dots, n$ . Построим ориентированный граф  $\mathbf{G}$ , каждая вершина которого соответствует одному из объектов. Соединим вершину  $i$  с вершиной  $j$  ориентированной дугой, если катастрофа объекта  $i$  может непосредственно индуцировать катастрофу на объекте  $j$ .

Пусть известно, что разрушительное внешнее (спонтанное) воздействие на объект  $i$  (то есть такое воздействие, что объект переходит из нормального в аномальное состояние), наступает в среднем один раз в  $T_i$  лет. Обозначим  $f_i = 1/T_i$  среднюю частоту (интенсивность) переходов объекта  $i$  в аномальное состояние под действием внешних факторов. Вектор  $\mathbf{f} = (f_1, \dots, f_n)$  мы будем называть вектором внешних воздействий на структуру. Вектор потенциалов структуры  $\mathbf{m} = (m_1, \dots, m_n)$  составим из частот (интенсивностей)  $m_i$  перехода объекта  $i$  в аномальное состояние (катастрофа на объекте  $i$ ). Отметим, что  $m_i$ , как и  $f_i$ , может принимать значения, большие 1. Если, например,  $m_i = 3$ , это означает, что за год в среднем объект  $i$  может перейти в аномальное состояние три раза. Соответственно, если  $m_i = 1/3$ , это означает, что данный объект может перейти в аномальное состояние в среднем один раз в три года.

Чтобы оценить интенсивность возможных катастроф на объекте  $j$ , нужно учесть влияние всех возможных внешних воздействий на все



объекты структуры и то, в какой степени эти внешние воздействия по цепочкам влияющих друг на друга объектов передадутся объекту  $j$ .

Сценарий развития влияния произошедшей на объекте  $i$  катастрофы на другие объекты мы будем абстрактно описывать процессом с дискретным временем. Заметим, что физически время передачи влияния катастрофы с одного объекта на другой различается в различных ситуациях и для различных объектов. В этом смысле время в данном случае описывает лишь шаги по цепочкам развивающихся катастроф.

Пусть  $p_{ij}$  — вероятность того, что перешедший в аномальное состояние объект  $i$  непосредственно повлияет на объект  $j$  таким образом, что объект  $j$  также перейдёт в аномальное состояние. Тогда величина  $f_i p_{ij}$  интерпретируется как степень возможного воздействия аномального объекта  $i$  (ставшего таким под влиянием воздействия извне) на объект  $j$ , точнее, как частота перехода объекта  $j$  в аномальное состояние лишь в результате прямого влияния объекта  $i$ .

Обозначим  $\mathbf{m}^{(t)} = (m_1^{(t)}, \dots, m_n^{(t)})$  — вектор частот  $m_i^{(t)}$  переходов в аномальное состояние объектов  $i = 1, \dots, n$  под влиянием как непосредственно внешнего воздействия, так и воздействий всех других объектов, дошедших до объекта  $i$  по цепочкам, количество шагов в которых не превосходит  $t - 1$ . Таким образом, для  $t = 1$  получаем

$$\mathbf{m}^{(1)} = \mathbf{f}.$$

В следующий момент времени,  $t = 2$ , мы учитываем влияние как внешних воздействий, так и непосредственных воздействий соседних объектов, то есть цепочки длины 1. Получаем:

$$m_j^{(2)} = f_j + \sum_{i \neq j} f_i p_{ij},$$

или, в векторном виде,

$$\mathbf{m}^{(2)} = \mathbf{f} + \mathbf{fP} = \mathbf{f}(\mathbf{E} + \mathbf{P}),$$

где  $\mathbf{E}$  — единичная матрица.

Далее, влияние «соседей порядка 2» скажется на следующем шаге развития сценария, то есть вектор

$$\mathbf{m}^{(3)} = \mathbf{f} + \mathbf{m}^{(2)}\mathbf{P} = \mathbf{f}(\mathbf{E} + \mathbf{P} + \mathbf{P}^2)$$

описывает возможности возникновения катастроф, пришедших по цепочке длины 2, длины 1 и непосредственно от внешнего воздействия.

Заметим, что векторы  $\mathbf{m}^{(t+1)}$  и  $\mathbf{m}^{(t)}$  отличаются тем, что  $\mathbf{m}^{(t+1)}$ , в отличие от  $\mathbf{m}^{(t)}$ , учитывает ещё и цепочки длины  $t$ , в то время как  $\mathbf{m}^{(t)}$  учитывает лишь воздействие цепочек длины  $t - 1$  и меньше. Следовательно, векторы  $\mathbf{m}^{(t+1)}$  и  $\mathbf{m}^{(t)}$  связаны рекуррентным соотношением

$$\mathbf{m}^{(t+1)} = \mathbf{f} + \mathbf{m}^{(t)}\mathbf{P}, \quad (8.11)$$

или, покомпонентно,

$$m_j^{(t+1)} = f_j + \sum_{i=1}^n m_i^{(t)} p_{ij},$$

где  $\mathbf{P} = (p_{ij})$  — матрица непосредственных воздействий. Заметим, что матрица  $\mathbf{P}$ , в отличие от задач теории вероятностей, не является стохастической, то есть сумма элементов матрицы по строке не равна 1. Однако матрица  $\mathbf{P}$  содержит много нулей, так как  $p_{ij} \neq 0$  только если граф непосредственных воздействий  $\mathbf{G}$  содержит ребро  $(i, j)$ . Кроме того,  $0 \leq p_{ij} \leq 1$  для всех  $i, j$ .

Мы предполагаем, что цепочки вторичных катастроф не могут быть очень длинными, иными словами, найдётся число  $K$ , такое, что  $\mathbf{P}^{K+1} = \mathbf{0}$ . Тогда предельный (или стационарный) вектор частот перехода компонент среды в аномальное состояние равен

$$\mathbf{m} = \mathbf{m}^{(K+1)} = \mathbf{f} \left( \mathbf{E} + \mathbf{P} + \mathbf{P}^2 + \dots + \mathbf{P}^K \right).$$

Этот вектор мы называем предельным (или стационарным) вектором потенциалов.

Заметим, что, формально говоря, можно отказаться от условия  $\mathbf{P}^{K+1} = \mathbf{0}$ . При условии, что ряд

$$\mathbf{E} + \mathbf{P} + \mathbf{P}^2 + \dots + \mathbf{P}^K + \dots$$

сходится<sup>1)</sup>, вектор предельных потенциалов задаётся формулой:

$$\mathbf{m} = \mathbf{f} \left( \mathbf{E} + \mathbf{P} + \mathbf{P}^2 + \dots + \mathbf{P}^K + \dots \right).$$

Это есть не что иное, как решение системы уравнений (8.9)

$$\mathbf{m} = \mathbf{f} + \mathbf{mP}.$$

Если известны величины ущербов  $u_i$  от перехода объекта  $i$  в аномальное состояние, то средний ожидаемый ущерб за год будет равен

$$U = \sum_{i=1}^n u_i m_i.$$

В нашей модели процесс развития катастроф в структуре может порождаться только внешним воздействием, которое всегда переводит соответствующую компоненту среды в аномальное состояние. Кроме того, в рассмотренной модели предполагается, что каждая компонента среды перед внешним или внутренним воздействиями всегда находится в нормальном состоянии. Из этого предположения следует, любая компонента среды  $i$  при воздействии на неё может перейти в аномальное состояние и индуцировать на воздействие, которое с вероятностью  $p_{ij}$

<sup>1)</sup> Этот ряд сходится в том и только в том случае, если  $P^n \rightarrow 0$  при  $n \rightarrow \infty$ , Гантмахер Ф.Р., 1967

переведёт в аномальное состояние компоненту  $j$ . Невыполнение этого предположения, очевидно, может только уменьшить частоту переходов компонент структуры в аномальное состояние. Следовательно, наши оценки вектора потенциалов  $\mathbf{m} = (m_1, \dots, m_n)$  и среднего ущерба  $U$  являются оценками сверху.

### 8.5.3. Распространение загрязнений с поверхностным стоком.

Рассмотрим поверхность некоторой ограниченной территории, которая подвержена загрязнению. Предполагается, что частицы загрязняющего вещества попадают тем или иным способом в дождевую либо талую воду и, находясь в воде в виде взвеси или раствора, распространяются по поверхности территории с поверхностным стоком. Некоторое количество воды с загрязняющим веществом может просочиться под землю. Задача состоит в описании распределения по пространству количества загрязняющего вещества.

Разобьём рассматриваемую область на ячейки равномерной координатной сетки, достаточно малые для того, чтобы корректно учесть необходимые характеристики рельефа территории. Пусть для удобства ячейки будут квадратными. Компонентами структуры в данной задаче являются указанные ячейки, а непосредственное влияние могут оказывать друг на друга соседние, то есть имеющие общую сторону, ячейки. Перенумеруем ячейки в произвольном порядке и мысленно поместим вершины графа  $\mathbf{G}$ , отражающего возможные непосредственные воздействия компонент пространственной структуры друг на друга, в центры ячеек. Тогда граф  $\mathbf{G}$  есть просто фрагмент квадратной решётки, сдвинутой относительно координатной сетки на половину длины ребра ячейки по обеим осям (так называемая двойственная решётка). Все рёбра графа  $\mathbf{G}$ , вообще говоря, двусторонние, что отражает двусторонность взаимного влияния соседних ячеек.

В качестве вектора потенциалов структуры  $\mathbf{m} = (m_1, \dots, m_n)$  естественно взять вектор,  $i$ -й компонентой которого является количество частиц загрязнений, находящееся в  $i$ -й ячейке в виде взвеси или раствора. Естественно, в разные моменты времени это количество будет разным, то есть вектор потенциалов структуры изменяется во времени:  $\mathbf{m} = \mathbf{m}^{(t)}$ . Заметим, что в отличие от предыдущего примера (цепочки катастроф), время в данном случае совпадает с физическим временем системы. Однако для простоты изложения мы будем рассматривать процесс распространения загрязнения в дискретном времени (вариант с непрерывным временем будет упомянут в конце изложения).

Обозначим через  $\mathbf{f}^{(t)} = (f_1^{(t)}, \dots, f_n^{(t)})$  внешний входной поток, то есть  $f_i^{(t)}$  есть количество загрязнений, поступающее с водой в момент времени  $t$  в ячейку  $i$  либо с дождём, талыми водами из самой ячейки  $i$ , либо в результате аварии коммуникаций. Мы предполагаем, что каждая частица загрязнения, находящаяся в момент времени  $t$  в ячейке  $i$ , может за единицу времени с некоторыми вероятностями  $p_{ij}$  переместиться со стоком воды в одну из соседних ячеек  $j$ . Кроме того, с некоторой вероятностью  $p_{i*}$  частица, находящаяся в ячейке  $i$ ,

может просочиться сквозь землю и, таким образом, покинуть систему. С некоторой вероятностью  $p_{ii}$  частица может остаться в ячейке  $i$ . Для описания возможности просочиться, то есть покинуть систему, построим граф  $\mathbf{G}$ , добавив в него ещё одну вершину, которую обозначим  $*$ , и построим из каждой вершины ребро, идущее в  $*$ . Рёбер, выходящих из вершины  $*$ , нет. Таким образом, вершина  $*$  описывает просачивание вглубь перемещающихся по поверхности частиц.

Естественно предположить, что вероятности переходов  $p_{ij}$  пропорциональны разности высоты рельефа в ячейках  $i$  и  $j$ , а вероятность просочиться из ячейки  $i$  вглубь определяется проницаемостью почвы в ячейке  $i$ . Однако здесь мы не будем подробно касаться проблемы определения вероятностей  $p_{ij}$ , подчеркнём лишь, что этот вопрос невозможно решить без участия эксперта.

Обозначим  $\mathbf{P} = (p_{ij})$  — матрицу переходов из ячеек (поглощающее состояние  $*$  здесь не учитывается). Подчеркнём, что в данной задаче величины  $p_{ij}$  следует задать как вероятности, причём

$$\sum_{j=1}^n p_{ij} + p_{i*} = 1, \quad (8.12)$$

и, следовательно,

$$\sum_{j=1}^n p_{ij} \leq 1 \quad (8.13)$$

для любого  $i = 1, \dots, n$ .

Пусть в момент времени  $t$  вектор потенциалов структуры равен  $\mathbf{m}^{(t)} = (m_1^{(t)}, \dots, m_n^{(t)})$ . Тогда к моменту времени  $t + 1$  ячейку  $i$  покинет количество вещества, равное

$$\sum_{j \neq i} m_i^{(t)} p_{ij} + m_i^{(t)} p_{i*},$$

а поступит в ячейку  $i$  количество вещества, равное

$$f_i^{(t+1)} + \sum_{j \neq i} m_j^{(t)} p_{ji}.$$

Мы можем записать уравнение баланса, которое даёт нам рекуррентное соотношение:

$$m_i^{(t+1)} = m_i^{(t)} - m_i^{(t)} p_{i*} - \sum_{j \neq i} m_i^{(t)} p_{ij} + f_i^{(t+1)} + \sum_{j \neq i} m_j^{(t)} p_{ji}. \quad (8.14)$$

Упростим уравнение (8.14). Из (8.12) следует, что

$$m_i^{(t)} (1 - p_{i*}) = m_i^{(t)} \sum_{j=1}^n p_{ij}. \quad (8.15)$$

Подставив (8.15) в (8.14) и приведя подобные члены, получаем:

$$m_i^{(t+1)} = m_i^{(t)} p_{ii} + f_i^{(t+1)} + \sum_{j \neq i} m_j^{(t)} p_{ji}, \quad (8.16)$$

или

$$m_i^{(t+1)} = \sum_{j=1}^n m_j^{(t)} p_{ji} + f_i^{(t+1)}. \quad (8.17)$$

В векторном виде уравнение (8.17) имеет вид

$$\mathbf{m}^{(t+1)} = \mathbf{f}^{(t+1)} + \mathbf{m}^{(t)}\mathbf{P}. \quad (8.18)$$

Естественно предположить, что если входной поток  $f_i^{(t)}$  не зависит от  $t$ , то на достаточно длинных временных интервалах система переходит в стационарный режим. Обозначим соответствующие стационарному режиму вектор потенциала  $\mathbf{m} = (m_1, \dots, m_n)$  и вектор входного потока  $\mathbf{f} = (f_1, \dots, f_n)$ . Перейдя к пределу  $t \rightarrow \infty$  в уравнении (8.18), мы получим уравнение, которому удовлетворяет предельный (стационарный) потенциал структуры:

$$\mathbf{m} = \mathbf{f} + \mathbf{m}\mathbf{P}. \quad (8.19)$$

Заметим, что, хотя интерпретация структуры и даже времени разная, мы получили то же самое уравнение, что и для взаимно влияющих катастроф. Как уже отмечалось выше, решение уравнения (8.19) имеет вид

$$\mathbf{m} = \mathbf{f}(\mathbf{E} - \mathbf{P})^{-1} = \mathbf{f}(\mathbf{E} + \mathbf{P} + \mathbf{P}^2 + \dots),$$

где мы предполагаем, что ряд  $\mathbf{E} + \mathbf{P} + \mathbf{P}^2 + \dots$  сходится.

### **III. ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ ПРОГНОЗ**

---

#### Глава 9

### **ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ АНАЛИЗ ПРОЦЕССОВ ПОДГОТОВКИ ГЕОЛОГИЧЕСКИХ КАТАСТРОФ**

#### **9.1. Введение**

При решении задач прогноза природных катастроф требуется оценить пространственно-временные и энергетические характеристики возможного события, такие как, например, время и координаты события, зону воздействия, распределение энергии и спектральные характеристики воздействия.

Прогноз геологических катастроф основан на том, что процесс подготовки сопровождается пространственно-временными аномалиями характеристик геологической среды в некоторой окрестности будущего события. Эти аномалии связаны с накоплением потенциальной энергии и постепенным переходом геологической среды от фазы упругих в фазу пластических деформаций. Аномалии проявляются в результате изменения физических свойств геологической среды и, прежде всего, в изменении её структуры и отклика на динамические воздействия. Для измерения динамических свойств геологической среды используются системы комплексного сейсмологического, геофизического, гидрогеологического, геохимического и геодезического мониторинга.

Одними из наиболее опасных геологических катастроф являются землетрясения (*Bolt, 1999*). В данной главе мы рассмотрим информационную технологию, которая разработана в ИППИ РАН в тесном сотрудничестве со специалистами ОИФЗ РАН для выполнения исследований по прогнозу землетрясений и, в частности, для выявления и анализа предвестников землетрясений. Экспериментальная апробация технологии связана с исследованием ряда землетрясений, произошедших в Китае, Греции и Центральной Азии. Можно полагать, что основы этой технологии могут быть предметно ориентированы на выявление и анализ предвестников других типов геологических катастроф.

В разделе 9.2 рассмотрены основные особенности предметной области прогноза землетрясений, в разделе 9.3 формулируются основы технологии пространственно-временного анализа наблюдений процесса подготовки землетрясения, в разделе 9.4 рассматривается аналитическая ГИС ГеоТайм. В разделах 9.5 и 9.6 рассматриваются реализованные в системе ГеоТайм методы обнаружения нестационарностей и методы выявления пространственно-временных аномалий.

## 9.2. Предметная область

При прогнозе геологических катастроф используются многодисциплинарные динамические наблюдения и статические данные о пространственном распределении сеймотектонических свойств геологической среды, не зависящих от времени. При прогнозировании землетрясений к динамическим наблюдениям в первую очередь относятся результаты сейсмологического, геофизического, геодезического, геохимического и гидрогеологического мониторинга. В некоторых случаях эти наблюдения дополняются метеорологическими данными, результатами атмосферных и космических наблюдений и информацией о планетарных ритмах. Стационарные свойства геологической среды представлены картами разломной тектоники, картами основных параметров сейсмического режима, а также набором карт, представляющих геологические, геофизические, геодезические и геохимические свойства земной коры.

Прогноз землетрясений основывается на предположении о существовании причинно-следственной связи регистрируемых аномалий с процессом подготовки землетрясения (*Keilis-Borok, and Soloviev, ed., 2003, Соболев 1993*). Тем не менее, несмотря на серьёзные усилия учёных разных стран, проблема надёжного прогноза землетрясений остаётся нерешённой. Во многом это связано со сложностью обнаружения предвестников катастроф на фоне шумов. Трудность решения этой задачи объясняется следующими основными причинами (*Соболев и Пономарёв, 2003*):

1. Недостаток информации, обусловленный:
  - а) отсутствием многолетних наблюдений, полученных с помощью плотной и равномерной по пространству сети комплексного сейсмического и геофизического мониторинга;
  - б) невозможностью экспериментальных данных по подготовке землетрясений с одними и теми же координатами гипоцентра, магнитудой и механизмом очага.
2. Сложность модели измерительных сигналов геомониторинга, обусловленная:
  - а) изменчивостью во времени и пространстве тензочувствительности геологической среды в местах расположения станций геомониторинга;
  - б) наличием в измеряемых сигналах нелинейного тренда и разнопериодных сезонных ритмов, поступающих от космических, метеорологических и антропогенных источников;
  - в) существованием триггерных эффектов, способствующих возникновению землетрясений;
  - г) неоднозначностью модели возникновения предвестников, предполагающей два принципиально различных типа предвестников: эпицентральных, порождённых в очаге землетрясения, и регио-

нальных, отражающих изменение поля напряжений в обширном районе.

3. Сложная форма выделяемого аномального сигнала, предшествующего землетрясению, в виде:

- а) положительных или отрицательных ступеней;
- б) положительных или отрицательных однополярных импульсов;
- в) двухполярных импульсов, начинающихся с любого знака.

К этому следует добавить, что предвестники могут неоднократно возникать и исчезать при подготовке одного и того же землетрясения. Условно они подразделяются на долговременные (годы до события), средневременные (месяцы до события), краткосрочные (недели, дни до события) и оперативные (дни, часы до события).

### **9.3. Технология поиска пространственно-временных предвестников геологических катастроф**

Информационная технология для обнаружения предвестников землетрясений, как и технология пространственного прогноза, ориентирована на комплексную обработку всех типов доступных данных и знаний об анализируемом явлении и в значительной степени использует традиционную методологию рассуждений экспертов при выполнении аналогичных исследований. В данном случае у пользователя имеются динамические и статические данные по изучаемому региону, знания о процессе подготовки землетрясения и гипотезы о моделях измеряемых данных и предвестниках землетрясений. Цель исследований состоит обычно в обнаружении, оценивании статистической значимости и анализе предвестников на основе всей доступной информации. Характер выполнения работы состоит в многократном выдвижении гипотез о наличии или отсутствии предвестников землетрясений и их проверке на фактическом материале. В этом смысле практический прогноз ничем не отличается от глубокого научного исследования. Пользователь прогнозирует возможные события, экспериментально ищет новые предвестники землетрясений, анализирует результаты прогноза и выдвигает новые гипотезы.

Для выполнения такой работы необходима среда информационного моделирования, которая должна обеспечить эксперта алгоритмической поддержкой для выполнения следующих операций:

- Формирование широкого класса гипотез по выделению предвестников землетрясений из исходных данных.
- Обнаружение предвестников землетрясений и оценивание их статистической значимости.
- Изменение наборов исходных данных для проверки устойчивости предвестников землетрясений.
- Сравнение версий прогноза.
- Анализ ошибок прогноза типа пропуска цели и ложной тревоги.



Большинство этих функций реализовано в ГИС ГеоТайм (*Gitis et al., Gitis et al., 1995, Gitis et al., 1997*). Основная идея системы ГеоТайм состоит в дополнении традиционного анализа отдельных временных рядов геомониторинга или временных рядов, состоящих из значений различных функционалов, вычисленных по приуроченным к заранее выбранным геологическим зонам подкаталогам землетрясений, анализом сеточных (растровых) динамических полей. Здесь под сеточным динамическим полем понимается функция двух пространственных и временной координат, заданная на регулярной сетке (трёхмерном растре). Динамические поля вычисляются по каталогам землетрясений и по временным рядам геофизических, гидрогеологических, геохимических и геодезических измерений. Представление данных в виде пространственно-временного растра позволяет более полно по сравнению с представлением динамики подготовки землетрясения в виде временных рядов наблюдать процессы, развивающиеся во времени и в пространстве. Естественно ожидать, что картографический и статистический анализ исходных и трансформированных динамических полей дополнят анализ временных рядов и дадут возможность полнее изучить процесс подготовки землетрясения и найти новые предвестники землетрясений или других геологических катастроф.

Выявление предвестников землетрясений базируется на предположении о том, что геологическая среда неоднородна в пространстве, но в обычном состоянии имеет стационарную динамику, которая нарушается при подготовке геологической катастрофы. Считается, что анализируемые последовательности стационарны, а в процессе подготовки землетрясения изменяются некоторые их статистические характеристики. Для обнаружения аномалий анализируется временная последовательность значений каждого элемента пространственного растра. При этом текущий временной интервал значений временного ряда в элементе растра разбивается на два последовательных подинтервала, длительности которых задаются пользователем, исходя из контекста решаемой задачи. Тем самым задача сводится к сравнению двух случайных выборок, относящихся к подинтервалам. Каждая выборка считается стационарной и однородной в вероятностном смысле. На каждом интервале выборки подчиняются некоторому вероятностному распределению. Проверка гипотезы о совпадении параметров этих распределений осуществляется с помощью некоторой статистики, вид которой зависит от статистической модели анализируемых динамических полей. Для выделения аномальных областей, состоящих из подмножества элементов пространственного растра, применяются методы оценивания и проверки статистических гипотез в пространственно-временном окне. При этом учитывается коррелированность соседних элементов.

В целом подход к анализу предвестников землетрясений развивает методологию и основные этапы построения информационных моделей в задачах геолого-геофизического прогноза с неполной информацией, которые были рассмотрены в разделе 2.3. Технология направлена

на исследование эффективности предвестников землетрясений, в том числе: вычисление по данным геомониторинга трёхмерных сеточных динамических полей параметров, косвенно характеризующих процесс подготовки землетрясения, преобразование этих полей, обнаружение трёхмерных динамических полей пространственно-временных аномалий и оценивание значимости аномалий. Наиболее важным в технологии анализа предвестников землетрясений являются следующие этапы.

**1. Формализация задачи.** В настоящее время известен ряд причинно-следственных моделей подготовки землетрясения (*Мячкин и др., 1975, Смирнов и Завьялов, 1999, Соболев, 1993, Соболев, 1999, Zschau, 1996, Keilis-Borok, and Soloviev, ed., 2003, Соболев и Пономарёв, 2003*). Они основаны на физике разрушения, многолетних наблюдениях междисциплинарных данных сейсмо-тектонического мониторинга, на данных мониторинга горных ударов и экспериментах по разрушению образцов горных пород. Эти модели используются для формализации объекта и признаков прогноза. Объектом прогноза является пространственно-временная аномалия в состоянии земной коры, связанная с формированием очага землетрясения. Аналогично задачам пространственного прогноза основная проблема здесь состоит в нахождении адекватного описания прогнозируемого пространственно-временного объекта и его свойств. На основании модели генерируются признаки, которые описывают пространственно-временные изменения энергетических, геометрических и статистических характеристик геологической среды в период подготовки землетрясения. Модели формализуются в виде алгоритмов преобразования исходных данных в признаки, в виде статистических гипотез и в виде методов обнаружения аномалий.

**2. Формализация информационного пространства задачи.** К этому этапу относятся два основных пункта:

- Предварительная обработка каталогов землетрясений, включая очистку каталогов землетрясений от афтершоков и роев землетрясений (*Молчан и Дмитриева, 1991*), анализ и оценивание полноты каталогов по минимальной представительной магнитуде землетрясений и построение трёхмерной сеточной динамической модели поля минимальной представительной магнитуды землетрясений.
- Предварительная обработка временных рядов геомониторинга. Сюда относится очистка измерений от высокочастотных шумов и квазилинейного тренда, выделение сезонных ритмов, нормировка и стандартизация измерений с учётом чувствительности измерительных приборов и свойств земной коры в точке мониторинга.

**3. Генерирование системы признаков прогноза.** Сюда относится построение трёхмерных растровых динамических полей по каталогам землетрясений и по данным геомониторинга.

Генерирование признаков по каталогам землетрясений включает в себя построение трёхмерных сеточных динамических полей классических параметров сейсмического режима, таких как поле сейсмической активности, поле наклона графика повторяемости землетрясений (b-параметр) и поле фрактальной размерности (d-параметр) (Садовский и Писаренко, 1991, Смирнов, 1995). Кроме того, известен ряд эффективных для прогноза функционалов от каталога землетрясений, относящихся к энергетическим и геометрическим свойствам сейсмического процесса (Соболев, 1993, Соболев, 1999, Соболев и Тюпкин, 1996).

Генерирование признаков по данным геомониторинга включает в себя пространственную интерполяцию наблюдений или оценивание параметров эпицентральной модели формирования предвестников землетрясений (Пономарёв и др., 1999, Соболев и Пономарёв, 2003).

К этому же этапу относится генерирование растровых динамических полей с помощью линейной и нелинейной фильтрации исходных полей, а также с помощью алгебраических и логических операций с несколькими полями.

**4. Нахождение решения.** Нахождение решения состоит в построении трёхмерных сеточных динамических полей отклонений от стационарности, обнаружении пространственно-временных аномалий и оценивании их статистической значимости и в распознавании типов пространственно-временных аномалий.

**5. Анализ и аргументация.** Этот этап состоит в анализе ретроспективных данных с целью сопоставления значимости обнаруженных пространственно-временных отклонений от стационарности с очаговыми зонами произошедших землетрясений. Этап поддерживается операциями картографического анализа с элементами анимации, а также анализом по аналогии с прецедентами, техника которого аналогична технике пространственного прогноза.

**6. Модификация версии решения.** Модификация версии решения может быть выполнена с помощью выбора альтернативной причинно-следственной модели, с помощью ввода дополнительных данных и за счёт изменения гипотез формирования признаков прогноза.

#### 9.4. Базовые функции системы GeoТайм

Первая версия аналитической ГИС GeoТайм была разработана в 1991 г. Она состоит из четырёх основных подсистем (рис. 9.1):

1. Подсистема препроцессинга векторных данных и генерирования трёхмерных сеточных моделей.
2. Подсистема преобразования трёхмерных сеточных моделей
3. Подсистема обнаружения пространственно-временных аномалий и оценивания их статистической значимости.
4. Подсистема визуального исследования данных.



Рис. 9.1. Блок-схема системы GeoТайм

На вход подсистемы препроцессинга векторных данных и генерирования трёхмерных сеточных моделей поступают каталоги землетрясений и временные ряды. Для анализа каталогов землетрясений разработан табличный процессор. Он позволяет просматривать каталоги, оформленные в виде произвольных таблиц, вычислять произвольные алгебраические и логические функции от строк таблицы, оценивать статистические параметры, получать произвольные подкаталоги, строить карты событий. Отдельный модуль, разработанный В. Б. Смирновым, позволяет очищать каталоги от афтершоков. В подсистему входят также модули очистки временных рядов, снятия тренда и сезонных ритмов, стандартизации и анализа. Подсистемой генерируются два типа первичных динамических полей.

- Динамические поля, вычисленные по каталогу землетрясений: поле плотности событий, поле плотности событий с весами, поле плотности энергии событий, поле расстояний до ближайшего

события, поле сейсмической активности, поле наклона графика повторяемости землетрясений с различными магнитудами, поле фрактальной размерности пространственного распределения сейсмичности.

- Динамические поля, полученные в результате обработки временных рядов наблюдений на нерегулярной сетке. Для построения могут быть использованы следующие методы пространственно-временной обработки: метод интерполяции на основе потенциалов, метод кригинга, метод зональной интерполяции, метод оценивания параметров модели очаговой зоны.

В подсистеме преобразования пространственно-временных полей выполняются следующие операции.

1. Пространственная фильтрация в скользящем окне произвольного радиуса: усреднение, медианное сглаживание, свёртка с произвольной весовой функцией, вычисление среднеквадратичного отклонения, максимума, минимума, разброса, модуля и азимута градиента, лапласиана, снятие тренда.
2. Вычисление полей, которые являются произвольными функциями нескольких пространственно-временных полей, конструируемыми с помощью набора элементарных функций и произвольных алгебраических и логических операций.
3. Интерполяция полей во времени для восстановления пропущенных значений, которые невозможно было статистически значимо оценить при пространственно-временном анализе данных.

*Подсистема обнаружения пространственно-временных аномалий и оценивания их статистической значимости* предназначена для выявления значимых нестационарных изменений динамических полей во времени. Выходом подсистемы является пространственно-временное динамическое поле отклонений. При выполнении условия стационарности значения поля отклонений представляют собой нормированные случайные величины, т.е. величины, имеющие нулевые математические ожидания и единичные дисперсии. Поэтому значение поля, существенно превосходящее единицу по абсолютной величине, свидетельствует о том, что произошло нарушение стационарности. В подсистему входят программные модули проверки статистических гипотез для нескольких моделей анализируемого динамического поля предвестников землетрясений. Возможны модели пуассоновского поля, гауссовского скалярного поля и векторного гауссовского поля.

*Подсистема визуального исследования* имеет доступ практически ко всем типам данных, с которыми манипулирует система. Подсистемой поддерживаются следующие функции:

- Оценивание элементарных статистических параметров полей и взаимных корреляционных матриц.
- Анимационное картографическое представление в основных географических проекциях многослойных моделей, состоящих из

трёхмерных динамических сеточных полей, пространственных зон, каталогов землетрясений, точек и линейных элементов.

- Произвольное изменение масштаба.
- Визуализация разрезов трёхмерных и двумерных сеточных полей по произвольным профилям и картографическое измерение.
- Комплексный анализ сеточных моделей по сходству с произвольной точкой региона.

### **9.5. Методы обнаружения нестационарностей в динамических полях**

**9.5.1. Обнаружение нестационарности как задача проверки статистических гипотез.** В настоящем разделе описаны методы обнаружения изменений характеристик динамических полей во времени. Для обнаружения нестационарности анализируется временная последовательность значений каждого элемента пространственного раstra. Предполагается, что обычно анализируемая последовательность стационарна, а при подготовке землетрясения изменяются некоторые статистические характеристики последовательности. Задача обнаружения сводится к сравнению двух случайных выборок, относящихся к двум последовательным интервалам. Предполагается, что выборки на каждом интервале являются стационарными, но могут иметь различные распределения вероятностей.

Проверка гипотезы о совпадении параметров этих распределений осуществляется с помощью некоторой статистики, вид которой зависит от выбранной модели. Первая модель относится к случаю полей сейсмических событий и основана на предположении о том, что эти поля имеют пуассоновское распределение. Вторая и третья модели основаны на предположении о том, что поля имеют гауссовское распределение. При этом третья модель описывает векторное гауссовское поле и предназначена для комплексной обработки нескольких динамических полей.

**9.5.2. Метод обнаружения для случая пуассоновских полей.** Предположим, что наблюдается пуассоновский поток событий. Разобьём текущий интервал наблюдения  $T$  на два последовательных временных подинтервала  $T_1$  и  $T_2$  и предположим, что на первом подинтервале  $T_1$  поток имеет интенсивность  $\lambda_1$ , а на втором  $T_2$  — интенсивность  $\lambda_2$ . Сформулируем задачу обнаружения изменения интенсивности пуассоновского потока как задачу проверки гипотезы  $H_0$  против альтернативной гипотезы  $H_A$ :

$$H_0 : \lambda_1 = \lambda_2, \quad H_A : \lambda_2 > \lambda_1.$$

Гипотеза  $H_0$  означает, что на интервале  $T$  наблюдаемый пуассоновский поток стационарен, гипотеза  $H_A$  означает, что интенсивность пуассоновского потока увеличивается на подинтервале  $T_2$ .

Можно показать, что подходящей статистикой для критерия проверки рассматриваемых гипотез является выражение:

$$\eta = \frac{(N_2 - Np_0)}{\sqrt{Np_0q_0}}, \quad (9.1)$$

где  $N$  — число событий на интервале  $T$ ,  $N_2$  — число событий на интервале  $T_2$ ,

$$p_0 = \frac{T_2}{T}, \quad q_0 = 1 - p_0.$$

При больших  $N$  и значениях  $p_0$ , не слишком близких к 0 или 1, распределение статистики  $\eta$  при гипотезе  $H_0$  близко к стандартному нормальному распределению. Критерий проверки гипотезы состоит в сравнении величины  $\eta$  с порогом  $d$ , который определяется уравнением

$$\alpha = \frac{1}{\sqrt{2}\pi} \int_d^{\infty} \exp\left(-\frac{t^2}{2}\right) dt, \quad (9.2)$$

где вероятность  $\alpha$ , называемая уровнем значимости критерия проверки гипотезы, выбирается достаточно малой величиной. Таким образом,  $d$  является  $(1 - \alpha)$ -квантилем стандартного нормального распределения, который может быть найден по таблицам нормального распределения.

**9.5.3. Метод обнаружения для случая гауссовских полей.** Будем считать, что в любой фиксированной точке динамическое поле описывается случайной последовательностью  $\xi_i$  с независимыми значениями, имеющими гауссовское распределение. Обозначив математические ожидания элементов последовательности на интервалах  $T_1$  и  $T_2$  через  $\mu_1$  и  $\mu_2$ , будем считать, что дисперсии подпоследовательностей одинаковы и равны  $\sigma^2$ . Три параметра рассматриваемой модели  $\mu_1$ ,  $\mu_2$ , и  $\sigma^2$  считаются неизвестными.

Соответствующие этой модели гипотезы — основная  $H_0$  и альтернативная  $H_A$  — имеют вид:

$$H_0 : \mu_1 = \mu_2, \quad H_A : \mu_2 > \mu_1.$$

Если принимается гипотеза  $H_A$ , то считается, что обнаружена нестационарность; если принимается гипотеза  $H_0$ , то считается, что процесс стационарен. Обозначим через  $n_1$  и  $n_2$  число отсчётов на интервалах  $T_1$  и  $T_2$ ,  $n = n_1 + n_2$ .

Введённая модель отвечает хорошо изученной в статистике схеме сравнения двух выборок из нормального распределения. Критерий проверки гипотезы  $H_0$  против гипотезы  $H_A$  основан на вычислении  $t$ -статистики, представляющей собой нормированную разность выборочных средних:

$$t = k \frac{\hat{\mu}_2 - \hat{\mu}_1}{\sqrt{s_1^2 + s_2^2}}, \quad (9.3)$$

где

$$k = \sqrt{\frac{n_1 + n_2 - 2}{1/n_1 + 1/n_2}}, \quad \hat{\mu}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} \xi_t, \quad \hat{\mu}_2 = \frac{1}{n_2} \sum_{t=n_1+1}^{n_1+n_2} \xi_t,$$

$$s_1^2 = \sum_{t=1}^{n_1} (\xi_t - \hat{\mu}_1)^2, \quad s_2^2 = \sum_{t=n_1+1}^{n_1+n_2} (\xi_t - \hat{\mu}_2)^2.$$

Если гипотеза  $H_0$  верна, то статистика  $t$  следует распределению Стьюдента с числом степеней свободы  $n_1 + n_2 - 2$ .

Выбор в пользу той или иной из гипотез осуществляется в соответствии со следующим правилом. Если статистика  $t < d$ , где  $d$  — некоторый порог, то принимается гипотеза  $H_0$  (т.е. считается, что динамическое поле в данной точке не претерпевает существенного изменения), если  $t \geq d$ , то принимается альтернативная гипотеза  $H_A$  (т.е. считается, что обнаружено значимое изменение динамического поля в данной точке).

Порог  $d$  для выбранного заранее уровня значимости  $\alpha$  находится из таблиц  $t$ -распределения. При гипотезе  $H_0$  и больших  $n$  распределение  $t$ -статистики приближается к стандартному нормальному распределению и для определения порога  $d$  можно использовать уравнение (9.2)

**9.5.4. Метод обнаружения для случая нескольких динамических полей.** Пусть одновременно наблюдаются несколько динамических полей. Совокупность динамических полей в любой фиксированной точке пространства будем описывать последовательностью векторных случайных величин  $\xi_i = (\xi_{i1}, \dots, \xi_{ip})$ , где  $i$  — момент времени,  $p$  — число наблюдаемых динамических полей.

Будем считать, что вектор  $\xi_i$  имеет гауссовское распределение, последовательность случайных векторов  $\xi_i$  некоррелирована во времени, ковариационная матрица компонент вектора  $\xi_i$  не зависит от времени и на каждом интервале наблюдения  $T_1$  и  $T_2$  соответствующие векторные подпоследовательности стационарны. Обозначая векторные математические ожидания указанных подпоследовательностей через  $\mu_1$  и  $\mu_2$ , сформулируем две гипотезы — основную  $H_0$  и альтернативную  $H_A$ :

$$H_0 : \mu_1 = \mu_2, \quad H_A : \mu_2 \neq \mu_1.$$

Опишем статистику, используемую для различения гипотез  $H_0$  и  $H_A$ .

Обозначим через  $\hat{\mu}_1$  и  $\hat{\mu}_2$  выборочные средние для векторных подпоследовательностей на интервалах  $T_1$  и  $T_2$  соответственно, через  $\mathbf{S}_1$  и  $\mathbf{S}_2$  — их матрицы рассеивания:

$$\hat{\mu}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} \xi_t, \quad \hat{\mu}_2 = \frac{1}{n_2} \sum_{t=n_1+1}^{n_1+n_2} \xi_t,$$

$$\mathbf{S}_1 = \sum_{t=1}^{n_1} (\xi_t - \hat{\mu}_1)(\xi_t - \hat{\mu}_1)^T, \quad \mathbf{S}_2 = \sum_{t=n_1+1}^{n_1+n_2} (\xi_t - \hat{\mu}_2)(\xi_t - \hat{\mu}_2)^T.$$



через  $\widehat{\mathbf{K}}$  — оценку матрицы ковариации по двум подпоследовательностям:

$$\widehat{\mathbf{K}} = \frac{\mathbf{S}_1 + \mathbf{S}_2}{n_1 + n_2 - 2},$$

через  $D$  — расстояние Махаланобиса между двумя подпоследовательностями:

$$D^2 = (\widehat{\mu}_1 - \widehat{\mu}_2)^T \widehat{\mathbf{K}}^{-1} (\widehat{\mu}_1 - \widehat{\mu}_2).$$

Запишем выражение, которое отличается от предыдущего лишь коэффициентом

$$F = \frac{(n_1 + n_2 - p - 1)n_1n_2}{p(n_1 + n_2 - 2)(n_1 + n_2)} D^2 \quad (9.4)$$

Как известно, эта статистика при гипотезе  $H_0$  следует  $F$ -распределению Фишера  $F(p, n_1 + n_2 - p - 1)$ .

Заметим, что так как выборочные средние  $\widehat{\mu}_1$  и  $\widehat{\mu}_2$  близки к соответствующим математическим ожиданиям  $\mu_1$  и  $\mu_2$ , то при гипотезе  $H_0$  разность  $\widehat{\mu}_1 - \widehat{\mu}_2$ , а также статистики  $D$  и  $F$  будут близки к нулю. Если же верна гипотеза  $H_A$ , то статистики  $D$  и  $F$  будут значимо отличаться от нуля. Следовательно, различение гипотез можно осуществить сравнением значения статистики  $F$  с некоторым порогом  $d$ . Если  $F < d$ , то принимается гипотеза  $H_0$  если  $F > d$ , то принимается альтернативная гипотеза  $H_A$ . Порог  $d$  зависит от выбранного уровня значимости  $\alpha$  и находится из таблиц  $F$ -распределения.

При больших  $n - p$   $F$ -распределение можно заменить на нормированное распределение  $\chi^2$  с  $p$  степенями свободы, т. е.  $\chi_p^2/p$ . Поэтому при больших  $n - p$  и достаточно большом  $p$  статистика  $(F - 1)\sqrt{p/2}$  приближённо имеет стандартное нормальное распределение.

Отдельно рассмотрим частный случай, когда компоненты случайного вектора  $\xi_i$  некоррелированы. В этом случае ковариационная матрица  $\mathbf{K}$  диагональна и можно считать её оценку  $\widehat{\mathbf{K}}$  также диагональной. Расстояние Махаланобиса (с поправкой на диагональность матрицы  $\widehat{\mathbf{K}}$ ) будет теперь иметь вид:

$$d^2 = \frac{(\widehat{\mu}_1 - \widehat{\mu}_2)^2}{\widehat{\sigma}_j^2}.$$

где

$$\widehat{\sigma}_j^2 = \frac{s_{1j}^2 + s_{2j}^2}{n_1 + n_2 - 2}, \quad j = 1, \dots, p$$

являются диагональными элементами матрицы  $\widehat{\mathbf{K}}$ ,

$$s_{1j}^2 = \sum_{i=1}^{n_1} (\xi_{ij} - \widehat{\mu}_{1j})^2, \quad s_{2j}^2 = \sum_{i=n_1+1}^{n_1+n_2} (\xi_{ij} - \widehat{\mu}_{2j})^2,$$

$\widehat{\mu}_{1j}$  и  $\widehat{\mu}_{2j}$  — компоненты векторов  $\widehat{\mu}_1$  и  $\widehat{\mu}_2$ , а  $\xi_{ij}$  — компоненты векторов  $\xi_i$ .

Отметим, что величина

$$t_j = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \frac{\hat{\mu}_{2j} - \hat{\mu}_{1j}}{\hat{\sigma}_j} \quad (9.5)$$

является  $t$ -статистикой Стьюдента с числом степеней свободы  $n_1 + n_2 - 2$ . Поэтому величина

$$T^2 = \frac{d^2 n_1 n_2}{n_1 + n_2}$$

является суммой квадратов  $t$ -статистик. Так как при больших  $n$  (например, 30 или более)  $t$ -статистика приближённо имеет стандартное нормальное распределение, то  $T^2$  приближённо имеет распределение  $\chi_p^2$ . Поэтому статистика

$$r = (T^2 - p) / \sqrt{2p}$$

при гипотезе  $H_0$  имеет нулевое математическое ожидание и единичную дисперсию и при достаточно большом  $p$  приближённо имеет стандартное нормальное распределение.

Если заранее известно, что компоненты вектора математического ожидания подчиняются соотношениям  $\mu_{1j} < \mu_{2j}$ , то естественно использовать вместо статистики  $t_j$  статистику  $t_j^+ = \max(t_j, 0)$ . Сумму квадратов статистик  $t_j^+$  обозначим  $T_+^2$ . Можно показать, что статистика  $T_+^2$  имеет математическое ожидание  $p/2$  и дисперсию  $5p/4$ . Для данного порога  $d$  статистика  $T_+^2$  имеет меньший уровень значимости  $\alpha$ , чем  $T^2$ . Иначе говоря, поле  $T_+^2$  имеет меньше «фоновых» выбросов, чем поле  $T^2$ .

## 9.6. Выделение пространственно-временных аномалий в динамических полях

**9.6.1. Задача выделения аномальных зон динамических сеточных полей.** Рассмотренные в предыдущем разделе методы обнаружения нестационарностей в динамических полях основаны на анализе временных последовательностей значений в каждом элементе пространственного растра в отдельности. Мы будем предполагать, что при отсутствии аномалий значения поля отклонений являются случайными величинами со стандартным гауссовским распределением. При наличии аномалий значения поля отклонений являются случайными величинами также с гауссовским распределением и единичной дисперсией, но с ненулевым математическим ожиданием.

Как известно аномалии геофизических параметров, предшествующие землетрясениям, проявляются одновременно в некоторой пространственной области, близкой к эпицентру. Поэтому изменения характеристик геологической среды во времени проявляются в динамических полях отклонений в виде пространственно-временных областей, состоящих из элементов, отличающихся от фона по значению мате-

математического ожидания. В связи с этим для выделения аномальных зон и оценки их значимости следует применить методы, которые одновременно учитывают информацию, относящуюся к некоторому множеству элементов раstra. В частности, аномальные области можно выделять, оценивая математическое ожидание по выборке элементов раstra в пространственно-временном окне. Можно также использовать для выделения аномалий и оценки их значимости методы проверки статистических гипотез. Задача осложняется тем, что значения динамического поля в близких точках могут быть коррелированы между собой.

**9.6.2. Оценивание математического ожидания.** Рассмотрим дискретную модель динамического поля, заданного на трёхмерном растре, образованном двумя пространственными и одной временной координатами. Рассмотрим пространственно-временное окно  $W$ , содержащее  $N$  элементов раstra. Будем считать, что в пределах окна  $W$  значения поля представлены случайным вектором  $x$ , компоненты которого  $x_i$  являются случайными величинами с равными математическими ожиданиями  $\mu$  и равными дисперсиями  $\sigma^2$ .  $N$ -мерную матрицу ковариаций вектора  $x$  обозначим через  $\mathbf{K}$ .

Рассмотрим оценку для  $\mu$  в виде взвешенного среднего

$$\hat{\mu} = \sum_{i=1}^N c_i x_i,$$

где  $c_i$  — весовые коэффициенты. Будем искать  $\hat{\mu}$  как оценку, имеющую минимальную дисперсию  $D\hat{\mu}$  в классе несмещённых оценок. Так как из несмещённости следует, что  $\sum_{i=1}^N c_i = 1$ , то весовые коэффициенты  $c_i$  можно найти из условия минимума функции

$$F = D\hat{\mu} + \lambda \left( \sum_{i=1}^N c_i - 1 \right),$$

где  $\lambda$  — неопределённый множитель Лагранжа.

Обозначим через  $\mathbf{e}$  единичный  $N$ -мерный вектор. Тогда вектор весовых коэффициентов  $\mathbf{c}$ , удовлетворяющий минимуму этой функции, имеет вид

$$\mathbf{c} = \frac{\mathbf{K}^{-1}\mathbf{e}}{\mathbf{e}^T \mathbf{K}^{-1} \mathbf{e}}.$$

а оптимальная оценка равна

$$\hat{\mu} = \frac{\mathbf{e}^T \mathbf{K}^{-1} \mathbf{x}}{\mathbf{e}^T \mathbf{K}^{-1} \mathbf{e}}.$$

Очевидно, что в частном случае, когда элементы поля некоррелированы, все  $c_i = 1/N$  — и оценкой для  $\mu$  является выборочное среднее

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Учитывая, что дисперсия взвешенного среднего

$$D\hat{\mu} = \mathbf{c}^T \mathbf{K} \mathbf{c} = (\mathbf{e}^T \mathbf{K}^{-1} \mathbf{e})^{-1},$$

а дисперсия выборочного среднего

$$D\bar{x} = \frac{\mathbf{e}^T \mathbf{K} \mathbf{e}}{N^2},$$

эффективность оптимальной линейной оценки  $\mu$  по сравнению с выборочным средним  $\bar{x}$ , определяемая как отношение соответствующих дисперсий, даётся выражением

$$\text{eff}(\hat{\mu}) = \frac{D\bar{x}}{D\hat{\mu}} = \frac{1}{N^2} \sum_{i,j} k_{ij} \sum_{i,j} k_{ij}^{-1},$$

где  $k_{ij}$  — элемент матрицы  $\mathbf{K}$ , а  $k_{ij}^{-1}$  — элемент матрицы  $\mathbf{K}^{-1}$ .

Можно показать, что оценка  $\hat{\mu}$  является также оценкой наименьших квадратов для случая линейной модели наблюдений с коррелированными наблюдениями. Изложенный выше способ оценки  $\mu$  в виде взвешенного среднего можно рассматривать как специальный случай сглаживания динамического поля с учётом пространственно-временной корреляции его элементов.

**9.6.3. Обнаружение аномалий.** Аномалии геофизических полей могут проявляться как локальные изменения статистических свойств динамического поля отклонений. Будем исходить из того, что при отсутствии аномалий значения динамического поля отклонений являются нормальными случайными величинами с нулевым математическим ожиданием и известной матрицей ковариаций  $\mathbf{K}$ . В аномальной области, т. е. в области содержащей предвестник, поле отклонений приобретает ненулевое математическое ожидание, но его матрица ковариаций не изменяется. Значения математических ожиданий элементов поля отклонений в аномальной области определяют форму предвестника. Алгоритм обнаружения аномалий основан на проверке статистических гипотез:

$$H_0 : \mu = 0, \quad H_A : \mu = \mathbf{a},$$

где  $\mu$  — вектор математических ожиданий поля отклонений,  $\mathbf{a}$  — фиксированное значение вектора  $\mu$ , определяемое положением и формой предвестника.

Отношение правдоподобия двух гауссовских распределений, отвечающих гипотезам  $H_0$  и  $H_A$  имеет вид

$$F_A(\mathbf{x}) = \exp\left(\mathbf{a}^T \mathbf{K}^{-1} \mathbf{x} - \frac{1}{2} \mathbf{a}^T \mathbf{K}^{-1} \mathbf{a}\right),$$

где  $\mathbf{K}^{-1}$  — матрица, обратная к ковариационной.

Известно, что сравнение отношения правдоподобия с порогом позволяет различить две гипотезы  $H_0$  и  $H_A$  так, чтобы вероятность ложной тревоги (ложного принятия гипотезы  $H_A$ ) имела заданное значение  $\alpha$ , а вероятность правильного принятия этой гипотезы была максимальной. В нашем случае отношение правдоподобия монотонно связано с линейной функцией  $\mathbf{a}^T \mathbf{K}^{-1} \mathbf{x}$ . Чтобы по данному  $\alpha$  подобрать порог для этой функции, удобно её нормировать. Поскольку дисперсия этой функции равна  $\mathbf{a}^T \mathbf{K}^{-1} \mathbf{a}$ , то нормированная статистика

$$f_A(\mathbf{x}) = \frac{\mathbf{a}^T \mathbf{K}^{-1} \mathbf{x}}{\sqrt{\mathbf{a}^T \mathbf{K}^{-1} \mathbf{a}}}$$

при гипотезе  $H_0$  имеет место стандартное нормальное распределение. При альтернативной гипотезе  $H_A$  математическое ожидание этой статистики положительно.

Выбор в пользу той или иной из гипотез осуществляется в соответствии со следующим правилом. Если статистика  $f_A(\mathbf{x}) < d$ , где  $d$  — порог, то принимается гипотеза  $H_0$ , если  $f_A(\mathbf{x}) \geq d$ , то принимается альтернативная гипотеза  $H_A$ . Поскольку величина  $f_A(\mathbf{x})$  при гипотезе  $H_0$  имеет стандартное нормальное распределение, то порог  $d$  для выбранного уровня значимости  $\alpha$  находится как и в разделе 9.5.

Конкретизируем эту схему. Значения любого из построенных в разделе 3 полей отклонений имеют дисперсию, равную 1. Предположим, что эти значения не коррелированы между собой. Величину  $a_i$  мы будем считать равной некоторой положительной постоянной на аномальной области  $W$  и нулю вне области  $W$ . Тогда

$$f_A(\mathbf{x}) = \frac{1}{\sqrt{N}} \sum_{i \in W} x_i,$$

где  $N$  — число элементов в  $W$ .

В качестве допустимых будем рассматривать области  $W$  нескольких форм и их сдвиги по растру, например, все круги с произвольным центром и любым радиусом из некоторого заданного интервала. Сравнив значения  $f_A(\mathbf{x})$  для всех рассматриваемых областей  $W$  с порогом  $d$ , мы выберем те  $W$ , для которых  $f_A(\mathbf{x})$  превосходит  $d$ , и будем интерпретировать эти  $W$  как аномальные области поля. Для каждой области выделяется точка, например, центр, к которой будем относить значение статистики  $f_A(\mathbf{x})$ , вычисленной для этой области. Отметим, что вероятность ошибочного обнаружения любой данной аномальной области  $W$  равна  $\alpha$ .

В результате применения описанного алгоритма в качестве аномальных могут оказаться выделенными несколько областей с общим центром, т.е. в одной точке выходного поля окажется несколько значений статистики  $f_A(\mathbf{x})$ . В этом случае естественно выбрать из них аномальную область с максимальным значением  $f_A(\mathbf{x})$ . В результате мы получаем поле максимальных значений  $f_A(\mathbf{x})$  и поле радиусов соответствующих аномальных областей.

Если к каждой точке раstra отнесено  $M > 1$  областей  $W$  (например, кругов с общим центром), то распределение максимального значения  $f_A(\mathbf{x})$  при гипотезе  $H_0$  не является стандартным гауссовским. Можно лишь утверждать, что для этой величины вероятность превзойти порог  $d$  меньше, чем  $M\alpha$ .

Если по каким-либо причинам точное значение  $\mathbf{x}$  поля динамических отклонений в точках раstra не известно, то алгоритм можно модифицировать так, чтобы для каждой точки  $i$  использовать только информацию о том, превосходит ли  $x_i$  некоторый порог или нет. Предположим, что порог  $A$  выбран так, что при гипотезе  $H_0$  вероятность того, что величине  $x_i$  превзойти порог  $A$  равна некоторому  $p_0$ . Допустим, что наличие сигнала  $a$  проявляется в том, что на какой-то из областей  $W$  значение  $p_0$  заменяется на  $p > p_0$ . Тогда в качестве статистики для проверки гипотезы  $H_A$  против  $H_0$  естественно взять сумму

$$L_A(\mathbf{x}) = \sum_{i \in W} \varphi(x_i),$$

где  $\varphi(x_i)$  — пороговая функция, равная 1, если  $x_i > A$ , и 0, если  $x_i \leq A$ . Удобно перейти к нормированной статистике

$$l_A(\mathbf{x}) = \frac{L_A(\mathbf{x}) - Np_0}{\sqrt{D}},$$

где  $N$  — число элементов в  $W$ , а  $D$  — дисперсия случайной величины  $L_A(\mathbf{x})$  при гипотезе  $H_0$

$$D = Np_0(1 - p_0) + \sum_{i, j \neq i \in W} c_{ij},$$

где  $c_{ij}$  — ковариация случайных величин  $\varphi(x_i)$  и  $\varphi(x_j)$ .

Величина  $c_{ij}$  может быть выражена через ковариацию величин  $x_i$  и  $x_j$  аналитически или может быть оценена статистически в предположении стационарности поля  $x_i$  при отсутствии аномалий.

При достаточно больших  $N$  (порядка нескольких десятков) величина  $l_A(\mathbf{x})$  имеет приближённо стандартное нормальное распределение, если верна гипотеза  $H_0$ . Решение о наличии аномалии принимается, если величина  $l_A(\mathbf{x})$  превзойдёт порог  $d$ , выбранный по заданному уровню значимости  $\alpha$ .

## Глава 10

# ИССЛЕДОВАНИЕ ПРОСТРАНСТВЕННО- ВРЕМЕННЫХ ПРОЦЕССОВ ПОДГОТОВКИ ЗЕМЛЕТРЯСЕНИЯ

### 10.1. Введение

При коровом землетрясении происходит разрыв и подвижка в земной коре, расходуется накопленная потенциальная энергия, перераспределяются напряжения. Можно считать, что землетрясения отображают вариации сейсмо-тектонических процессов: процесс выделения накапливаемой тектонической энергии в земной коре, процесс перераспределения напряжений и процесс изменения структурных свойств земной коры. Каталоги землетрясений обычно содержат координаты эпицентра землетрясения, время события, оценку его энергии (магнитуду землетрясения, энергетический класс или интенсивность в эпицентре). Иногда эти данные сопровождаются моделями механизма землетрясения и параметрами сейсмограмм, оценками точности определения координат и величины события, списком сейсмических станций, по данным которых производилось оценивание, макросейсмическими данными, именами географических пунктов, к которым приурочено землетрясение и другой атрибутивной информацией. Известно, что анализ сейсмологических каталогов позволяет обнаруживать пространственно-временные области активизации и затиший, области изменения соотношений между частотой сильных и слабых землетрясений, области с изменением типа геометрических характеристик эпицентров и т. д. Сейсмологический мониторинг ведётся практически для всех регионов земного шара. Поэтому каталоги землетрясений представляют собой наиболее доступное пространственно-временное описание сейсмического процесса.

При исследовании сейсмо-тектонических процессов по сейсмическим наблюдениям разработан ряд функционалов, которые преобразуют каталоги во временные ряды (*Соболев, 1993, Соболев и Пономарёв 2003, Zhang Zhaocheng et al. 1992, Keilis-Borok and Soloviev, editors, 2003*). Прогноз землетрясений основан на выявлении аномалий в этих временных рядах. Аномалии, которые предшествуют землетрясениям, называются предвестниками.

В разделе 10.2 мы рассмотрим алгоритмы пространственно-временного анализа каталогов землетрясений, которые реализованы в ГИС GeoТайм. Анализ каталога землетрясений производится в скользящем окне наблюдений, размеры которого адаптивно подбираются в зависи-

мости от плотности потока событий. Один из возможных алгоритмов адаптации параметров окна рассматривается в разделе 10.2.1. В разделах 10.2.2, 10.2.3 кратко излагаются некоторые алгоритмы оценивания параметров сейсмического потока, которые используются при вычислении пространственно-временных сеточных моделей. Данная технология применялась для анализа каталога землетрясений Кавказа, северо-восточного Китая, Греции и Центральной Азии. В разделе 10.2.4 мы рассмотрим некоторые результаты анализа каталога землетрясений, относящегося к Алматынскому прогностическому полигону Института сейсмологии Министерства образования и науки Республики Казахстан. Эти результаты получены совместно с сотрудниками этого института и Института Физики Земли РАН.

Ряд исследований по выявлению предвестников землетрясений связан с анализом данных геофизического, гидрогеологического, геохимического и геодезического мониторинга. Как указывалось в предыдущей главе, одна из основных трудностей анализа этого типа данных связана с редкой сетью наблюдений и малым количеством длинных временных рядов синхронных наблюдений, содержащих различные параметры. Наличие таких данных могло бы существенно продвинуть понимание физической природы аномалий, предшествующих землетрясениям, и облегчить решение задачи выделения предвестников землетрясений на фоне шумов.

По-видимому, самые представительные по количеству и длительности материалы геомониторинга внутриконтинентальных землетрясений накоплены в Китае. Раздел 10.3 посвящён развитию и экспериментальному исследованию рассмотренного в гл. 9 подхода к анализу пространственно-временных предвестников землетрясений применительно к временным рядам геомониторинга (*Пономарёв и др., 1999*). В качестве тестовых примеров для анализа краткосрочных предвестников землетрясений по геофизическим измерениям выбраны два землетрясения северо-восточного Китая: Тяньшанское (28 июля 1976 года  $M = 7,8$ ) и Датонгское (19 октября 1989 года  $M = 6,1$ ). В разделах 10.3.2, 10.3.3 и 10.3.4 дано описание исходных данных и предложена общая схема анализа. Разделы 10.3.5 и 10.3.6 представляют результаты моделирования и информационные модели предвестников Тяньшанского и Датонгского землетрясений. В разделах 10.5.7 и 10.5.8 рассмотрен метод обнаружения эпицентральных аномалий, который экспериментально исследован на примере процесса подготовки Тяньшанского землетрясения.

## **10.2. Исследование процесса подготовки землетрясения по каталогам**

**10.2.1. Адаптивный алгоритм подбора пространственно-временного окна оценивания.** Для оценивания пространственно-временных параметров сейсмического процесса применяются статистиче-



ские методы. Точность статистических оценок зависит от количества событий, попадающих в скользящее окно, с которого осуществляется сканирование пространственно-временной области анализа. При оценивании обычно предполагается, что в пределах окна анализируемая выборка статистически однородна. Очевидно, что для этого окно должно покрывать тектонически однородную пространственную область и относиться к такому временному интервалу, на котором не происходит существенных изменений параметров сейсмического потока. Таким образом, увеличение размеров пространственно-временного окна может привести к сомнительным оценкам из-за смещения наблюдений, принадлежащих различным вероятностным распределениям, а уменьшение размеров окна приводит к неточностям оценок из-за малого объёма выборки. Компромиссное решение состоит в адаптивном подборе размеров окна, которые выбираются пропорциональными интенсивности сейсмического потока.

В технологии GeoТайм принят следующий подход к оцениванию параметров. Параметр оценивается для каждого узла сетки трёхмерного динамического поля, которое задано интервалами по каждой пространственной и временной координатам и координатной привязкой. При оценивании задаются минимальное число событий  $N_{\min}$  и номинальное число событий  $N$ . Пространственно-временное окно оценивания представляет собой цилиндр, ось которого направлена вдоль временной координаты (рис. 10.1).

Один центр основания цилиндра совмещён с точкой раstra, к которой приписывается оценка. Второе основание цилиндра смещено назад по оси времени. Размеры цилиндра подбираются адаптивно в соответствии с попадающим внутрь его числом событий. Заранее задаются минимальный  $R_{\min}$  и максимальный  $R_{\max}$  радиусы оснований цилиндра в пространственной области и его высоты — временные интервалы  $T_{\min}$  и  $T_{\max}$ . Вначале для оценивания берутся события, попадающие в цилиндр с минимальными размерами  $R_{\min}$  и  $T_{\min}$ . Если число событий в «минимальном» цилиндре  $n \geq N$ , то производится оценивание и выбирается следующая точка раstra. В противном случае рассматриваются события, попадающие в цилиндр с максимальными размерами. Если число событий в этом цилиндре  $n < N_{\min}$ , то производится отказ от оценивания и выбирается следующая точка раstra. Если число событий в «максимальном» цилиндре  $N_{\min} \leq n \leq N$ , то производится оценивание и выбирается следующая точка раstra. Если число событий в цилиндре с максимальными размерами  $n > N$ , то включается

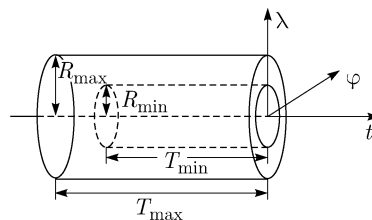


Рис. 10.1. Пространственно-временное окно оценивания статистических параметров сейсмического процесса по каталогу землетрясений

процедура выбора  $N$  ближайших событий к оцениваемой точке растра. Расстояния от точки растра  $\mathbf{a} = (x_a, y_a, t_a)$  до эпицентров землетрясений  $\mathbf{b} = (x_b, y_b, t_b)$  измеряются в метрике  $\rho(\mathbf{a}, \mathbf{b}) = \max(\alpha r_{ab}^2, t_{ab})$ , где  $\alpha$  — параметр алгоритма, фиксированный для всего динамического поля,  $r_{ab} = [(x_a - x_b)^2 + (y_a - y_b)^2]^{1/2}$  — радиус подбираемого цилиндра,  $R_{\min} < r_{ab} < R_{\max}$ ,  $t_{ab} = t_a - t_b$  — временной интервал  $T_{\min} < t_{ab} < T_{\max}$  (длина образующей цилиндра). Таким образом выбирается подмножество из  $N$  ближайших к оцениваемой точке растра событий.

**10.2.2. Динамическое поле минимальных представительных магнитуд.** В сейсмологии широко используется предположение об экспоненциальном распределении магнитуд землетрясений. Экспериментальным подтверждением этого предположения является то, что эмпирическая зависимость логарифма числа землетрясений от магнитуды, известная как зависимость Гутенберга–Рихтера или график повторяемости землетрясений, является линейной (Садовский и Писаренко, 1991). Диапазон магнитуд землетрясений, в котором эта зависимость достаточно хорошо аппроксимируется прямой линией, ограничен слева некоторой магнитудой. Этой магнитудой определяется чувствительность сети сейсмологических наблюдений. Землетрясения с меньшими магнитудами регистрируются с пропусками. Из-за этого на малых магнитудах точки эмпирической зависимости лежат ниже теоретической прямой. Минимальная представительная магнитуда землетрясений  $m_0$  является левой границей интервала, в котором выполняется предположение о линейности графика повторяемости. Величина  $m_0$  зависит от количества, географического расположения и чувствительности станций сейсмического мониторинга.

Предложенное в системе ГеоТайм представление значений  $m_0(\lambda, \varphi, t)$  в виде динамического поля позволяет наглядно отобразить совершенствование и деградацию сейсмической сети в пространстве и во времени. Кроме того, трёхмерная растровая модель пространственно-временного распределения значений  $m_0(\lambda, \varphi, t)$  необходима для оценивания практически всех других динамических полей параметров сейсмического потока. Метод статистического оценивания  $m_0$  разработан в работах Молчан и Дмитриева, 1991, Писаренко, 1989, алгоритм оценивания разработан в работе Смирнов, 1995 и совместно с ним модифицирован для ГИС ГеоТайм.

Алгоритм оценивания представительной магнитуды в каждой точке растра динамического поля использует гистограммы распределения землетрясений по магнитуде (или энергетическому классу) в диапазоне от  $M_{\min}$  до  $M_{\max}$ , построенные с шагом  $\Delta M$ , который определяется чувствительностью сейсмической сети. В каждой точке поля с достаточным объёмом сейсмических наблюдений ищется минимальное значение  $m_0(\lambda, \varphi, t)$ , при котором график повторяемости в области  $M \geq m_0(\lambda, \varphi, t)$  ещё линеен. Для этого на каждом шаге, начиная с  $m_0(\lambda, \varphi, t) \equiv M_{\min}$ , проверяется гипотеза о линейности эмпирическо-

го графика повторяемости. Гипотеза принимается, если эмпирическое распределение землетрясений по магнитудам на выбранном уровне значимости согласовано с теоретическим распределением Гутенберга–Рихтера, и текущее значение  $m_0(\lambda, \varphi, t)$  назначается оценкой представительной магнитуды. В противном случае текущее значение  $m_0(\lambda, \varphi, t)$  увеличивается на  $\delta m$  и процедура повторяется.

**10.2.3. Параметры сейсмического режима.** Классическими параметрами сейсмического режима являются параметры зависимости Гутенберга–Рихтера  $\lg N(m) = a - b(m - m_0)$ , где  $N$  — число событий, а  $m$  — магнитуда события,  $m_0$  — минимальная представительная магнитуда регистрируемых землетрясений,  $m_0 < m \leq M_{\max}$ ,  $M_{\max}$  — предельная магнитуда. Легко видеть, что параметр  $a$  связан с логарифмом числа землетрясений с магнитудой  $m_0$ . Обычно вместо параметра  $a$  используется нормированная по пространственно-временному окну и приведённая к заданной магнитуде землетрясений величина, которая называется сейсмической активностью. Сейсмическая активность характеризует интенсивность сейсмического процесса. Наклон графика повторяемости определяется параметром  $b$  ( $b$ -value). Этот параметр характеризует распределение землетрясений по энергиям. Поскольку энергия землетрясения тесно связана с размерами очага, то можно считать, что наклон графика повторяемости характеризует распределение сейсмогенных разрывов по их размерам, а следовательно, характеризует и соотношение между размерами структурных неоднородностей различного ранга в геологической среде.

Параметры закона Гутенберга–Рихтера не отражают геометрические особенности распределения эпицентров землетрясений, а следовательно, и геометрические особенности структуры геологической среды. Для этой цели, в частности, используется распределение расстояний между парами землетрясений, которое предполагается экспоненциальным. Параметр этого распределения принято называть корреляционной или фрактальной размерностью и обозначать буквой  $D$ , ( $D$ -value).

В среде GeoТайм используются следующие общепринятые оценки параметров сейсмического режима.

— Оценка максимального правдоподобия для параметра наклона графика повторяемости землетрясений ( $b$ -value):

$$b = \beta \lg e \approx 0,43\beta,$$

где

$$\beta(\lambda, \varphi, t) = \left( \frac{1}{N_{m_0}} \sum_{m_i \geq m_0} (m_i - m_0(\lambda, \varphi, t)) \right)^{-1}.$$

— Оценка максимального правдоподобия для параметра сейсмической активности — нормированное по пространственно-временному

окну число землетрясений, приведённое к магнитуде

$$m \in \left( m_A - \frac{\delta m}{2}, m_A + \frac{\delta m}{2} \right)$$

$$A(\lambda, \varphi, t) = k N_{m_0}(\lambda, \varphi, t) \left[ \exp \left( -\beta(\lambda, \varphi, t) \left( m_A - \frac{\delta m}{2} - m_0(\lambda, \varphi, t) \right) \right) - \exp \left( -\beta(\lambda, \varphi, t) \left( m_A + \frac{\delta m}{2} - m_0(\lambda, \varphi, t) \right) \right) \right],$$

где:  $N_{m_0}(\lambda, \varphi, t)$  — число землетрясений с магнитудой

$$m \in (m_0(\lambda, \varphi, t), M_{\max}(\lambda, \varphi, t)), \quad k = \frac{1}{ST},$$

$S$  — площадь наблюдаемой области,  $T$  — временной интервал, для которых производится оценивание.

— Оценка максимального правдоподобия для параметра  $D$  ( $D$ -value):

$$D(\lambda, \varphi, t) = \left( \ln r_{\max}(\lambda, \varphi, t) - \frac{1}{2S(\lambda, \varphi, t)} \sum_{i=1}^{S(\lambda, \varphi, t)} \ln r_i^2(\lambda, \varphi, t) \right)^{-1},$$

где:

$$S(\lambda, \varphi, t) = \frac{N_{m_0}(\lambda, \varphi, t)(N_{m_0}(\lambda, \varphi, t) - 1)}{2},$$



Рис. 10.2. Модель освещённости поверхности рельефа, активные разломы и эпицентры землетрясений с энергетическим классом  $K \geq 12$ . Эпицентр Байсорунского землетрясения,  $\lambda = 77,93^\circ$ ,  $\varphi = 42,93^\circ$ ,  $K = 14,6$ , 12.11.1990 г., показан звёздочкой

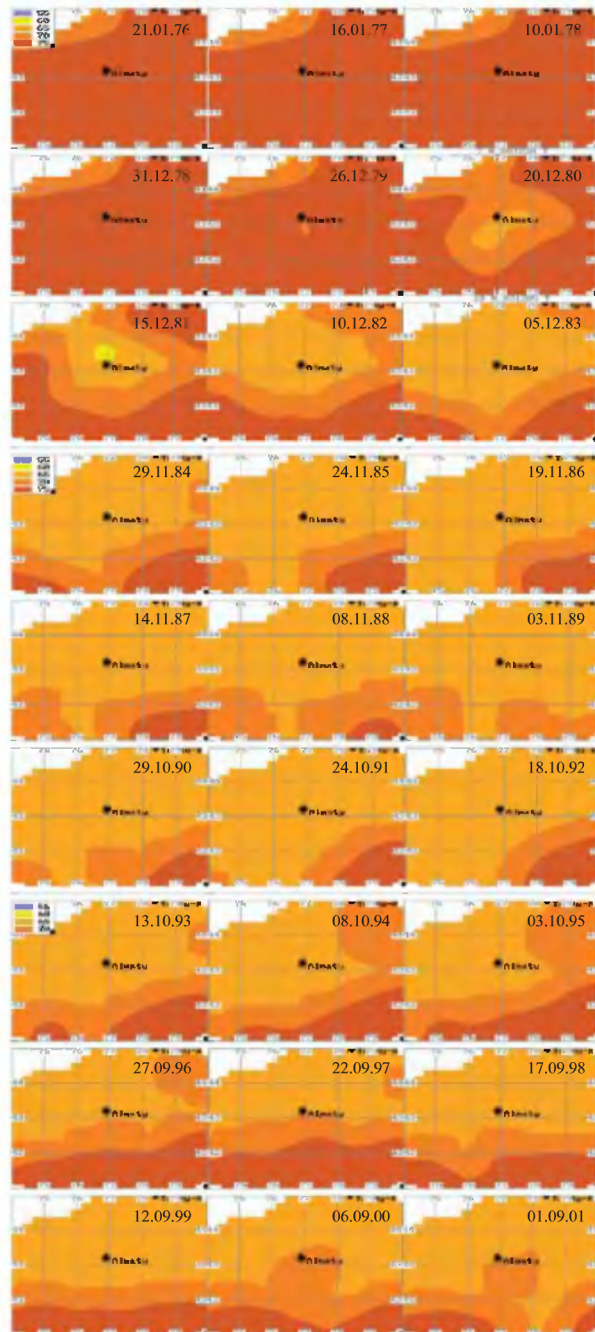


Рис. 10.3. Оценка пространственно-временных изменений минимальных значений представительного энергетического класса регистрируемых землетрясений  $K_{\min}$  с 1975 по 2001 гг. Значения  $K_{\min}$  в легенде умножены на 10. Даты указаны на каждой карте

$N_{m_0}(\lambda, \varphi, t)$  — число землетрясений с магнитудой

$$m \in (m_0(\lambda, \varphi, t), M_{\max}(\lambda, \varphi, t)),$$

$r_i(\lambda, \varphi, t)$  — расстояние между двумя эпицентрами,  $r_{\max}(\lambda, \varphi, t)$  — максимальное расстояние между парой землетрясений, которое выбирается как некоторая фиксированная доля от радиуса пространственной области.

В работе *Соболев и Тюпкин, 1996* предложен функционал, отражающий динамику энергетических свойств сейсмического потока и связанных с ним структурных свойств геологической среды, названный критерием *RTL*. Близкий к *RTL* функционал реализован в системе ГеоТайм в виде:

$$F(\lambda, \varphi, t) = \sum_n E_n^\alpha \exp\left(-\frac{t-t_n}{T}\right) \exp\left(-\frac{r_n(\lambda, \varphi, t)}{R}\right),$$

где:  $t_n$  — момент от  $n$ -го землетрясения,  $r_n(\lambda, \varphi, t)$  — расстояние от эпицентра  $n$ -го землетрясения до точки раstra  $(\lambda, \varphi, t)$ ,  $T$  — декремент затухания экспоненты по времени, если  $t_n(\lambda, \varphi, t) = T$ ,  $r_n(\lambda, \varphi, t) = 0$ , то вклад  $n$ -го события в сумму берётся с весом 0,36,  $R$  — декремент затухания экспоненты по пространству; если  $t_n(\lambda, \varphi, t) = 0$ ,  $r_n(\lambda, \varphi, t) = R$ , то вклад  $n$ -го события в сумму берётся с весом 0,36,  $E_n$  — энергия  $n$ -го землетрясения,  $\lg E_n = c_1 + c_2 m_n$ ,  $m_n$  — магнитуда (или энергетический класс) землетрясения,  $c_1$  и  $c_2$  — параметры корреляционной зависимости между энергией и магнитудой землетрясения, например,  $c_1 = 4$ ,  $c_2 = 1,8$ ,  $\alpha$  — степенной параметр, выбор которого можно пояснить на следующих примерах: при  $\alpha = 1$  по критерию *RTL* суммируются энергии землетрясений с весами, экспоненциально убывающими с увеличением расстояния и временного интервала от точки раstra  $(\lambda, \varphi, t)$  до эпицентра и момента времени землетрясения, при  $\alpha = 0,33$  суммируются взвешенные размеры очагов землетрясений.

**10.2.4. Пространственно-временной анализ каталога землетрясений перед Байсорунским землетрясением.** Для анализа использовался каталог землетрясений Института сейсмологии Министерства образования и науки Республики Казахстана. В каталоге сняты афтершоки и рои землетрясений. Каталог покрывает территорию  $41^\circ$ – $45^\circ$  СШ и  $74^\circ$ – $80^\circ$  ВД и относится к периоду 1975–2002 гг. В каталоге представлены события от энергетического класса  $K$  от 3,3 до 15,3.

Байсорунское землетрясение энергетического класса  $K = 14,6$ , координаты эпицентра  $\lambda = 77,93^\circ$  по долготе, и  $\varphi = 42,93^\circ$  по широте, произошло 12.11.1990 г. Эпицентр Байсорунского землетрясения показан на рис. 10.2. На рисунке показана также модель освещённости поверхности рельефа, активные разломы и эпицентры землетрясений с энергетическим классом  $K \geq 12$ .

На рис. 10.3 представлены результаты пространственно-временного анализа чувствительности сети сейсмического мониторинга с 1975 по 2001 г. Минимальное представительное значение энергетического класса регистрируемых землетрясений  $K_{\min}$  оценивалось по сетке с шагом по долготе  $\Delta\lambda = 0,3^\circ$ , по широте  $\Delta\varphi = 0,2^\circ$  и по времени  $\Delta T = 365$  дней. При оценивании использовались следующие параметры адаптивного скользящего окна: минимальный радиус  $R_{\min} = 100$  км, максимальный радиус  $R_{\max} = 200$  км, минимальный временной интервал  $T_{\min} = 730$  дней, максимальный временной интервал  $T_{\max} = 1825$  дней, минимальное число событий в окне  $N_{\min} = 100$ , номинальное число событий временной интервал  $N_n = 200$ . В некоторых пространственно-временных областях значения  $K_{\min}$  заменялись символами пропущенных значений, так как объём выборки оказывался недостаточным для оценивания (в окне  $R_{\max} = 200$  км и  $T_{\max} = 1825$  дней оказывалось  $N < 100$ ). Пропуски в оценках заполнялись с помощью линейной интерполяции во времени. После этого каждый временной срез был пространственно сглажен с  $R = 30$  км.

Для анализа пространственно-временных аномалий, предшествующих Байсорунскому землетрясению, использовались динамические поля с шагом по долготе  $\Delta\lambda = 0,3^\circ$ , по широте  $\Delta\varphi = 0,2^\circ$  и по времени  $\Delta T = 30$  дней. Анализ выполнялся по следующей схеме:

1. Оценка минимального значения представительного энергетического класса  $K_{\min}$  с параметрами:  $R_{\min} = 100$  км,  $R_{\max} = 175$  км,  $T_{\min} = 730$  дней,  $T_{\max} = 1460$  дней,  $N_{\min} = 75$ ,  $N_n = 150$ , линейная интерполяция во времени.
2. Оценка полей сейсмической активности ( $x_1$ ) для энергетического класса  $K_0 = 10$  наклона графика повторяемости ( $x_2$ ) и фрактальной размерности с параметрами ( $x_3$ ):  $R_{\min} = 75$  км,  $R_{\max} = 150$  км,  $T_{\min} = 360$  дней,  $T_{\max} = 730$  дней,  $N_{\min} = 50$ ,  $N_n = 100$ , фрактальная размерность оценивалась в скользящем цилиндре  $R_{\max} = 100$  км, эпицентры, находящиеся на расстоянии менее 10 км при оценивании не учитывались.
3. Оценка полей аномалий  $y_1(x_1)$ ,  $y_2(x_2)$ ,  $y_3(x_3)$  для Гауссовской модели (9.3) с параметрами:  $T_1 = 3600$  дней (интервал в 30 дней, умноженный на 120 временных срезов динамического поля),  $T_2 = 90$  дней (интервал в 30 дней, умноженный на 120 временных срезов динамического поля), линейная интерполяция во времени, сглаживание  $R = 30$  км.
4. Оценка поля аномалий по трём характеристикам сейсмического процесса. При этом на основании моделей подготовки землетрясений (Соболев, 1993) предполагалось, что перед землетрясением сейсмическая активность понижается, наклон графика повторяемости увеличивается и фрактальная размерность очаговой зоне уменьшается. С учётом этих предположений комплексное решение имеет вид:

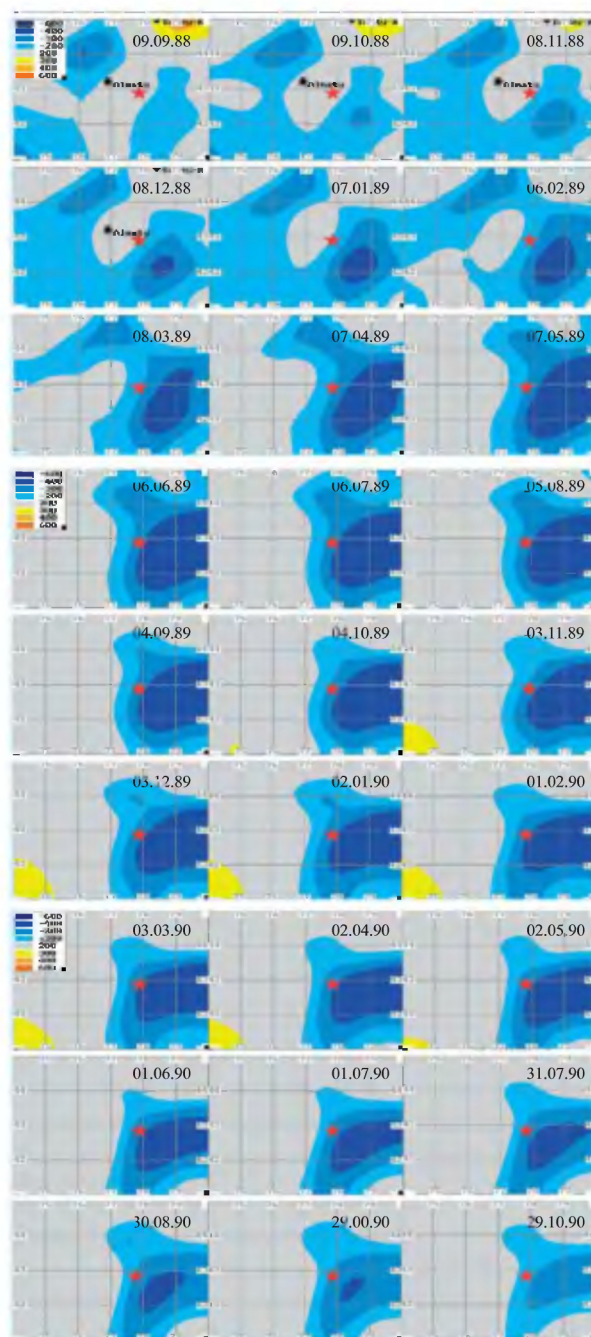


Рис. 10.4. Оценка пространственно-временных изменений сейсмической активности перед Байсорунским землетрясением 12.11.1990 г. Значения аномалий, статистики (9.3), умножены на 100, звёздочкой показан эпицентр землетрясения. Даты указаны каждой карте



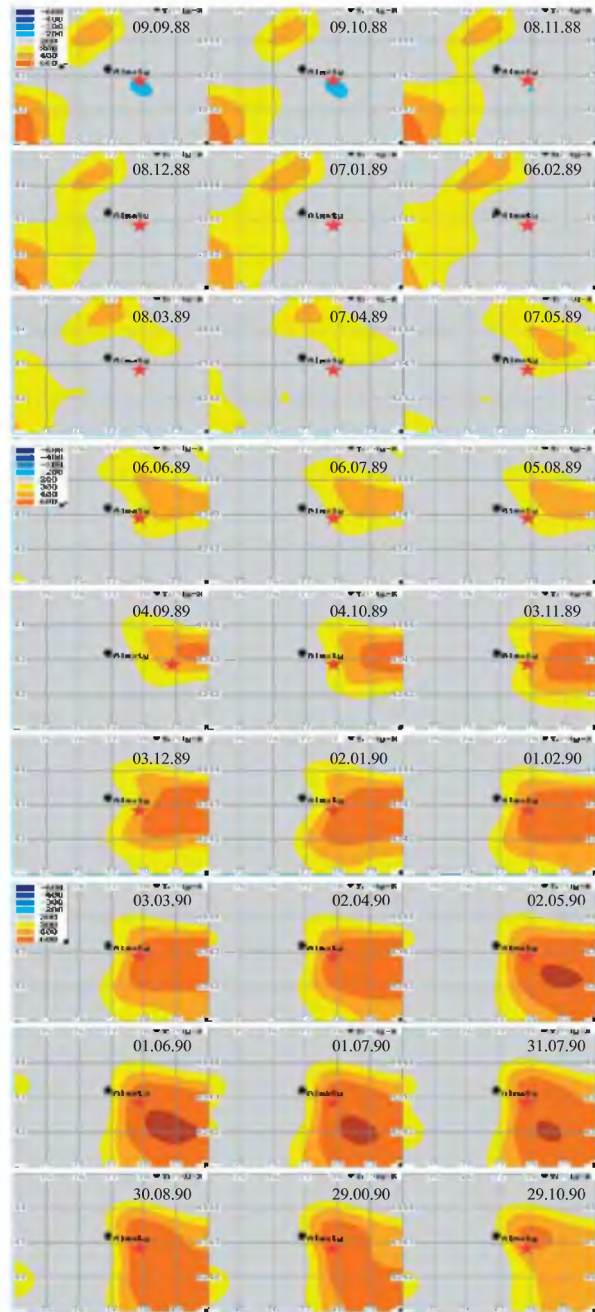


Рис. 10.5. Оценка пространственно-временных изменений аномалий наклона графика повторяемости перед Байбуринским землетрясением 12.11.1990 г. Значения аномалий статистики (9.3) умножены на 100, звездочкой показан эпицентр землетрясения. Даты указаны на каждой карте

$$F(z_1, z_2, z_3) = \min \{z_1, z_2, z_3\},$$

где:

$$z_1 = \begin{cases} -y_1, & \text{если } y_1 < 0, \\ 0, & \text{если } y_1 \geq 0. \end{cases}$$

$$z_2 = \begin{cases} 0, & \text{если } y_2 < 0, \\ y_2, & \text{если } y_2 \geq 0. \end{cases}$$

$$z_3 = \begin{cases} -y_3, & \text{если } y_3 < 0, \\ 0, & \text{если } y_3 \geq 0. \end{cases}$$

На рис. 10.4, 10.5, 10.6 показаны временные срезы динамических полей аномалий сейсмической активности, наклона графика повторяемости и фрактальной размерности. Значения статистики (9.3), указанные в легенде, умножены на 100. Карты соответствуют временным срезам динамического поля с шагом 30 дней. Указанная на карте дата среза соответствует концу интервала  $T_2$ . Например, первая карта с датой 9.09.1988 соответствует положению интервала  $T_2$  от 12.06.1988 до 9.09.1988 гг. (два года и два месяца до Байсорунского землетрясения), а последняя карта с датой 29.10.1990 соответствует положению интервала  $T_2$  от 10.09.1990 до 29.10.1990 гг. (две недели до Байсорунского землетрясения).

Из рис. 10.4 видно, что за два года землетрясения начинает формироваться отрицательная аномалия, западный край которой захватывает эпицентр готовящегося землетрясения. Отрицательное значение статистики (9.3) означает, что среднее значение сейсмической активности в окне  $T_2$  меньше среднего значения сейсмической активности в предыдущем окне длительностью  $T_1$  (около десяти лет), которое можно рассматривать как фоновое. То есть сейсмическая активность в зоне ожидаемого землетрясения в среднем уменьшилась. Можно видеть, что за полтора года до землетрясения отрицательная аномалия стала значимой (статистика (9.3) достигла величины  $4\sigma$ ) и приобрела устойчивую форму. Аномалия достигает максимума примерно за год до землетрясения, затем аномалия очень медленно убывает, и, наконец, за последние полтора-два с половиной месяца до землетрясения скорость убывания заметно увеличивается.

На рис. 10.5. показаны срезы динамического поля аномалий наклона графика повторяемости. Значения статистики (9.3), указанные в легенде, умножены на 100. Из рисунка видно, что аномалии динамического поля наклона графика повторяемости положительные. Это свидетельствует о том, что отношение числа слабых землетрясений к сильным увеличилось. С точностью до знака характер развития аномалии напоминает изменение поля сейсмической активности. Сначала в регионе появляются нерегулярные положительные аномалии, за

полтора года до землетрясения формируется устойчивая форма положительной аномалии, западная часть которой захватывает эпицентр Байсорунского землетрясения, за полгода до землетрясения аномалия достигает максимума (статистика (9.3) достигает величины  $6\sigma$ ) и затем аномалия заметно убывает.

На рис. 10.6 показаны срезы динамического поля аномалий фрактальной размерности. Значения статистики (9.3), указанные в легенде, умножены на 100. Из рисунка видно, что аномалии динамического поля фрактальной размерности имеют более сложный характер. К северу от эпицентра Байсорунского землетрясения преобладают положительные аномалии, а в эпицентральной зоне и к востоку от неё — отрицательные аномалии. Это свидетельствует об уменьшении фрактальной размерности в эпицентральной зоне, т. е. об изменении характера сейсмичности в эпицентральной зоне и к востоку от неё с диффузного к более линеаризованному.

На рис. 10.7 показаны срезы поля аномалий, найденного по трём характеристикам сейсмического процесса:

$$F(z_1, z_2, z_3) = \min \{z_1, z_2, z_3\}.$$

Значения статистики (9.3), указанные в легенде, умножены на 100. Легко видеть, что комплексное решение позволяет уменьшить аномалии вне эпицентральной зоны ожидаемого землетрясения.

### 10.3. Исследование процесса подготовки землетрясения по данным геомониторинга

**10.3.1. Исходные данные.** В анализируемом районе расположено около 50 станций, на которых в течение нескольких десятков лет регистрируются геофизические, гидрогеологические, гидрохимические наблюдения. Для анализа выбраны все временные ряды с ежедневными наблюдениями от 1 января 1972 года для Тяньшанского землетрясения и от 1 января 1981 года для Датонгского землетрясения, которые не содержали существенных пропусков, вызванных сбоями измерительной и регистрирующей аппаратуры. В целом оказалось, что этим условиям удовлетворяют только 16 временных рядов, относящихся к различным геофизическим характеристикам.

Полный список наблюдений представлен в табл. 10.1. Станции с 1 по 10 отобраны для анализа Тяньшанского землетрясения, а станции с 7 по 16 отобраны для анализа Датонгского землетрясения. Для анализа временные ряды были ограничены за сутки перед соответствующим землетрясением. Таким образом, для Тяньшанского землетрясения использовались наблюдения от 1 января 1972 г. до 27 июля 1976 г., а для анализа Датонгского землетрясения — от 1 января 1981 г. до 18 октября 1989 г.

На рис. 10.8 (верхняя диаграмма, график *a*) показан пример временного ряда наблюдений изменений наклона на станции FS. Видны

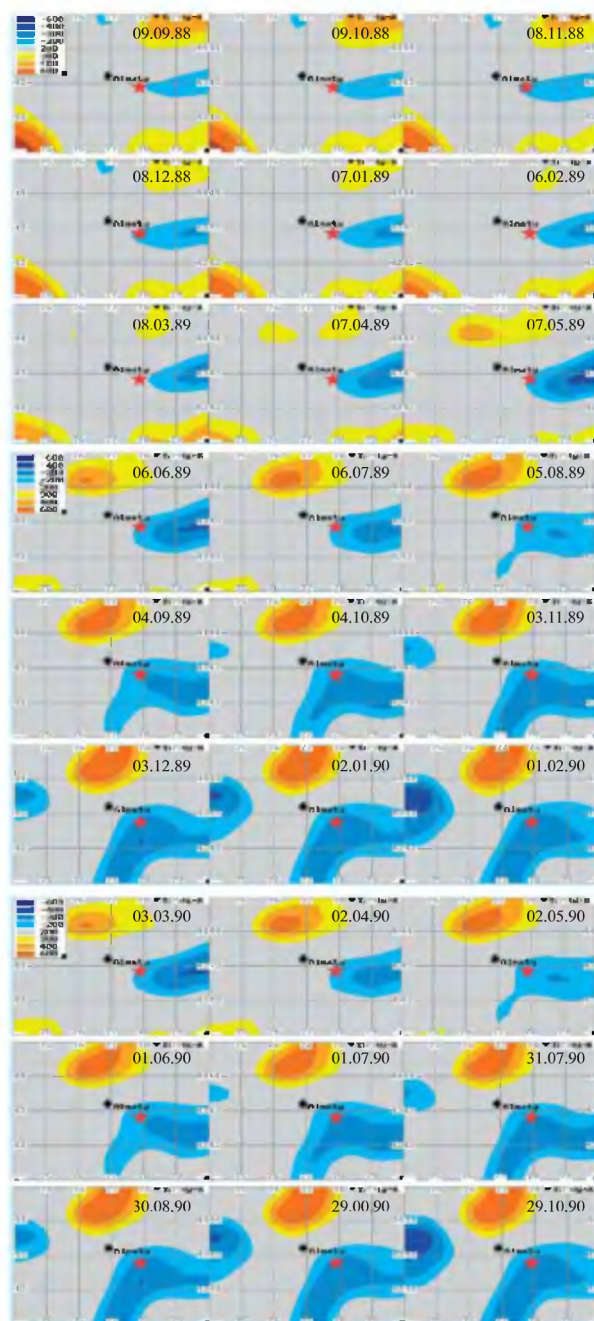


Рис. 10.6. Оценка пространственно-временных изменений аномалий фрактальной размерности перед Байсорунским землетрясением 12.11.1990 г. Значения аномалий статистики (9.3) умножены на 100, звёздочкой показан эпицентр землетрясения. Даты указаны на каждой карте

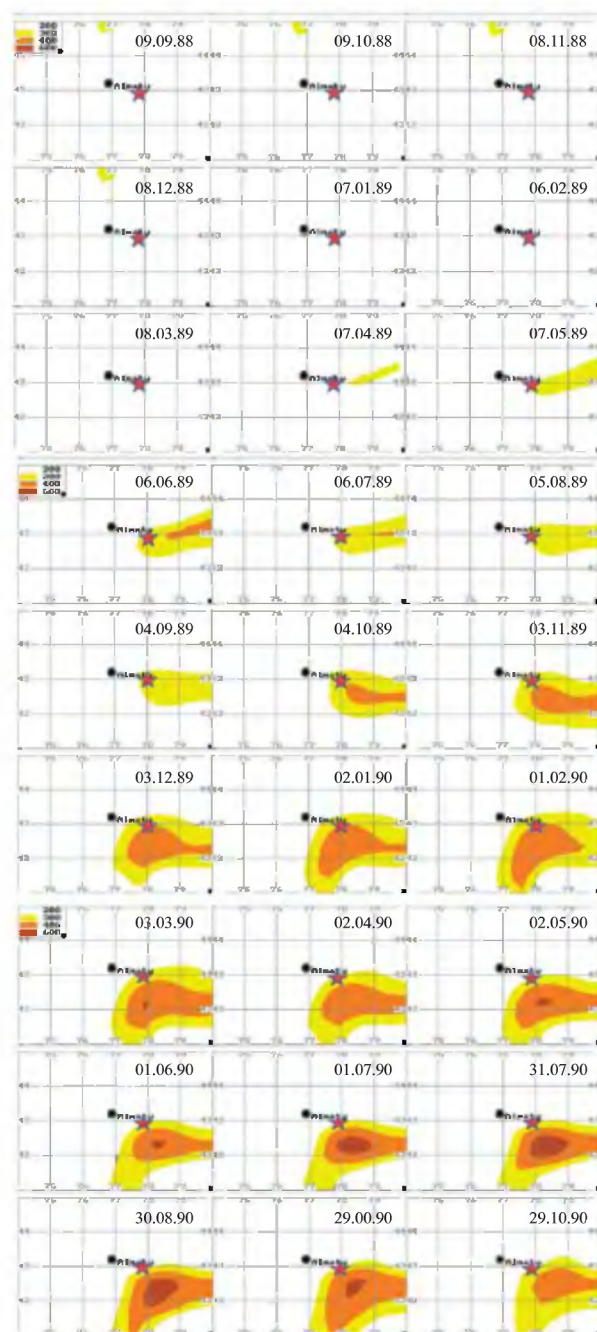


Рис. 10.7. Оценка пространственно-временных изменений функции  $F$  от аномалий сейсмической активности, наклоне графика повторяемости и фрактальной размерности. Значения функции  $F$  умножены на 100, звездочкой показан эпицентр землетрясения. Даты указаны на каждой карте

сильные сезонные вариации с годовым циклом, которые усложняются высокочастотными пиками в июле, связанными с интенсивными дождями. Спектральный анализ рядов показывает наличие максимума на периоде в 1 год. На рис. 10.8 (нижняя диаграмма, график *a*) представлен пример изменений уровня воды в скважине на станции TS.

Таблица 10.1. Список временных рядов, выбранных для исследования предвестников Тяньшанского и Датонского землетрясений

№	Название станции	Период наблюдений	Параметр
1	Джин Ксинь (JH)	1972.01.01–1982.12.31	Наклон
2	Дахуичанг (DH)	1970.01.01–1990.12.31	Наклон
3	Байшан (BS)	1971.01.01–1979.12.31	Уровень воды
4	Венжуан (BH)	1971.01.01–1979.12.31	Уровень воды
5	Нингхе (TS)	1970.07.21–1979.12.31	Уровень воды
6	Дачанг (DC)	1972.01.01–1979.12.31	Уровень воды
7	Фангшан (FS)	1970.01.01–1990.12.31	Наклон
8	Шуангкяо (SQ)	1972.01.01–1990.12.31	Уровень воды
9	Чангли (CL)	1970.01.01–1995.07.17	Сопротивление
10	Баоди (BD)	1970.01.01–1995.07.17	Сопротивление
11	Дахуичанг (DH)	1970.01.01–1990.12.31	Нивелирование
12	Гуанзхуанг (GZ)	1969.01.01–1990.12.31	Радон
13	Гаоцун (GC)	1980.01.01–1990.12.31	Уровень воды
14	Хуайдай (HL01)	1972.03.01–1990.12.31	Радон
15	Ютиан (YT)	1981.01.01–1992.12.31	Уровень воды
16	Тайюань (TY)	1979.02.03–1990.12.31	Наклон

В этой точке наблюдений периодичность сезонных вариаций менее выражена, в то же время имеются многочисленные выбросы, которые можно объяснить как наличием техногенных возмущений, так и случайными ошибками оператора.

Использовалась упрощённая аддитивная модель сигнала. Предполагалось, что сигнал  $X(t)$  состоит из отклика на геодинамический процесс подготовки землетрясения  $S(t)$ , периодической компоненты сезонных ритмов  $U(t)$  и высокочастотного шума с нулевым средним  $V(t)$ :

$$X(t) = S(t) + U(t) + V(t).$$

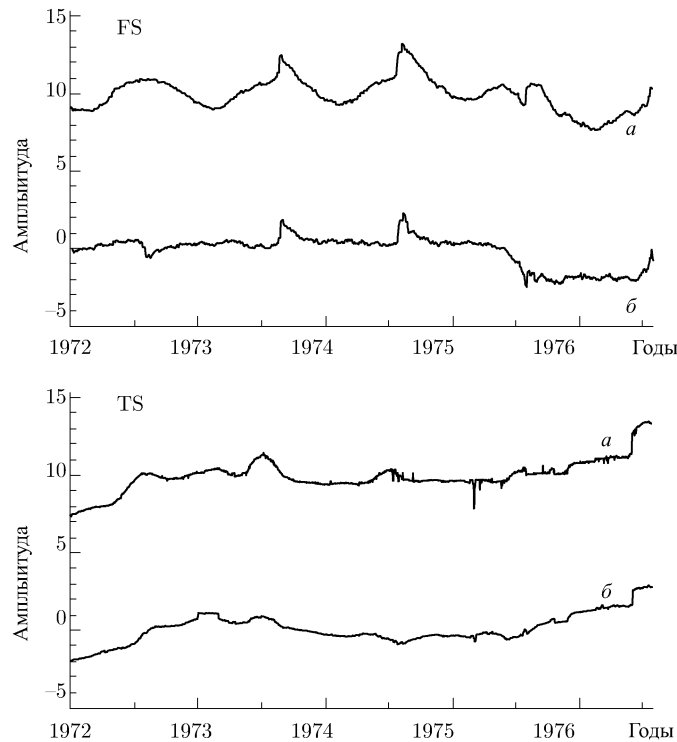


Рис. 10.8. Временные ряды до (а) и после (б) снятия сезонного ритма. Верхние графики относятся к данным по изменению наклона на станции FS, нижние графики относятся к данным по изменению уровня воды в скважине на станции TS

В разделе 9.5.4 описан метод, в котором задача обнаружения аномалий для разнотипных временных рядов решается с использованием многомерной статистической модели. Однако в нашем случае этот подход использовать не удаётся из-за очень малого числа станций наблюдения. В связи с этим приходится выбрать компромиссное решение, при котором все временные ряды независимо от физической природы измерений обрабатываются как однотипные скалярные переменные. Для этого все временные ряды стандартизируются и нормализуются по дисперсиям.

После стандартизации и нормализации обработка временных рядов выполнялась в три этапа: (I) очистка временных рядов от сезонных ритмов, (II) вычисление динамических полей измерительных сигналов, (III) вычисление динамических полей аномалий.

**10.3.2. Предварительная обработка временных рядов.** Относительно короткие временные ряды наблюдений около 5 лет перед Тянь-

шанским и около 9 лет перед Датонгским землетрясениями позволяют анализировать только среднесрочные и краткосрочные предвестники. Относительно большие и сложные по форме вариации сезонных ритмов в значительной степени затруднили поиск предвестников продолжительностью более одного года. В связи с этим целью анализа на данном этапе исследований было нахождение аномалий продолжительностью от одного до нескольких месяцев.

Для оценивания образа годового сезонного ритма использовались временные ряды за предыдущие 2 года. Так, например, для Таньшанского землетрясения оценка ритма за 1972 и 1973 годы принималась за образ сезонного ритма в 1974 году и при очистке вычиталась из временного ряда. Аналогично обрабатывались данные с 1974 по 1976 годы. Чтобы не потерять статистику за первые 2 года, ритм за 1972 год оценивался на статистике 1973 и 1974 годов, а ритм для 1973 — по статистике за 1974 и 1975 годы.

На рис. 10.8 (верхняя и нижняя диаграммы, графики б) показаны временные ряды станций FS и TS после вычитания сезонных ритмов. Нетрудно видеть, что сигналы после обработки содержат не только аномалии, обусловленные неполным снятием сезонного хода, но и отклонения неизвестной природы, среди которых вероятно есть и предвестники землетрясений.

Известно, что предвестники землетрясений независимо от типа измеряемого физического параметра могут иметь вид как положительных, так и отрицательных аномалий. В 1975–1976 годах для станции FS заметна аномалия понижения сигнала, в то время как для станции TS наблюдается повышение уровня сигнала. Обе аномалии могут быть отнесены к предвестникам Таньшанского землетрясения. На основании этих соображений полученные временные ряды были квадрированы. Полученный сигнал можно интерпретировать как энергетический отклик геологической среды на тектонические процессы.

**10.3.3. Пространственная интерполяция.** По временным рядам была вычислена сеточная модель трёхмерного динамического поля с шагом 0,1 градуса по параллелям, 0,67 градуса по меридианам и 5 дней по временной оси. Интерполяция производилась по формуле

$$Z(\lambda, \varphi, t) = \frac{\sum_{n=1}^{10} Y_n(t)/r_n}{\sum_{n=1}^{10} 1/r_n},$$

где  $Z(\lambda, \varphi, t)$  — значение в точке раstra  $(\lambda, \varphi, t)$ ,  $Y_n(t)$  — значение временного ряда (после первичной обработки)  $n$ -й станции в момент  $t$ ,  $r_n$  — расстояние от  $n$ -й станции до точки  $(\lambda, \varphi, t)$ , при  $r_n = 0$ ,  $Z(\lambda, \varphi, t) = Y_n(t)$ .



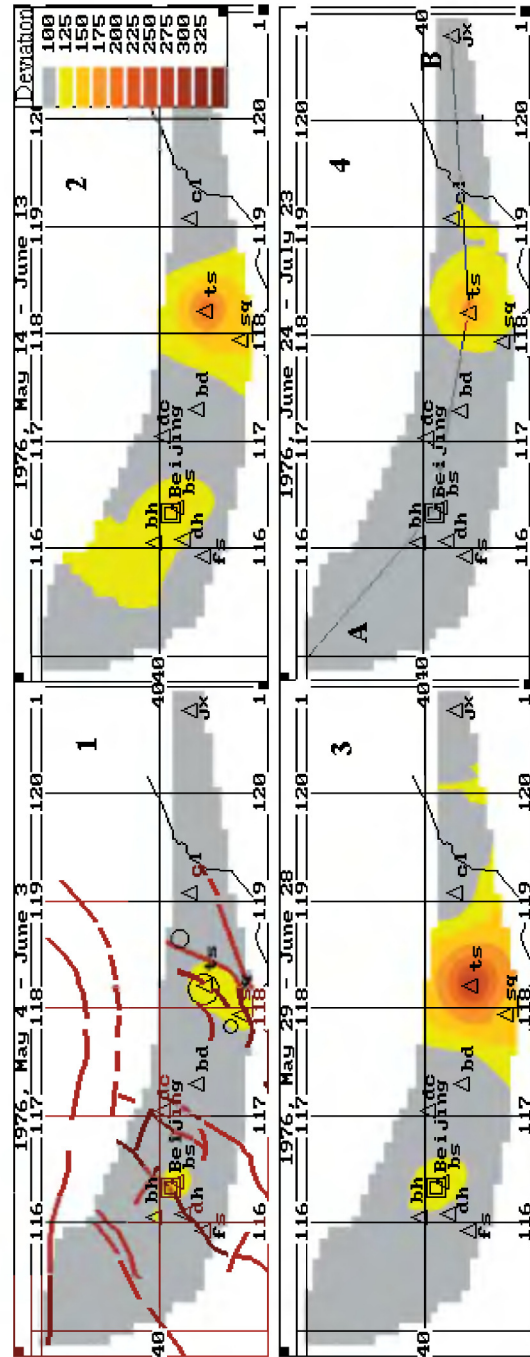


Рис. 10.9. Временные срезы динамического поля  $u(\lambda, \varphi, t)$  перед землетрясением в мае-июле 1976. Значения  $u(\lambda, \varphi, t)$  в легенде умножены на 100 (в верхнем углу справа). Beijing — Пекин

**10.3.4. Обнаружение аномалии.** На этом этапе для каждого элемента раstra проверяются две статистические гипотезы: гипотеза  $H_0$  — математические ожидания значений в первом и втором окнах продолжительностью  $T_1$  и  $T_2$  равны и гипотеза  $H_A$  — математическое ожидание значений во втором окне больше, чем в первом. Если принимается гипотеза  $H_0$ , то процесс считается стационарным, в противном случае считается, что имело место аномальное отклонение от стационарности. Критерий проверки гипотезы  $H_0$  против гипотезы  $H_A$  основан на вычислении статистики, представляющей разность выборочных средних значений:

$$G(\lambda, \varphi, t) = m_2(\lambda, \varphi, t) - m_1(\lambda, \varphi, t),$$

где

$$m_1(\lambda, \varphi, t) = \frac{1}{T_1} \sum_{n=1}^{T_1} Z(\lambda, \varphi, t - T_1 - T_2 + n),$$

$$m_2(\lambda, \varphi, t) = \frac{1}{T_2} \sum_{n=1}^{T_2} Z(\lambda, \varphi, t - T_2 + n),$$

$T_1 = 73$ ,  $T_2 = 6$  отсчётам с пятидневным шагом, что соответствует 365 и 30 дням.

Очевидно, что при выполнении гипотезы  $H_0$  статистика  $G(\lambda, \varphi, t)$  близка к 0. Целесообразно нормировать статистику  $G(\lambda, \varphi, t)$  по среднеквадратичному отклонению. Однако при оценивании дисперсии необходимо учесть, что наблюдения  $Z(\lambda, \varphi, t)$  сильно коррелированы. Будем предполагать, что дисперсия  $\sigma^2(\lambda, \varphi, t)$  является медленно изменяющейся функцией времени и одинакова на интервалах  $T_1$  и  $T_2$ . Тогда в качестве оценки приходится использовать верхнюю границу среднеквадратичного отклонения случайной переменной  $G(\lambda, \varphi, t)$ , которая равна  $2\sigma(\lambda, \varphi, t)$ , где

$$\sigma^2(\lambda, \varphi, t) = \frac{1}{T_1 - 1} \sum_{n=1}^{T_1} (Z(\lambda, \varphi, t - T_1 - T_2 + n) - m_1(\lambda, \varphi, t))^2.$$

Следовательно, для стационарного процесса при выполнении гипотезы  $H_0$  случайная величина

$$u(\lambda, \varphi, t) = \frac{G(\lambda, \varphi, t)}{2\sigma(\lambda, \varphi, t)}$$

имеет среднее 0 и среднеквадратичное отклонение меньше 1.

**10.3.5. Информационная модель прогноза предвестников Тяньшанского землетрясения.** При выбранных продолжительностях окон  $T_1$  и  $T_2$  весь период 1972 года, включая январь 1973 года, не мог быть использован для поиска аномалий. Для остального периода динамическое поле статистики  $u(\lambda, \varphi, t)$  состоит из 257 срезов динамического поля с пятидневным шагом. Последний пространственный

растр относится к периоду, когда окно  $T_2$  покрывает интервал от 24 июня до 23 июля 1976 года.

Анализ аномалий июля–августа показывает неполное удаление сезонных ритмов. Аномалии в интервале от сентября по июль менее обусловлены сезонными причинами и, в частности, обусловлены геодинамическими причинами. В сентябре 1975 года имела место мощная аномалия, покрывающая практически всю изучаемую область. Аномалия затухала и практически полностью исчезла в апреле 1976 года. В районе будущего Тяньшанского землетрясения остались только две локальные аномалии в районе станций NS и SQ.

На рис. 10.9 показаны срезы динамического поля  $u(\lambda, \varphi, t)$  за 1–2 месяца перед Тяньшанским землетрясением и станции наблюдения. Хорошо заметно, что площадь и значимость аномалии сначала увеличивается, а затем уменьшается. В мае 1976 года аномалии начинают возрастать как это видно из рис. 10.9, (1). В то же время в окрестности Пекина аномалия менее выражена. Для удобства на рис. 10.9, (1) нанесены тектонические разломы, эпицентр Тяньшанского землетрясения и два его сильнейших автershока с магнитудами 6,9 и 7,0.

Следующий срез на рис. 10.9, (2) относится к периоду с 14 мая по 13 июня. В этот период аномалия в окрестности Пекина достигает максимума, а в области Тяньшанского землетрясения продолжает нарастать. На срезе, относящемся к периоду от 29 мая до 28 июня, рис. 10.9, (3), представлено распределение максимума аномалии в районе Тяньшаня. На следующем срезе видно, что аномалия непосредственно перед землетрясением ослабевает. Динамика развития аномалий вдоль профиля А–В, изображённого на рис. 10.9, (4), показана

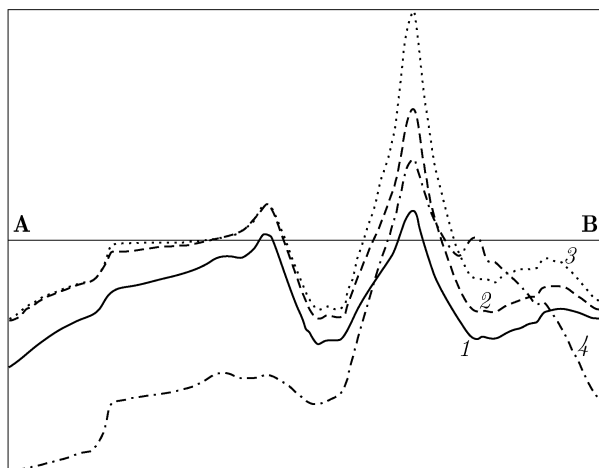


Рис. 10.10. Динамика развития аномалий вдоль профиля А–В, изображённого на рис. 10.9 (4)

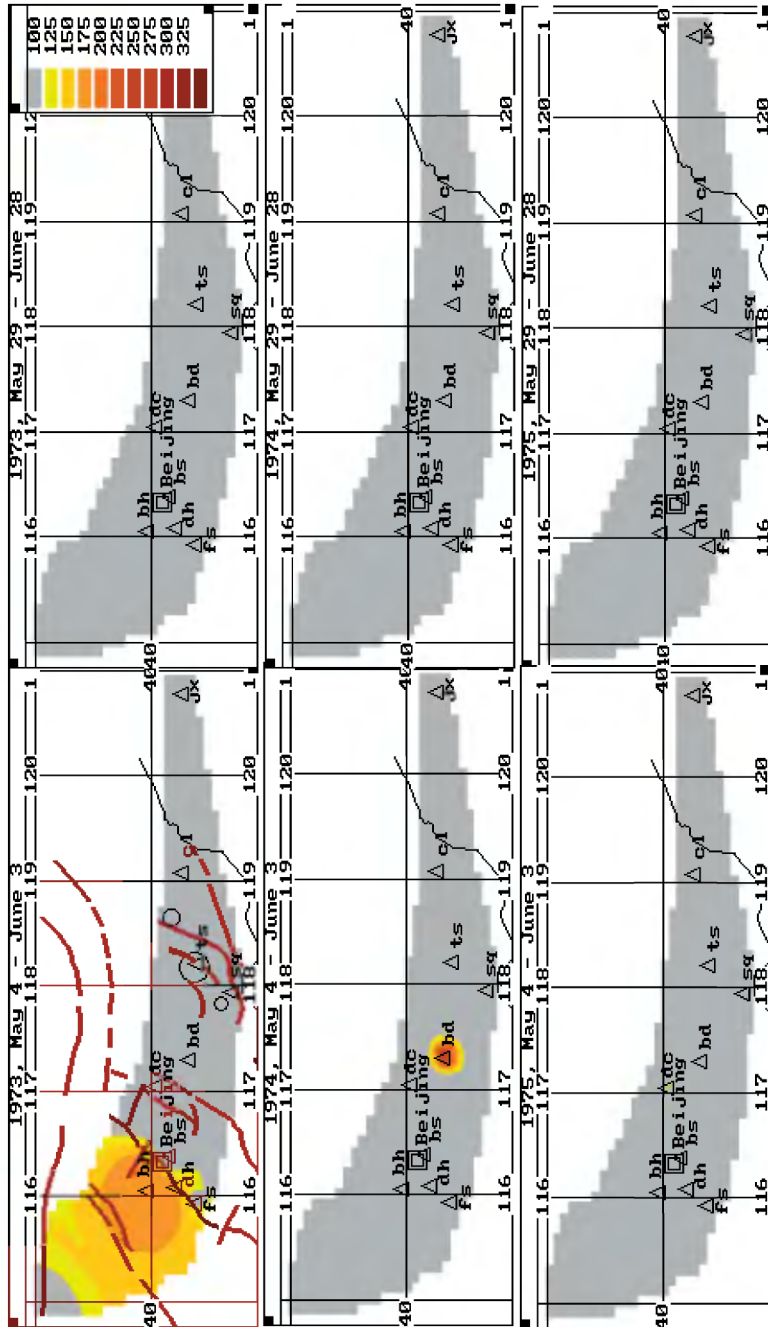


Рис. 10.11. Временные срезы динамического поля  $u(\lambda, \varphi, t)$  в 1973–1975 гг. Значения  $u(\lambda, \varphi, t)$  в легенде умножены на 100 (в верхнем углу справа), Beijing — Пекин

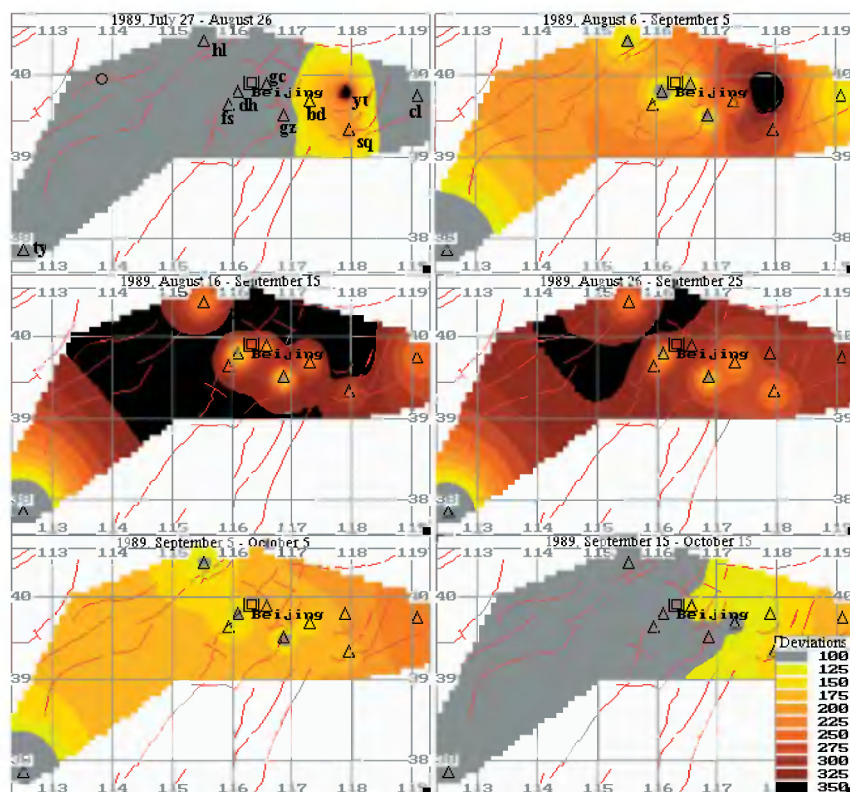


Рис. 10.12. Временные срезы динамического поля  $u(\lambda, \varphi, t)$  перед землетрясением в августе–октябре 1989 г. Значения  $u(\lambda, \varphi, t)$  в легенде умножены на 100 (в нижнем углу справа), Beijing — Пекин

на рис. 10.10. Кривые 1, 2, 3, 4 соответствуют периодам наблюдений соответствующим срезам на рис. 10.9, (1, 2, 3, 4).

Для сравнения на рис. 10.11 показаны 6 срезов динамического поля  $S(\lambda, \varphi, t)$  за предыдущие 3 года, относящиеся к периодам начала аномалии (4 мая – 3 июня) и максимума Тяньшанской аномалии (29 мая – 28 июня). Легко видеть, что в течение этих лет значительных аномалий в изучаемой области не наблюдалось.

Следует заметить, что полученный результат был подтверждён на новых данных с изменённым набором станций.

**10.3.6. Информационная модель прогноза предвестников Датонгского землетрясения.** Аналогичная процедура была применена для ретроспективного прогноза Датонгского землетрясения, 19 октября 1989 г.,  $M = 6,1$ . В этом случае первый срез динамического поля отклонений от стационарности относится к январю 1981 года. Как

и в предыдущем случае, срезы шли с пятидневным шагом. Всего было вычислено 568 срезов. Последний срез относится к периоду, когда окно с продолжительностью  $T_2$  покрывает интервал от 15 сентября до 14 октября 1989 г.

Поведение динамического поля отклонений перед Датонгским землетрясением значительно отличается от поведения для поля отклонений перед Тяньшанским землетрясением. Это легко видеть из срезов динамического поля, представленных на рис. 10.12. В период от 21 июля до 29 августа 1989 года появляется значимая аномалия, интенсивность, которой растёт и распространяется с запада на восток. Затем аномалия достигает максимума в период от 20 августа до 19 сентября и затем сокращается.

Для сравнения на рис. 10.13 показаны срезы динамического поля  $S(\lambda, \varphi, t)$  за предыдущие 1983–1988 годы, относящиеся к периоду максимума Датонгской аномалии для августа–сентября. В течение всех этих лет в изучаемой области наблюдались только локальные аномалии, относящиеся к отдельным станциям.

**10.3.7. Метод обнаружения эпицентральных аномалий.** Сравнение срезов динамических полей отклонений, представленных на рис. 10.9 и рис. 10.12 показывает, что поведение пространственно-временной аномалии перед Тяньшанским землетрясением существенно отличается от поведения аномалии, предшествующей Датонгскому землетрясению. В первом случае аномалия покрывает район будущего землетрясения, в то время как во втором случае аномалия покрывает практически всю область, где расположены станции геофизического мониторинга. Можно предположить, что тип поведения аномалий отражает различный механизм генерирования предвестников землетрясений. Тяньшанская аномалия показывает, что генерирование предвестников имеет локальный характер и связано с эпицентром ожидаемого землетрясения. Датонгская аномалия демонстрирует региональный характер процесса подготовки землетрясения, при котором предвестники не связаны прямо с эпицентром будущего землетрясения, а скорее отражают изменение тектонических напряжений на обширной территории.

Особенности эпицентрального (локального) проявления предвестников неоднократно обсуждались в научной литературе. В частности, по материалам многочисленных наблюдений предвестников для сильных землетрясений Китая установлены зависимости между относительным числом станций, которые зарегистрировали предвестники различной природы, и эпицентрными расстояниями (*Chen Yong et al., 1992, Zhang Guomin and Zhang Zhaocheng, 1992, Zhang Zhaocheng et al., 1992*). Эти данные представлены в табл. 10.2.

Если предположить, что амплитуда предвестника уменьшается примерно так же, как и число станций, а на расстояниях порядка 1000 км

Таблица 10.2. Зависимость числа станций, зарегистрировавших предвестники, от эпицентральных расстояний будущих землетрясений

Число станций, зарегистрировавших предвестники, %	76	47	38
Среднее эпицентральное расстояние, км	50	150	350

предвестники не наблюдаются, то зависимость затухания амплитуды предвестника от расстояния может быть аппроксимирована функцией

$$AF(R) = A(\exp(-0,02R) - 0,27), \quad (10.1)$$

где  $R$  — расстояние до источника сигнала (эпицентра) в километрах,  $A$  — амплитуда предвестника.

Сделанное предположение подкреплено большим количеством эмпирических данных и используется как первое приближение при построении модели предвестниковой аномалии и выявления её пространственно-временного поведения для локализации будущего эпицентра.

Обозначим временные ряды, полученные на  $n$  станциях наблюдения, как  $u(n, t)$  и предположим, что в течение любого фиксированного момента времени центр сигнала-предвестника совпадает с одной из точек растра  $k = 1, 2, \dots, K$ , а значения временного ряда описываются соотношением

$$u_n = A \cdot F(R_{nk}) + \varepsilon_n,$$

где  $A$  — амплитуда сигнала,  $R_{nk}$  — расстояние от сигнала, распространяющегося из точки  $k$ , до точки приёма  $n$ -й станцией,  $\varepsilon_n$  — независимая гауссовская переменная с нулевым математическим ожиданием. Алгоритм оценивания параметров сигнала строится на минимизации функционала

$$\varphi(A, k) = \sum_{n=1}^N (u_n - AF(R_{nk}))^2.$$

Для каждой точки растра получим оценки амплитуды  $A(k)$

$$\hat{A}(\hat{k}) = \frac{\sum_{n=1}^N u_n F(R_{n\hat{k}})}{\sum_{n=1}^N F^2(R_{n\hat{k}})} \quad (10.2)$$

и остаточные суммы квадратов

$$S_k^2 = \sum_{n=1}^N (u_n - \hat{A}(k)F(R_{nk}))^2.$$

Далее, для оценки положения центра сигнала используется отношение сигнал/шум

$$\alpha(k) = \frac{E_s(k)}{E_n(k)},$$

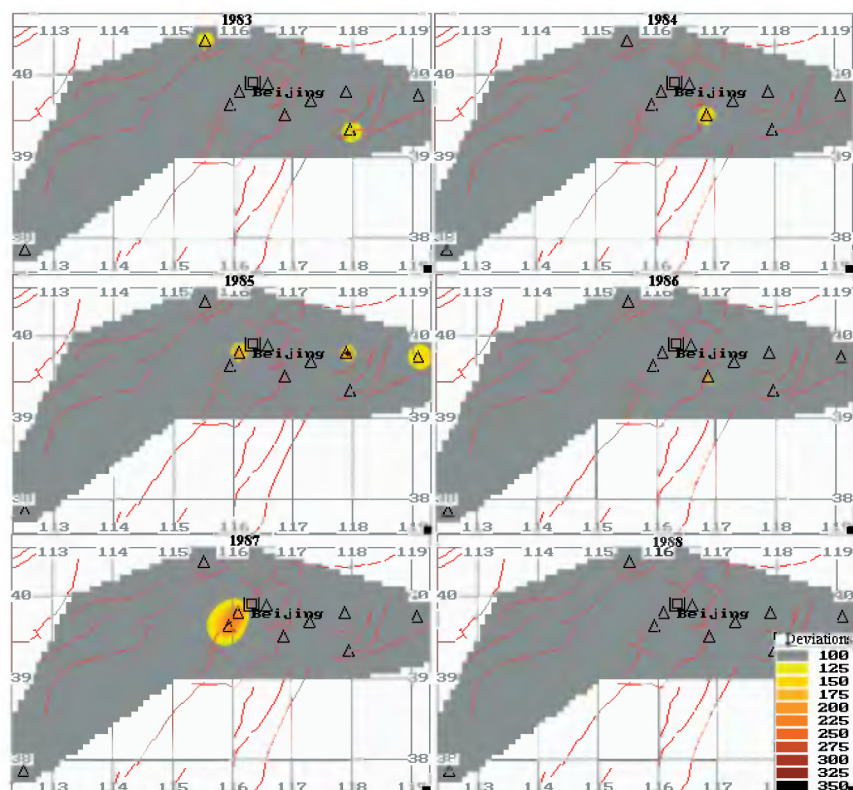


Рис. 10.13. Временные срезы динамического поля  $u(\lambda, \varphi, t)$  в сентябре 1981–1988 годов. Значения  $u(\lambda, \varphi, t)$  в легенде умножены на 100 (в нижнем углу справа), Beijing – Пекин

где

$$E_s(k) = (\hat{A}(k))^2 \sum_{n=1}^N F^2(R_{nk}) - \text{энергия сигнала,}$$

$$\begin{aligned} E_n(k) &= \sum_{n=1}^N S_n^2 = \sum_{n=1}^N u_n^2 - (\hat{A}(k))^2 \sum_{n=1}^N F^2(R_{nk}) = \\ &= \frac{\sum_{n=1}^N u_n^2 \sum_{m=1}^N F^2(R_{mk}) - (\sum_{n=1}^N u_n F(R_{nk}))^2}{\sum_{n=1}^N F^2(R_{nk})} - \text{энергия шума.} \end{aligned}$$



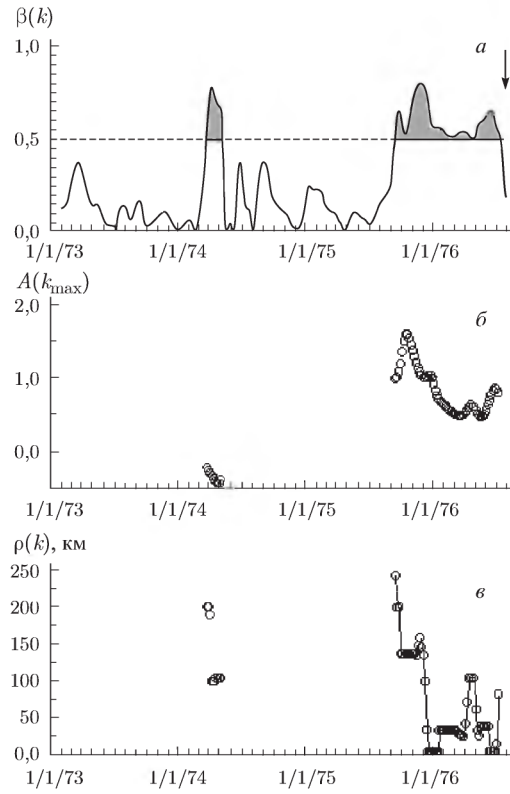


Рис. 10.14. Графики изменения параметров эпицентральной модели перед Тяньшанским землетрясением (момент землетрясения обозначен стрелкой): *a* — график критерия существования сигнала предвестника, *б* — график оценки амплитуды предвестника, *в* — график оценки расстояния от центра модели сигнала до эпицентра Тяньшанского землетрясения

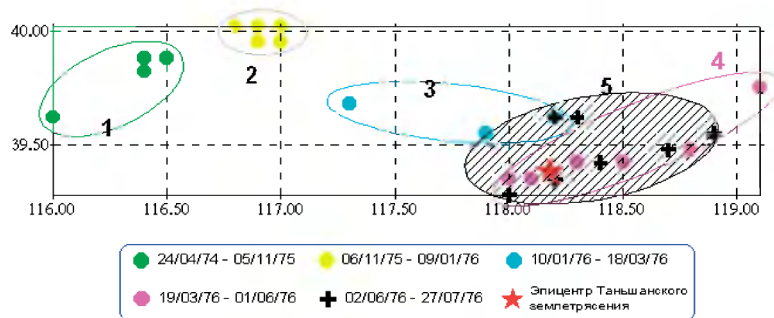


Рис. 10.15. Миграция оценок центра аномалии перед Тяньшанским землетрясением

Более удобна для рассмотрения функция  $\beta(k)$ , монотонно связанная с  $\alpha(k)$

$$\beta(k) = \frac{\alpha(k)}{1 + \alpha(k)} = \frac{\left( \sum_{n=1}^N u_n F(R_{nk}) \right)^2}{\sum_{n=1}^N u_n^2 \sum_{m=1}^N F^2(R_{mk})}.$$

Значение  $\beta(k)$  изменяется в диапазоне от 0 до 1, причём значение 1 отвечает ситуации, когда  $u_n = A_k \cdot F(R_{nk})$ , то есть, когда отношение сигнал/шум очень велико. При  $\beta(k) = 0$  сигнал отсутствует, а при  $\beta(k) \leq 0,5$  неопределённость оценки амплитуды  $A_k$  больше, чем собственно значение амплитуды. В последнем случае мы можем полагать, что в анализируемой области сигнал, соответствующий предложенной эпицентральной модели, не наблюдается.

Таким образом, оценка положения источника сигнала в пространстве выбирается из условия максимума функции  $\beta(k)$ ,  $k = 1, 2, \dots, K$ , то есть,

$$\hat{k} = \arg \max_{\langle k \rangle} \beta(k), \quad (10.3)$$

а значение  $A(\hat{k})$  принимается как оценка амплитуды сигнала (10.2).

**10.3.8. Информационная модель прогноза эпицентральных предвестников землетрясений Тяньшанского землетрясения.** Моделирование эпицентральной аномалии катастрофического Тяньшанского землетрясения было выполнено с использованием среднесуточных временных рядов длительностью около 4 лет, полученных на 1–10 станциях Пекинского прогностического полигона. Предполагалось, что предвестники землетрясения генерируются одним источником, расположенном в пределах полигона, а изменение сигнала описывается уравнением (10.1).

Предварительная обработка включала обычные процедуры стандартизации рядов, медианного сглаживания, удаления трендов, простого подавления сезонных ритмов. Далее вычислялись временные ряды отклонений от стационарности по формулам раздела 10.2.4 с заменой в них координат  $\lambda, \varphi$  на номера станций  $n$ . Параметры окон оценивания были выбраны равными  $T_1 = 365$  дней и  $T_2 = 30$  дней. Последний этап, осуществляющий переход к эпицентральной пространственно-временной модели, состоял в оценивании параметров модели  $\hat{k}$  и  $\hat{A}(\hat{k})$  с интервалом 5 дней по алгоритму (10.2), (10.3).

На рис. 10.14 показаны графики изменения критерия наличия сигнала  $\beta(\hat{k})$ , оценки его амплитуды  $A(\hat{k})$  и оценки расстояния  $\rho(\hat{k}, C)$  между центром модели сигнала и реального эпицентра землетрясения. Совместное рассмотрение полученных зависимостей показывает, что вплоть до августа 1975 г.  $\beta(\hat{k}) < 0,5$ . Это означает, согласно нашей гипотезе, что эпицентральный предвестниковый сигнал либо не проявлялся, либо не соответствовал предложенной модели (кратковременное

появление низкоамплитудного сигнала фиксируется в апреле 1974 г. на значительном удалении от будущего эпицентра и, согласно модели, возможно, обусловлено подготовкой другого землетрясения). С сентября 1975 г. и почти до момента землетрясения значения  $\beta(\hat{k})$  увеличились и устойчиво превысили 0,5. Другими словами, на этом временном интервале предложенная модель наилучшим образом отвечала набору реальных данных. Одновременно было зарегистрировано существенное увеличение амплитуды сигнала. Наконец, начиная с ноября 1975 г., центр модельной аномалии переместился в эпицентральную область землетрясения со средним отклонением от будущего эпицентра не более 50 км.

Это хорошо иллюстрируется на рис. 10.15, на котором показана миграция оценок центра аномалии перед Тяньшанским землетрясением. Каждая выделенная область, от 1-й до 5-й, появляется последовательно, одна за другой, и представлена равным числом оценок моделируемых аномалий. На рисунке отчетливо прослеживается ускоряющаяся миграция предвестника в район эпицентра. Возникновение аномалии в период подготовки землетрясения, близость её пространственного максимума к эпицентру Тяньшанского землетрясения и отсутствие подобной аномалии в предыдущие годы подтверждают гипотезу об её тектоническом происхождении.

Предложенный подход демонстрирует возможность выделения обусловленных геодинамическими процессами (детерминированных) пространственно-временных вариаций, описываемых эмпирическими моделями. По существу, в работе сделана попытка выделить детерминированную часть сигнала, которая описывается моделью, построенной по эмпирическим закономерностям, обусловленным геодинамическими факторами, в частности, подготовкой сейсмических событий.

Предложенный подход позволяет исключить случайные факторы, обусловленные весьма высоким уровнем шумов в исходных наблюдениях, неполным удалением сезонного хода и тренда. Подход позволяет также исключить эпицентральные аномалии, распространяющиеся от источников, которые находятся за пределами полигона, и аномалии, которые не имеют выраженного эпицентрального характера.

## IV. ПРОГНОЗИРОВАНИЕ И АНАЛИЗ В СЕТЕВЫХ ГИС

---

### Глава 11

## СЕТЕВЫЕ АНАЛИТИЧЕСКИЕ ГЕОИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ

### 11.1. Введение

Значительные успехи в телекоммуникационных технологиях и создание единого информационного пространства сети Интернет открыли новые возможности публикации и доступа к географической информации. Для доступа к ГИ в сетях Интернет и Интранет созданы специализированные приложения. Но в ряде случаев конечная цель потребителя ГИ состоит не столько в просмотре изображений карт и получении данных как таковых, сколько в интерактивном картографическом анализе и комплексной обработке ГИ для решения своих конкретных задач. Это привело к интеграции сетевых средств хранения и поиска ГИ с сетевыми ГИС, которые поддерживают картографическое представление ГИ в виде электронных карт, интерактивную ГИС-публикацию, комплексную обработку ГИ и её анализ (*Kraak and Brown, 2001*).

Основные отличия сетевых аналитических ГИС от настольных ГИС обусловлены, главным образом, изменением типа пользователя *от эксперта в области ГИС технологий к эксперту в предметной области*. Это обстоятельство накладывает дополнительные требования к основным операциям сетевых аналитических ГИС. С одной стороны, система должна включать достаточный набор функций, чтобы обеспечить решение задач эксперта предметной области. С другой стороны, эти функции должны реализовываться с помощью удобных и достаточно понятных для эксперта операций. Так, операции визуального исследования должны обеспечивать неограниченную интерактивность при картографическом анализе данных. Аналитические преобразования должны включать в себя как расширенный класс заранее подготовленных операторов, ориентированных на предметную область, так и интерпретатор, позволяющий пользователю самому конструировать необходимые для исследования ГИ операторы преобразований. Правдоподобный вывод в сетевых аналитических ГИС должен быть основан на интуитивно понятных методах интеллектуальной обработки пространственно-временных данных.

Сетевые ГИС имеют два типа пользователей:

(А) *Поставщики ГИ*, которыми могут быть правительственные и общественные организации, ведомства, учебные и научные учреждения, электронные библиотеки и т. д.

(Б) *Потребители ГИ*, к которым относятся сотрудники научных учреждений, федеральных и региональных администраций, комитетов по ЧС и предприятий, учащиеся учебных заведений и непрофессиональные пользователи сети Интернет (ГИС-зрители).

Для поставщиков ГИ сетевые аналитические геоинформационные технологии обеспечивают следующие возможности:

- наглядная публикация ГИ с возможностью анализа;
- предоставление ГИ пользователям для детального анализа, но без потери права собственности.

Для потребителей ГИ поддерживаются следующие возможности:

- доступ картографический просмотр и анализ ГИ отовсюду, для всех и в любое время;
- доступ и анализ данных геологического, геофизического, экологического; социологического и геополитического мониторинга и их анализ;
- активное изучение электронных коллекций ГИ.

В данной главе мы опишем основы сетевых геоинформационных технологий. В разделе 11.2 даётся спецификация требований основных групп потребителей сетевой ГИ, структура данных и основные операции сетевых аналитических ГИС, в разделе 11.3 кратко рассмотрены два основных типа архитектурных решений сетевых ГИС. Разделы 11.4 и 11.5 посвящены проблеме комплексного анализа пространственных данных в сетевых аналитических ГИС. В разделе 11.6 рассматриваются функции, реализованные в сетевых аналитических ГИС ГеоПроцессор и КОМПАС.

## 11.2. Потребители сетевых аналитических ГИС

Потребителей сетевых аналитических ГИС можно разделить на следующие группы:

- Непрофессиональные пользователи сети Интернет (ГИС-зрители).
- Администраторы всех уровней.
- Учащиеся школ, колледжей и университетов.
- Эксперты и исследователи.

Наиболее многочисленной является группа непрофессиональных потребителей ГИ. Запросы этой группы ГИС-зрителей относятся в основном к получению справочной, мало изменяющейся во времени информации из достаточно больших баз данных либо к просмотру постоянно обновляемой динамической информации в реальном времени. Этой группе в основном необходимо обеспечить дружественный интерфейс к средствам поиска релевантной ГИ и наглядное представление

ГИ. Поддержка запросов этой массовой группы пользователей очень важна для построения информационного общества.

Запросы остальных групп потребителей ГИ связаны с обучением, поддержкой принятия решений, производством и исследовательской работой. Обеспечение этих запросов требует существенно более детального анализа ГИ. Поэтому эти группы должны иметь с одной стороны возможность настройки аналитических ресурсов системы под свои задачи вплоть до подключения своих программных средств, а с другой — возможность подключения своих личных данных для их совместной обработки с данными сервера (в том числе и с данными, предназначенными только для абонированных пользователей).

Рассмотрим более подробно запросы основных групп пользователей и системные средства для их поддержки (*Гитис, Вайншток, 2001*).

Задачи группы *непрофессиональных пользователей сети Интернет* состоят в просмотре справочной ГИ, в получении ГИ в реальном времени и её экспресс-анализе (рис. 11.1). Для поддержки этих запросов достаточно обеспечить доступ к средствам картографии, когнитивной графики и мультимедиа.

Задачи группы *администраторов* добавляют к этому списку запросы по детальному анализу ГИ, по подготовке аналитических отчётов и по поддержке принятия решений. Для этого требуется подключение к ГИС средств визуального исследования ГИ с помощью картографии и когнитивной графики, а также средств вычисления новых свойств ГИ с помощью аналитических преобразований.

Задачи группы *учащихся* университетов, колледжей и школ могут относиться к изучению пространственно-временных процессов и явлений, происходящих в природе и обществе. Вместе с тем, задачи этой группы включают в себя и рассмотренные выше запросы. Необходимые дополнительные средства должны поддерживать комплексный анализ свойств ГИ и классификацию географических объектов, а также возможность подключения данных пользователя.

Задачи группы экспертов в различных предметных областях являются наиболее сложными. В дополнение к остальным группам пользователей задачи экспертов могут включать исследование закономерностей в географических данных, подготовку новых данных, а также разработку предметно-ориентированных технологий анализа ГИ. Эти запросы поддерживаются средствами оценивания связей между географическими понятиями, средствами изменения структуры ГИ, моделирования пространственно-временных процессов и средствами подключения программных модулей пользователя.

### 11.3. Схемы сетевых ГИС

Рассмотрим две архитектуры сетевых ГИС, которые не требуют установки у пользователя дополнительных программ: серверные системы и системы клиент–сервер.

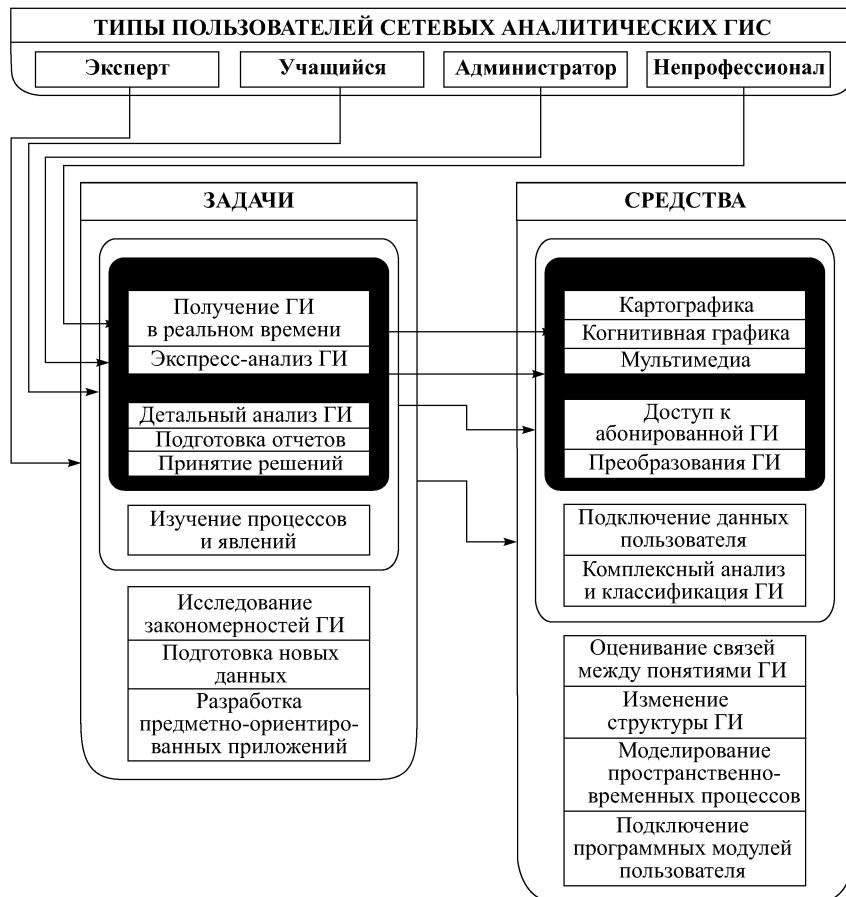


Рис. 11.1. Запросы и средства поддержки основных групп пользователей сетевых аналитических ГИС

Структура серверной схемы приведена на рис. 11.2. По запросу пользователя ГИ выбирается из удалённой БД и обрабатывается на ГИС-сервере. Затем картографический результат преобразуется в растровое изображение и пересылается пользователю. Просмотр данных осуществляется с помощью стандартных Web-браузеров. Поэтому у пользователя не требуется установка дополнительных программ.

Эта технология, по-видимому, обеспечивает наиболее простой подход к публикации ГИ в сети Интернет: на сервере можно использовать настольные ГИС. Поскольку отсутствует прямой доступ к базе данных, то данные достаточно просто защитить от несанкционированного доступа. Вместе с тем имеется ряд принципиальных ограничений: значительные времена ожидания на каждой транзакции при передаче

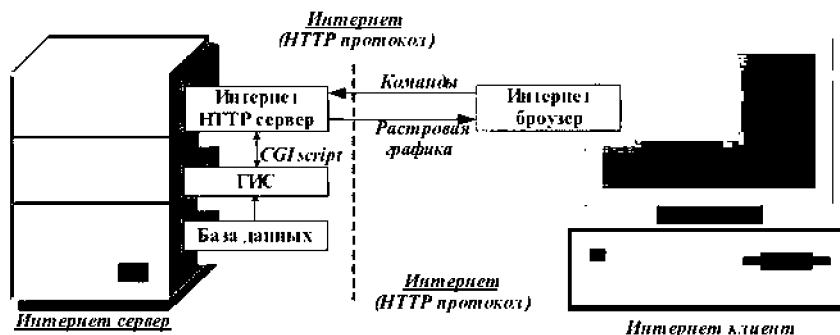


Рис. 11.2. Взаимодействие клиента с геосервером

большого растрового изображения от сервера к клиенту, низкая интерактивность при обслуживании динамических запросов и при выполнении операций, ограниченный интерфейс пользователя, ограниченное качество графики, обусловленное передачей данных в наиболее типовых форматах GIF и JPG.

Схема клиент-сервер (рис. 11.3) позволяет обеспечить неограниченную интерактивность картографического исследования, пространственно-временной обработки и анализа ГИ. Часто клиентская часть сетевой ГИС реализуется в виде программы на языке Java и включается в Web-страницу. При загрузке такой страницы система автоматически загружается на компьютер пользователя и начинает работу. При этом ГИС может использовать данные, расположенные на различных серверах. Такие программы называются апплетами. Апплеты могут работать на всех платформах, на которых установлены стандартные средства сети Интернет. Быстродействие выполнения всех ГИС-функций определяется только параметрами компьютера пользователя.

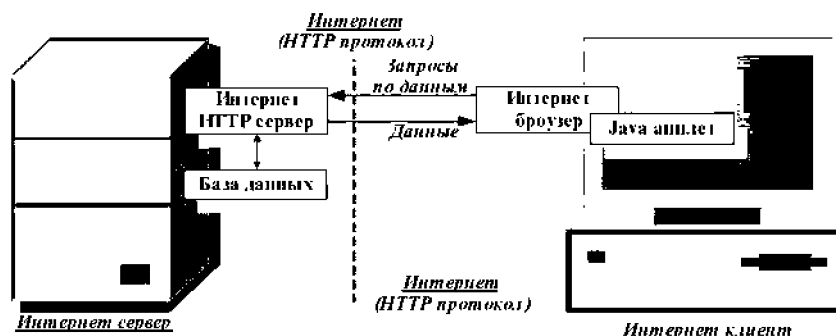


Рис. 11.3. Взаимодействие по схеме клиент-сервер



Данная архитектура наиболее приемлема для создания аналитических сетевых ГИС. Её существенными достоинствами являются высокая интерактивность анализа данных, динамическое управление средствами картографической и графической визуализации, высококачественная графика. Недостатки по сравнению с серверной схемой состоят в необходимости однократной затраты времени на загрузку Java апплета. Кроме того, поскольку данные передаются на компьютер клиента, то нельзя гарантировать их защиту от несанкционированного доступа.

Следует отметить также, что Java апплеты могут быть установлены на машине пользователя и работать в локальном режиме, обеспечивая доступ как к удалённым информационным ресурсам, так и к данным пользователя. Локальный режим работы позволяет разрабатывать новые ГИС приложения.

Примерами аналитических ГИС в такой архитектуре являются CommonGIS (Descartes) (Andrienko and Andrienko, 1999, Andrienko et al., 1999, Andrienko and Andrienko 2002), Lava Magma (Berg et al., 1999), GeoПроцессор и КОМПАС (Gitis et al., 1998, Gitis et al., 2000).

## 11.4. Вывод и объяснение при сетевом анализе ГИ

**11.4.1. Правила аналогии по прецеденту и логические выражения.** Нахождение эмпирических закономерностей по комплексу данных является одним из наиболее сложных этапов анализа ГИ в задачах прогноза. Желательно, чтобы в сетевых аналитических ГИС эта функция реализовалась с помощью методов, принципы которых интуитивно были бы понятны для специалистов предметных областей, а результаты применения этих методов позволяли бы специалистам интерпретировать полученное правило прогноза.

Задачи вывода закономерностей при прогнозировании с неполной информацией обладают следующими общими свойствами:

- наличие причинно-следственной связи между значением прогноза и значениями переменных, отображающих свойства прогнозирующей величины;
- отсутствие готовой формулы или математической модели для нахождения прогноза;
- неполнота и зашумленность данных: свойства измерены не все, часть свойств нельзя измерить инструментально, и они заменяются экспертными решениями.

Одним из примеров таких задач является рассмотренный в гл. 5 прогноз максимальных энергий ожидаемых землетрясений ( $M_{\max}$ ). В этой задаче требуется построить и обосновать карту максимальных магнитуд ( $M_{\max}$ ) ожидаемых землетрясений. В основе решения задачи лежат две гипотезы: 1) гипотеза о повторяемости землетрясений и 2) гипотеза о наличии зависимости  $M_{\max}$  землетрясения от свойств геологической среды. Анализ показывает, что степень неполно-

ты информации при построении прогнозных карт  $M_{\max}$  землетрясений слишком велика для применения чисто математических методов или чисто экспертных методов. Подход, предложенный в гл. 5, направлен на полноценное использование обеих гипотез. На основании первой гипотезы создаётся выборка прецедентов, в которую входят наиболее изученные пункты региона. Затем, в соответствии со второй гипотезой по выборке прецедентов и экспертным знаниям, находится зависимость  $M_{\max}$  от свойств геологической среды. После этого зависимость экстраполируется на весь регион. Поскольку из-за отсутствия математической модели прогноза точность решения нельзя оценить статистически, то на последнем шаге ищется объяснение и обоснование зависимости и прогноза на языке специалистов предметной области.

В данной задаче, как и во многих других задачах прогнозирования с неполной информацией, эксперты используют два типа рассуждений:

1. Рассуждения по аналогии с прецедентом: объекты со сходными характеристиками одинаковы по некоторому свойству.
2. Рассуждения в виде логических выражений: объекты, характеристики которых подчиняются некоторому логическому выражению, одинаковы по некоторому свойству.

Рассуждения по аналогии с прецедентом могут применяться, если знаний для формирования логических выражений недостаточно. Полученные правила имеют следующий вид: если  $S(a, b) \geq C$ , то  $F(a) = F(b)$ , где  $a$  и  $b$  объекты,  $S(\cdot)$  — функция сходства,  $S(a, b) \in [0, 1]$ ,  $S(a, a) = 1$ ,  $C = \text{const}$ ,  $F(\cdot)$  — целевое свойство объекта. Правила обладают следующими свойствами:

- результат не вербализован (нуждается в вербализации);
- результат интуитивно понятен;
- результат имеет эмпирическое подтверждение;
- результат существенно зависит от выбранных характеристик, по которым определяется сходство, и от вида функции сходства.

Рассуждения в виде логических выражений имеют следующий вид: если  $\{L(a) = L(b)\} = \text{true}$ , то  $F(a) = F(b) = F$ , где  $L(\cdot)$  — логическая функция от свойств объекта,  $F$  — значение целевого свойства. Логические правила могут применяться, если прецеденты отсутствуют. Полученные правила имеют следующие свойства:

- результат вербализован;
- результат понятен специалистам предметной области;
- результат не имеет эмпирического подтверждения (нуждается в подтверждении).

Можно видеть, что свойства правил, полученных в результате индуктивного обучения по прецеденту, и правил, построенных в виде логических выражений, дополняют друг друга. Поэтому при решении задачи желательно использовать оба типа правил.

Для решения рассмотренных задач возможно применение следующих сценариев анализа данных.

1. Исследование на примерах:
  - нахождение правила прогноза по принципу прецедента;
  - экстраполяция правила на весь регион;
  - проверка правила на непротиворечивость со знанием предметной области.
2. Исследование на основе знаний предметной области:
  - построение логического правила на основе знания предметной области;
  - экстраполяция правила на весь регион;
  - проверка правила на совместимость со значениями прогноза и оценка значимости решения на выборке прецедентов.

**11.4.2. Элементы технологии.** Введём следующие обозначения:  $R^I$  — пространство признаков, нормированных по среднеквадратичным отклонениям, т. е. в  $R^I$  среднеквадратичные отклонения  $\sigma_i \equiv 1$  для всех  $i = 1, \dots, I$ ;  $\rho(\mathbf{x}, \mathbf{x}^{(n)})$  — метрика для  $\mathbf{x} \in R^I$ .

*Функции сходства с прецедентом.* Рассмотрим элементарную функцию сходства  $S(\mathbf{x}, \mathbf{x}^{(n)})$  точки  $\mathbf{x}$  с точкой  $\mathbf{x}^{(n)}$ . Она является неотрицательной для любого  $\mathbf{x} \in R^I$ , равной 1 в точке  $\mathbf{x}^{(n)}$  и убывающей при удалении от точки  $\mathbf{x}^{(n)}$ . Функцию сходства можно трактовать как степень уверенности решения: чем больше сходство признаков данного географического объекта или точки растра с прецедентом  $\mathbf{x}^{(n)}$ , тем больше уверенности в том, что данный элемент обладает прогнозируемым свойством прецедента.

Функции сходства легко конструируются на основе измерения расстояний. Расстояние в метрике Минковского имеет вид:

$$\rho(\mathbf{x}, \mathbf{x}^{(n)}) = \left[ \sum_{i=1}^I (w_i |x_i - x_i^{(n)}|)^p \right]^{\frac{1}{p}}, \quad w_i > 0, \quad p > 1. \quad (11.1)$$

В частности, полагая  $p = 1, 2$  и  $\infty$ , получим метрики  $l_1, l_2$  и  $c$  соответственно:

$$\rho_{l_1}(\mathbf{x}, \mathbf{x}^{(n)}) = \sum_{i=1}^I w_i |x_i - x_i^{(n)}|, \quad (11.2)$$

$$\rho_{l_2}(\mathbf{x}, \mathbf{x}^{(n)}) = \left[ \sum_{i=1}^I (w_i |x_i - x_i^{(n)}|)^2 \right]^{\frac{1}{2}}, \quad (11.3)$$

$$\rho_c(\mathbf{x}, \mathbf{x}^{(n)}) = \max_{\langle i \rangle} \{w_i |x_i - x_i^{(n)}|\}. \quad (11.4)$$

Элементарную функцию сходства можно задать, например, выражением

$$S(\mathbf{x}, \mathbf{x}^{(n)}) = \exp \left\{ -\rho(\mathbf{x}, \mathbf{x}^{(n)}) \right\}. \quad (11.5)$$

Выражение (11.5) можно обобщить для построенного вокруг точки  $\mathbf{x}^{(n)}$  интервала  $\mathbf{g}^{(n)} = (\mathbf{x}^{(n)} - \Delta, \mathbf{x}^{(n)} + \Delta)$ :

$$S(\mathbf{x}, \mathbf{g}^{(n)}) = \exp \left\{ -\rho(\mathbf{x}, \mathbf{g}^{(n)}) \right\}, \quad (11.6)$$

где:

$$\rho(\mathbf{x}, \mathbf{g}^{(n)}) = \left\{ \sum_{i=1}^I \left( \frac{w_i}{2} \left[ |x_i - (x_i^{(n)} - \Delta_i)| + |x_i - (x_i^{(n)} + \Delta_i)| - 2\Delta \right] \right)^p \right\}^{\frac{1}{p}},$$

$$w_i > 0, \quad p > 1, \quad \Delta_i \geq 0.$$

Функция сходства к множеству точек  $\{\mathbf{x}^{(n)}\}$  равна элементарной функции сходства к ближайшей точке этого множества:

$$S(\mathbf{x}, \{\mathbf{x}^{(n)}\}) = \exp \left( -\min_{\langle n \rangle} \left\{ \rho(\mathbf{x}, \mathbf{x}^{(n)}) \right\} \right) = \max_{\langle n \rangle} \left\{ \exp \left( -\rho(\mathbf{x}, \mathbf{x}^{(n)}) \right) \right\}. \quad (11.7)$$

Рассмотрим функцию сходства для метрики  $s$ . Для этого перепишем (11.6) с учётом (11.4)

$$S_c(\mathbf{x}, \mathbf{x}^{(n)}) = \exp \left( -\rho_c(\mathbf{x}, \mathbf{x}^{(n)}) \right) = \exp \left( -\max_{\langle i \rangle} \left\{ w_i |x_i - x_i^{(n)}| \right\} \right) =$$

$$= \min_{\langle i \rangle} \left\{ \exp w_i \left( -|x_i - x_i^{(n)}| \right) \right\}. \quad (11.8)$$

Можно видеть, что функция сходства с прецедентом для метрики  $s$  (11.8) представляет конъюнкцию в нечёткой логике. Тогда для метрики  $s$  функция сходства к множеству точек  $\{\mathbf{x}^{(n)}\}$  (11.7) представляет дизъюнкцию конъюнкций. Это обстоятельство легко позволяет строить текстовые объяснения в виде логических конструкций для решающих правил, выведенных на основании сходства с прецедентами в метрике  $s$ .

*Монотонные решающие функции.* Рассмотрим теперь класс решающих правил, которые построены на монотонных решающих функциях (Гитис В.Г. 1974, Юрков Е.Ф. 1974). Эти правила используются в задачах, в которых знания предметной области дают основание предполагать, что для некоторого интервала значений зависимость прогнозируемого свойства (или степени уверенности в принадлежности рассматриваемых объектов некоторому классу) от признаков монотонно изменяется при изменении значения признака. Без ограничения общности будем считать, что прогнозируемое свойство  $S(\mathbf{x})$  возрастает или не изменяется (но не уменьшается) при увеличении значения любой из характеристик объекта, т. е.

$$\frac{\partial S(\mathbf{x})}{\partial x_i} \geq 0, \quad i = 1, \dots, I.$$

Например, сейсмоопасные зоны часто характеризуются неоднородностью геологической среды, близостью к зонам активным геологическим

разломов, аномальными значениями скоростей современных вертикальных движений, высокими значениями градиентов гравитационных аномалий и т. д. (Schenk et al., 1994). Можно предполагать, что уверенность в том, что точка карты принадлежит к сейсмоопасной зоне, тем больше, чем, при прочих равных условиях, больше значение каждого из перечисленных признаков.

Этим же свойством монотонности обладает эмпирическая функция распределения

$$F(\mathbf{x}) = \nu_{\mathbf{x}}/N, \quad (11.9)$$

где  $\nu_{\mathbf{x}}$  обозначает число прецедентов, для которых  $x_i^{(n)} < x_i$  по всем  $i = 1, \dots, I, n = 1, \dots, N$ .

Во многих задачах классификации выборки прецедентов слишком малы для корректного оценивания многомерного вероятностного распределения. Обычно для классификации по выборке прецедентов оценивается граница множества  $A = \{\mathbf{x} : F(\mathbf{x}) > 0\}$ . В этой задаче желательно также построить логическое объяснение правила классификации.

Рассмотрим алгоритм решения данной задачи.

1. Выбрать эталонные точки класса (выборка или прецеденты)

1.1. Варианты экспертного решения:

- картографический ввод точек учебной выборки, основанный на выборе элементов географического слоя прецедентов;
- картографический ввод обучающего полигона, из которого затем точки выборки выбираются автоматически;
- картографический выбор обучающего географического объекта, из буферной зоны которого затем автоматически выбираются точки выборки.

1.2. Варианты автоматического решения:

- выбрать все точки слоя по некоторому условию (например, эпицентры с магнитудой больше некоторого заданного порога);
- выбрать в качестве прецедентов все узлы сетки некоторого полигона;
- выбрать в качестве прецедентов все точки растрового слоя по некоторому условию (например, со значениями больше некоторого порога).

2. Построить функцию (11.9).

3. Построить таблицу значений признаков  $x_i$  в точках выборки  $\{\mathbf{x}^{(n)}\}, n = 1, 2, \dots, N$ .

4. Задать коэффициент  $\mu$ , определяющий точность аппроксимации области  $A = \{\mathbf{x} : F(\mathbf{x}) > 0\}$ .

5. Построить матрицу принадлежности  $P = (P_{mn})$ . Строки матрицы обозначают покрытия, каждое из которых связано с точкой выборки  $m$ , а столбцы — точки выборки. Элемент матрицы  $P_{mn} = 1$ , если точка  $n$  принадлежит покрытию точки  $m$ , то есть, если

$x_i^{(m)} - \Delta x_i \leq x_i^{(n)}$  для всех  $i = 1, 2, \dots, I$ , где  $\Delta x_i = \mu \sigma_i$ , где  $\sigma_i$  — среднее квадратичное отклонение (СКО) по  $i$ -му признаку (легко видеть: если  $\mu = 0$ , то область  $A$  аппроксимируется без ошибки, если  $\mu = 1$ , то область  $A$  аппроксимируется с точностью  $\sigma_i$ ). Чем больше значение  $\mu$ , тем меньше точность аппроксимации области  $A$  покрытием  $B$  и тем меньше число конъюнкций необходимо для конструирования  $B$ .

6. Найти по матрице принадлежности минимальное покрытие точек выборки: найти подмножество из минимального числа покрытий (строк матрицы), для которых в каждом столбце матрицы оставалась хотя бы одна единица (условие, при котором каждая точка выборки имеет хотя бы одно покрытие). Для этого могут быть применены стандартные методы минимизации дизъюнктивных нормальных форм булевых функций (Закревский, 1966). Эти алгоритмы являются итеративными и состоят из трёх типовых операций.

А) поглощение столбцов матрицы.

Рассмотрим 2 столбца. Пусть второй из них имеет единицы везде, где имеет первый, а там, где у первого нули, второй столбец может иметь и единицы и нули (сравнимые столбцы). Тогда первый столбец поглощает второй. Второй столбец можно вычеркнуть. Таким образом, столбец  $k$  исключается столбцом  $n$ , если  $P_{mk} \geq P_{mn}$  для всех строк матрицы.

Пример:

	$a$	$b$	$c$	$d$
$A$	1	0	0	1
$B$	1	1	0	1
$C$	0	1	1	0

Столбец  $c$  исключает столбец  $b$ .

Б) поглощение строк матрицы.

Рассмотрим 2 строки. Пусть вторая из них имеет нули везде, где имеет первая, а там, где у первой единицы, вторая строка может иметь и единицы и нули (сравнимые строки). Тогда первая строка поглощает вторую. Вторую строку можно вычеркнуть. Таким образом, строка  $k$  исключается строкой  $m$ , если  $P_{mn} \geq P_{kn}$  для всех столбцов матрицы.

Пример:

	$a$	$b$	$c$	$d$
$A$	1	0	0	1
$B$	1	1	0	1
$C$	0	1	1	0

Строка  $B$  исключает строку  $A$ .

- В) расщепление матрицы выполняется, если все строки и столбцы являются несравнимыми (пункты А и Б выполнять бесполезно) и нет ни одного столбца, где имеется одна единица

(строка с таким столбцом выбирается безальтернативно). Расщепление матрицы является эвристической и неоптимальной операцией. При расщеплении в минимальное покрытие выбирается строка с наибольшим числом единиц. После этого все столбцы, для которых в выбранной строке находятся единицы, вычёркиваются. Для оставшейся матрицы применяем пп. А) и Б).

Пример:

	$a$	$b$	$c$	$d$
$A$	1	0	1	1
$B$	1	1	0	1
$C$	0	1	1	0

Мы выбираем строку  $A$ . После этого столбцы  $a$ ,  $c$  и  $d$ , покрываемые строкой  $A$ , исключаются.

7. Для каждой из отобранных в п. 6 строк  $m$  построить функцию принадлежности с помощью конъюнкции

$$\Phi^{(m)}(\mathbf{x}) = \min_{\langle i \rangle} \{\varphi_i^{(m)}(x_i)\}, \quad (11.10)$$

где:

$$\varphi_i^{(m)}(x_i) = \begin{cases} 0, & x_i \leq (x_i^{(m)} - \Delta x_i), \\ \frac{0,75(x_i - (x_i^{(m)} - \Delta x_i))}{\Delta x_i}, & (x_i^{(m)} - \Delta x_i) < x_i \leq x_i^{(m)}, \\ 0,75 + \frac{0,25(x_i - x_i^{(m)})}{\Delta x_i}, & x_i^{(m)} < x_i \leq (x_i^{(m)} + \Delta x_i), \\ 1, & (x_i^{(m)} + \Delta x_i) < x_i. \end{cases}$$

8. Построить область  $B = \{x : \psi(\mathbf{x}) > 0\}$ , которая определяется как дизъюнкция всех строк (покрытий), отобранных в п. 4,  $\psi(\mathbf{x}) = \max_{\langle m \rangle} \{\Phi^{(m)}(\mathbf{x})\}$  и которая аппроксимирует область  $A$  с точностью  $\Delta x_i$ , при этом  $A \subset B$ .

Область  $B$  аппроксимирует область  $A$  с точностью  $\Delta x_i$ , при этом  $A \subset B$ . Можно видеть, что функция  $\varphi_i^{(m)}(x_i) = 0$ , если  $x_i \leq (x_i^{(m)} - \Delta x_i)$ , затем она линейно возрастает до 0,75 в точке  $x_i = x_i^{(m)}$ , затем линейно возрастает до 1 в точке  $x_i = (x_i^{(m)} + \Delta x_i)$  и после этого остаётся равной 1. Выражение  $\Phi^{(m)}(\mathbf{x})$  является конъюнкцией в нечёткой логике, а  $\psi(\mathbf{x})$  является дизъюнкцией конъюнкций (Kaufman, 1975). Значение функции  $\psi(\mathbf{x})$  можно интерпретировать как степень уверенности  $C$  в корректности нашего вывода. Таким образом, можно определить область  $B_C = \{\mathbf{x} : \Psi(\mathbf{x}) \geq C\}$ , для которой вывод принимается корректным с уверенностью  $C$ . В этом случае значение  $x_i^*$ ,

соответствующее уравнению  $\Psi(\mathbf{x}) = C$ , находится из  $\varphi_i^{(m)}(x_i^*) = C$ . В частном случае  $C = 0,75$  имеем  $x_i^* = x_i^{(m)}$ .

Для конструирования текстового объяснения логического выражения  $\Psi(\mathbf{x}) = C$  можно использовать заранее заготовленные лингвистические переменные и шаблоны. Простейшая текстовая конструкция может выглядеть следующим образом:

ЕСЛИ  $\langle$ имя атрибута $\rangle$  ПРИНИМАЕТ ЗНАЧЕНИЯ ОТ  $\langle$ числовое значение $\rangle$  ДО  $\langle$ числовое значение $\rangle$   
 И/ИЛИ  $\langle$ имя атрибута $\rangle$  ПРИНИМАЕТ ЗНАЧЕНИЯ ОТ  $\langle$ числовое значение $\rangle$  ДО  $\langle$ числовое значение $\rangle$   
 .....  
 ТО  $\langle$ имя выходного атрибута $\rangle$  БУДЕТ ПРИНИМАТЬ ЗНАЧЕНИЯ ОТ  $\langle$ числовое значение $\rangle$  ДО  $\langle$ числовое значение $\rangle$ .

### 11.5. Методы правдоподобного вывода ГИС ГеоПроцессор и КОМПАС

Методы правдоподобного вывода предназначены для нахождения операторов аналитических преобразований на основании имеющихся данных и знаний. Рассмотрим методы, реализованные в ГИС ГеоПроцессор, Декарт и КОМПАС (*Gitis и др., 1999*). В этих системах для вывода используются только непараметрические процедуры. Основанием для выбора непараметрических методов является их относительная простота использования и простота понимания их принципов для широкого круга возможных пользователей.

Введём следующие обозначения:  $R^I$  — пространство признаков, нормированных по среднеквадратичным отклонениям, для  $\mathbf{x} \in R^I$  среднеквадратичные отклонения  $\sigma_i \equiv 1$ ,  $\rho(\mathbf{x}, \mathbf{x}^{(n)})$  — метрика в  $R^I$ ,  $C$  — константа.

*Функция сходства.* Сходство с точкой  $\mathbf{x}^n$  определено, как

$$\begin{aligned} S(\mathbf{x}, \mathbf{x}^{(n)}) &= 1 - \frac{1}{C} \rho(\mathbf{x}, \mathbf{x}^{(n)}) \quad \text{для} \quad \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) \leq C, \\ S(\mathbf{x}, \mathbf{x}^{(n)}) &= 0 \quad \quad \quad \text{для} \quad \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) > C, \end{aligned} \quad (11.11)$$

а функция сходства с выборкой прецедентов  $\{\mathbf{x}^{(n)}\}$ ,  $n = 1, \dots, N$ , имеет вид:

$$\begin{aligned} S(\mathbf{x}, \{\mathbf{x}^{(n)}\}) &= 1 - \frac{1}{C} \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) \quad \text{для} \quad \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) \leq C, \\ S(\mathbf{x}, \{\mathbf{x}^{(n)}\}) &= 0 \quad \quad \quad \text{для} \quad \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) > C. \end{aligned} \quad (11.12)$$

Сходство по экспертному решению определено следующим образом. Пусть  $\mathbf{g}^{(n)} = (a_i^{(n)}, b_i^{(n)})$ ,  $i = 1, \dots, I$ , обозначает  $I$ -мерный интервал на



$R^I$ , который представляет экспертное высказывание типа конъюнкции, а  $\{\mathbf{g}^{(n)}\}$ ,  $n = 1, \dots, N$  обозначает множество интервалов, обозначающее дизъюнкцию. Функция сходства задаётся следующим образом

$$S(\mathbf{x}, \{\mathbf{g}^{(n)}\}) = 1 - \frac{1}{C} \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{g}^{(n)}) \quad \text{для} \quad \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{g}^{(n)}) \leq C, \quad (11.13)$$

где  $\rho(\mathbf{x}, \mathbf{g}^{(n)})$  — расстояние от точки  $\mathbf{x}$  в  $R^I$ .

*Функция принадлежности для двух классов.* Пусть  $\{\mathbf{x}^{(q,n)}\}$  и  $\{\mathbf{x}^{(\bar{q},m)}\}$ ,  $n = 1, \dots, N$ ,  $m = 1, \dots, M$ , обозначают два множества прецедентов в  $R^I$ . Функция принадлежности для двух классов имеет вид

$$f(\mathbf{x}, \{\mathbf{x}^{(q,n)}\}, \{\mathbf{x}^{(\bar{q},m)}\}) = -\frac{\rho(\mathbf{x}, \mathbf{x}^{(q,u)}) - \rho(\mathbf{x}, \mathbf{x}^{(\bar{q},v)})}{\rho(\mathbf{x}^{(q,u)}, \mathbf{x}^{(\bar{q},v)})} \quad \text{для} \quad \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) \leq C, \quad (11.14)$$

где

$$\mathbf{x}^{(q,u)} = \arg \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(q,n)}), \quad \mathbf{x}^{(\bar{q},v)} = \arg \min_{\langle m \rangle} \rho(\mathbf{x}, \mathbf{x}^{(\bar{q},m)}).$$

*Функция непараметрической регрессии для множества прецедентов.* Пусть  $\{\mathbf{x}^{(n)}, y^{(n)}\}$ ,  $n = 1, \dots, N$ , обозначает множество прецедентов, состоящее из векторов и оценок функции. Функция регрессии имеет вид

$$f(\mathbf{x}, \mathbf{x}^{(n)}, y^{(n)}) = \frac{\sum_{n=1}^N h(\mathbf{x} - \mathbf{x}^{(n)}) y^{(n)}}{\sum_{n=1}^N h(\mathbf{x} - \mathbf{x}^{(n)})}, \quad (11.15)$$

где  $h(\mathbf{x} - \mathbf{x}^{(n)}) = 1 - \frac{\rho(\mathbf{x}, \mathbf{x}^{(n)})}{C}$  для  $\rho(\mathbf{x}, \mathbf{x}^{(n)}) < C$  и  $h(\mathbf{x} - \mathbf{x}^{(n)}) = 0$  для  $\rho(\mathbf{x}, \mathbf{x}^{(n)}) \geq C$ .

*Эмпирическая функция распределения.* Пусть  $\{\mathbf{x}^{(n)}\} \subset R^I$ ,  $n = 1, \dots, N$ , обозначает множество прецедентов. Эмпирическая функция распределения имеет вид

$$F(\mathbf{x}) = \nu_{\mathbf{x}}/N, \quad (11.16)$$

где  $\nu_{\mathbf{x}}$  обозначает число прецедентов, для которых  $x_i^{(n)} < x_i$  по всем  $i = 1, \dots, I$ .

### 11.6. Сетевые ГИС ГеоПроцессор и КОМПАС

Разработанные в ИППИ РАН сетевые аналитические ГИС ГеоПроцессор и КОМПАС (*Gitis et al., 1998, Gitis et al., 2000*) являются одними из первых реализаций быстро развивающихся сетевых аналитических геоинформационных технологий.

Сетевая ГИС ГеоПроцессор предназначена для моделирования и комплексного анализа пространственно-временных свойств ГИ. Области применения — изучение геологической среды и поддержка принятия решений в сложных задачах, таких как мониторинг экологического состояния и оценка опасности природной среды, прогноз природно-техногенных катастроф, оценка и прогноз природных ресурсов.

Версия 2.01 системы ГеоПроцессор реализована как система клиент-сервер на языке Java и доступна по адресу <http://www.iitp.ru/projects/geo>. Система поддерживает следующие функции:

1. Интерактивное картографическое представление и анализ цифровых сеточных моделей, линейных и точечных данных.
  - Композиция и представление карт, состоящих из нескольких сеточных, векторных и точечных слоёв.
  - Интерактивное управление визуализацией слоёв карты.
  - Изменение размеров пространственной области и сдвиг карты с интерполяцией и без интерполяции сеточных значений.
  - Анализ растровой информации с помощью изменения закрашки.
  - Построение моделей освещённости сеточного слоя.
  - Чтение значений всех сеточных слоёв карты в произвольных точках.
  - Построение разрезов сеточных слоёв карты по произвольному профилю с возможностью измерения значений по разрезу.
  - Формирование учебных выборок в виде совокупностей единичных точек или полигонов.
  - Построение карт сходства с прецедентом по произвольному набору сеточных слоёв.
2. Преобразование данных.
  - Генерирование растров с помощью операций над линейными слоями.
  - Генерирование растров с помощью операций над точечными слоями.
  - Генерирование растров с помощью операций растровой фильтрации.

- Генерирование растров с помощью алгебраических и логических операций над несколькими растрами, а также с помощью набора элементарных функций.
3. Правдоподобный вывод прогнозных карт по комплексу сеточных слоёв (Приложение 11.1).
- Оценивание границы для монотонных зависимостей.
  - Оценивание функций сходства.
  - Оценивание функций принадлежности к классу.
  - Оценивание непараметрической регрессии.

На рис. 11.4 приведён пример представления системой ГеоПроцессор данных <sup>1)</sup> из высокосейсмичных районов Болгарии в окрестности города Кресна, где 4 марта 1904 года произошло землетрясение с магнитудой 7,8. В верхнем ряду слева представлена сеточная цифровая модель рельефа с нанесёнными слоями активных разломов, рек, эпицентров землетрясений с  $m > 5,0$  после 1900 года и эпицентров землетрясений без афтершоков с  $m > 1,8$  после 1987 года. Под ней показана вычисленная модель освещённости рельефа. В середине в верхнем ряду показана сеточная модель суммарной длины разломов, вычисленная со скользящим окном  $R = 5$  км. Под ней представлена сеточная модель плотности эпицентров, вычисленная со скользящим окном  $R = 30$  км. Вверху справа представлена сеточная цифровая модель модуля градиента рельефа. Под ней представлена версия сеймотектонического районирования, выполненная по следующим сеточным признакам: высота рельефа, модуль градиента рельефа, плотность разломов и плотность эпицентров. На карте коричневые тона соответствуют областям повышенного сейсмического потенциала, синие тона — областям пониженного сейсмического потенциала, треугольники и квадраты — точки учебной выборки сейсмичной и асейсмичной зон соответственно.

Сетевая аналитическая ГИС КОМПАС (COMPASS — Cartography Online Modeling, Presentation and Analysis System) предназначена для публикации региональной пространственно-временной экономической, социологической и демографической информации, анализа комплексных свойств объектов и поддержки принятия политических, экономических и административных решений. Система оперирует с географическими объектами, представленными на карте в виде слоёв полигонов, линий и точек. Система реализована на языке Java в технологии клиент-сервер и доступна по адресу <http://www.iitp.ru/projects/geo>.

Информация системы структурирована по информационным слоям. Система может выполнять операции со всем слоем, с группой объ-

<sup>1)</sup> Данные подготовлены участниками проекта ASPELEA, грант Inco-Sorernicus, контракт № ERBIC 15CT 970200.

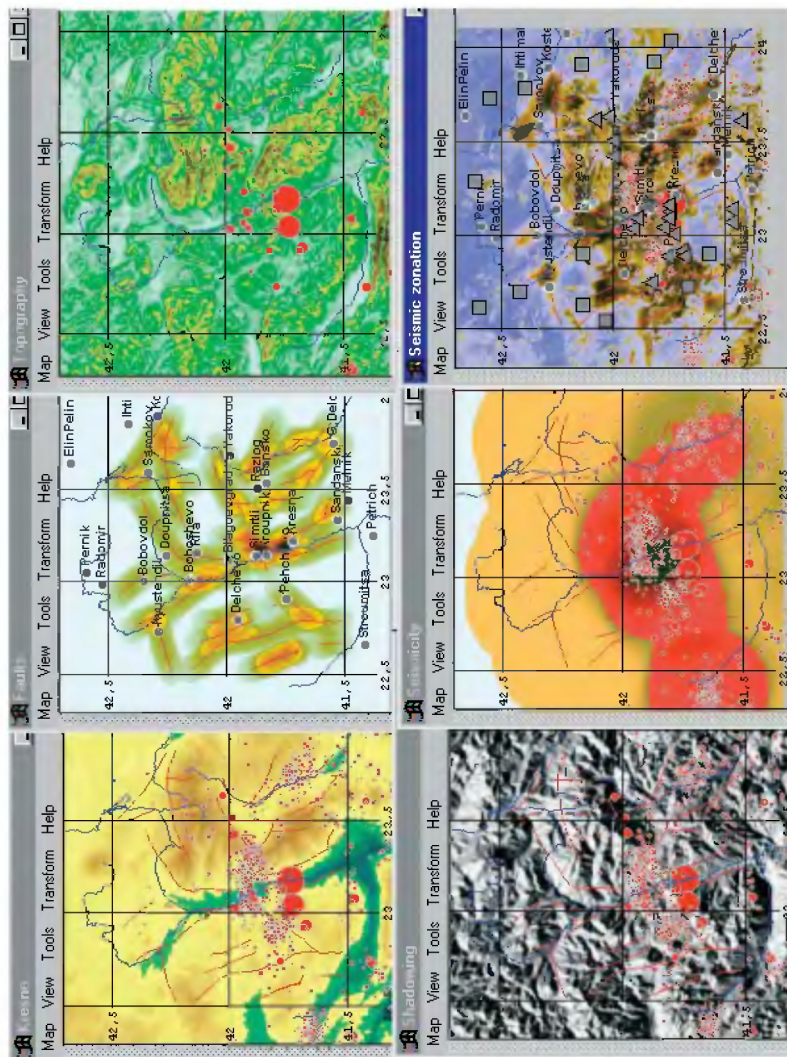


Рис. 11.4. Пример операций системы GeoПроцессор для района Креснинского землетрясения, 4.04.1904,  $m = 7,8$

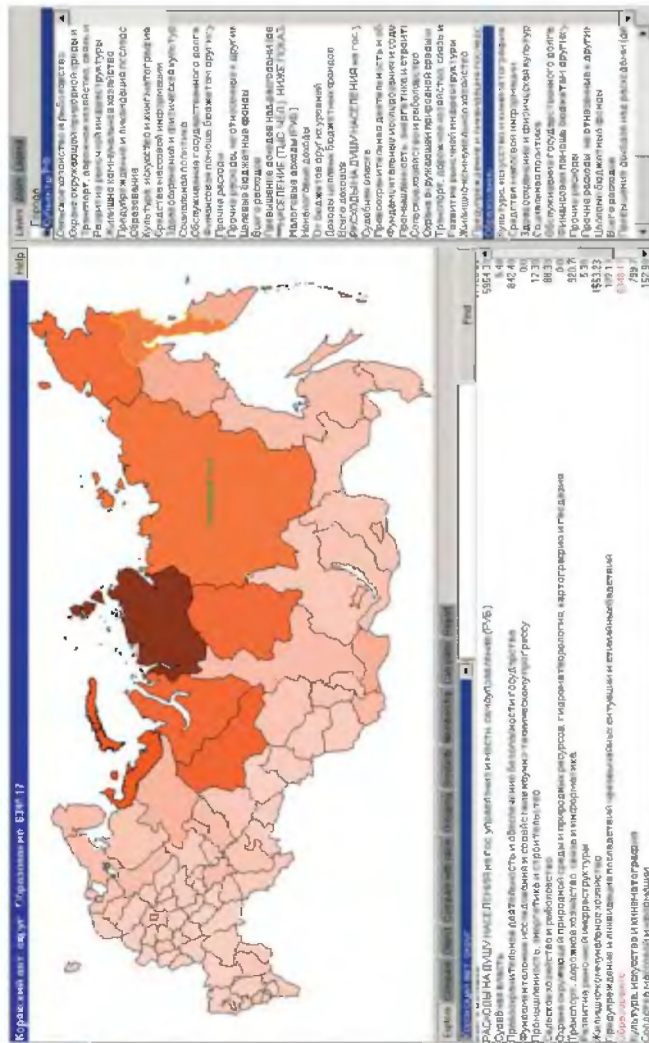


Рис. 11.5. Ответ на вопросы: Что, Где и Сколько? Картограмма расходов в руб. на душу населения по статье «Образование» отчёта по исполнению консолидированного бюджета субъектами РФ на 1.01.2001. Более интенсивный красный цвет закраски Северных субъектов РФ показывает, что для них эти расходы существенно увеличены. В окне карты подсветкой выделен Корякский автономный округ, в верхней строке окна видно, что расходы составили 6348,17 руб. на душу населения. В окне режимных вкладок показана часть списка значений индикаторов для Корякского автономного округа, анализируемый индикатор «Образование» выделен красным

ектов слоя и с отдельным объектом слоя. При этом анализ может выполняться над одним или несколькими атрибутами слоя. Система поддерживает следующие функции.

1. Визуальное исследование данных.

- Интерактивное управление визуализацией слоёв карты.
- Интерактивное масштабирование размеров и сдвиг карты.
- Интерактивное построение картограммы тематического свойства.
- Чтение значения тематического свойства при установке курсора на объекте слоя.
- Интерактивная группировка значений индикатора в произвольные интервалы, построение картограмм интервальных значений и одновременная подсветка групп объектов, относящихся к одному интервалу, выбираемому курсором.
- Интерактивное построение картограмм отклонений значений индикатора от порога, задаваемого пользователем с помощью ввода численного значения, установки значения с помощью движущегося указателя или с помощью выбора объекта на карте.
- Интерактивное построение гистограммы значений индикатора.
- Интерактивное построение круговых и столбчатых диаграмм для одного индикатора по нескольким географическим объектам.
- Интерактивное построение круговых и столбчатых диаграмм для нескольких индикаторов по одному объекту.
- Представление цифровой таблицы всех индикаторов выбранного объекта.
- Подключение к Web-сайту выбранного объекта (при работе в сети Интернет).

2. Аналитические преобразования данных.

- Генерирование нового тематического свойства с помощью произвольной функции от нескольких исходных свойств, интерактивно конструируемой пользователем из алгебраических и логических операций и элементарных функций.
- Генерирование вектора тематических свойств с помощью интерактивно конструируемого пользователем произвольного оператора от нескольких векторов исходных свойств.

3. Правдоподобный вывод многомерных эмпирических закономерностей.

- Нахождение функции сходства с выборкой примеров по комплексу свойств.
- Нахождение функции принадлежности к двум классам по комплексу свойств.

На рис. 11.5–11.9 приведены примеры применения системы КОМПАС.

Рассмотрим более подробно один из примеров применения ГИС КОМПАС для анализа структуры использования сельскохозяйственных земель России.

В примере использована база данных по природопользованию для субъектов РФ (Думнов, Рыбальский и др., 2001), в частности данные за 1998 г. по площадям сельских земель, находящимся в частной собственности: площади фермерских ( $\xi_1$ ), приусадебных ( $\xi_2$ ) и дачных ( $\xi_3$ ) земель. Задача состояла в том, чтобы произвести региональное районирование России по структуре сельских земель, находящихся в частной собственности.

Анализ структуры земель производился в два этапа:

1. Вычисление долей каждого типа земель  $x_i$  с помощью преобразования **Векторные слои**  $\Rightarrow$  **Атрибуты векторного слоя**:

$$x_i = \frac{\xi_i}{\xi_1 + \xi_2 + \xi_3}, \quad i = 1, 2, 3.$$

2. Выбор эталонных объектов классов структуры земель, оценка функции принадлежности к классам и построение картограммы. Для классификации использован вывод по методу ближайшего соседа. Реализованный в ГИС КОМПАС алгоритм сводится к следующему. Пусть  $R^I$  — пространство признаков, нормированных по среднеквадратичным отклонениям, тогда для точек  $\mathbf{x} \in R^I$  среднеквадратичные отклонения  $\sigma_i \equiv 1$ , пусть  $\rho(\mathbf{x}, \mathbf{x}^{(n)})$  — метрика в  $R^I$ ,  $C$  — константа. Классы заданы выборками прецедентов  $\{\mathbf{x}^{(q,n)}\}$  и  $\{\mathbf{x}^{(\bar{q},m)}\}$ ,  $n = 1, \dots, N$ ,  $m = 1, \dots, M$ . Функция принадлежности для двух классов имеет вид

$$f\left(\mathbf{x}, \{\mathbf{x}^{(q,n)}\}, \{\mathbf{x}^{(\bar{q},m)}\}\right) = -\frac{\rho(\mathbf{x}, \mathbf{x}^{(q,u)}) - \rho(\mathbf{x}, \mathbf{x}^{(\bar{q},v)})}{\rho(\mathbf{x}^{(q,u)}, \mathbf{x}^{(\bar{q},v)})}$$

для  $\min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(n)}) \leq C$ ,

где

$$\mathbf{x}^{(q,u)} = \arg \min_{\langle n \rangle} \rho(\mathbf{x}, \mathbf{x}^{(q,n)}), \quad \mathbf{x}^{(\bar{q},v)} = \arg \min_{\langle m \rangle} \rho(\mathbf{x}, \mathbf{x}^{(\bar{q},m)}).$$

В качестве эталона Первого класса выбрана Московская область, а в качестве эталонов Второго класса — Краснодарский край и Оренбургская область. В верхней части рис. 11.9 показаны диаграммы значений признаков для выбранных эталонов. В нижней части показана картограмма функции принадлежности.

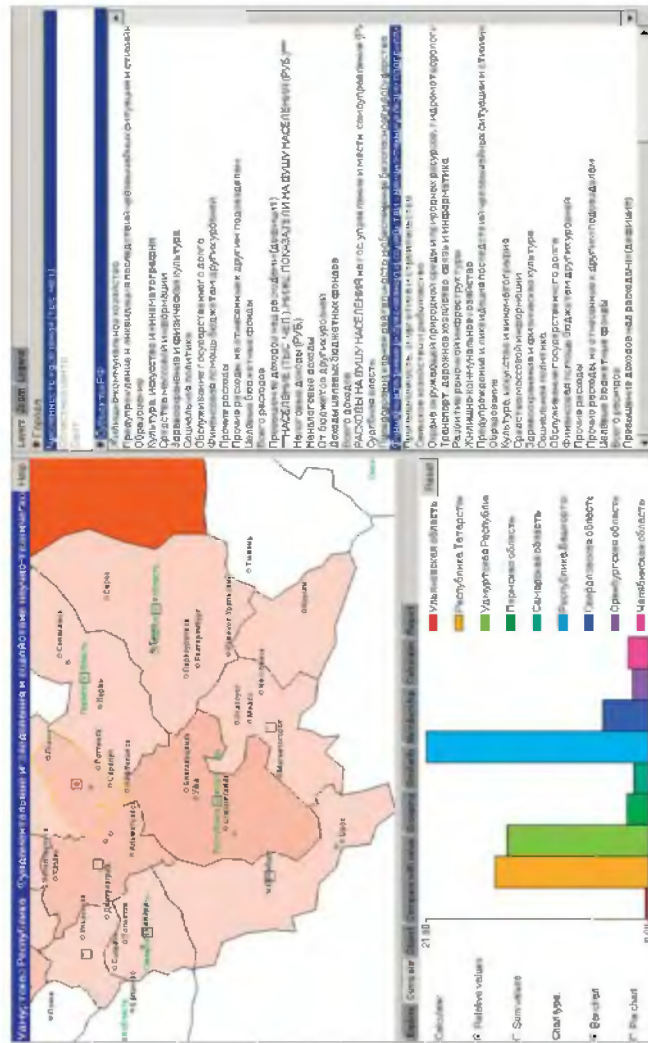


Рис. 11.6. Ответ на вопрос: Как распределён индикатор по нескольким объектам? Сравнение расходов в руб. на душу населения по статье «Фундаментальные исследования и содействие научно-техническому прогрессу» отчёта об исполнении консолидированного бюджета субъектами РФ на 1.01.2001. Анализируемые субъекты выделены в окне карты квадратами. В окне управления цветом выделен анализируемый индикатор. В окне режимных вкладок — столбчатая диаграмма. На диаграмме видно, что максимальные расходы составляют 21,8 руб. Цвета столбиков диаграммы указаны справа



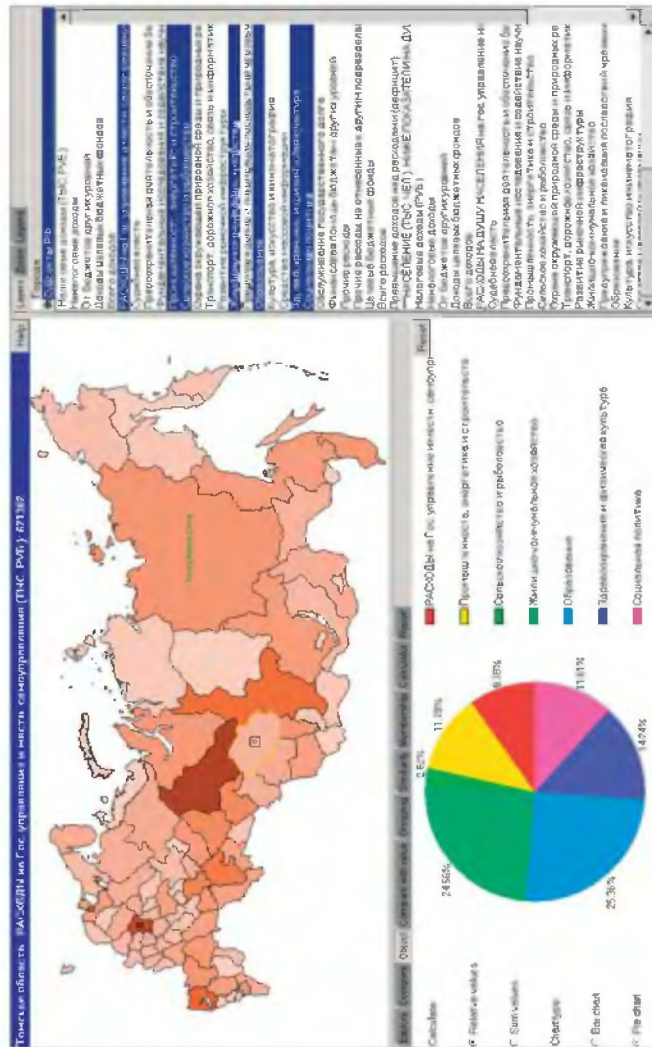


Рис. 11.7. Ответ на вопрос: Как распределены индикаторы для одного объекта? Сравнение нескольких статей расходов консолидированного бюджета субъектов РФ на 1.01.2001 для Томской области. В окне режимных вкладок показана круговая диаграмма, на которой указаны проценты каждой статьи расходов по отношению к сумме выбранных. Цвета секторов диаграммы указаны справа. В верхней строке окна карты можно видеть, что расходы по статье «Государственное и местное самоуправление» для Томской области составляют 621.362 тыс. руб., а на диаграмме видно, что это составляет 9,78 % от расходов по анализируемым статьям

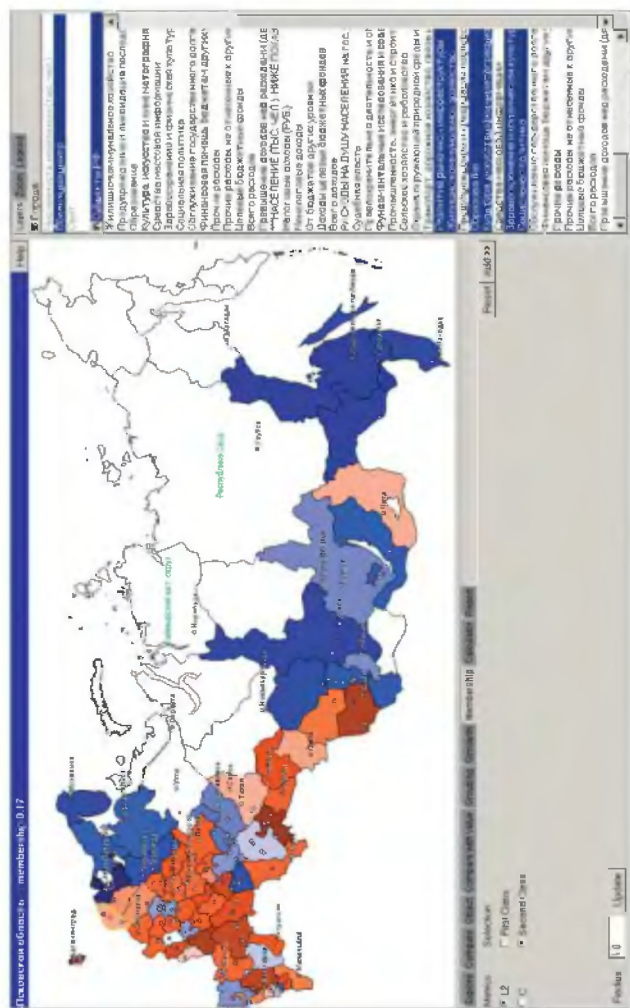


Рис. 11.8. Ответ на вопрос: На что больше похоже? Картограмма функции принадлежности субъектов РФ к Калининградской области и к Ленинградской области по следующим статьям расходов консолидированного бюджета субъектов РФ на душу населения: «Развитие рыночной инфраструктуры», «Жилищно-коммунальное хозяйство», «Образование», «Культура, искусство и кинематография», «Социальная политика». Красные тона выделяют субъекты более похожие по комплексу показателей на Калининградскую область, синие тона — на Ленинградскую область. Белым выделены субъекты, на которых функция принадлежности равна 0 (несравнимые и равноудалённые от эталонных объектов классов). В окне карты подсвечена Псковская обл., в верхней строке окна можно видеть, что функция принадлежности на ней 0,17

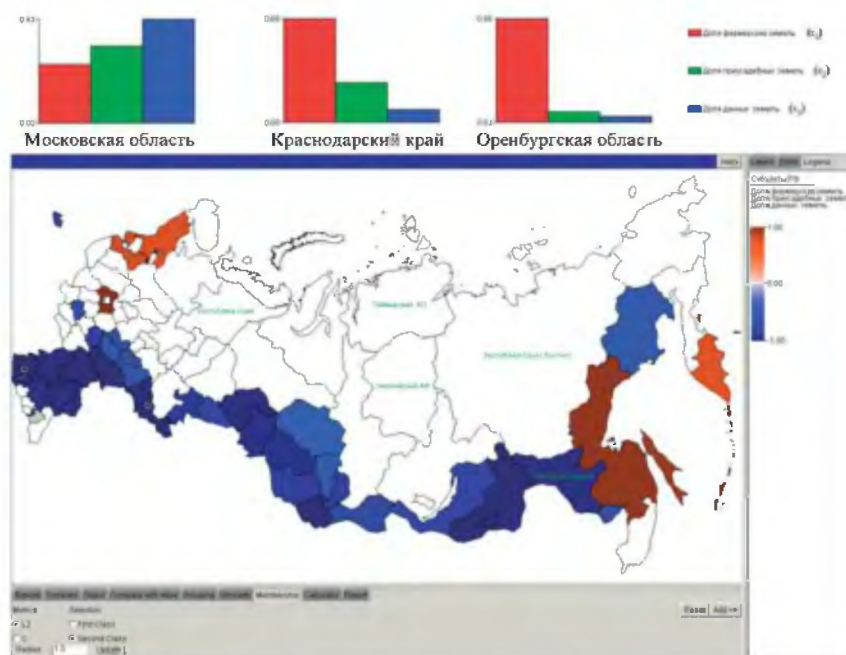


Рис. 11.9. Диаграммы значений признаков для выбранных эталонов и картограмма функции принадлежности. Объекты, принадлежащие Первому и Второму классам, выделены коричневыми и синими тонами соответственно. Белым цветом показаны непоставимые объекты, для которых расстояния до эталонов в нормированном пространстве признаков больше 1

На диаграмме видно, что структуры долей фермерских, приусадебных и дачных земель для эталонов разных классов существенно отличаются. В Московской области преобладают доли дачных и личных земель, а в эталонах 2-го класса доминирует доля фермерских земель. На картограмме видно, что, кроме Московской области, по структуре использования земель к 1-му классу относятся Ленинградская область, Республика Карелия, Хабаровский край, Сахалинская и Камчатская области. Структура использования земель 2-го класса характерна в основном для Южных регионов России.

## Глава 12

# КОМПЛЕКСНЫЙ АНАЛИЗ В СЕТЕВЫХ ГИС

### 12.1. Введение

В данном разделе излагаются методы, разработанные с целью применения сетевых аналитических систем CommonGIS (*Andrienko and Andrienko, 1999*) и ГеоПроцессор (*Gitis et al., 2000*), для поддержки принятия решений в задачах оценки последствий землетрясений и анализа зон возможного возникновения землетрясений (*Гумис и др., 2004*). Рассмотренные решения являются некоторым приближением к технологиям, разработанным для оценки сейсмического риска (*Антикаев и др., 1997*) и построения карт  $M_{\max}$  землетрясений (*Gitis et al., 1992*). Особенность предлагаемых в главе методов состоит в том, что они позволяют найти комплексные решения на основе данных и с помощью инструментальных средств, доступных любому пользователю в сети Интернет.

### 12.2. Оценка возможного ущерба от землетрясений

Метод применения сетевых аналитических ГИС для поддержки принятия решений по оценке прямого ущерба от возможных сильных землетрясений основан только на применении различного рода аналитических преобразований, рассмотренных в гл. 1. Иллюстративный пример выполнен с помощью системы CommonGIS для региона Западной Турции (*Andrienko et al., 2003*). В примере использованы исходные данные, взятые в сети Интернет. Этих данных явно недостаточно для постановки задачи по оценке прямого и косвенного сейсмического риска. Вместе с тем, пример указывает на возможность использования системы CommonGIS для поддержки принятия решений по планированию мероприятий по снижению сейсмического риска. Пример выполнен только для одного типа застройки (сооружения типа А) и для одного типа разрушений: сильные повреждения, не подлежащие восстановлению. Обобщение предложенного метода на все типы сооружений и разрушений является очевидным.

Решение строится на применении аналитических преобразований над сеточной цифровой моделью пиковых ускорений  $a$  на поверхности Земли, полученной в рамках проекта GSHAPE (*Giardini et al., 1999*), и векторными данными по инфраструктуре: главных городах и административных районах Западной Турции.

При решении по сеточной модели пиковых ускорений  $a$  была вычислена сеточная модель доли  $D$  строений типа А, которые должны получить сильные повреждения, не подлежащие восстановлению. Для этой цели применялись два преобразования типа **Сеточные слои**  $\Rightarrow$  **Сеточный слой** (раздел 1.3). Сначала выполнялась операция вычисления максимальной интенсивности землетрясений  $I$  из уравнения  $\log(a) = 0,3I + 0,014$  (Trifunac and Brady, 1975). Затем выполнялась операция линейной интерполяции зависимости  $D(I)$ , которая опубликована в работе (Coburn, 1995). Зависимость  $D(I)$  интерполировалась по следующим точкам:  $D(I < 5) = 0$ ,  $D(I = 6) = 0,05$ ,  $D(I = 7) = 0,5$ ,  $D(I = 8) = 0,75$ ,  $D(I > 9) = 0,9$ .

Средняя доля сильных разрушений строительных сооружений типа А для крупных городов и административных районов оценивалась в предположении о равномерном распределении застройки в радиусе 10 км вокруг центра города и в пределах административных районов. Для оценивания использовались преобразования **Сеточные слои** и **Векторный слой**  $\Rightarrow$  **Атрибуты векторного слоя** и интерактивные средства визуального анализа данных. Полученные результаты картографически представлены в ГИС Декарт (рис. 12.1). На рисунке можно видеть многослойную карту исследуемого региона и её легенду. Карта представляет цифровую модель рельефа, на которую нанесён полупрозрачный слой доли тяжёлых разрушений застройки типа А. Жирными линиями показаны границы административных районов. Высоты столбиков соответствуют доле разрушений в городах, а размеры кружков — доле разрушений в административных районах. Наибольшая доля разрушений среди городов равна 33% и относится к г. Маниса (доля разрушений для этого административного района равна 26%). Наибольшая доля разрушений среди административных районов равна 37% и относится к району Кутайа.

### 12.3. Оценка сейсмической опасности географического объекта по каталогу землетрясений <sup>1)</sup>

**12.3.1. Алгоритм.** Задача состоит в оценивании вероятности  $P_n(k > 0 | M \geq m_0)$  того, что в течение  $T$  лет на заданном географическом объекте произойдёт хотя бы одно землетрясение с магнитудой  $M \geq m_0$ .

Метод основан на двух принятых в сейсмологии предположениях (Бунэ и Горшков, ред., 1980):

1. Магнитуды землетрясений подчиняются экспоненциальному распределению:

$$f(m) = \beta \exp(-\beta(m - m_{\min})), \quad (12.1)$$

<sup>1)</sup> Работа выполнялась совместно с Г. Андриенко и Н. Андриенко в рамках проекта SPIN!

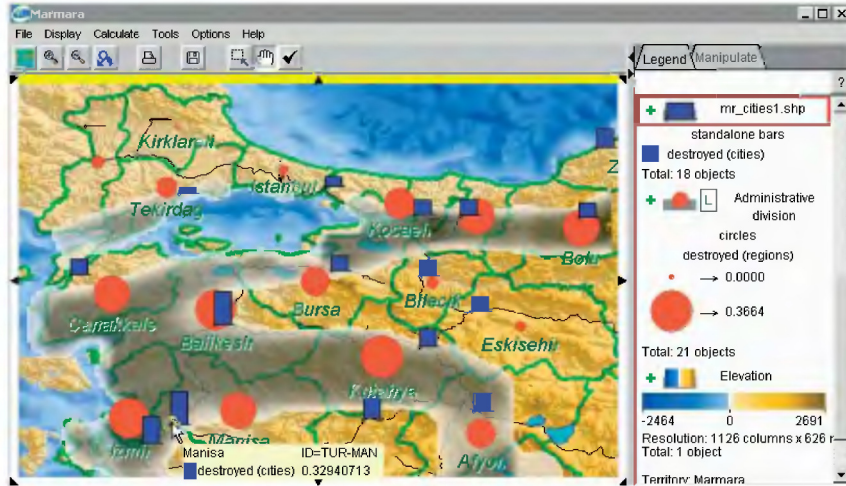


Рис. 12.1. Возможные сильные разрушения застройки типа А в Западной Турции: Высоты столбиков соответствуют доле разрушений в городах, а размеры кружков — доле разрушений в административных районах

где:  $\beta$  — параметр распределения магнитуд землетрясений;  $m_{\min}$  — минимальная представительная магнитуда регистрации землетрясений (которая зависит от чувствительности сейсмической сети наблюдений).

2. Поток землетрясений подчиняется распределению Пуассона, согласно которому вероятность  $k$  событий в области  $S$  за  $T$  лет определяется выражением:

$$P(k | M \geq m_0) = \frac{S(\lambda(M \geq m_0)T)^k}{k!} \exp(-\lambda(M \geq m_0)T), \quad (12.2)$$

$$\lambda(M \geq m_0) = \lambda(M \geq m_{\min}) \exp(-\beta(m_0 - m_{\min})), \quad (12.3)$$

$\lambda(M \geq m)$  — пространственная плотность среднего числа землетрясений с магнитудой  $M \geq m$  в области  $S$  за год.

Алгоритм решения состоит из следующих операций:

1. Определение представительной минимальной магнитуды  $m_{\min}$ .
2. Оценивание сеточной модели пространственной плотности  $\lambda(M \geq m_{\min})$ .
3. Оценивание регионального (среднего по всему региону) значения  $b$ -value:

$$\beta = \left( \frac{1}{N_{m_{\min}}} \sum_{m_i \geq m_{\min}} (m_i - m_{\min}) \right)^{-1},$$

где  $N_{m_{\min}}$  — число землетрясений с магнитудой больше  $m_{\min}$ ; параметр наклона графика повторяемости землетрясений ( $b$ -value) определяется как  $b = \beta \lg e \approx 0,43\beta$ .

4. Оценивание интенсивности потока землетрясений с магнитудой  $M \geq m_0$  и с эпицентрами в буферной зоне географического объекта  $n$  с помощью операции вычисления атрибутов векторных объектов по сеточным и векторным данным:

$$\Lambda_n(M \geq m_0) = \int_{G_n} \lambda(M \geq m_0) ds,$$

где:  $G_n$  — множество узлов координатной сетки, находящихся внутри буферной зоны  $n$ -го векторного объекта.

5. Вычисление вероятности того, что в течение  $T$  лет на заданном географическом объекте произойдет хотя бы одно землетрясение с магнитудой  $M \geq m_0$ .

$$P_n(k > 0 | M \geq m_0) = 1 - \exp(-\Lambda_n(M \geq m_0)T).$$

**12.3.2. Моделирование.** Пример выполнен с помощью системы CommonGIS для региона Западной Турции. Использовался очищенный от афтершоков каталог землетрясений <sup>1)</sup> с 01.01.1976 по 12.31.1999, содержащий 10 560 событий с магнитудами от 2,9 до 7,3.

Минимальная представительная магнитуда землетрясений  $m_{\min}$  оценивалась экспертом на основе качественного анализа графиков распределения землетрясений.

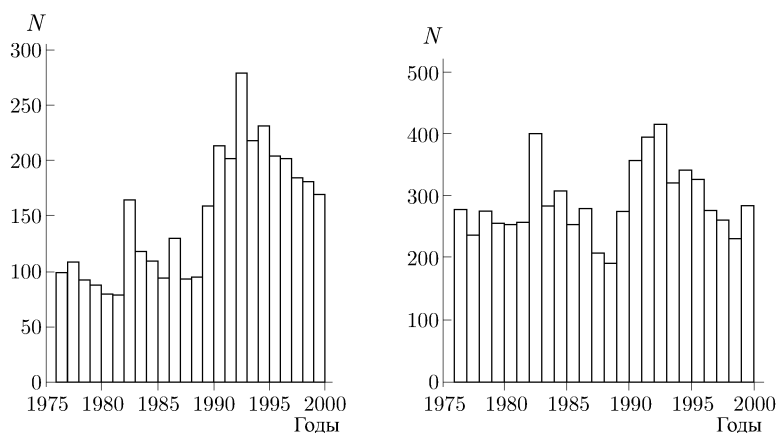


Рис. 12.2. Гистограммы распределения землетрясений по годам: слева  $m < 3,1$  (3588 событий), справа  $m \geq 3,1$  (6972 события)

На рис. 12.2 показаны гистограммы распределения землетрясений по годам с магнитудами  $m < 3,1$  (слева) и  $m \geq 3,1$  (справа). Можно видеть, что в 1990 году число зарегистрированных землетрясений

<sup>1)</sup> Каталог землетрясений был подготовлен в Геофизическом центре (GFZ), г. Потсдам, Германия в рамках Проекта SPIN!

с магнитудами  $m < 3,1$  существенно увеличилось. Это свидетельствует об изменении чувствительности сейсмической сети. В то же время число землетрясений с магнитудами  $m \geq 3,1$  менее изменчиво.

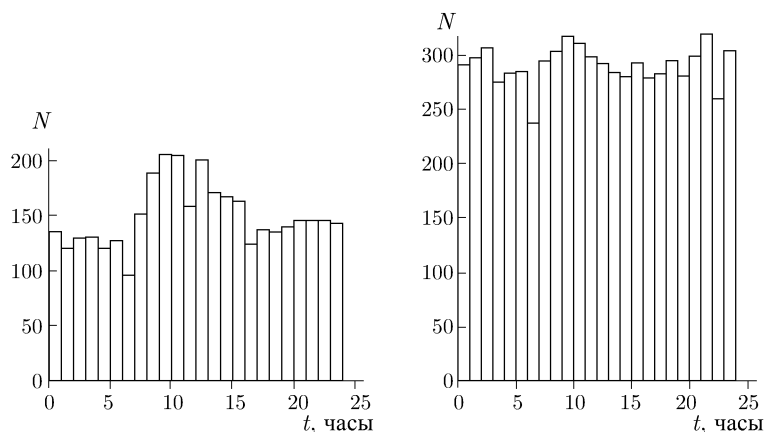


Рис. 12.3. Гистограммы распределения землетрясений по времени суток: слева  $m < 3,1$  (3588 событий), справа  $m \geq 3,1$  (6972 события)

На рис. 12.3 показаны гистограммы распределения землетрясений по времени суток с магнитудами  $m < 3,1$  (слева) и  $m \geq 3,1$  (справа).

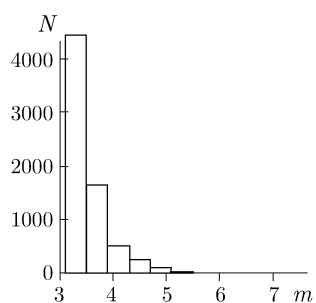


Рис. 12.4. Гистограмма распределения числа событий с  $m \geq 3,1$  по магнитуде (шаг 0,4). Значение  $b$ -value для событий с  $m \geq 3,1$  оказалось равным  $b = 1,08$

Можно видеть, что днём в рабочее время число событий с  $m < 3,1$  существенно увеличивается, в то время как для событий с магнитудами  $m \geq 3,1$  такие вариации не наблюдаются. Скорее всего, изменчивость числа событий с  $m < 3,1$  объясняется техногенными шумами. Это предположение косвенно подтверждается относительным уменьшением числа событий в 6, 11 и 16 часов по Гринвичу, что соответствует 8, 13 и 18 часам местного времени (начало работы, обеденный перерыв и конец работы).

Гистограмма распределения числа событий с  $m \geq 3,1$  по магнитуде показана на рис. 12.4. Гистограмма напоминает экспоненциальное распределение. На основании этого принята гипотеза о том, что мини-

мальная представительная магнитуда  $m_{\min} = 3,1$ .

С помощью системы CommonGIS найдена сеточная модель оценки плотности эпицентров землетрясений с магнитудой  $m \geq 5,0$ . После этого на основании этой модели и полигонов, представляющих адми-



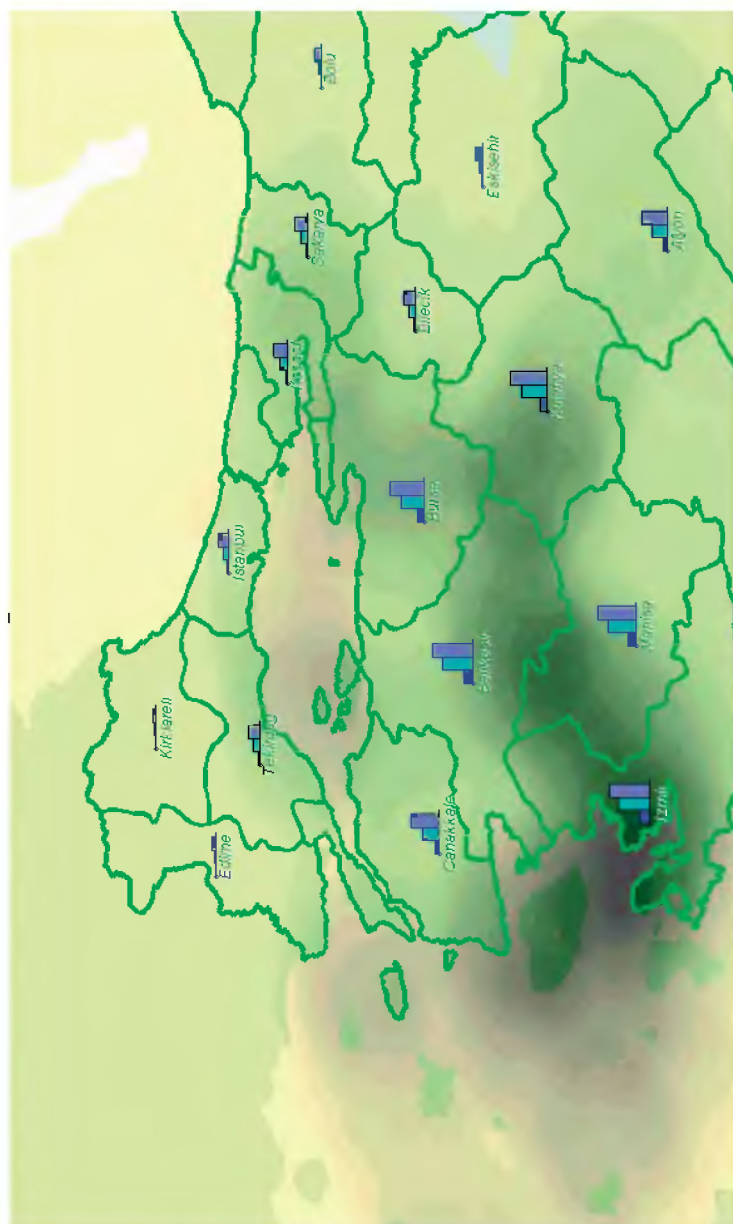


Рис. 12.5. Сеточная модель плотности эпицентров и диаграммы вероятностей того, что в округе произойдет хотя бы одно землетрясение с магнитудой  $m \geq 5,0$  за 1,5 и 10 лет

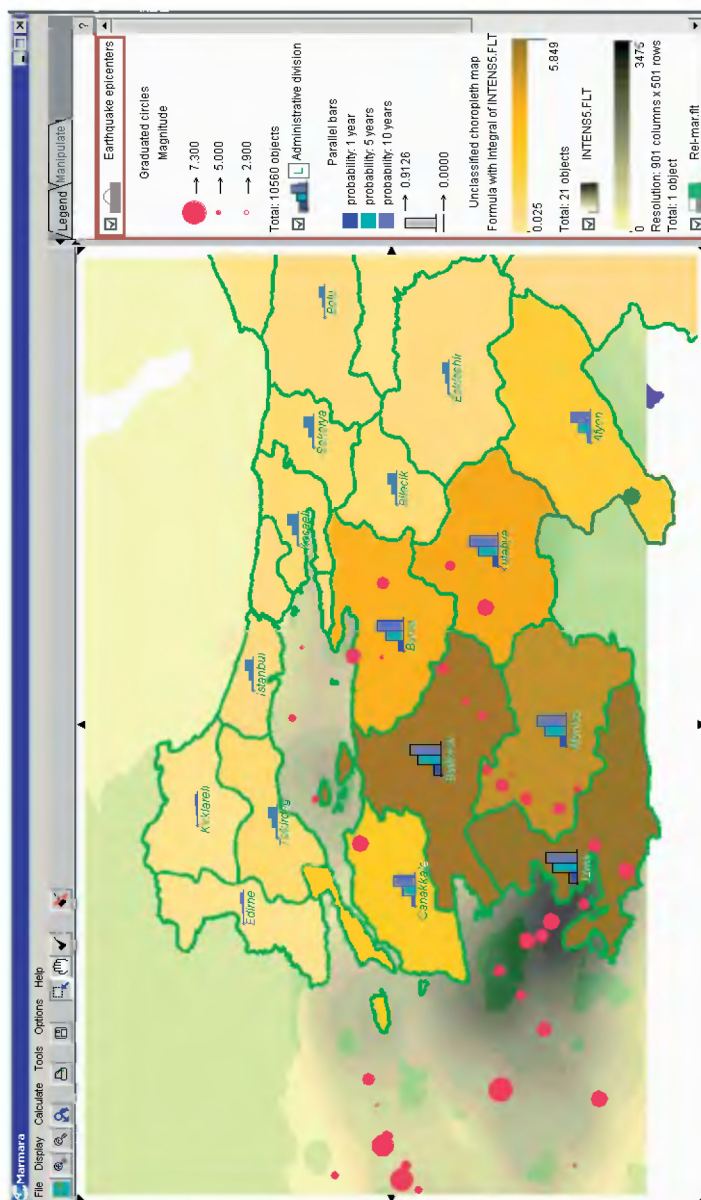


Рис. 12.6. Диаграммы вероятностей того, что в округе произойдет хотя бы одно землетрясение с магнитудой  $m \geq 5,0$  за 1, 5 и 10 лет, сеточная модель оценки числа событий с  $m \geq 5,0$  с 1.01.1976 до 31.12.1999

Таблица 12.1. Атрибуты административных областей Западной Турции

Название района	Ожидаемое число событий с $m \geq 5,0$ за 24 года	Вероятность хотя бы одного события с $m \geq 5,0$ за 5 лет	Вероятность хотя бы одного события с $m \geq 5,0$ за 1 год	Ожидаемое число событий с $m \geq 5,0$ за 24 года
Edirne	0,0832	0,0425	0,0086	0,208472
Canakkale	0,6194	0,3831	0,0921	2,318308
Izmir	0,9126	0,7043	0,2163	5,849015
Kirklareli	0,0520	0,0264	0,0053	0,128260
Zonguldak	0,0103	0,0052	0,0010	0,024889
Tekirdag	0,2552	0,1370	0,0290	0,707152
Kocaeli	0,3262	0,1791	0,0387	0,947537
Sakarya	0,2842	0,1540	0,0329	0,802531
Bolu	0,1336	0,0692	0,0142	0,344273
Ankara	0,0223	0,0112	0,0023	0,054215
Istanbul	0,2297	0,1223	0,0258	0,626370
Balikesir	0,9056	0,6928	0,2103	5,665105
Bursa	0,7725	0,5230	0,1376	3,552984
Bilecik	0,2526	0,1355	0,0287	0,698627
Eskisehir	0,1622	0,0847	0,0175	0,424766
Kutahya	0,8159	0,5710	0,1557	4,061706
Manisa	0,8678	0,6365	0,1832	4,856915
Afyon	0,5631	0,3390	0,0795	1,987089
Konya	0,1076	0,0553	0,0113	0,273194

нистранивные округа Западной Турции, были вычислены вероятности того, что в округе произойдет хотя бы одно землетрясение с магнитудой  $m \geq 5,0$  за 1,5 и 10 лет, а также ожидаемое число землетрясений с  $m \geq 5,0$  за 24-летний период наблюдений. Результаты представлены в табл. 12.1 и на рис. 12.5, 12.6.

#### 12.4. Выделение возможных зон возникновения землетрясений

Пример выделения зон возможного возникновения землетрясений выполнен с помощью системы ГеоПроцессор для области Болгарии, в которой 4.4.1904 произошло сильное землетрясение  $M = 7,8$  с эпицентром вблизи г. Кресна (*Гитис и др., 2004*). Для этого района мы располагали данными, подготовленными участниками проекта IC 15 CT97 0200 «Assessment of Seismic Potential in European Large Earthquake Areas (ASPELEA)»:

1. Рельеф поверхности Земли с детальностью  $30'' \times 30''$  (рис. 12.7 а).
2. Рельеф поверхности Мохоровичича с детальностью  $4' \times 4'$  (рис. 12.7 б).
3. Гравитационные аномалии в редукции Буге с детальностью  $4' \times 4'$  (рис. 12.7 в).
4. Разломы, ранжированные по степени активности в четвертичное время (рис. 12.7 г).
5. Каталог землетрясений с 1900 по 1997 годы, представительный, начиная с магнитуды  $M = 1,8$ , начиная с 1985 года.

Имеющиеся данные являются весьма неполными. Каталог землетрясений относится к очень короткому интервалу наблюдений по сравнению с частотой возникновения сильных землетрясений. В геолого-геофизических данных отсутствует геодинамическая информация о скоростях современных движений, тепловом потоке, кинематических свойствах разломов и т. д. При решении были приняты следующие предположения:

1. Повторяемость землетрясений: землетрясения могут повторяться в тех местах, где они уже были. Потому зарегистрированные землетрясения задают нижнюю границу максимальных магнитуд землетрясений. Однако выделение областей только по максимальным зарегистрированным землетрясениям даёт заниженную оценку из-за короткого интервала наблюдений. В каталоге зарегистрированы сильные землетрясения с магнитудами  $M$  до 7,8. Следовательно, вблизи очаговых зон зарегистрированных землетрясений возможны землетрясения с такими же магнитудами в будущем.

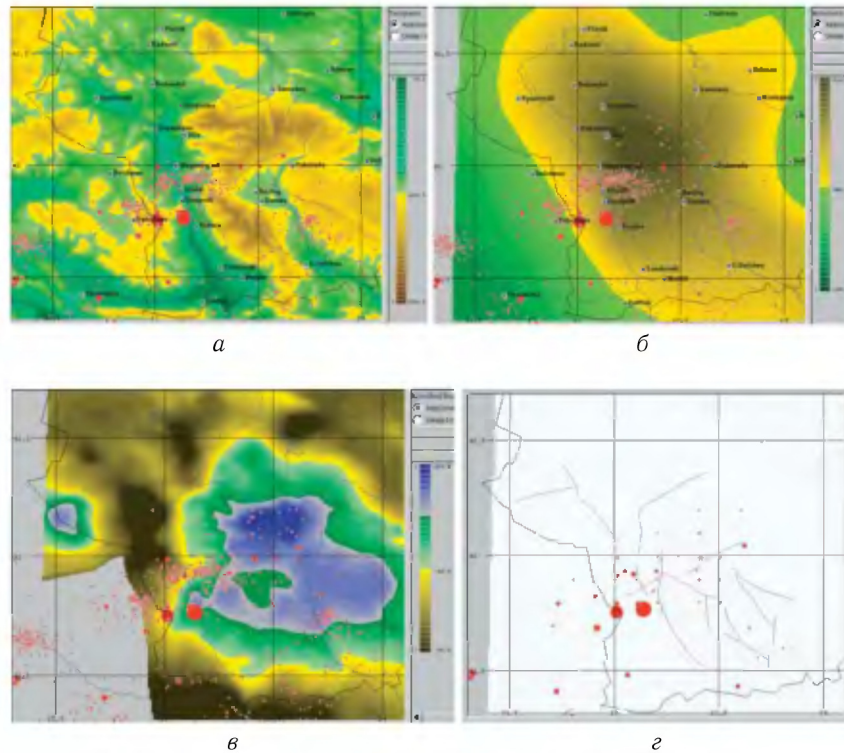


Рис. 12.7. Карты исходных данных: а) Рельеф поверхности Земли и эпицентры землетрясений  $M > 1,8$ , 1900–1997 гг.; б) Рельеф поверхности Мохоровичича; в) Гравитационные аномалии в редукции Буге; г) Разломы, активные и частично активные в четвертичное время и эпицентры землетрясений с магнитудой  $M \geq 4,5$  с 1900 г по 1997 гг., афтершоки удалены

2. Квазистационарность сейсмического процесса: максимальная магнитуда землетрясений зависит от тектонических свойств геологической среды, которые медленно изменяются во времени.
3. Геологическая обусловленность: очаги землетрясений с магнитудами  $M = 7-8$  приурочены к глубинным неоднородностям земной коры, а с магнитудами  $M = 5-6$  приурочены к менее крупным неоднородностям, которые могут проявиться в рельефе поверхности Земли. Пересечение зон неоднородности земной коры с зонами влияния активных в четвертичное время разломов выделяет области с наибольшими тектоническими напряжениями.

Признаки прогноза искались в соответствии с предположением 3. Зоны глубинных неоднородностей коры искались по сеточным моделям рельефа поверхности Земли, поверхности Мохоровичича и аномалиям силы тяжести с помощью преобразований типа **Сеточные**

слои  $\Rightarrow$  **Сеточный слой**, таких как вычисление производных по нескольким направлениям, вычисление модулей градиентов, вычисление максимального разброса значений и среднеквадратичного отклонения (СКО) в скользящих окнах с радиусами  $R = 2,5, 5, 7,5$  и  $10$  км. Зоны влияния активных разломов находились с помощью преобразований **Векторный слой**  $\Rightarrow$  **Сеточный слой**, таких как вычисление плотности линий, близости к линиям и суммарной длины линий в скользящих окнах с радиусами  $R = 5, 10, 15, 20$  и  $25$  км. Были отобраны следующие 3 признака:

1. Сеточная модель среднеквадратичного отклонения (СКО) значений поверхности Мохоровичича в скользящем окне радиуса  $R = 7,5$  км ( $x_1$ ). Из рис. 12.8а видно, что признак  $x_1$  выделяет

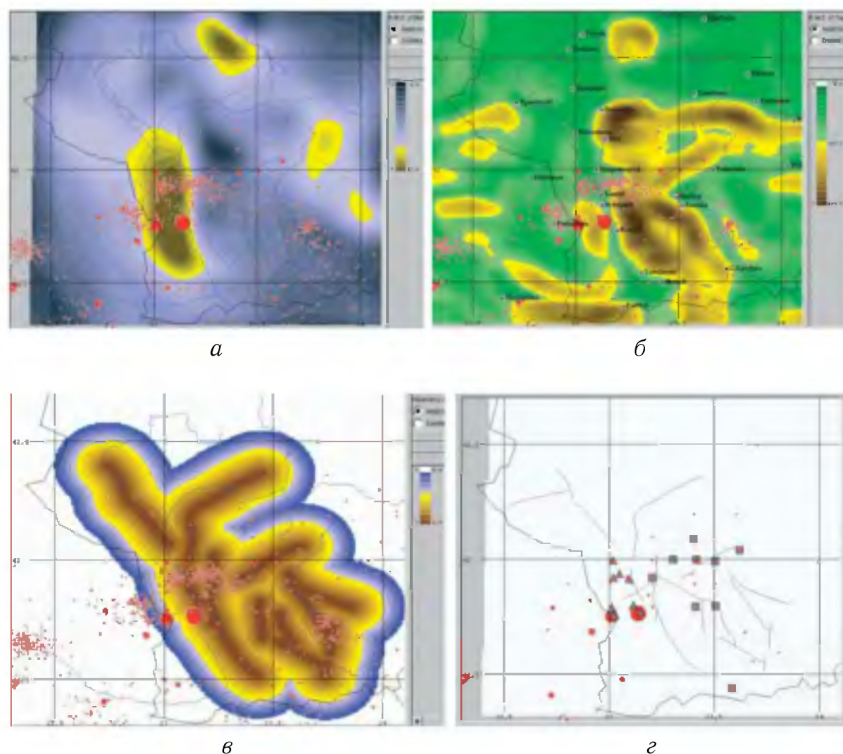


Рис. 12.8. Геолого-геофизические признаки и выборки прецедентов для областей 1 и 2: а) карта среднеквадратичных отклонений (СКО) значений поверхности Мохоровичича в текущем окне  $R = 7,5$  км; б) карта СКО значений рельефа поверхности Земли в текущем окне  $R = 7,5$  км; в) карта близости к разломам, частично активным в четвертичное время с  $R = 20$  км; г) выборки прецедентов 1 (треугольники) и 2 (квадраты)

- несколько областей, в одной из которых происходили землетрясения с магнитудами  $M = 7-7,8$ .
2. Сеточная модель СКО поверхности рельефа Земли в окне радиуса  $R = 7,5$  км ( $x_2$ ). Из рис. 12.8 б видно, что признак  $x_2$  является менее избирательным, но он достаточно хорошо выделяет область, где происходили землетрясения с магнитудами  $M = 5-5,5$ .
  3. Сеточная модель близости к разломам, частично активным в четвертичное время (рис. 12.8 в), вычисленная с окном  $R = 20$  км ( $x_3$ ). Близость узла сетки к линии определяется как  $S = 1 - \rho(u, l)/R$ , при  $\rho(u, l) < R$ ,  $S = 0$  при  $\rho(u, l) \geq R$ , где  $\rho(u, l)$  — евклидово расстояние от узла сетки  $u$  до множества линий.

Решающее правило искалось по двум выборкам, составленным по эпицентрам землетрясений с магнитудами  $M \geq 4,5$  (рис. 12.8 г). Можно предположить, что увеличение неоднородности геологической среды по СКО поверхности Мохоровичича или поверхности рельефа Земли или большая близость к середине зоны активного разлома при прочих равных условиях увеличивают уверенность в возможности накопления и разрядки высокого сейсмического потенциала. Таким образом, предполагается, что отобранные признаки обладают свойством положительной монотонности, состоящей в том, что увеличение значения любого из признаков при прочих равных условиях увеличивает или не изменяет (но не уменьшает) степень уверенности в возможности того, что данный набор признаков относится к очаговой зоне сильных землетрясений. Можно видеть, что функция уверенности совпадает с эмпирической функцией распределения (11.9).

На рис. 12.9 а, 12.9 б показаны Западная область, выделенная по функции уверенности  $F_1(\mathbf{x})$  на выборке 1 и на признаках  $x_1, x_3$ , и Восточная область, выделенная по функции уверенности  $F_2(\mathbf{x})$  на выборке 2 и на признаках  $x_2, x_3$ .

При принятии решения на относительно небольших обучающих выборках наиболее важен детерминистский аспект функции уверенности: граница области признаков  $A = \{\mathbf{x} : F(\mathbf{x}) > 0\}$ , на которой функция принимает положительные значения. Для объяснения решения необходима сейсмо-тектоническая интерпретация этой границы в виде не очень сложного логического высказывания. Метод построения такого высказывания описан в разделе 11.4.2.

Рассмотрим пример построения логического правила для Западной области рассматриваемого региона. Значения признаков в точках выборки 1 для признаков  $x_1, x_3$  приведены в табл. 12.2.

Матрица принадлежности  $P$  точек выборки покрытиям представлена в табл. 12.3 ( $\mu = 1$ , тогда  $\Delta x_1 \approx 1$ ,  $\Delta x_3 \approx 0,05$ ).

После минимизации выбраны строки матрицы 5 и 7, тогда

$$\psi_1(\mathbf{x}) = \max \left( \Phi^{(5)}(\mathbf{x}), \Phi^{(7)}(\mathbf{x}) \right),$$

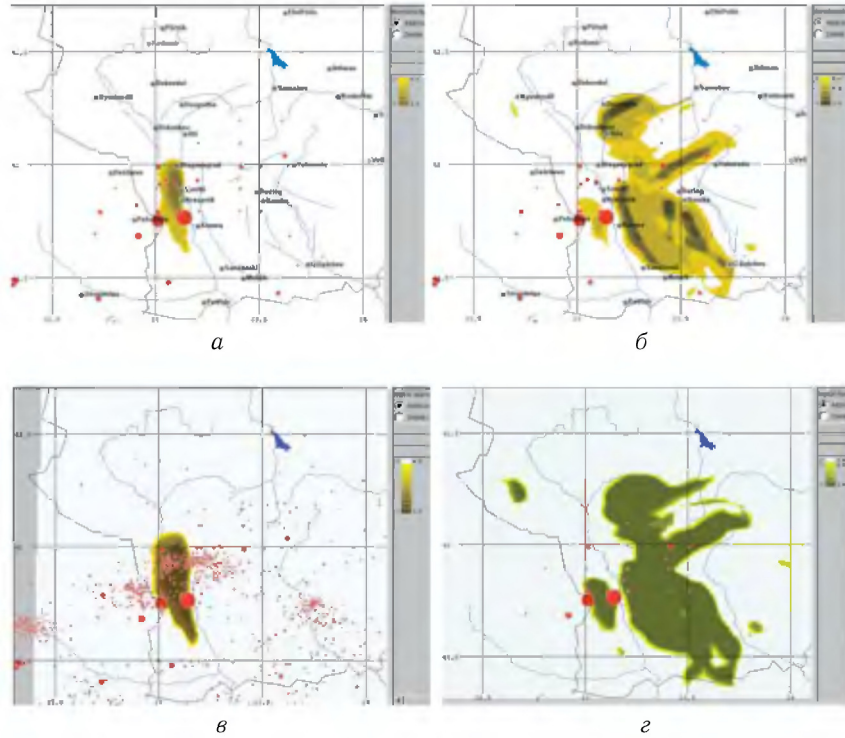


Рис. 12.9. Результаты районирования по прецеденту и функциям нечёткой логики: а) и б) области, выделенные по функциям уверенности  $f_1(\mathbf{x})$  и  $f_2(\mathbf{x})$  соответственно; в) и г) области, выделенные по функциям нечёткой логики для  $\Psi_1(\mathbf{x})$  и  $\Psi_2(\mathbf{x})$  соответственно

где:

$$\Phi^{(5)}(\mathbf{x}) = \min(\varphi_1^{(5)}(x_1), \varphi_3^{(5)}(x_3)), \quad \Phi^{(7)}(\mathbf{x}) = \min(\varphi_1^{(7)}(x_1), \varphi_3^{(7)}(x_3)),$$

$$\varphi_1^{(5)}(x_1) = \begin{cases} 0, & x_1 \leq 15,52, \\ 0,75(x_1 - 15,52), & 15,52 < x_1 \leq 16,52, \\ 0,75 + 0,25(x_1 - 16,52), & 16,52 < x_1 \leq 17,52, \\ 1, & 17,52 < x_1; \end{cases}$$

$$\varphi_3^{(5)}(x_3) = \begin{cases} 0, & x_3 \leq 0,34, \\ \frac{0,75(x_3 - 0,34)}{0,05}, & 0,34 < x_3 \leq 0,39, \\ 0,75 + \frac{0,25(x_3 - 0,39)}{0,05}, & 0,39 < x_3 \leq 0,44, \\ 1, & 0,44 < x_3; \end{cases}$$



Таблица 12.2. Значения признаков  $x_1$ ,  $x_3$  в точках выборки 1

№	Долгота	Широта	$x_1$ (км)	$x_3$
1	23,019	41,984	15,81	0,62
2	23,047	41,934	16,40	0,75
3	23,019	41,915	16,06	0,63
4	23,096	41,911	17,40	0,94
5	23,016	41,786	16,52	0,39
6	23,121	41,786	16,64	0,74
7	23,030	41,754	16,94	0,33
8	23,135	41,763	16,03	0,66

Таблица 12.3. Значения функции принадлежности  $P$  на точках выборки 1

	1	2	3	4	5	6	7	8
1	1	1	1	1		1		1
2		1		1		1		
3	1	1	1	1		1		1
4				1				
5	1	1	1	1	1	1		1
6		1		1		1		
7		1	1	1	1	1	1	1
8	1	1	1	1		1		1

$$\varphi_1^{(7)}(x_1) = \begin{cases} 0, & x_1 \leq 15,94, \\ 0,75(x_1 - 15,94), & 15,94 < x_1 \leq 16,94, \\ 0,75 + 0,25(x_1 - 16,94), & 16,94 < x_1 \leq 17,94, \\ 1, & 17,94 < x_1; \end{cases}$$

$$\varphi_3^{(7)}(x_3) = \begin{cases} 0, & x_3 \leq 0,28 \\ \frac{0,75(x_3 - 0,28)}{0,05}, & 0,28 < x_3 \leq 0,33, \\ 0,75 + \frac{0,25(x_3 - 0,33)}{0,05}, & 0,33 < x_3 \leq 0,38, \\ 1, & 0,38 < x_3. \end{cases}$$

Приведённые выше формулы выглядят достаточно громоздкими, но в рассматриваемых ГИС операция взятия произвольной кусочно-линейной функции задаётся просто перечислением абсцисс и ординат точек излома.

Аналогичное построение для Восточной области по выборке примеров 2 дало следующий результат:

$$\psi_2(\mathbf{x}) = \Phi^{(9)}(\mathbf{x}),$$

где:

$$\Phi^{(9)}(\mathbf{x}) = \min \left( \varphi_2^{(9)}(x_2), \varphi_3^{(9)}(x_3) \right),$$

$$\varphi_2^{(9)}(x_2) = \begin{cases} 0, & x_2 \leq 300, \\ \frac{0,75(x_2 - 300)}{30}, & 300 < x_2 \leq 330, \\ 0,75 + \frac{0,25(x_2 - 330)}{30}, & 330 < x_2 \leq 360, \\ 1, & 360 < x_2; \end{cases}$$

$$\varphi_3^{(9)}(x_3) = \begin{cases} 0, & x_3 \leq 0,2 \\ \frac{0,75(x_3 - 0,2)}{0,05}, & 0,2 < x_3 \leq 0,25, \\ 0,75 + \frac{0,25(x_3 - 0,25)}{0,05}, & 0,25 < x_3 \leq 0,3, \\ 1, & 0,3 < x_3. \end{cases}$$

Теперь можно составить текстовое объяснение выявленных двух областей с возможными очагами ожидаемых сильных землетрясений в исследуемом регионе Болгарии (рис. 12.9 в, 12.9 г).

Земная кора в Западной области 1 характеризуется большими неоднородностями поверхности Мохоровичича и близостью к активным в четвертичное время разломам. Решению по Западной области со степенью уверенности 0,75 соответствует следующее правило:

*если* (СКО поверхности Мохоровичича ( $R = 7,5$  км) больше 16,52 км)  
*и* (близость к разломам ( $R = 20$  км) больше 0,39)  
*или* (СКО поверхности Мохоровичича ( $R = 7,5$  км) больше 16,94 км)  
*и* (близость к разломам при  $R = 20$  км больше 0,33),  
*то* возможны землетрясения с магнитудами  $M = 7-8$ .

Восточная область 2 характеризуется большими неоднородностями поверхности рельефа Земли и близостью к активным в четвертичное время разломам. Решению по Восточной области соответствует следующее правило:

*если* (СКО поверхности рельефа Земли ( $R = 7,5$  км) больше 330 м)  
*и* (близость к разломам при  $R = 20$  км больше 0,25),  
*то* возможны землетрясения с магнитудами  $M = 5-6$ .

Интерпретация полученного решения состоит в следующем. Можно предположить, что выделенная Западная область 1 характеризует на-

пряжённое состояние верхней мантии, которое выражено высокой дисперсией поверхности Мохоровичича. Эта область могла генерировать активные разломы и являться причиной наличия очагов с большими энергиями землетрясений. Выделенная Восточная область 2 относится к очаговым зонам землетрясений с меньшими энергиями. Эти очаги могут быть связаны с неоднородностями верхней части Земной коры, которые выражены в рельефе поверхности Земли и зонах активных разломов.

## Список литературы

- Авсюк Ю.Н.* 1996. Приливные силы и природные процессы. — М.: ОИФЗ РАН. — 188 с.
- Андерсон Т.* 1963. Введение в многомерный статистический анализ. — М.: ГИФМЛ.
- Антикаев Ф.Ф., Гитис В.Г., Кофф Г.Л., Фролова Н.И.* 1997. Оценка сейсмической опасности и сейсмического риска (пособие для должностных лиц). — М.: БСТС Центр. — 53 с.
- Артемьев М.Е.* 1976. Изостазия территории СССР. — М.: Наука. — 215 с.
- Атлас Народна Республика България, 1983. София: ГУГК. 168 л.
- Берлянт А.М.* 1996. Геоиконика. — М.: МГУ, АЕН РФ, «Астрей». — 208 с.
- Банди В.* 1988. Методы оптимизации. — М.: Радио и связь.
- Богданов Б., Димитров Д., Стайков М. и др.* 1978. Рудные формации Карпато-Балканской области. М., Недра. — 239 с.
- Бончев Е.К., Бунэ В.И., Карагюлева Ю. и др.* 1983. Методика составления прогнозной карты сейсмического районирования территории Народной Республики Болгарии // Оценка сейсмической опасности. — М.: Наука. С. 5–35. (Вопросы инженерной сейсмологии; Вып. 24).
- Боровков А.А.* 1984. Математическая статистика. — М.: Наука. — 472 с.
- Буне В.И., Гитис В.Г., Каленик В.Н., Щукин Ю.К.* 1984. Методика и результаты прогноза максимальных магнитуд землетрясений региона Крым-Кавказ-Западная Туркмения // Исследование земной коры и верхней мантии сейсмоопасных зон. — М.: Наука. С. 148–174.
- Буне В.И., Гитис В.Г., Каленик В.Н., Щукин Ю.К.* 1986. Применение метода аппроксимации экспертных оценок для построения карты  $M_{\max}$  землетрясений Кавказа // Сейсмологические исследования. № 9. — М.: ВИНТИ. С. 49–77.
- Бунэ В.И., Горшков Г.П.* ред. 1980. Сейсмическое районирование территории СССР. — М.: Наука. — 307 с.
- Буне В.И., Турбович И.Т., Борисов Б.А. и др.* 1974. О методе выявления связи магнитуды землетрясений с тектоническими параметрами района // Докл. АН СССР Т. 214, № 3. С. 553–556.
- Буне В.И., Турбович И.Т., Борисов Б.А. и др.* 1975. Метод прогнозирования максимальной магнитуды землетрясений // Изв. АН СССР Физика Земли. № 10. С. 31–43.
- Вапник В.Н., Червоненкис А.Я.* 1974. Теория распознавания образов. — М.: Наука.
- Вахтанова А.Н., Эсенов Э.М., Непесов Р.Д.* 1970. О сейсмическом микрорайонировании городов ТССР // Изв. АН СССР. Сер. физ.-техн., хим. и геол. наук, № 3.

- Вольфсон Ф.Н., Дружинин Н.М. 1975. Главнейшие типы рудных месторождений. — М.: Недра. — 392 с.
- Воронин Ю.Я., Черемисина Е.Н. 1971. Использование ЭВМ для эффективного опробования участков и выделения месторождений // Математические проблемы геофизики. Новосибирск. С. 321–341.
- Габриелов А.М., Дмитриева О.Д., Кейлис-Борок В.И. и др. 1986. Долгосрочный прогноз землетрясений. — М.: ИФЗ АН СССР. — 125 с.
- Гвишиани А.Д., Горшков А.И., Ранцман Е.Я. и др. 1988. Прогнозирование мест землетрясений в регионах умеренной сейсмичности. — М.: Наука. — 174 с.
- Гитис В.Г. 1974. Алгоритмы прогнозирования и синтеза признаков с использованием одномерных кусочно-линейных функций // Нелинейные и линейные методы в распознавании образов. — М.: Наука. С. 20–30.
- Гитис В.Г. 1977. Об одном классе задач прогнозирования // Распознавание образов. — М.: Наука. С. 5–11.
- Гитис В.Г. 1987. Метод аппроксимации функциональных зависимостей по балльным экспертным оценкам // Проблемы передачи информации. том XXIII. Вып. 3. С. 94–100.
- Гитис В.Г., Андриенко Г.Л., Андриенко Н.В. 2004. Исследование сейсмологической информации в сетевых аналитических ГИС // Физика Земли № 3. С. 43–53.
- Гитис В.Г., Буне В.И., Вычев В.Т., Миронов М.А. 1982. Построение карты  $M_{\max}$  землетрясений на основе метода аппроксимации интервальных экспертных оценок // Физика Земли. № 4. С. 31–44.
- Гитис В.Г., Вайншток А.П. 2001. Сетевые аналитические ГИС, часть 1 // ГИС обозрение. № 2. С. 14–16.
- Гитис В.Г., Вайншток А.П. 2001. Сетевые аналитические ГИС, часть 2. — ГИС обозрение, № 3. С. 8–12.
- Гитис В.Г., Вайншток А.П., Кофф Г.Л., Карагодина М.В. 1997. Геоинформационная технология оценки сейсмического риска и ущерба от землетрясений // Прикладная геоэкология, чрезвычайные ситуации, земельный кадастр и мониторинг. Выпуск 2. Издательско-полиграфический комплекс РЭФИА. С. 38–42.
- Гитис В.Г., Вайншток А.П., Сидорин А.Я., Татевосян Р.Э. 1997. Построение карты  $M_{\max}$  землетрясений Ставропольского края в инструментальной среде ГЕО // Тезисы II Национальной конференции по сейсмостойкому строительству и сейсмическому районированию. Сочи. — М.: РЭФИА. С. 23–24.
- Гитис В.Г., Деарт Д.А., Ошер Б.В., Руденко С.Л. 1989а. ГЕО — экспертная система для геолого-геофизического прогноза // Экспертные системы: состояние и перспективы. — М.: Наука. С. 119–130.
- Гитис В.Г., Добрев Т.В., Ермаков Б.В. и др. 1989б. Применение экспертной системы ГЕО для регионального прогноза свинцово-цинковых месторождений Болгарии // Геофизический журнал. Т. 11, № 4. С. 36–48.
- Гитис В.Г., Добрев Т.В., Ермаков Б.В. и др. 1989в. Региональный прогноз медных месторождений Болгарии по геофизическим данным // Карпато-Балканская геологическая ассоциация. Труды XIV Конгресса СВГА. София. С. 1247–1250.

- Гитис В. Г., Миронов М. А., Буне В. И., Вычев В. Т. 1982. Применение метода аппроксимации интервальных экспертных оценок в построении карты  $M_{\max}$  землетрясений Болгарии // *Geologica Balcanica*. V. 12, № 2. Sofia. С. 49–70.
- Гитис В. Г., Миронов М. А., Буне В. И. и др. 1986. Прогноз  $M_{\max}$  землетрясений на основе аппроксимации интервальных экспертных оценок // *Физика Земли*. № 4. С. 24–31.
- Гитис В. Г., Нагорнов В. С., Турбович И. Т., Юрков Е. Ф. 1974. Опыт решения задач медицинского и сейсмического прогнозирования на основе использования одномерных нелинейных преобразований // *Нелинейные и линейные методы распознавания образов*. — М.: Наука. С. 34–42.
- Гитис В. Г., Петрова Е. Н., Пирогов С. А. 2004. Модель локального взаимодействия компонент геозкологической структуры // *Информационные процессы*. Т. 4, № 1. С. 1–7.
- Григоров Б. 1961. Магнитно поле в НР България // *Изв. Геофиз. Ин-т Бълг. АН*. Т. 2. С. 109–115.
- Дикенштейн Г. Х. ред. 1975. Тектоническая карта юга СССР, масштаб 1:1000000. — М.: ГУГК.
- Димитрова Е. ред., 1978. Карта магматических формаций Карпато-Балканской области // 1:1 000 000 София. КИПП по картография. 7 л.
- Добрев Т. Б., Щукин Ю. К. 1974. Геофизические поля и сейсмичность восточной части Карпато-Балканского региона. — М.: Наука. — 170 с.
- Дрейпер Н., Смит Г. 1973. Прикладной регрессионный анализ. — М.: Статистика.
- Думнов А. Д., Рыбальский Н. Г., Самотесов Е. Д., Максимов Ю. И. и др. 2001. Аналитический доклад «ПРИРОДНЫЕ РЕСУРСЫ И ОКРУЖАЮЩАЯ СРЕДА РОССИИ». — М.: НИА-Природа, РЭФИА, — 572 с.
- Загоруйко Н. Г. 1999. Прикладные методы анализа данных и знаний // Новосибирск. Изд. Института математики. — 268 с.
- Закревский А. Д. 1966. Оптимизация покрытия множеств // *Сб. Логический язык для преобразования алгоритмов синтеза релейных устройств*. — М.: Наука.
- Закс Ш. 1975. Теория статистических выводов. — М.: Мир.
- Землетрясения в СССР в 1975–1988 гг. 1978–1990. — М.: Наука.
- Зоненшайн Л. П., Савостин Л. А., Вардапетян А. И., Володина В. И. 1979. Кайнозойская тектоника плит и геологическая история срединного отрезка Альпийско-Гималайского пояса // *Строение литосферных плит* М. Наука.
- Ибрагимов И. А., Хасьминский Р. З. 1978. Асимптотическая теория оценивания. — М.: Наука.
- Йовчев И. С. 1960. Полезны ископаемы на НР България // *София: Техника*. — 105 с.
- Карта Сейсмического микрорайонирования г. Иркутска, 1989. Составлена Восточно-Сибирским трестом инженерно-строительных изысканий Госстроя РСФСР. Утверждена постановлением Госстроя РСФСР от 27.08.89 № 40.
- Кендалл М. Дж., Стьюарт А. 1973. Статистические выводы и связи. — М.: Наука.
- Кокс Д., Хинкли Д. 1978. Теоретическая статистика. — М.: Мир.
- Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Наука, 1972.

- Кондорская Н.В., Шебалин Н.В. ред. 1977. Новый каталог сильных землетрясений на территории СССР с древнейших времён до 1975 г. — М.: Наука. — 535 с.
- Конторович А.Э., Фотиади Э.Э., Дёмин В.И. и др. 1981. Прогноз месторождений нефти и газа. — М.: Недра. — 350 с.
- Копп М.Л. 1982. Некоторые вопросы позднеальпийской геодинамики Юго-Восточного Кавказа, Талыша и Нижнекуринской впадины // Проблемы геодинамики Кавказа. — М.: Наука.
- Косыгин Ю.А., Кулиш Е.А. ред., 1984. Основные типы рудных формаций. Терминологический справочник // М., Недра. — 216 с.
- Кофф Г.Л., Гитис В.Г., Баранников В.Г. и др., 1995. Геоинформационный подход к оценке сейсмического риска урбанизированных территорий // Прикладная геоэкология, чрезвычайные ситуации, земельный кадастр и мониторинг. Вып. 1. — М.: С. 34–37.
- Кузнецов Н.А., 2001. Информационное взаимодействие в технических и живых системах // Информационные процессы. Т. 1. № 1. С. 1–9.
- Кульбак С. 1967. Теория информации и статистика. -М.:Наука.
- Ларищев О.И. 2000. Теория и методы принятия решений, а также хроника событий в волшебных странах. — М.: Логос. 295с.
- Лихачёва Э.А., Гитис В.Г., Бахирева Л.В. и др. 1993. Комплексное районирование городской территории по степени устойчивости к динамическому воздействию // Геоморфология. № 4. С. 41–47.
- Медведев С.В. ред. 1977 Сейсмическое микрорайонирование. — М.:Наука. — 248 с.
- Молчан Г.М. 1975. Статистическое оценивание эффектов сейсмичности: Диссертация на соискание учёной степени доктора физ.-мат. Наук. — М.: ИФЗ АН СССР.
- Молчан Г.М., Дмитриева О.Е. 1991. Идентификация афтершоков: обзор и новые подходы // Вычислительная сейсмология. Вып. 24. С. 19–50.
- Мячкин В.И., Костров Б.В., Соболев Г.А., Шамина О.Г. 1975. Основы физики очага и предвестники землетрясений // Физика очага землетрясений. — М.: Наука. С. 6–29.
- Никитин А.А. 1986. Теоретические основы обработки геофизической информации. — М.: Наука. — 232 с.
- Осипов Г.С. 1997. Приобретение знаний интеллектуальными системами: Основы теории и технологии. — М.: Наука, Физматлит. — 112 с.
- Писаренко В.Ф. 1989. О частотно-магнитудной зависимости землетрясений // Дискретные свойства геологической среды. — М.: Наука. С. 47–60.
- Полканова Л.П. ред. 1971. Карта новейшей тектоники юга СССР, масштаб 1:1000000. — М.: ГУГК.
- Полякова Т.П., Медведева Н.С., Степанова М.Б. 1993. Оценка  $M_{\max}$  и сейсмологической обстановки методом сканирования // Сейсмичность и сейсмическое районирование Северной Евразии. Вып. 1. — М.: ИФЗ РАН. С. 51–56.
- Пономарёв А.В., Соболев Г.А., Гитис В.Г. Чжан Чжаочэн, Ван Гусюан, Чин Синси. 1999. Комплексный анализ геофизических полей для обнаружения пространственно-временных предвестников землетрясений // Электронный научно-информационный журнал Вестник ОГГГН РАН,

- 4 (10). URL: [http://www.scgis.ru/russian/cp1251/h\\_dgggms/4-99/comp-an.zip](http://www.scgis.ru/russian/cp1251/h_dgggms/4-99/comp-an.zip).
- Ратнер М.Я., Гельфанд И.М., Варшавский В.А. и др.* 1984. Морфоклинические варианты хронического гломерулонефрита и их значение для оценки тяжести болезни // Архив патологии. Т. 16, № 11. С. 46–51.
- Рейснер Г.И.* 1980. Геологические методы оценки сейсмической опасности. — М.: Недра. — 176 с.
- Рейснер Г.И., Иогансон Л.И.* 1993. Сейсмический потенциал Западной России, других стран СНГ и Балтии // Сейсмичность и сейсмическое районирование Северной Евразии. Вып. 1. — М.: ИФЗ РАН. С. 24–44.
- Ризниченко Ю.В.* 1985. Проблемы сейсмологии. — М.: Наука. — 408 с.
- Рябой В.З.* 1979. Структура верхней мантии территории СССР по сейсмическим данным. — М.: Недра. — 246 с.
- Садовский М.А., Писаренко В.Ф.* 1991. Сейсмический процесс в блоковой среде. — М.: Наука. — 96 с.
- Сергеев Е.М.* 1978. Инженерная геология. — М.: МГУ. С. 165–167.
- Сидоренко А.В.* ред. 1980. Карта разломов территории СССР и сопредельных стран, масштаб 1:2500000 // М., Аэрогеология.
- Симонов Ю.Г., Кружалин В.И.* 1989. Инженерная геоморфология. — М.: МГУ, ч. 1. — 99 с.
- Смирнов В.Б.* 1995. Повторяемость землетрясений и параметры сейсмического режима // Вулканология и сейсмология. № 3. С. 59–70.
- Смирнов В.Б., Завьялов А.Д.* 1999. Концентрационный критерий разрушения с учётом фрактального распределения разрывов // Вулканология и сейсмология. № 4–5. С. 75–80.
- Соболев Г.А.* 1993. Основы прогноза землетрясений. — М.: Наука. — 313 с.
- Соболев Г.А.* 1999. Стадии подготовки сильных камчатских землетрясений // Вулканология и сейсмология, № 4–5. С. 63–72.
- Соболев Г.А., Тюпкин Ю.С.* 1996. Аномалии в режиме слабой сейсмичности перед сильными землетрясениями Камчатки // Вулканология и сейсмология. № 4. С. 64–74.
- Соболев Г.А., Пономарёв А.В.* 2003. Физика землетрясений и предвестники. — М.: Наука, 2003. — 270 с.
- Стоянов Р.Д.* 1979. Металлогения Родопского срединного массива. — М.: Недра. — 180 с.
- Турбович И.Т., Гитис В.Г., Маслов В.К.* 1971. Опознавание образов (детерминированно-статистический подход). — М.: Наука. — 246 с.
- Уломов В.И.* 1999. Сейсмогеодинамика и сейсмическое районирование Северной Евразии // Вулканология и сейсмология. № 4–5. С. 6–22.
- Шалдун Т.Н., Добровольская М.Г.* 1976. Рудные формации свинцово-цинковых месторождений // в книге Рудные формации эндогенных месторождений. — М.: Наука. С. 149–297.
- Юрков Е.Ф.* 1974. Нахождение одномерных нелинейных преобразований на основе одномерных статистических характеристик при прогнозировании // Нелинейные и линейные методы распознавания образов. — М.: Наука. С. 13–18.



- Юрков Е. Ф. 1983. Оценка апостериорной вероятности в задачах распознавания образов и прогнозирования // Проблемы передачи информации. Т. 17, № 2. С. 85–92.
- Amendola A., Bayer J., Ermoliev Y. et al. 2000. System Approach to Modeling Catastrophic Risk and Insurability // Natural hazards. 21. Kluwer Academic Publishers. Netherlands: P. 381–393.
- Andrienko G., Andrienko N. 1999a. Interactive Maps for Visual Data Exploration // International Journal of Geographical Information Science. 13. P. 355–374.
- Andrienko G., Andrienko N., Voss H., Carter J. 1999b. Internet mapping for dissemination of statistical information // Computer Environment and Urban Systems. 23. P. 425–441.
- Andrienko G., Andrienko N. 1999c. Knowledge-Based Visualization to Support Spatial Data Mining // In: Hand, D.J., Kok, J.N., and Berthold, M.R., Advances in Intelligent Data Analysis. IDA-99, Amsterdam, Berlin, Springer. P. 149–160.
- Andrienko N., Andrienko G. 2002. Intelligent Support for Geographic Data Analysis and Decision Making // Journal of Geographic Information and Decision Analysis. V. 5 (2). P. 115–128.
- Andrienko G., Andrienko N., Gitis V. 2003. Interactive Maps for Visual Exploration of Grid and vector geodata. — ISPRS Journal of Photogrammetry & Remote Sensing 1257 1–10.
- Berg C., Tuinman F., Vijbrief T. et al. 1999. Multiserver Internet GIS: standardization and practical experiments // Goodchild M., Egenhofer M., Feageas R., Kottman C. (eds) Interoperating Geographical Information Systems. — Boston: Kluwer Academic Publishers. P. 365–377.
- Bolt B.A. 1999. Earthquakes // N. Y.: W. H. Freeman and Company. — 366 p.
- Bonchev Ek., Bune V.I., Christoskov L., Karaguleva J.A. et al. 1982. Method for Compilation of Seismic Zoning Prognostic Maps for the Territory of Bulgaria // Geologica Balcanica. V. 12.2. Sofia: P. 3–48.
- Bradley R.A., Terry M.E. 1952. Rank Analysis of incomplete Block Designs. I. The method of paired comparisons // Biometrika. № 39. P. 324–345.
- Burrough P.A. 1992. Are GIS data structure too simple minded? // Computers and Geosciences 18 (4); 395–400.
- Burrough P., McDonnel R. 1998. Principle of Geographical Information Systems // Oxford University Press, Oxford.
- Burton P. W. 1979. Seismic risk in Southern Europe through to India examined using Gumbel's third distribution of extreme values // Geophys. J. R. Astr. Soc. 59. P. 249–280.
- Carrara A., Guzzetti F. eds. 1995. Geographical Information Systems is Assessing Natural Hazards // Kluwer Academic Publishers. — 306 p.
- Chen Yong, Wang Wei, Zhu Yueqing, Ji Ying. 1992. Multidisciplinary approach used in expert system for earthquake prediction in China // Journal of Earthquake Prediction Research. V. 1, № 1. P. 107–113.
- Coburn A. 1995. Disaster Prevention and Mitigation in Metropolitan Areas: Reducing Urban Vulnerability in Turkey // Parker, R., Kreimer, A., Munasinghe, M., Editors, Informal Settlements, Environmental Degradation, and Disaster

- Vulnerability. The Turkey Case Study. World Bank and International Decade for Natural Disaster Reduction. P. 66–94.
- Colombo D., Gitis V., de Franco R. 1997. Application of pattern recognition techniques to long-term earthquake prediction in central Costa Rica // *Engineering Geology* 48, 7–18.
- Cox D.R. 1970. The analysis of binary data. - London, Methuen.
- Frank A.U., Raubal M., Vlugt M. editors. 2000. Panel-GI Compendium. A guide to GI and GIS // *European Communities*. P. 141.
- Fukunaga K. 1972. Introduction to statistical pattern recognition // *ACADEMIC PRESS*, N. Y. and London. — 321 p.
- Giardini D., Grunthal G., Shedlock K., Zhang P. 1999. Global Seismic Hazard Map assembled by, GSHAPE.
- Gitis V. 1995. GIS Technology for the Design of Computer-Based Models in Seismic Hazard Assessment // *Geographical Information Systems in Assessing Natural Hazards*, A. Carrara and F. Guzzetti (eds.). Kluwer Academic Publishers. P. 219–233.
- Gitis V.G. 1999. Information Technologies and systems for Complex Analysis of Space-Time Properties of Geological Environment // *Национальный отчёт Международной ассоциации и физики недр Земли Международного геодезического и геофизического Союза, 1995–1998*. — М.: Национальный геофизический комитет РАН. С. 55–62.
- Gitis V.G., Alvares L., Chuy T. et al. 1989. Estimation and Analysis of  $M_{\max}$  of the Earthquakes of Caribbean and Middle America Region with the Help of GEO Expert System // *Proc. of the 4th International Symposium. on the Analysis of Seismicity and Seismic Risk, Czechoslovakia, September 4–9*. P. 128–137.
- Gitis V., Dougyallo A., Osher B., Gergely T. 1998. GeoNet: an information technology for WWW on-line intelligent Geodata analysis // *Proceedings of the 4<sup>th</sup> EC-GIS Workshop. Hungary. Joint Research Centre of European Commission*. P. 124–135.
- Gitis V.G., Ermakov B.V., Ivanovsnaia L.V. et al. 1992. The GEO expert system application to seismic hazard analysis of the Caucasus region // *Cahiers du Centre Europeen de Geodynamique et de Seismologie*. V. 6. P. 117–131.
- Gitis V.G., Ermakov B.V., Ivanovsnaia L.V. et al. 1993. The Information Technology of the GEO system for Prediction  $M_{\max}$  of Earthquake // *Journal of Earthquake Prediction Research*. V. 2. n. 2. P. 221–228.
- Gitis V.G., Ermakov B.V., Semov V.N. et al. 1994. GEO Expert System Application to Oil and Gas Resources Forecasting by the Deep Criteria (Example for West-Siberian Platform) // *A. G. Fabbri et J. J. Royer (eds) 3rd CODATA Conference on Geomathematics and Geostatistics. Sci. de la Terre, Ser. Inf. Nancy*. 32. P. 653–666.
- Gitis V.G., Jurkov E.F., Osher B.V. 1991. GEO - an expert system for geological and geophysical prognosis // *Proc. of the 4th International Conference on Seismic Zonation, August 25th-29th, California, USA*. P. 59–64.
- Gitis V., Jurkov E., Osher B. et al. 1995. Dynamic fields analysis system. *Cahiers du Centre Europeen de Geodynamique et de Seismologie*. V. 9. P. 129–140.

- Gitis V., Jurkov E., Osher B. et al.* 1997. Information technology for forecasting geological processes and phenomena // *Journal Artificial Intelligence in Engineering*. 11. P. 41–48.
- Gitis V., Osher B., Dougiallo A., Vainshtok A.* 2000. COMPASS: Cartography On-line Modeling, Presentation and Analysis System // *Proc. of the 5<sup>th</sup> EC-GIS Workshop*. Stresa. Italy, EC JRC. P. 487–497.
- Gitis V.G., Osher B.V., Pirogov S.A. et al.* 1994. A System for Analysis of Geological Catastrophe Precursors // *Journal of Earthquake Prediction Research*. V. 3. P. 540–555.
- Gitis V.G., Petrova E.N., Pirogov S.A.* 1994. Catastrophe Chains: Hazard Assessment // *Natural Hazards*. 10. P. 117–127.
- Gitis V.G., Petrova E.N., Pirogov S.A.* 1996. Expert knowledge approach to catastrophe Chains // *Cahiers du Centre Europeen de Geodynamique et de Seismologie*. V. 12. P. 67–72.
- Gitis V.G., Shulcloper J.R., Gomez L.A.* 1992. Reconcimiento de elementos de estructuras espaciales. Editorial Academia, La Habana. — 120 p.
- Gitis V., Vainchok A., Tatevosjan R.* 1998. Maximum expected magnitude assessment in GEO computer environment: case study // *Natural Hazards* 17. Kluwer Academic Publishers. Netherlands. P. 225–250.
- Gomez-Fernandez F.* 2000. Contribution of Geographical Information systems to the Management of Volcanic Crises // *Natural hazards*. 21, Kluwer Academic Publishers, Netherlands. P. 347–360.
- Hayes-Roth F., Waterman D.A., Lenat D.B.* 1983. Building Expert Systems // Addison-Wesley. P. 444.
- Hirtle S.C., Frank A.U.* eds. 1997. Spatial Information Theory: A theoretical Basis for GIS, Lecture Notes in Computer Science, 1329, Springer, Berlin,.
- Inselberg A.* 1998. Visualization multi-dimensional structure using parallel coordinates // *Computer Science and Statistics: Proceedings of the 25 Symposium of the Interface*. P. / 175–190.
- Kaufmann A.* 1975. Introduction to the Theory of Fuzzy Subsets. Volume I // Academic Press. — 416 p.
- Keilis-Borok V.I.* 1996. Intermediate-term earthquake prediction // *Proc. Nat. l. Acad. Sci. USA*. V. 93. April P. 3748–3755.
- Keilis-Borok V.I., Soloviev A.A.* editors. 2003. Nonlinear dynamics of the lithosphere and earthquake prediction // Springer, 2003, 337p.
- Klosgen W., Zytkow J.M.* editors. 2002. Handbook of Data mining and Knowledge Discovery // Oxford university press. — 908 p.
- Kraak M., Brown F.* 2001. Web Cartography // Taylor & Francis. — 214 p.
- Kullback S., Leibler R.A.* 1951. On information and sufficiency // *Ann. Math. Statist.* V. 22. P. 79–86.
- Larionov V.I. et al.* 2000. Real Time Damage and Losses Assessment due to Strong Earthquakes // *Proc. XXVII ESC General Assembly*, Lisbon, Portugal.
- MacEachren A.M.* 1994. Visualization in modern cartography: setting the agenda // *Visualisation in Modern Cartography* (NY: Elsevier Science Inc). P. 1–12.
- MacEachren A.M., Kraak M.-J.* 1997. Exploratory cartographic visualization: advancing the agenda». *Computers and Geosciences*, 23 (4). P. 335–344.

- MacEachren A.M., Kraak M.-J.* editors. 2001. Cartography and Geographic Information Science // Special Issue «Research Challenges in Geovisualization» Edited by V. 28 (1).
- Malczewski J.* 1999. GIS and Multicriteria decision analysis // Jon Willey & Sons. INC. — 389 p.
- Miller H.J., Han J.* editors. 2001. Geographical data mining and knowledge discovery // Taylor & Francis. — 367 p.
- Openshow S., Openshow C.* 1997. Artificial Intelligence in Geography. John Wiley & Sons. — 329 p.
- Osher B.V.* 1996. Statistical estimation of maximum magnitude and its uncertainty from the catalogue including magnitude errors // in V. Schenk (ed.), Earthquake hazard and risk, Kluwer Academic Publishers, Netherlands. P. 25–38.
- Philip H., Cisternas A., Gvishiani F., Gorshkov A.* 1989. The Caucasus -an actual example of the initial stages of continental collision // Tectonophysics. V. 161. P. 1–21.
- Pyle D.* 1999. Data preparation for data mining // Morgan Kaufman Published, Inc., San Francisco, California. — 540 p.
- Schenk V., Gitis V., Schenkova Z. et al.* 1990. Maximum Earthquake Assessment and the GEO 1.2 Expert System: Example of the Central Europe // Proc. of the XXII General Assembly European Seismol. Commission, Barcelona, Generalitat de Catalunya, V. II. P. 571–578.
- Schenk V., Gitis V.G., Schenkova Z. et al.* 1991. Maximum Earthquake Prediction in Central Europe Given by the GEO 1.2 Expert System // Proceedings of 4th Intern. Conference on Seismic Zonation, Stanford, Vol. III. P. 83–91.
- Schenk V., Gitis V.G., Schenkova Z.* 1992. Aspects of Artificial Intelligence Techniques in Earthquake Hazard Assessment and Seismic Risk Mitigation // CONSEIL DE L'EUROPE. V. 6. Proceedings of the Workshop on Application of Artificial Intelligence Techniques in Seismology and Engineering Seismology, Luxembourg, 23–25 March. P. 145–152.
- Schenk V., Schenkova Z., Gitis V.G.* 1994. Characteristic Features of Geonomic Forecasting Functions for the Maximum Possible Earthquake // Natural Hazards. 10. P. 97–115.
- Sibson R.H.* 1974. Frictional constraints on thrust, strike slip and normal faults. Nature. V. 249. P. 542–544.
- Theus M.* 2002. Highly Multivariate Interaction Techniques // in Handbook of Data mining and Knowledge Discovery. W. Klossgen, J.M. Zytkow, editors. Oxford university press. P. 233–241.
- Trifunac M.D., Brady A.G.* 1975. On the correlation of seismic intensity with peaks of recorded strong ground motion // Bull. Seismol. Soc. Amer., 65. P. 139–162.
- Vainshtok A., Gitis V., Andrienko G. et al.* 1998. RESOURCE: knowledge based GIS on oil and gas resource monitoring // Proceedings of the 4<sup>th</sup> EC-GIS Workshop, Hungary, Joint Research Centre of European Commission. P. 141–145.
- Wills G.J., Kein D.* 2002. Interactive Statistical Graphics // in Handbook of Data mining and Knowledge Discovery. W. Klossgen, J.M. Zytkow, editors. Oxford university press. P. 226–232.

- 
- Worrall L.* 1991. Spatial analysis and spatial policy using geographical information systems. — London: Belhaven Press.
- Zhang Guomin, Zhang Zhaocheng.* 1992. The study of multidisciplinary earthquake prediction in China // Journal of Earthquake Prediction Res. V. 1, № 1. P. 71–86.
- Zhang Zhaocheng, Zheng Dalin, Luo Yongsheng and Jia Qing.* 1992. Studies on earthquake precursors and the multidisciplinary earthquake prediction in China mainland. Journal of Earthquake Prediction Research. V. 1. № 2. P. 191–205.
- Zhuravlev Yu.I.* 1998. An algebraic approach to recognition or classification problems // Pattern Recognition and Image Analysis. M., № 8 (10), 59–100.
- Zschau J. et al.* 1996. SEISMOLAP: A new approach to prediction // Proceeding of the International Conference on Earthquake prediction: state of art. Council of Europe, Strasbourg. P. 444–453.

Gitis V.G., Ermakov B.V.

**FUNDAMENTALS OF SPATIO-TEMPORAL FORECASTING  
IN GEOINFORMATICS**

*FIZMATLIT*

Moscow, 2004, 256 pages

The book lays foundations of geoinformation approach to the study of spatio-temporal processes and phenomena. The purpose of the book is to supply the experts with a guide, both for the development of analytical GIS and for creation of the GIS applications intended to solve spatio-temporal forecast problems in basic research and applied subject domains. The book combine the methodology, techniques and algorithms for geoinformation analysis and forecasting, and offers examples of solutions in seismology, geophysics, geology, ecology and economy. The book is based on unique results obtained in the Section for Geoinformation Technologies and Systems of the Institute for Information Transmission Problems.

The book is intended for scientists and experts in the field of geoinformatics, analysis of spatio-temporal data and Earth sciences and can be used as a manual for universities and other higher educational institutions.

## CONTENTS

Foreword

### Part 1. Principles of geoinformation forecasting

#### 1. Introduction to geoinformation analysis

- 1.1. Geographical information
- 1.2. Problem types
- 1.3. Analytical GIS tools

#### 2. Methodology of spatio-temporal forecasting

- 2.1. Introduction
- 2.2. Solution quality criteria of problems with incomplete information
- 2.3. Information model of problem solution under uncertainty conditions
- 2.4. Elements for solving the spatio-temporal forecast problems
- 2.5. Components of the computer information modeling environment

### Part 2. Spatial forecasting

#### 3. Forecasting of stationary properties of geological environment

- 3.1. Introduction
- 3.2. Design technique of spatial forecast information model
- 3.3. Analytical GIS GEO 2.5

#### 4. Inference of empirical relationships under uncertainty

- 4.1. Introduction
- 4.2. Interval expert evaluations method
- 4.3. Score expert evaluations method
- 4.1. L likelihood ratio approximation method

#### 5. Spatial forecasting of expected maximal earthquake energies

- 5.1. Introduction
- 5.2. Basic approaches to the maximal earthquake energy forecasting
- 5.3. Composition of the forecast map of expected maximal earthquake magnitudes for the Caucasus region

#### 6. Development of the forecast maps of the oil-and-gas content

- 6.1. Introduction
- 6.2. Regional separated forecast of the oil-and-gas content for the West Siberian platform
- 6.3. Forecast of the oil-and-gas accumulations based on electrical and shooting data (Hungary)
- 6.4. Forecast of hydrocarbon strength areas in an oil-field (East China)

#### 7. Development of the forecast maps of inorganic resources

- 7.1. Introduction
- 7.2. Regional forecast of the plumbum-zinc deposits in Bulgaria
- 7.3. Regional forecast of the copper deposits in Bulgaria

## **8. Spatial forecast based on expert knowledge**

- 8.1. Introduction
- 8.2. Local oil-and-gas forecast
- 8.3. Estimation of urban soil stability with respect to dynamic effects
- 8.4. Estimation of earthquake consequences
- 8.5. Forecast of spatial process under local interaction

### **Part 3. Spatio-temporal forecast**

## **9. Spatio-temporal analysis of processes preceding a geological catastrophe**

- 9.1. Introduction
- 9.2. Problem domain search technique of detection of spatio-temporal precursors of geological catastrophes
- 9.3. Basic features of GIS GeoTime
- 9.4. Techniques for anomaly detection
- 9.5. Detection of spatio-temporal anomalies in dynamic fields

## **10. Analysis of spatio-temporal earthquake preparation processes**

- 10.1. Introduction
- 10.2. Analysis of earthquake preparation process based on the catalogue
- 10.3. Analysis of earthquake preparation process based on the geomonitoring data

### **Part 4. Web-GIS forecasting and analysis**

## **11. Network analytical geoinformation technologies and systems**

- 11.1. Introduction
- 11.2. Users of analytical web-GISs
- 11.3. Web-GIS schemes
- 11.4. Inference and explanation under network analysis of GI
- 11.5. Plausible inference techniques in GIS GeoProcessor and COMPASS
- 11.6. Web-GIS GeoProcessor and COMPASS

## **12. Comprehensive analysis in web-GISs**

- 12.1. Introduction
- 12.2. Assessment of earthquake consequences
- 12.3. Seismic hazard assessment of geographical objects by the earthquake catalogue
- 12.4. Seismic sources zone delineation

References