

А.В.КАЖДАН О.И.ГУСЬКОВ

МАТЕМАТИЧЕСКИЕ  
МЕТОДЫ  
В ГЕОЛОГИИ



ВЫСШЕЕ  
ОБРАЗОВАНИЕ

**А.Б.КАЖДАН О.И.ГУСЬКОВ**

# **МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОЛОГИИ**

Допущено Государственным комитетом СССР  
по народному образованию в качестве учебника  
для студентов геологических специальностей вузов



**МОСКВА "НЕДРА" 1990**

ББК 26.3

К 12

УДК 518:55

Рецензенты: кафедра месторождений полезных ископаемых и их разведки  
Университета дружбы народов им. П. Лумумбы, д-р геол.-минер. наук  
*A. B. Каждан*

**Каждан А. Б., Гуськов О. И.**

К 12 Математические методы в геологии: Учебник для  
вузов.—М.: Недра, 1990.—251 с.: ил.  
ISBN 5-247-00857-X

Изложены основы геолого-математического моделирования состава, строения и других свойств горных пород и полезных ископаемых в иедрах применительно к решению важнейших геологических задач. Рассмотрены сущность и условия применения одномерных, двумерных и многомерных статистических моделей, методы математического описания пространственных геологических закономерностей, факторы, определяющие выбор и эффективность использования математических методов в геологии, с учетом возможностей ЭВМ. Большое внимание уделено использованию ЭВМ для моделирования геологических процессов, эффективности применения ЭВМ в различных областях геологии.

Для студентов геологических специальностей вузов.

К 1804010000—433  
043(01)—90 36—90

ББК 26.3

ISBN 5-247-00857-X

© А. Б. Каждан, О. И. Гуськов, 1990

## ПРЕДИСЛОВИЕ

---

Курс «Основы геолого-математического моделирования» изучается студентами геологоразведочного факультета МГРИ уже в течение почти двадцати лет. В настоящее время он имеет название «Математические методы в геологии» и включен в типовой учебный план подготовки инженеров-геологов по специальности «Геологическая съемка, поиски и разведка месторождений полезных ископаемых».

Согласно учебному плану данный курс изучается на пятом семестре, после завершения курсов «Высшей математики» и «Вычислительной техники с основами программирования», т. е. после освоения студентами элементарных навыков обращения с ЭВМ. К началу его прохождения студенты должны овладеть знаниями по общей геологии, палеонтологии, минералогии, частично по структурной геологии и петрографии. Это обстоятельство определяет круг возможных геологических задач и конкретных примеров, рассматриваемых в отдельных разделах курса. Однако поскольку полученные знания должны использоваться студентами и в последующие годы, при прохождении специальных дисциплин, здесь рассматриваются и некоторые примеры из области поисков и разведок месторождений полезных ископаемых.

В основу учебника положены лекции по данному курсу, которые были прочитаны в 1973—1975 гг. профессором А. Б. Кажданом, а начиная с 1976 года — доцентом О. И. Гуськовым. При составлении учебника использованы методические разработки П. И. Кушнарева, В. И. Пахомова, В. Г. Будрика, Е. А. Сидоркова, а также доцента Иркутского политехнического института А. А. Шиманского, который является соавтором четвертой главы учебника.

## ВВЕДЕНИЕ

---

На современном этапе развития естественных наук, под влиянием научно-технического прогресса происходят существенные изменения методов научных экспериментов, анализа и обобщения получаемых результатов. Этому способствуют не только расширяющиеся возможности фундаментальных наук, но также бурное развитие электронно-вычислительной техники и комплексной автоматизации самых разнообразных видов человеческой деятельности. В последние десятилетия наблюдается глубокое проникновение математических методов исследования во все отрасли естественных наук, что способствовало исключительным успехам некоторых из них, например биологии, метеорологии и др. Для успешного развития геологических наук необходимо также использовать полный арсенал существующих прогрессивных научных и технических средств, включая математические методы и ЭВМ.

Современная геология уже не может ограничиваться изучением лишь качественных сторон явлений и процессов, а должна выявлять их количественные характеристики, обеспечивая тем самым более высокий научный уровень исследования земных недр. Необходимость применять математические методы обработки, анализа и обобщения данных все острее ощущается не только при прогнозировании, поисках, разведках и оценках месторождений полезных ископаемых, но и вообще при проведении любых геологических исследований. Так, например, палеонтологические, стратиграфические, структурно-геологические, литологические, петрографические, минералогические, геохимические, геоморфологические и другие геологические исследования, которые в недавнем прошлом ограничивались чисто описательными приемами, требуют в настоящее время использования меры и числа.

Ежегодно в геологических организациях страны накапливается колоссальный эмпирический материал — миллионы количественных определений химического состава различных минералов и их агрегатов, химического и минерального составов горных пород и полезных ископаемых, их физических, горно-технологических и других свойств, требующих применения ЭВМ для обработки и обобщений с целью более полного извлечения содержащейся в них полезной информации.

Острую необходимость внедрения математических методов в практику геологоразведочных работ испытывают производственные геологические организации в связи с возросшими требованиями промышленности к конкретности и достоверности геологоразведочных данных. Так, в соответствии с действующими положениями количественные оценки прогнозных ресурсов полезных ископаемых должны быть обоснованы уже по данным геологических съемок с уточнениями цифр прогнозных ресурсов (а затем запасов) на каждой из последующих стадий геологоразведочных работ.

Резкое увеличение количественной информации, получаемой в процессе геологической съемки, поисков и разведки полезных ископаемых, вызвало необходимость разработки принципиально новых способов ее хранения, поиска, обработки и анализа с помощью ЭВМ. К настоящему времени в организациях Министерства геологии СССР

и других министерств уже накоплен опыт машинного хранения, поиска и обработки геологической, геофизической и геохимической информации, однако геологам предстоит еще огромный труд по созданию автоматизированных отраслевых систем управления геологоразведочным процессом на всех его иерархических звеньях — от геологического отряда и партии до геологической службы страны.

В основных направлениях экономического и социального развития СССР на 1986—1990 годы и на период до 2000 года сформулированы задачи интенсификации народного хозяйства на основе научно-технического прогресса, компьютеризации и комплексной автоматизации производства. В частности, в них принято решение о расширении комплексных научных исследований строения и эволюции Земной коры, результаты которых позволят обеспечить глубокие качественные изменения в производственных силах. Для достижения этих целей предусмотрено планомерное техническое переоснащение геологоразведочной службы, будут созданы и внедрены специализированные высокомеханизированные и автоматизированные средства поисков и разведки полезных ископаемых, в том числе высокопроизводительные ЭВМ всех классов — от суперЭВМ до персональных компьютеров, вычислительные центры, интегрированные банки данных и развитые сети обработки и передачи информации.

С учетом все возрастающей роли математических методов и широкого использования ЭВМ во всех отраслях геологической науки становится очевидным значение данной дисциплины в образовании современного геолога, специалиста по геологической съемке, поискам и разведке месторождений полезных ископаемых.

Курс «Математические методы в геологии» имеет своей целью ознакомить студентов с особенностями геологических образований и процессов, как объектов математического изучения и моделирования, со спецификой геологических задач, решаемых с помощью математических методов, с возможностями различных математических методов и факторами, влияющими на эффективность их использования.

*Задачи изучения дисциплины определяются требованиями геологоразведочной службы страны и квалификационной характеристикой молодого специалиста. После изучения курса он должен:*

- 1) знать основные принципы геолого-математического моделирования, главные типы моделей и особенности их применения в различных областях геологии;
- 2) владеть методами математической обработки геологической, геохимической и геофизической информации;
- 3) уметь формулировать геологические задачи в виде, пригодном для их решения математическими методами, и выбирать наиболее эффективные методы их решения.

К настоящему времени накоплен значительный опыт применения математических методов в геологии.

Еще в 30-х годах прошлого столетия Ч. Ляйель использовал выборочный статистический метод оценки количественных соотношений раковин ископаемых и современных видов в разрезах третичных отложений с целью их стратиграфического расчленения. К эоцену он отнес отложения, содержащие менее 3,5% ныне живущих видов моллюсков, к миоцену и плиоцену — соответственно от 3,5 до 17% и от 17 до 50%, а к верхнему плиоцену — более 50%.

В конце прошлого века геологи начали описывать осадки и осадочные породы с учетом состава, размеров и форм отдельных зерен

или минералов, а также их количественных соотношений с применением вероятностно-статистических методов. Примерно в те же годы Ф. Ю. Левинсон-Лессингом, П. Ниггли и другими петрографами по результатам статистической обработки силикатных химических анализов и микроскопических исследований шлифов были установлены границы семейств различных изверженных пород, парагенетические ассоциации породообразующих минералов, а также закономерные связи между химическим и минеральным составами пород.

Попытки применения математического (в том числе и статистического) анализа к решению задач поисков и разведки полезных ископаемых имеются в трудах И. А. Корзухина (1908), С. Ю. Деборжинского (1910—1911), Б. И. Бокия (1914) и др. В частности, С. Ю. Деборжинским были предложены решение задачи об отыскании средних элементов залегания пласта, путем аппроксимации его плоскостью по способу наименьших квадратов, а также способ оценки вероятности обнаружения тела полезного ископаемого заданной формы и размеров равномерной сетью поисково-разведочных выработок.

В 30-х годах нашего столетия статистические методы постепенно проникли во многие отрасли геологии, особенно в геохимию, минералогию, палеонтологию, петрографию, литологию и опробование полезных ископаемых. Их использование преследовало главным образом описательные цели и ограничивалось статистической обработкой геологических наблюдений с целью разумной свертки информации или проверки каких-либо гипотез (выявление функций статистических распределений, оценки математических ожиданий, дисперсий, коэффициентов корреляций или других статистических характеристик оцениваемых совокупностей). Так, например, М. А. Журавскому (1932) принадлежит первая работа по применению вероятностных характеристик результатов количественного минералогического анализа шлифов под микроскопом, а С. Н. Иванову (1936) — методика определения густоты сети опробования методами вариационной статистики.

В середине XX в. математические методы проникли практически во все отрасли геологии, чему в значительной степени способствовало появление электронно-вычислительной техники. Кроме теории вероятностей и математической статистики случайных величин, в теорию и практику геологических исследований стали постепенно проникать методы многомерного статистического анализа, теории случайных функций и гармонического анализа, теории множеств, линейной алгебры и других разделов математики, в связи с чем расширился круг задач, решаемых с применением математических методов. Помимо статистической обработки наблюдений, они стали применяться для целей геологической классификации, а также математического моделирования геологических объектов и процессов. К этому периоду относятся широкие статистические исследования по изучению закономерностей распределения содержаний элементов в горных породах, проводившиеся Н. К. Разумовским, Д. А. Родионовым и Л. Аренсом; исследования А. Б. Вистелиуса по корреляции немых красноцветных толщ и анализу пористости нефтеносных пород с помощью аппарата случайных функций; работы В. Н. Бондаренко по статистическому изучению петрохимических особенностей вулканических пород, А. В. Канцеля — по математическому моделированию процессов рудоотложения, Б. М. Роненсона — по корреляции немых метаморфических толщ при крупномасштабном геологическом картировании, М. Ф. Миличинка и В. П. Бухарцева — по выявлению корреляционными методами

структурных соотношений геологических комплексов Волго-Уральской области и многие другие работы. В послевоенные годы во многих исследованиях отечественных и иностранных геологов свойства геологических образований стали рассматриваться как пространственные переменные. При описании геофизических полей их характеристики подразделялись на региональные и локальные составляющие, а при оценках и подсчетах запасов полезных ископаемых — на пространственно коррелированные и случайные (Д. А. Казаковский, П. Л. Каллистов, З. Д. Низгурецкий, Д. Г. Криге, Ж. Матерон и др.).

Пространственные статистические модели широко использовались при анализе структурных и фациальных карт (У. Крамбейн и Ф. Грейбилл, Р. Миллер и Д. Кан), а методы картирования векторных полей — в работах Ф. П. Агдеберга, Д. А. Родионова и др.

На современном этапе развития естественных наук, к числу которых относится геология, одним из показателей их прогресса служит степень использования математических методов при проведении научных экспериментов, анализе и обобщении данных. Среди многочисленных отраслей естественных наук и народного хозяйства вряд ли найдутся такие отрасли, в которых потенциальная необходимость применения математических методов и ЭВМ была бы столь широкой, как в геологии. Уже в настоящее время, в связи с резко возросшими возможностями электронно-вычислительной техники, применение математических методов в геологии способствует совершенствованию теоретических, логически непротиворечивых представлений и получению более строгих практических выводов по совокупности исходных теоретических данных.

Внедрение математики в геологию и практику геологоразведочных работ происходит по нескольким взаимосвязанным направлениям, в числе которых главное значение имеют:

- 1) статистические методы обработки числовых результатов наблюдений;
- 2) методы математического моделирования геологических объектов;
- 3) методы классификации геологических объектов.

В настоящее время для решения геологических задач применяется весьма широкий набор математических методов с использованием многих разделов современной математики. Среди них отчетливо доминируют методы теории вероятностей и математической статистики. Помимо статистической обработки исходных геологических данных наибольшие успехи достигнуты в применении математических методов для решения задач прогнозирования, поисков, разведки и оценки месторождений полезных ископаемых, а также для информационного обеспечения перечисленных геологических задач.

*Перспективы дальнейшего развития математических методов в геологии* зависят от общего прогресса геологических наук и тесно связаны с компьютеризацией и комплексной автоматизацией геологической отрасли. Несомненно, что создание и освоение новых поколений ЭВМ всех классов обеспечат мощный импульс для расширения возможностей известных и появления новых математических методов в геологии, а также для повышения их эффективности.

Весьма перспективно использование непараметрических методов описания геологических признаков и способов математического описания пространственных переменных. По мере совершенствования математических методов в геологии, в число важнейших направлений

их использования войдут методы математического моделирования геологических процессов и методы оптимизации процессов сбора, хранения, поиска, обработки и передачи геологической информации.

Методы математического моделирования геологических процессов входят в задачу специальной научной дисциплины «Математической геологии», содержание и научные основы которой сформулированы А. Б. Вистелиусом (1944—1969).

Методы оптимизации, сбора, хранения, поиска, обработки и передачи геологической информации основаны на автоматизации этих процессов с помощью современных ЭВМ. Предпосылкой создания эффективных автоматизированных информационно-поисковых систем является разработка формального геологического языка для создания банков данных. Основой такого языка служат представления о структуре многостадийного геологоразведочного процесса, как о системе взаимосвязанных и взаимообусловленных операций, создающих «информационный контур» геологоразведочного производства.

# ГЛАВА 1

## ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ГЕОЛОГИИ

---

### 1.1. СПЕЦИФИКА ГЕОЛОГИЧЕСКИХ ОБРАЗОВАНИЙ И ПРОЦЕССОВ КАК ОБЪЕКТОВ ИЗУЧЕНИЯ

Геологические процессы и образования обладают специфическими особенностями, в значительной мере определяющими методику их изучения:

— геологические процессы представляют собой совокупность физических, химических и биологических природных явлений, между которыми существуют сложные причинно-следственные связи, поэтому свойства геологических образований зависят от множества факторов, характеризуются сильной изменчивостью, а сами объекты, как правило, имеют весьма сложное строение;

— геологические процессы длительны, а геологические образования имеют значительные размеры и скрыты в недрах, что исключает возможность их полного и всестороннего изучения путем непосредственного наблюдения.

Сложность, неоднородность объектов изучения наук о Земле заставляет рассматривать их как *природные системы*. Под *системой* понимается совокупность элементов, находящихся в отношениях и связях друг с другом, которая образует определенную целостность, единство.

Геологические процессы относятся к классу *динамичных систем*, т. е. систем, изменяющих свое состояние во времени, а геологические образования, ввиду медленного протекания геологических процессов, в большинстве случаев могут рассматриваться как *статичные системы*, свойства которых постоянны во времени.

В качестве *элементов динамичных геологических систем* выступают либо отдельные факторы (параметры), влияющие на ход геологического процесса, либо относительно простые природные явления (процессы), которые рассматриваются как составные части более сложного процесса. Так, элементами (параметрами) процесса гидротермального рудообразования будут температура, давление, химический состав гидротермальных растворов, физические свойства и состав вмещающих пород и т. д. А в процессе формирования осадочных рудных месторождений можно выделить следующие последовательные или совмещенные во времени «элементарные» процессы:

1) мобилизацию (переход в подвижное состояние) полезных минералов или химических элементов за счет разрушения пород в области питания бассейна осадконакопления;

2) перенос мобилизованного вещества различными агентами;

3) распределение привнесенного вещества в бассейне осадконакопления;

4) преобразование и перераспределение рудного вещества в результате превращения осадка в горную породу и т. д.

Элементами горных пород и руд являются слагающие их минеральные агрегаты и минералы, а в толщах осадочных пород — это различные по составу горизонты и слои.

В большинстве случаев каждый элемент изучаемой геологической системы, в свою очередь, сам является системой, а любая система может рассматриваться как элемент более сложной системы. Каждый из процессов, рассмотренных выше в качестве элементов осадочного рудогенеза, сам по себе сложен и зависит от множества факторов. Так, например, процесс разрушения пород является совокупностью разнообразных и довольно сложных физических, химических, биологических и других процессов, ход которых зависит от состава пород, климата, рельефа местности и других факторов.

В сложных геологических образованиях обычно можно выделить несколько разномасштабных уровней. Так, в учении о поисках и разведке месторождений полезных ископаемых рассматривается двенадцать разномасштабных уровней минерализованных участков недр. Каждый из них определяется совокупностью минеральных скоплений, которые на данной стадии изученности рассматриваются как элементы неоднородности. В зависимости от масштабов исследований это могут быть: элементы строения кристаллических решеток минералов; зерна или обломки полезных минералов; агрегаты минеральных зерен; пространственно обособленные локальные скопления полезных ископаемых — их тела, залежи, зоны, месторождения; более крупные таксономические единицы — рудные поля, узлы и районы, металлогенические области и провинции. Таким образом, представления о сложности строения природных скоплений полезных ископаемых зависят от детальности проведенных исследований. Строение (структуря) геологических объектов определяется формами, размерами и взаиморасположением элементов неоднородности каждого структурного уровня, а также числом выявленных структурных уровней.

Такой подход не только обеспечивает возможность количественного описания особенностей строения и состава изучаемых объектов по совокупности признаков, наблюденных на данной стадии исследования, но и позволяет количественно описать новые представления об их строении, получаемые по мере детализации наблюдений. *Системный подход* выражает один из важнейших методологических принципов изучения материального мира, который широко используется в диалектической теории познания.

Свойства геологических образований и параметры геологических процессов могут меняться как скачкообразно, так и плавно, постепенно. Поэтому выделить отдельные элементы в геологических системах не всегда легко. Чаще всего такое разделение производится в известной степени условно, поскольку элементы не имеют четких границ. Взаимозависимость различных геологических процессов приводит к тому, что и геологические системы обычно не имеют реальных границ. Поэтому их следует отнести к классу *открытых систем*, при всестороннем изучении которых необходимо учитывать не только

связи между составляющими ее элементами, но и связи между самой системой и окружающей средой. Вместе с тем, при решении большинства конкретных практических задач внешними связями можно пренебречь и рассматривать геологические объекты как *закрытые системы*.

У. Крамбейн такие системы называет *геологическими совокупностями*. Примерами последних могут служить геохимические и геофизические поля, совокупности интрузий определенного возраста или состава, совокупности шлифов определенной породы, совокупности минералов в шлиховой пробе и т. п.

Как правило, любая геологическая совокупность может быть разделена по какому-либо признаку на множество самостоятельных совокупностей. Например, совокупность эфузивных пород п-ова Камчатка подразделяется на ряд совокупностей по возрасту, составу или приуроченности к конкретным вулканическим аппаратам. Вулканогенные породы одного вулканического аппарата, в свою очередь, могут рассматриваться как совокупности отдельных лавовых потоков.

Принцип выделения геологической совокупности зависит от поставленной задачи. Результатами изучения геологической совокупности можно пользоваться только в ее границах, а правомерность перенесения полученных выводов на более широкий или более узкий круг объектов или явлений должна обосновываться специальными исследованиями.

Так, например, при изучении россыпного месторождения золота геологической совокупностью для решения задач, связанных с его промышленной оценкой — подсчетом запасов, определением среднего содержания золота в песках и т. п., будут только те объемы золотоносных песков, где содержание золота превышает значение, обеспечивающее рентабельность отработки. Одновременно с этим для определения области сноса золота и поисков его коренных источников проводится изучение формы и размеров отдельных золотин в разных участках россыпи и за ее пределами. Объектом изучения при этом являются все аллювиальные отложения. Очевидно, что среднее содержание золота, установленное в контурах промышленных залежей, нельзя распространять на весь объем потенциально золотоносных отложений, а результаты исследования размеров золотин по всему объему аллювиальных отложений нельзя использовать для характеристики гранулометрического состава золота в промышленной россыпи.

## 1.2. МЕТОДЫ ИЗУЧЕНИЯ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

Недоступность геологических образований и процессов для непосредственного наблюдения обусловила широкое распространение в практике геологических исследований *выборочных методов изучения* с помощью естественных и искусственных обнажений, в пределах которых отбираются образцы и пробы для различных исследований и анализов. Локальные площади наблюдений и отбираемые пробы несопоставимо

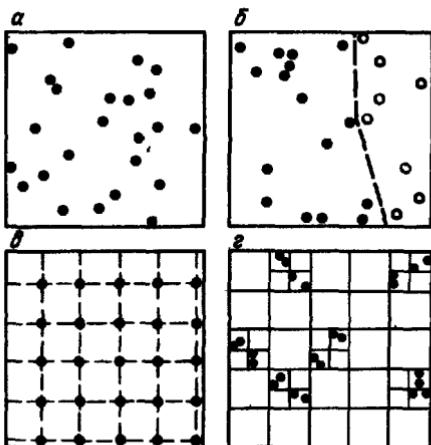


Рис.1. Системы расположения точек наблюдений. По У. Крамбейну:  
а—случайное; б—случайное стратифицированное (объект разделен на две подсовокупности); в—равномерное; г—многостадийное

малы, по сравнению с площадями и объемами недр, на которые распространяются наблюденные данные. В связи с этим возникают проблемы пространственного размещения пунктов локальных наблюдений, систематизации выборочных данных и их распространения на прилегающие объемы недр.

О свойствах всей геологической совокупности геолог судит по какой-то ее части, доступной для наблюдения и опробования, которую М. Розенфельд предложил назвать *опробуемой совокупностью*. Степень соответствия свойств опробуемой совокупности и изучаемой геологической совокупности зависит от расположения, густоты и общего количества точек наблюдений, а также от размеров, ориентировки, формы, объема отбираемых проб или способа измерения данного свойства.

Выделяют три основные системы расположения точек наблюдения: равномерное, случайное и многостадийное опробование (рис. 1).

Наибольшее распространение имеет *равномерное опробование*, при котором точки наблюдений в плоскости изучаемого объекта распределяются по правильной геометрической сети (см. рис. 1, в). Такое опробование позволяет с одинаковой детальностью изучить все части изучаемого объекта, поэтому оно является основным при поисках и разведке месторождений полезных ископаемых.

*Случайное опробование* (см. рис. 1, а, б) обычно применяется в тех случаях, когда исследователя не интересуют закономерности изменения изучаемого свойства в пространстве или достоверно известно, что таких закономерностей нет, а также тогда, когда невозможно или затруднительно создать сеть равномерных наблюдений. Так, например, при геологическом картировании в гористой местности пробы берутся преимущественно из естественных обнажений, размещение которых в пределах изучаемой площади близко к случайному. Случайный способ рекомендуется также при отборе проб для контрольных анализов.

*Многостадийное опробование* (см. рис. 1, г) применяется для изучения свойств сложных геологических объектов на разных масштабных уровнях их строения. Для этого объект разделяется на участки,

соответствующие элементам его неоднородности, в которых, в свою очередь, выделяются более мелкие элементы неоднородности и т. д. В пределах каждого участка опробуется только определенная часть элементарных участков более высокого порядка. За счет этого общее количество наблюдений при многостадийном опробовании существенно сокращается по сравнению с равномерным. Многостадийное опробование применяется при составлении ландшафтных карт. Сначала по результатам дешифрирования космоснимков масштабов 1:500 000—1:200 000 производится районирование территории по типам ландшафтов, затем в пределах каждого из этих типов выделяются ландшафты водоразделов, склонов, речных долин и т. п.

Для определения границ элементарных ландшафтов используются аэрофотоснимки масштаба 1:50 000, а их основные характеристики — состав и мощность рыхлых отложений, тип почвы и растительности — оцениваются путем изучения так называемых ключевых участков, т. е. относительно небольших по площади участков, где проявлены все особенности данного ландшафта.

Выборочный характер экспериментальных числовых данных о свойствах геологических объектов исключает возможность полного их совпадения с истинными количественными характеристиками изучаемых свойств. При проведении любых замеров или изучении проб неизбежны технические погрешности, определяющие значения *погрешностей измерений*, а при распространении результатов экспериментальных данных или анализов проб на окружающие объемы недр возникают *погрешности геологических аналогий*.

При хорошем качестве геологических исследований значения погрешностей измерения изучаемых свойств геологических объектов невелики и имеют случайный характер. Погрешности отдельного замера, как правило, не превышают несколько относительных процентов. Погрешности анализов отдельных проб могут достигать 25—30%, поскольку они складываются из технических погрешностей их отбора, обработки и анализов, однако значения средних содержаний компонентов, вычисленные по ряду проб, значительно ниже и редко превышают 5—10%. При низком качестве геологических исследований значения случайных технических погрешностей возрастают и наряду с ними часто появляются систематические погрешности, свидетельствующие о существенных дефектах в методике проведения замеров или анализов. Выявление и устранение этих дефектов способствуют уменьшению систематических погрешностей замеров и анализов проб, а повышение качества геологических исследований способствует снижению уровней случайных погрешностей измерения.

*Погрешности геологических аналогий* возникают при распространении результатов измерений или анализов проб, полученных по отдельным обнажениям, на прилегающие к ним объемы недр. Из-за отсутствия сведений о значениях изучаемых признаков в этих объемах обычно допускается их линейная интерполяция между смежными обнажениями, хотя заведомо известно, что значения признаков изменяются в недрах не по линейным, а по значительно более сложным законам.

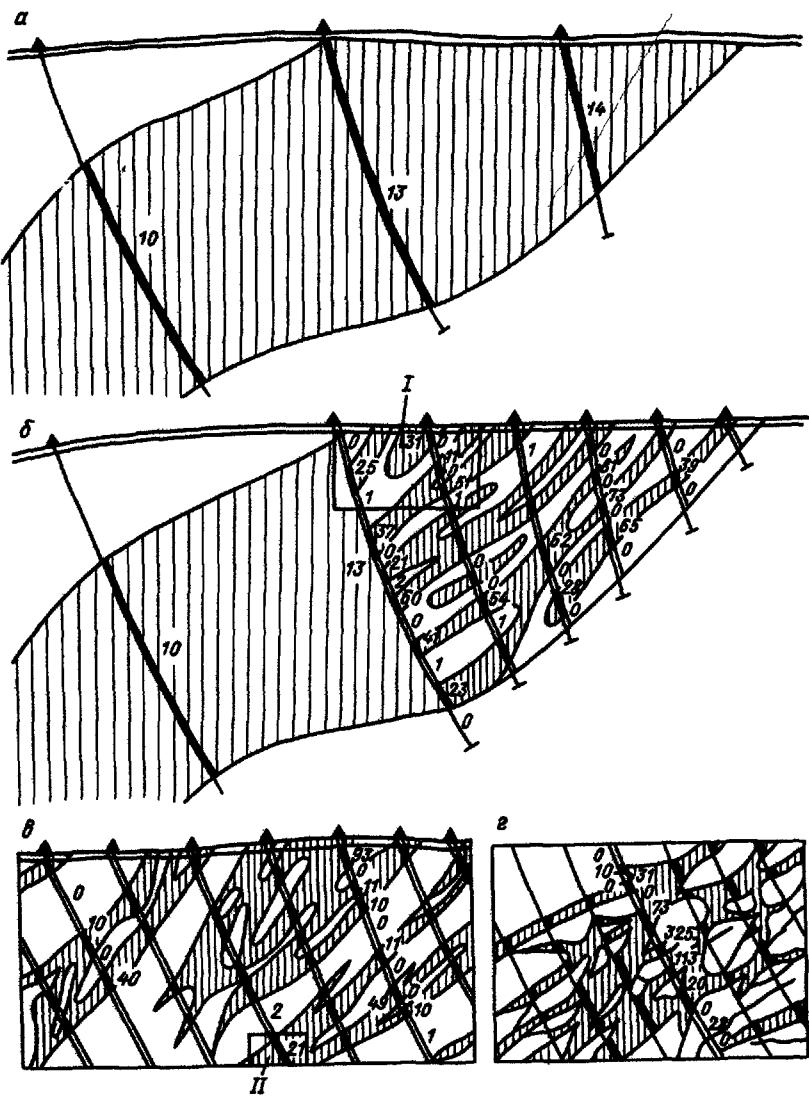


Рис. 2. Уровни строения золоторудного месторождения (цифрами указаны содержания золота в условных единицах):

*a*—по сквозным разведочным пересечениям; *b*—по секционным пробам при ступенчатом разведочном сети; *c*—по участку детализации I; *d*—по участку детализации II

По редкой сети наблюдений невозможно выявить элементы неоднородности высоких порядков, поэтому форма и внутреннее строение геологических объектов представляются относительно простыми. Сгущение сети, как правило, приводит к усложнению контуров геологических образований и усложнению наших представлений об их внутренней структуре (рис. 2).

Заметное влияние на представление об изменчивости свойств

природных геологических образований оказывают также размеры отбираемых проб. Так, например, при опробовании золоторудной зоны пробами микроскопических размеров изменчивость содержания золота окажется исключительно высокой, так как отдельные весьма редкие пробы будут состоять целиком из золота, некоторые пробы — из золотосодержащих сульфидов или кварца, а в остальных пробах золота вообще не будет. Если же увеличить размеры проб на порядок по сравнению с размерами золотосодержащих минеральных агрегатов, то изменчивость содержаний золота заметно снизится, хотя содержание в отдельных пробах с самородками будет все же резко отличаться от содержаний в остальных пробах. Дальнейшее увеличение размеров проб настолько снижает наблюдаемую изменчивость содержаний в них золота, что оно будет представляться практически равномерным (например, среднее содержание золота в загруженных рудой железнодорожных вагонах).

Зависимость характеристик изменчивости свойств геологических объектов от условий экспериментальных наблюдений свидетельствует о том, что последние не могут объективно отражать их природную изменчивость. Природная изменчивость свойств геологических объектов — это явление столь сложного и многопланового характера, что полное его количественное описание практически невозможно.

В зависимости от целей геологических исследований из всех возможных признаков и видов их изменчивости изучаются только те, которые способствуют решению поставленных задач, и лишь на тех уровнях строения объектов, которые сопоставимы с масштабами исследования.

Поэтому принципиально важно отличать понятия *природной изменчивости* свойств геологических объектов от понятия их *наблюдаемой изменчивости* по результатам приведенных исследований.

Наблюдаемая изменчивость свойств отражает детальность наших представлений об изменчивости реально существующего объекта в зависимости от природной сложности его строения и совокупности экспериментальных наблюдений, положенных в основу ее характеристики.

Количественная характеристика наблюдаемой изменчивости любого свойства геологического объекта лишена практического смысла без указания тех конкретных условий наблюдений, по которым она получена, — в частности, размеров проб и густоты сети наблюдений.

Широкое распространение косвенных и дистанционных методов изучения — аэрокосмических, геофизических и др. — также вносит определенные искажения в наши представления о природных свойствах геологических объектов. Эти методы основаны на использовании ряда зависимостей между интересующими нас свойствами геологических образований и некоторыми легко поддающимися измерению физическими свойствами. Например, породы различного состава часто обладают неодинаковыми магнитными и электрическими свойствами, имеют разную плотность и отражательную способность, что позволяет широко использовать при геологическом картировании космо- и аэрофотоснимки, а также результаты площадных геофизических съемок.

Однако на характер физических полей помимо состава пород оказывают заметное влияние и другие факторы — аппаратурные погрешности, высота полета, состояние атмосферы (при аэрометодах) и т. д. К тому же различия между физическими свойствами некоторых пород не всегда значительны, а зависимость между их составом и физическими свойствами непостоянна. Поэтому результаты косвенных методов изучения часто не поддаются однозначной интерпретации без выборочной заверки прямыми методами.

### 1.3. ХАРАКТЕР ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Многообразие геологических объектов и методов их изучения приводит к тому, что результатом геологических исследований является весьма разнородная по характеру информация — *словесная* (*описательная*), *графическая* (*картографическая*), *цифровая*.

Недоступность геологических объектов для непосредственного наблюдения служит причиной того, что геология, как теоретическая дисциплина, развивалась в условиях практически полного отсутствия экспериментальных данных и на протяжении многих лет считалась чисто описательной наукой.

До недавнего времени геологическая информация имела в основном *качественный характер*, т. е. она заключалась в *словесном описании* и *зарисовках*, в то время как число и мера играли довольно скромную роль, выполняя главным образом иллюстративные функции. Теоретические выводы геологов, основанные на личном опыте и интуиции, отражали не только реальные свойства природных образований и явлений, но и, в определенной степени, субъективные представления авторов. Это привело к тому, что существующие в геологии понятия и определения часто неоднозначны, неконкретны, сформулированы на языке, полном образных выражений, сравнений, аналогий. В геологической литературе имеется несколько десятков определений понятий «минерал», «горная порода», «формация» и более ста определений понятия «фация».

Весьма распространенной формой обобщения знаний о свойствах геологических объектов являются *классификации* и *группировки*. Однако в основу большинства из них положены качественные признаки, причем набор этих признаков и количество групп в классификациях неодинаковы. Например, для разделения изверженных пород по минеральному и химическому составам используется, как минимум, пять различных классификаций, предложенных С. Мишель-Леви, Г. Розенбушем, Ф. Ю. Левинсоном-Лессингом, П. Ниггли и А. Н. Заварицким.

Неоднозначно определенные геологические понятия берутся за основу условных обозначений при составлении *графических геологических документов* — зарисовок, разрезов, планов, карт. В результате этого картографическая геологическая информация также является неоднозначной, и нередко геологические карты, составленные в одном и том же масштабе на одну и ту же территорию, но в разные годы и различными исследователями, существенно отличаются друг от друга.

*Количественная (цифровая)* геологическая информация, объем которой резко возрос в последние годы, также имеет некоторые специфические особенности. Ввиду выборочного метода изучения и сложности геологических объектов она отражает их свойства не полностью, а из-за технических погрешностей измерения — не всегда достаточно точно.

Определенная неоднозначность возникает также за счет того, что некоторые свойства геологических объектов иногда могут быть выражены различными числовыми характеристиками. Так, например, изучение степени окатанности песчаных зерен и галек позволяет судить о характере их транспортировки и расстояниях до источника сноса. Однако в качестве оценки степени окатанности могут быть использованы следующие величины: частное от деления радиуса кривизны самого острого конца песчинки или гальки на ее средний радиус; отношение среднего радиуса максимальных окружностей, описывающих вершины всех углов границы в ее проекции на плоскости, к радиусу наибольшего круга, вписанного в эту проекцию; и т. д.

При изучении полезных ископаемых могут анализироваться валовые содержания химических элементов, содержания их оксидов, сульфидов или других химических соединений, содержания минералов-носителей полезных компонентов или другие количественные показатели качества руд. Для большинства рудных месторождений чаще всего используются содержания химических элементов, для россыпных месторождений — содержания полезных минералов, а для некоторых месторождений — содержания различных соединений металлов, обладающих резко контрастными технологическими свойствами. Так, при переработке оловянных руд значительно легче извлекаются в концентраты оксиды олова по сравнению с сульфидами, в металлургических процессах железных руд силикаты железа не выплавляются, а уходят в шлаки и т. д., поэтому для выбора наиболее подходящего вида числовых измерений прежде всего следует установить, какая из возможных количественных характеристик наиболее полно выражает изменения интересующего нас свойства.

Каждой геологической совокупности может быть поставлен в соответствие набор числовых характеристик, полученных в результате измерения или анализа каких-либо свойств геологических объектов. Такие наборы числовых характеристик называются *выборочными (статистическими) совокупностями*.

Для правильного решения поставленных геологических задач принципиальное значение имеет однозначное и четкое определение соотношений геологической и выборочной совокупностей.

Для определения конкретной геологической совокупности необходимо, прежде всего, установить ее элементарные составляющие (т. е. изучаемые объекты), границы и виды последующих числовых измерений.

Объекты (элементарные составляющие) и границы геологических совокупностей устанавливаются геологом в зависимости от целей и задач исследований. По мнению У. Крамбейна, элементарные

составляющие геологических совокупностей можно разделить на две большие группы: образованные первичными индивидами (объектами) или наборами исходных объектов.

К совокупностям, образованным первичными индивидами (объектами), относятся совокупности ископаемых организмов, минералов в шлихах или шлифах и др. По каждому из таких объектов измеряется одно свойство, несколько свойств или оцениваются средние значения свойств в группировках изучаемых объектов. К совокупностям, образованным наборами исходных объектов, относятся совокупности образцов или проб, по которым определяют физико-химические свойства, их гранулометрический состав, содержания полезных или вредных компонентов и др. В таких наборах свойства каждого исходного объекта не измеряются, а оцениваются средние значения тех или иных свойств в объемах проб или образцов. Отличительной особенностью этой группы совокупностей является зависимость числовых характеристик свойств от размеров и объемов проб.

В геологической практике применяются четыре шкалы измерений: номинальная, порядковая, интервальная и относительная.

*Номинальная шкала* используется для классификации объектов по принципу равенства их свойств. Число в данном случае служит кодом качественных признаков объектов (например, цвета пород: красный — 1, черный — 2 и т. д.).

*Порядковая шкала* измерений применяется в тех случаях, когда изучаемые объекты или пробы можно расположить в некотором порядке в зависимости от изменения какого-либо свойства, однако разницу в изменении этого свойства нельзя оценить количественно. Порядковая шкала используется, например, при визуальной оценке степени метасоматического изменения пород (сильно-, средне-, слабоизмененные породы), для фиксации положения каждого отдельного слоя в толще осадочных пород по порядку от нижнего к верхнему, для записи результатов полукачественного спектрального анализа проб и т. п.

*Интервальная шкала* применяется в тех случаях, когда интервалы перехода от одного класса объектов к другому равны, но не указана точка абсолютного нуля (например, температурная шкала с различными точками нулей — по Цельсию, Реомюру и др.).

*Относительная шкала* используется для измерения свойств геологических объектов в случаях, когда они могут быть оценены количественно. Она является наивысшей шкалой измерений, требует точного знания нулевой точки и равенства отношений применительно к измеряемому количеству. Числами относительной шкалы характеризуются замеры мощностей тел полезных ископаемых, результаты химических анализов проб, замеры элементов залегания пород и др.

## 1.4. МОДЕЛИРОВАНИЕ В ГЕОЛОГИИ

Материальные системы, как объекты изучения, принято разделять на хорошо и плохо организованные.

*Хорошо организованные системы* состоят из ограниченного количества элементов, между которыми существуют строго определенные

и однозначные зависимости. К этим системам можно отнести простейшие химические и физические процессы, механизмы, приборы и т. п. Их свойства и состояния могут быть количественно описаны с помощью законов физики и химии.

К плохо организованным системам относятся сложные природные объекты и явления, на состояние и свойства которых влияет множество факторов различной природы. Типичными плохо организованными системами являются живые организмы и их сообщества, а также большинство объектов, изучаемых науками о Земле. При изучении этих систем в их структуре удается установить лишь отдельные закономерности, т. е. тенденции, не поддающиеся строгому количественному выражению.

Основным методом изучения плохо организованных систем является моделирование, когда непосредственный объект изучения заменяется его упрощенным аналогом — моделью.

По характеру моделей выделяют предметное и знаковое (информационное) моделирование.

Предметным называется моделирование, в ходе которого исследование ведется на модели, воспроизводящей определенные геометрические, физические, динамические либо функциональные характеристики объекта.

При знаковом моделировании в качестве моделей выступают схемы, чертежи, формулы, мысли, высказанные или записанные на каком-либо языке.

В зависимости от того, какие особенности объекта изучаются, различают модели его структуры и поведения (функционирования). Первые используются для изучения статичных систем (т. е. свойств материальных предметов), а вторые — для исследования динамичных систем (т. е. процессов).

Рассмотренные выше свойства геологических образований и процессов исключают возможность широкого применения предметного моделирования в геологии, хотя в последние годы все чаще предпринимаются попытки воспроизведения в лабораторных условиях отдельных элементов геологических процессов. Появились такие научные направления, как экспериментальная геотектоника, петрология, геохимия. Большие успехи экспериментальной минералогии привели к разработке технологических процессов получения синтетического кристаллосыря в промышленных масштабах.

При промышленной оценке месторождений предметное моделирование применяется для изучения технологических свойств руд по лабораторным и полупромышленным пробам. При этом лабораторные установки, имитирующие процесс переработки руды, являются действующими моделями оборудования будущей обогатительной фабрики.

Однако ведущую роль в науках о Земле играют различные методы знакового (информационного) моделирования. По характеру информации их можно разделить на словесные, графические и математические.

К словесным моделям можно отнести многочисленные классификации, понятия и определения, которыми изобилуют все геологические дисциплины.

К *графическим* моделям следует отнести все разнообразные графические геологические документы — карты, планы, разрезы, проекции и т. п., в связи с тем, что они отражают свойства реальных объектов недр упрощенно и приблизительно.

В качестве *математических* моделей в геологии используются числа и формулы, описывающие взаимосвязи и закономерности изменений свойств геологических образований или параметров геологических процессов.

В последние годы в связи с широким внедрением в практику геологических исследований моделирования на ЭВМ с использованием разнородной геологической информации границы между этими видами моделей становятся в известной степени условными. Картографическая информация с помощью номинальной шкалы измерений переводится в цифровую, а результаты замеров при геохимических и геофизических съемках с помощью граffопостроителей или графических дисплеев изображаются в виде карт изолиний.

## 1.5. ТИПЫ ГЕОЛОГО-МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

По принципу построения математической модели различают статическое и динамическое моделирование.

*Статическое моделирование* заключается в математическом описании свойств исследуемых объектов по результатам их изучения выборочным методом на основе индуктивного обобщения эмпирических данных.

*Динамическое моделирование* использует приемы дедуктивного метода, когда свойства конкретных объектов выводятся из общих представлений о его структуре и законах, определяющих его свойства.

В настоящее время в практике геологических исследований применяются главным образом статические модели. Это обусловлено сложностью и разнообразием геологических объектов и трудностью описания геологических процессов даже в самых общих чертах.

Статическое моделирование сводится к:

- преобразованию геологической информации в вид, удобный для анализа;
- выявлению закономерностей в массовых и в известной степени случайных замерах свойств изучаемых объектов;
- математическому описанию выявленных закономерностей (составлению математической модели);
- использованию полученных количественных характеристик для решения конкретных геологических задач — проверки геологических гипотез, выбору методов дальнейшего изучения объекта и т. п.;
- оценке вероятности возможных ошибок в решении поставленной задачи за счет выборочного метода изучения объекта.

Порядок решения геологических задач на основе динамического моделирования иной. Исходя из общих соображений о генезисе изучаемого объекта строится теоретическая математическая модель процесса его образования, учитывающая основные факторы, влияющие на конечный результат этого процесса, т. е. на свойства объекта.

Такая модель обычно может быть предложена лишь в самом общем виде, поскольку параметры процесса неизвестны. Эти параметры определяют путем перебора различных вариантов и сравнения теоретических реализаций процесса с фактическими свойствами изучаемого объекта, установленными эмпирическим путем. Динамическое моделирование сопряжено с большим объемом довольно сложных вычислений и возможно лишь на базе ЭВМ.

По характеру связи между параметрами и свойствами изучаемых объектов математические модели разделяются на детерминированные и статистические.

*Детерминированные модели* выражают функциональные связи между аргументом и зависимыми переменными. Они записываются в виде уравнений, в которых определенному значению аргумента соответствует только одно значение переменной. При моделировании геологических объектов детерминированные модели используются редко. Это объясняется тем, что они плохо согласуются с реальными явлениями, в которых функциональные связи сохраняются лишь в узких, весьма ограниченных областях.

*Статистическими моделями* называются математические выражения, содержащие, по крайней мере, одну случайную компоненту, т. е. такую переменную, значение которой нельзя предсказать точно для единичного наблюдения. Их весьма широко используют для целей математического моделирования, поскольку они хорошо учитывают случайные колебания экспериментальных данных.

Многообразие геологических задач и объектов изучения вызвало необходимость использования при геолого-математическом моделировании методов из разных разделов математики: теории вероятностей и математической статистики, теории множеств, теории групп, теории информации, теории графов, теории игр, матричной и векторной алгебры, дифференциальной геометрии и др. При этом одна и та же задача может быть решена разными методами, а в некоторых случаях для решения одной задачи необходимо использовать комплекс методов из разных разделов математики. Это создает определенные трудности при систематизации математических методов, применяемых в геологии.

Вместе с тем по типу решаемых задач, набору используемых для этого математических методов и главным допущениям относительно свойств геологических объектов все геолого-математические модели отчетливо разделяются на две группы.

В первую группу объединяются модели, использующие главным образом математический аппарат теории вероятностей и математической статистики. В них геологические объекты предполагаются внутренне однородными, а изменения их свойств в пространстве — случайными, не зависящими от места замера. Такие модели можно условно назвать статистическими. В зависимости от количества одновременно рассматриваемых свойств они разделяются на одномерные, двумерные и многомерные.

Статистические модели обычно используются для:

— получения по выборочным данным наиболее надежных оценок свойств геологических объектов;

- проверки геологических гипотез;
- выявления и описания зависимостей между свойствами геологических объектов;
- классификации геологических объектов;
- определения объема выборочных данных, необходимого для оценки свойств геологических объектов с заданной точностью.

Во вторую группу можно объединить модели, рассматривающие свойства геологических объектов как *пространственные переменные*. В этих моделях предполагается, что свойства геологических объектов зависят от координат точки замера, а в изменении этих свойств в пространстве существуют определенные закономерности. При этом, наряду с некоторыми вероятностными методами (случайные функции, временные ряды, дисперсионный анализ), применяются также приемы комбинаторики (полиномы), гармонического анализа, векторной алгебры, дифференциальной геометрии и других разделов математики.

Для изучения пространственных геологических переменных используются приемы как статического, так и динамического моделирования.

Модели пространственных геологических переменных используются для решения задач, связанных с:

- проверкой гипотез о закономерностях размещения геологических объектов относительно друг друга;
- проверкой гипотез о характере процессов формирования геологических образований;
- выделением аномалий в геологических и геофизических полях;
- классификацией геологических объектов по особенностям их внутреннего строения;
- разработкой приемов интерполяции и экстраполяции при оконтуривании геологических объектов;
- выбором оптимальной густоты и формы сети наблюдений при изучении геологических объектов.

## 1.6. ПРИНЦИПЫ И МЕТОДЫ ГЕОЛОГО-МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Применение математического моделирования в геологии сопряжено с рядом трудностей.

Математическая модель, как и любая другая, является упрощенным аналогом исследуемого объекта. Из-за сложности геологических объектов ни одна математическая модель не может воспроизвести все их свойства. Поэтому для описания различных свойств одного и того же объекта часто приходится использовать различные математические модели. При этом необходимо убедиться в том, что выбранная модель достаточно полно отражает именно те свойства объекта, которые непосредственно влияют на решение поставленной задачи.

Математические модели не могут исчерпывающе полно характеризовать изучаемые свойства. Они основаны на определенных допущениях о характере свойств объекта моделирования. Поэтому необходи-

мо следить за тем, чтобы эти допущения не приводили к принципиальному искажению реальных свойств объекта в рамках поставленной задачи. В связи с тем, что встречающиеся в практике геологических исследований задачи также весьма разнообразны, может возникнуть ситуация, когда для моделирования одного и того же свойства объекта необходимо использовать различные модели.

Определенные сложности иногда возникают также из-за отсутствия четких границ геологических совокупностей и рассмотренных выше особенностей их изучения.

Итак, решение геологических задач на основе математического моделирования представляет собой довольно сложный процесс, в котором можно выделить следующие этапы:

- 1) формулировка геологической задачи;
- 2) определение геологической совокупности, т. е. установление границ геологического объекта или временного интервала геологического процесса;
- 3) выявление главных свойств объекта или параметров процесса в рамках поставленной задачи;
- 4) переход от геологической совокупности к опробуемой и выборочной с учетом особенностей методов исследования;
- 5) выбор типа математической модели;
- 6) формулировка математической задачи в рамках выбранной математической модели;
- 7) выбор метода решения математической задачи;
- 8) решение математической задачи на основе вычисления параметров математической модели объекта;
- 9) интерпретация полученных результатов применительно к геологической задаче;
- 10) оценка вероятности и величины возможной ошибки за счет неадекватности модели и объекта.

Таким образом, этапу собственно математического моделирования предшествуют этапы создания геологической модели (опробуемой и выборочной геологической совокупности). Поэтому модели, используемые для решения геологических задач математическими методами, можно назвать *геолого-математическими*.

Справедливость конечного вывода при решении задач на основе геолого-математического моделирования зависит от правильности решений, принимаемых на каждом этапе. Нетрудно заметить, что решения на большинстве этапов принимаются исходя из особенностей геологических задач и свойств геологических объектов, поэтому они полностью находятся в компетенции геолога. Консультант математик может оказать существенную помощь геологу лишь при выборе метода решения математической задачи. Как показал многолетний опыт, большинство ошибок, допускавшихся при использовании математических методов в геологии, было обусловлено не слабой математической подготовкой геологов, а тем, что не учитывалась специфика геологических объектов и задач. Поэтому при изложении дальнейшего материала на эти аспекты геолого-математического моделирования обращено особое внимание.

## 1.7. МЕСТО И РОЛЬ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ПРОЦЕССЕ ИЗУЧЕНИЯ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

Необходимость внедрения математических методов в геологию признается в настоящее время практически всеми геологами, однако по поводу путей и возможностей их использования возникают различные мнения.

Некоторые геологи (Ю. А. Воронин, Э. Э. Фотиади и др.) считают, что для полноценного использования всех возможностей современной математики и вычислительной техники необходимо предварительно формализовать существующие геологические понятия и представления, т. е. перевести их с естественного разговорного языка на формализованный машинный язык. По сути дела это равносильно созданию новой геологической науки, так как, по мнению упомянутых исследователей, современные теоретические представления геологии не могут служить основой для эффективного внедрения математических методов и ЭВМ в геологии.

Многие геологи считают возможным использовать математические методы обработки и обобщения экспериментальных данных в рамках сложившихся геологических представлений и понятий, т. е. главным образом в описательных целях.

Мы будем придерживаться третьего мнения — мнения о том, что при использовании математических методов в геологии они должны совершенствоваться применительно к специфике решаемых геологических задач, а формализации должна подвергаться не вся геологическая наука, а лишь объекты непосредственных наблюдений, в соответствии с поставленной задачей исследования.

Таким образом, математические методы изучения земных недр следует использовать в комплексе с другими традиционными методами, для расширения и углубления возможностей геологического анализа, поскольку при качественном подходе не удается в равной мере учитывать и интерпретировать всю исходную информацию. В зависимости от опыта и интуиции геолог неизбежно преувеличивает роль тех или иных факторов, оставляя другие факторы без достаточного внимания. Использование математических методов способствует более глубокому познанию изучаемых объектов, количественному учету и оценкам не только двумерных, но и многомерных связей между наблюдаемыми явлениями, исключая в ряде случаев возможности различного толкования одних и тех же фактических данных.

В процессе изучения геологических объектов математическим методам отводится важная роль, поскольку их применение обеспечивает возможность перехода от словесных, часто субъективных определений сложных геологических объектов к их объективным количественным оценкам, часто приводящим к качественно иным интерпретациям наблюдаемых явлений.

### *Контрольные вопросы*

1. Чем обусловлены трудности изучения геологических объектов?
2. Почему геологические образования и процессы целесообразно рассматривать как природные системы?
3. Что такое «геологическая совокупность»?

4. Почему геологические объекты изучаются выборочным методом?
5. Что такое «опробуемая совокупность»?
6. Какие системы расположения точек наблюдений используются в геологии?
7. Чем различаются погрешности измерений и погрешности аналогий?
8. Что такое «наблюдаемая изменчивость» свойств геологических образований и от чего она зависит?
9. Какие виды информации используются в геологии?
10. Что такое «выборочная совокупность»? Чем она отличается от «геологической» и «опробуемой совокупности»?
11. Какие шкалы измерений используются в геологии?
12. Какие виды моделирования применяются в геологии?
13. Какие типы математических моделей применяются в геологии?
14. Какие трудности возникают при математическом моделировании геологических объектов?
15. На какие этапы можно разделить процесс решения геологических задач математическими методами?
16. Для чего используются математические методы в геологии?

## ГЛАВА 2

### ОДНОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

---

#### 2.1. СУЩНОСТЬ И УСЛОВИЯ ПРИМЕНЕНИЯ

Геологические исследования в большинстве случаев основаны на изучении свойств геологических образований путем замеров в отдельных точках непосредственно на месте их залегания или путем анализа образцов и проб, отобранных в отдельных участках. При этом выборочные наблюдения относятся к элементарно малым пространственно разобщенным объемам недр (в искусственном или естественном обнажении), а выводы, полученные по ним, распространяются на весь изучаемый объем.

Изучая сложные природные объекты выборочными методами геологи всегда учитывали возможность получения ошибочных результатов по ограниченному количеству наблюдений. Поэтому первые попытки математического моделирования в геологии связаны с использованием математического аппарата теории вероятностей и математической статистики, который обеспечивает возможность получения наиболее надежных выводов по выборочным данным и позволяет оценить точность этих выводов.

В основе статистического моделирования лежат два понятия: *о генеральной совокупности* — множестве возможных значений определенного признака изучаемого объекта или явления и *о выборке* — совокупности наблюденных значений этого признака. Оба понятия фактически совпадают с рассмотренными в разд. 1.2 понятиями геологической и опробуемой совокупности.

При использовании статистической модели геологические объекты рассматриваются как совокупности бесконечно большого количества элементарных участков, каждый из которых соответствует по размеру отдельной пробе или месту единичного замера изучаемого свойства. Такой подход вполне правомерен, поскольку размеры проб или

сечения искусственных обнажений — скважин и горных выработок обычно неизмеримо малы по сравнению с изучаемыми геологическими объектами.

При статистическом моделировании предполагается, что выборочная совокупность отвечает требованиям массовости, однородности, случайности и независимости.

*Условие массовости* вызвано тем, что статистические закономерности проявляются лишь в массовых явлениях, поэтому объем выборочной совокупности должен быть достаточно большим. Эмпирическим путем установлено, что надежность статистических оценок резко снижается при уменьшении объема выборки в диапазоне от 60 до 30—20 значений, а при меньшем количестве наблюдений применять статистические методы в большинстве случаев вообще не имеет смысла. При проведении геологических, геохимических и геофизических съемок количество наблюдений, как правило, велико и условие массовости соблюдается. Однако, в процессе разведки месторождений полезных ископаемых, когда для каждого наблюдения обычно требуется проходка специальной горной выработки или скважины, геологу часто приходится иметь дело с малыми выборками. Это вызывает определенные трудности, так как свойства многих статистических оценок в данных условиях изучены недостаточно.

В связи с этим вопрос о минимально допустимом объеме выборки в дальнейшем будет рассматриваться применительно к решению отдельных конкретных задач.

*Условие однородности* заключается в том, что выборочная совокупность должна состоять из наблюдений, принадлежащих одному объекту и выполненных одинаковым способом, т. е. при постоянном размере проб и методе анализа или измерения. Нарушения этого условия могут быть связаны с ошибками при определении границ изучаемой геологической совокупности или техническими и организационными сложностями проведения исследований.

Границы геологической совокупности обычно задаются исходя из поставленной задачи до получения массовых результатов исследования. При этом предполагается, что все объекты, включенные в геологическую совокупность, аналогичны и внутренне однородны. Однако это предположение подтверждается не всегда. Схожие по качественным признакам объекты могут иногда существенно отличаться по количественным характеристикам. К тому же большинство реальных геологических образований имеет сложное внутреннее строение, обусловленное зональностью и наличием элементов неоднородности разного масштаба.

При обобщении результатов геологических исследований часто приходится иметь дело с данными, полученными в разные годы с помощью различных технических средств.

В связи с тем, что условие однородности в практике геологических исследований соблюдается далеко не всегда, применение статистических методов должно сопровождаться анализом возможных последствий за счет нарушения данного условия. Для этого необходимо учитывать характер решаемой геологической задачи, а в ряде случаев

использовать также специальные методы для проверки гипотезы об однородности выборки.

*Условие случайности* предусматривает непредсказуемость результата единичного выборочного наблюдения. Сложность и изменчивость геологических объектов, как правило, исключают возможность точной оценки их свойств до проведения наблюдения. Поэтому элемент случайности присутствует во всех геологических исследованиях. Однако условие случайности строго выполняется лишь тогда, когда расположение мест отбора проб или проведения замеров изучаемого свойства вообще не будет каким-либо образом связано с величиной, характеризующей это свойство. В практике проведения геологоразведочных работ это обычно достигается за счет проведения наблюдений по равномерной сети, когда все места наблюдений намечаются заранее до проведения работ и в процессе их выполнения не корректируются. Однако при изучении геологических образований по естественным обнажениям это условие может нарушаться. Так, например, на территориях со слабо расчлененным рельефом естественные обнажения преимущественно располагаются в бортах речных долин, которые, в свою очередь, часто совпадают с разрывными нарушениями или выходами пород, наиболее легко поддающихся процессам зрозии. В то же время прочностные свойства пород связаны с их текстурными особенностями и минеральным составом. Поэтому статистическая обработка результатов петрографических исследований или испытаний их физико-механических свойств по образцам, отобранным только из естественных обнажений, может дать искаженное представление о свойствах пород изученной территории в целом.

Условие случайности может нарушаться за счет субъективности при проведении замеров или отборе проб. Если при отборе образцов из толщи гнейсов один исследователь будет отдавать предпочтение прослойям более светлой окраски, а другой — прослойям более темной окраски, то полученные ими выборки будут существенно отличаться по среднему минеральному составу как друг от друга, так и от истинного среднего состава изучаемой толщи.

В процессе проведения геологоразведочных работ часто возникает необходимость в сгущении сети наблюдений на наиболее интересных или перспективных участках. Свойства геологических объектов в пределах этих участков и на всей остальной изучаемой территории могут существенно отличаться. Поэтому при статистической обработке данных для соблюдения условия случайности результаты наблюдений по участку детализации должны быть выделены в самостоятельную выборочную совокупность.

*Условие независимости* предполагает, что результаты каждого наблюдения не зависят от результатов предыдущих и последующих наблюдений, а при проведении наблюдений на площади или в объеме результаты не зависят от координат пространства. Для большинства геологических процессов и образований это условие не соблюдается. В изменчивости свойств геологических образований в пространстве и параметров геологических процессов

во времени обычно наблюдаются определенные закономерности. Ввиду этого область применения статистических моделей ограничена объектами, для которых характерно отсутствие каких-либо закономерностей изменения в пространстве или во времени, или задачами, при решении которых эти закономерности можно не учитывать.

В геологической практике одномерные статистические модели используются для решения двух типов задач: оценки средних параметров геологических объектов и статистической проверки гипотез.

В связи с тем, что геологу часто приходится сталкиваться с ситуацией, когда допущения математической статистики не могут быть приняты безоговорочно, применение статистических методов в геологии должно всегда базироваться на всестороннем анализе характера решаемой задачи, особенностей изучаемых объектов и методики получения информации об их свойствах. Подробно вопросы, связанные с выбором наиболее эффективной геолого-математической модели, рассмотрены в гл. 7.

## 2.2. ПРОСТЕЙШИЕ ПРЕОБРАЗОВАНИЯ КОЛИЧЕСТВЕННОЙ ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

В связи с возможными отклонениями условий изучения геологических объектов от строгих требований, предъявляемых к статистическому эксперименту, статистический анализ геологических данных целесообразно разделять на два этапа — разведочный и подтверждающий.

Цель первого этапа — преобразование наблюдаемых данных в более компактную и наглядную форму, позволяющую выявить имеющиеся в них закономерности. Это дает возможность более обоснованно подходить к выбору традиционных статистических методов решения геологических задач на втором этапе.

На первом этапе целесообразно применять методы, свободные от каких-либо априорных допущений относительно свойств выборочной совокупности и не требующие трудоемких расчетов. Предпочтение следует отдавать методам, где числовая информация преобразуется в графическую. Как отмечает Дж. Тьюки, график часто «вынуждает» нас заметить то, что мы совсем не ожидали увидеть.

Некоторые преобразования такого типа рассмотрим на следующем примере.

Для определения петрографического типа пород из горизонта неогеновых лав отобрано и проанализировано на содержание  $\text{SiO}_2$  30 проб (табл. 1). Содержание  $\text{SiO}_2$  в отдельных пробах меняется от 56,6 (андезитобазальт) до 73,2% (риолит), что не позволяет оценить состав лав горизонта в целом по единичному наблюдению. Для получения усредненных характеристик приведенные в табл. 1 данные необходимо каким-то способом преобразовать. Нетрудно заметить, что содержания  $\text{SiO}_2$  по некоторым пробам близки и различаются лишь на десятые доли процента. Это позволяет заменить таблицу числовой диаграммой, которую американский

статистик Дж. Тьюки образно назвал «стеблем с листьями». При построении диаграммы часто повторяющиеся части чисел, в данном случае целые проценты содержаний  $\text{SiO}_2$ , записываются в порядке возрастания в виде вертикального столбца («стебля»), а оставшиеся части, в нашем примере десятые доли процента, записываются около соответствующей начальной части в горизонтальную строку («листья»). Такая запись более компактна, наглядна, позволяет оценить частоту встречаемости значений в разных диапазонах, а также легко найти некоторые важные характеристики совокупности выборочных данных.

Таблица 1. Содержание  $\text{SiO}_2$  (в %) в неогеновых лавах

№ пробы	$\text{SiO}_2$						
1	59,5	9	73,2	17	69,3	24	61,1
2	66,8	10	64,6	18	64,6	25	63,8
3	60,5	11	62,9	19	67,8	26	67,5
4	63,7	12	62,4	20	56,6	27	65,3
5	72,5	13	71,6	21	71,4	28	69,9
6	69,2	14	65,8	22	67,6	29	73,2
7	61,2	15	63,1	23	63,6	30	60,7
8	66,3	16	61,2				

По рис. 3, *a* легко определить, что из 30 проб по содержанию  $\text{SiO}_2$  одна соответствует андезитобазальту, шесть — андезиту, две —

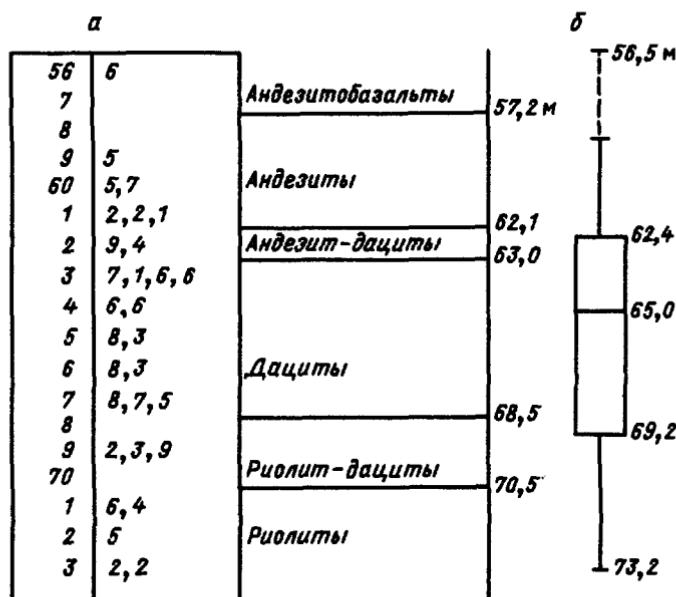


Рис. 3. Изображение содержаний  $\text{SiO}_2$  (в %) в неогеновых лавах:

*a* — числовая диаграмма «стебель с листьями»; *б* — схематическая диаграмма «ящик с усами»

андезит-дациту, тринадцать — дациту, три — риолит-дациту и пять — риолиту. Следовательно, определения, соответствующие дациту, явно преобладают. По числовой диаграмме легко находится также центр распределения — медиана. Для этого с любого конца «стебля» нужно отсчитать количество чисел («листьев»), равное половине общего количества замеров изучаемого свойства. В данном примере количество проб четное, поэтому центр распределения попадает в интервал между значениями 64,6 и 65,3%, т. е. соответствует примерно 65%, что отвечает содержанию  $\text{SiO}_2$  в даците. Таким образом, по числовой диаграмме можно сделать вывод о том, что средний состав лав изученного горизонта соответствует дациту.

Основные особенности числовых массивов могут быть также отображены на схематической диаграмме, названной Дж. Тьюки «ящиком с усами». На эти диаграммы в определенном масштабе выносятся минимальные и максимальные значения признака, медиана и так называемые «гибы» — середины распределений в интервалах от медианы до минимального и максимального значения. Положение «гибов» определяется по числовой диаграмме путем отсчета от медианы в сторону больших и меньших значений количества «листьев», равного  $1/4$  от их общего количества. Интервал между гибами изображается в виде прямоугольника («ящика»), а интервалы от гибов до максимального и минимального значения — в виде прямых линий («усов»). Иногда единичные значения на числовой диаграмме отделены от основной совокупности значительными интервалами — участками «стебля» без «листьев». Такие интервалы на схематической диаграмме целесообразно показывать пунктиром. По длине «ящика» и «усов» можно судить о характере разброса данных относительно медианы, оценивать степень асимметричности распределения, выявлять резкие отклонения по отдельным замерам от общей совокупности, оценивать однородность числового массива.

В приведенном примере в «ящик» попадают все определения, соответствующие дациту, а пробы с содержанием  $\text{SiO}_2$ , характерным для других пород, приходятся на «усы» (рис. 3, б). При этом проба № 20, где содержание  $\text{SiO}_2$  соответствует андезитобазальту, может рассматриваться как резкое отклонение, нетипичное для горизонта в целом.

Визуальное сравнение числовых и схематических диаграмм, построенных для нескольких геологических объектов, нередко позволяет произвести их предварительную группировку по принципу близости средних значений и степени изменчивости изучаемых свойств, а также выявить отличительные индивидуальные особенности каждого объекта.

*Пример.* На месторождении редких металлов с целью количественного описания морфологии рудных тел были замерены их площади по разведочным разрезам, ориентированным вкrest простирания рудовмещающих структур. По результатам замеров, приведенным в табл. 2, можно заключить, что для всех изученных рудных тел характерна сильная изменчивость данного параметра. Это затрудняет сравнение рудных тел непосредственно по табличным данным.

Таблица 2. Площади рудных тел в поперечных разрезах ( $\text{в м}^2$ )

№ разреза	№ рудного тела				№ разреза	№ рудного тела			
	525	518	501	509		525	518	501	509
1	232	119	137	25	14		451	260	77
2	293	120	177	56	15		627	68	195
3	87	130	98	101	16		597	254	234
4	121	417	25	634	17		726	211	219
5	422	355	115	340	18		686	254	75
6	1580	198	360	195	19		683	82	43
7	835	567	195	158	20		525	100	
8	204	504	493	24	21		605	9	
9	218	574	487	210	22		1042	30	
10	243	404	379	50	23		504		
11	146	502	247	228	24		648		
12	49	697	116	335	25		220		
13	174	579	629	153					

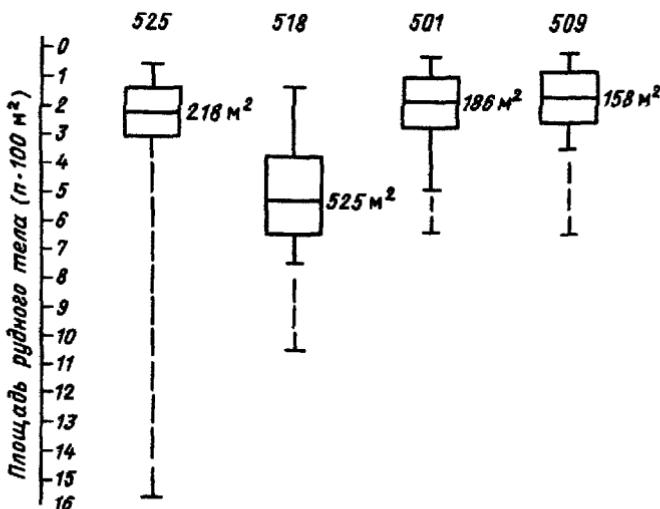


Рис. 4. Схематические диаграммы площадей рудных тел по разрезам

В то же время числовые и схематические диаграммы (табл. 3, рис. 4) позволяют увидеть некоторые интересные особенности:

— рудные тела 501 и 509 весьма схожи как по среднему значению площади в поперечном сечении, так и по степени изменчивости этого параметра;

— рудное тело 525 отличается от рудных тел 501 и 509 лишь наличием резких увеличений площади по единичным разрезам;

— рудное тело 525 явно отличается от остальных рудных тел большей площадью в поперечных сечениях.

Приемы предварительного анализа результатов массовых наблюдений подробно изложены в работе Дж. Тьюки.

Таблица 3. Числовые диаграммы площадей рудных тел в поперечных разрезах

Сотни, м <sup>2</sup>	м <sup>2</sup>			
0	87, 49		98, 25, 68, 82, 09, 30	25, 56, 24, 50, 77, 75, 43
1	21, 46, 74	19, 20, 30, 98	37, 77, 15, 95, 16, 00	01, 95, 58, 53, 95
2	32, 93, 04, 18, 43	20	47, 60, 54, 11, 54	10, 28, 34, 19
3		55	60, 79	40, 35
4	22	17, 04, 51	93, 87	
5		67, 04, 74, 02, 79, 97, 25 04		
6		97, 27, 86, 83, 05, 48	29	34
7		26		
8		35		
9				
10		42		
11				
12				
13				
14				
15	80			
№ руд- ного тела	525	518	501	509

### 2.3. СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ, ИСПОЛЬЗУЕМЫЕ В ГЕОЛОГИИ

В основе статистического моделирования лежит понятие о *вероятности случайного события*, которая представляет собой численную меру степени объективной возможности его проявления. Вероятность достоверного события равна единице, а вероятность невозможного события — нулю. Таким образом, вероятность случайного события выражается числом, лежащим в пределах от 0 до 1, или в процентах — от 0 до 100%.

*Случайные величины* могут быть прерывистыми (дискретными) и непрерывными. Случайная величина, принимающая в испытаниях только определенные числовые значения, называется *дискретной*. Примером дискретной случайной величины может служить количество зерен минерала при изучении шлифа под микроскопом или количество скважин, вскрывших рудное тело при разведке месторождения, так как эти величины могут быть только целыми числами. *Непрерывная* случайная величина может принимать в некотором интервале любые значения. Например, содержание свинца в рудах полиметаллического месторождения, полученное по результатам анализа проб малого объема, колеблется от нуля до величины концентрации свинца в чистом галените.

Специфическими свойствами обладают широко распространенные в геологии *случайные угловые величины* — замеры элементов залегания контактов пород и разрывных нарушений, ориентировка осей складок,

элементов текстуры пород и руд, направление остаточного магнитного поля и т. п. Если замеры обычных (скалярных) случайных величин можно рассматривать как точки на прямой, соответствующей интервалу ее возможных значений, то угловые измерения удобно трактовать как точки на окружности или сфере (вспомним лимб теодолита или компаса). В виде точек на окружности можно рассматривать и замеры параметров процессов, периодически изменяющихся во времени: суточные или сезонные колебания температуры воздуха, уровня грунтовых вод, дебита естественных и искусственных источников водоснабжения, водопритоков в горные выработки и т. п. Аналогия этих параметров с угловыми измерениями становится очевидной, если вспомнить о сходстве циферблата часов с лимбом компаса.

Число появлений события в серии испытаний называется его *частотой*, а отношение числа появлений события к общему числу опытов в серии — *частостью*. При увеличении числа опытов (объема выборочной совокупности) частость события сходится по вероятности к его вероятности.

Соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями, называется *законом* или *функцией распределения*.

*Функция распределения*  $F(x)$  выражает вероятность того, что выборочное значение случайной величины  $\xi$  окажется меньше некоторого предела, ограниченного  $x$ , где  $x$  — заданная переменная, т. е. вероятность события  $\xi \leq x$ .

*Функция плотности распределения*  $f(x)$  характеризует вероятность попадания выборочного значения случайной величины  $\xi$  в заданный интервал от  $x$  до  $x + \Delta x$ , т. е. вероятность события  $x < \xi < x + \Delta x$ .

Эти функции связаны отношением  $F(x) = \int_{-\infty}^x f(x)dx$ , причем

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Функции распределения можно изобразить графически (рис. 5).

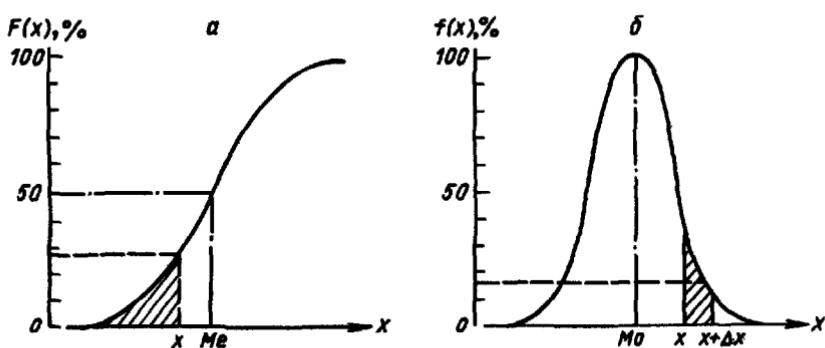


Рис. 5. Графики функций распределения:  
а — интегральная функция распределения; б — функция плотности распределения (дифференциальная функция распределения)

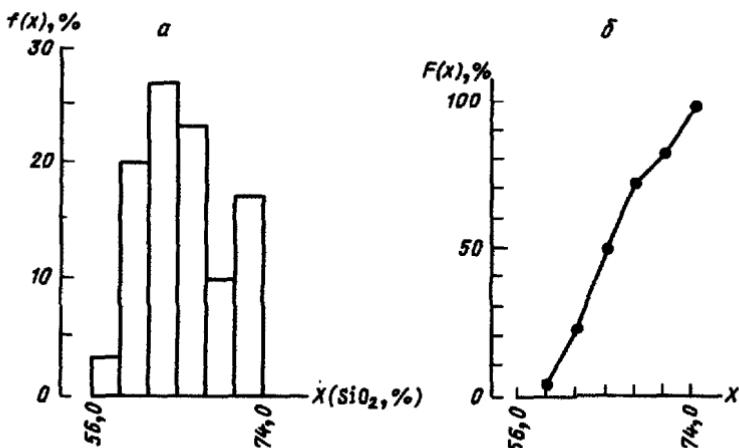


Рис. 6. Графики частотного распределения содержания  $\text{SiO}_2$  в неогеновых лавах:  
а — гистограмма; б — кумуляты

По выборочным данным вероятности рассмотренных выше случайных событий оцениваются по их частостям, рассчитанным для определенных интервалов (классов) значений изучаемой случайной величины. При построении эмпирических графиков функции плотности распределения  $f(x)$  (гистограммы) по оси ординат откладываются частоты, соответствующие каждому классу значений случайной величины, а при построении графиков функции  $F(x)$  (кумуляты) — накопленные частоты, т. е. суммы частостей по всем классам, где  $\xi$  меньше заданной переменной  $x$ .

На рис. 6 изображены гистограмма и кумулята для рассмотренного в разд. 2.2 распределения  $\text{SiO}_2$  в неогеновых лавах. Для построения графиков распределений выборочные данные первоначально целесообразно представить в виде числовой диаграммы «стебель с листьями». Это позволяет правильно выбрать длину классового интервала и быстро подсчитать количество значений в каждом классе (табл. 4).

Изображение распределений выборочных данных в виде кумулят используется при решении задач, связанных с поиском оптимальных границ геологических совокупностей. Так при оценке рудных месторождений, где скопления полезных ископаемых не имеют четких геологических границ, контуры промышленного оруденения определяются по заданному минимальному содержанию полезного компонента в руде. Выбор оптимального значения этого показателя обычно производится на основе геолого-экономического анализа путем сравнения нескольких вариантов. Кумулята содержаний полезного компонента в пробах примерно одинаковых размеров позволяет оперативно оценить как будут меняться общие запасы месторождения при изменении требований к минимальной концентрации полезного компонента в руде.

Таблица 4. Частотное распределение содержаний  $\text{SiO}_2$  в неогеновых лавах

Содержание SiO <sub>2</sub> , %		Класс, от—до	Частота	Частость, %	Накопленная частость, %
целые числа	десятые доли				
56	6	56,0—58,9	1	3	3
57					
58					
59	5				
60	5, 7				
61	2, 2, 1	59,0—61,9	6	20	23
62	9, 4				
63	7, 1, 6, 8				
64	6, 6	62,0—64,9	8	27	50
65	8, 3				
66	8, 3				
67	8, 7, 5	65,0—67,9	7	23	73
68					
69					
70	2, 3, 9	68,0—70,9	3	10	83
71	6, 4				
72	5				
73	2, 2	71,0—73,9	5	17	100

На рис. 7, а изображена кумулята содержаний молибдена по результатам опробования молибденового штокверка. По ней легко определить, что при минимальном его количестве, равном 20 условным единицам ( усл. ед.), в пределах штокверка практически не будет безрудных участков, при увеличении этого показателя до 40 усл. ед. их доля составит 11,5%, а при минимально допустимом содержании, равном 80 усл. ед., безрудные участки будут составлять больше половины объема штокверка.

Гистограммы можно применять для решения задач, связанных с разделением геологических совокупностей на несколько самостоятельных совокупностей по величине изучаемого свойства. Так, например, изображенная на рис. 7, б гистограмма распределения содержания молибдена позволяет определить соотношение различных сортов руд в штокверке. На данном месторождении выделяют три типа руд: богатые — с содержанием молибдена больше 200 усл. ед., рядовые — с содержанием от 100 до 200 усл. ед., бедные — с содержанием ниже 100 усл. ед. По гистограмме легко определить, что на месторождении преобладают последние два типа, а первый составляет лишь 4% от общего объема штокверка.

Аналогом гистограммы для угловых измерений является *диаграмма розы наблюдений*. Для ее построения угловые измерения разбиваются на классы — равновеликие секторы круга, а частоты или

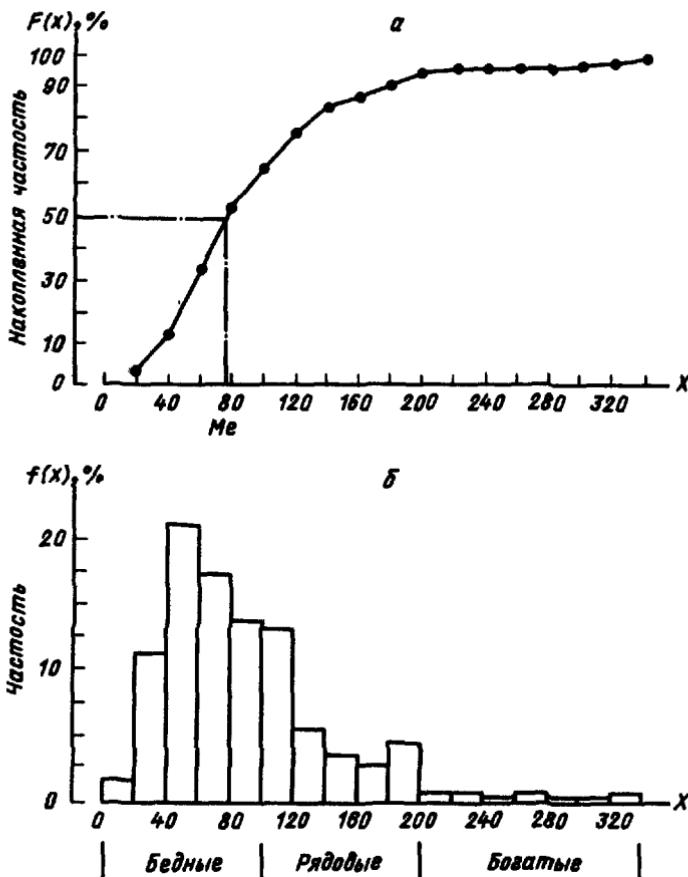


Рис. 7. Графики частотного распределения содержания молибдена в руде:  
а—кумулята; б—гистограмма

частоты попадания наблюдений в эти классы откладываются в виде радиусов этих секторов (рис. 8). Площади секторов на диаграмме розы наблюдений пропорциональны квадратам частот или частостей.

Наиболее существенные особенности распределения случайной величины могут быть выражены с помощью числовых *характеристик положения и разброса*.

К важнейшим характеристикам положения относятся *математическое ожидание, мода и медиана*.

*Математическое ожидание* ( $Mx$ ) дискретной случайной величины есть сумма произведения всех ее возможных значений ( $x_i$ ) на соответствующие им вероятности ( $p_i$ ):

$$Mx = \sum_{i=1}^n x_i p_i. \quad (1)$$

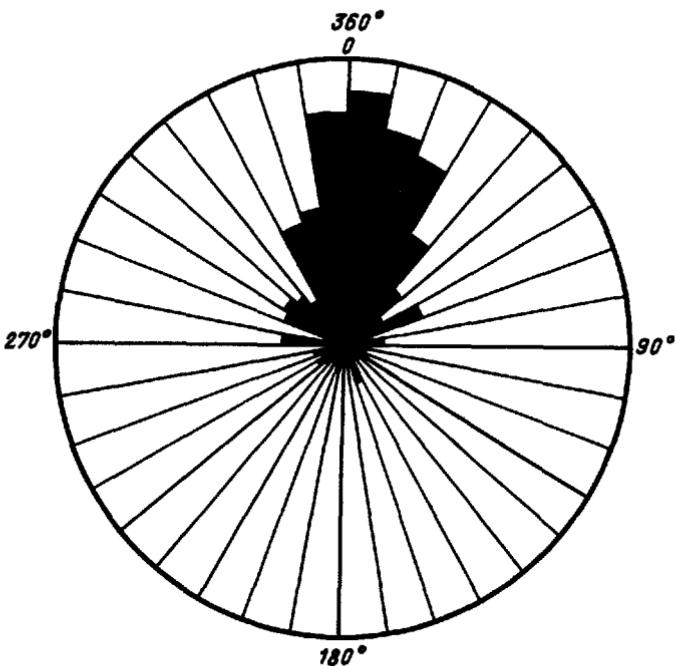


Рис. 8. Диаграмма розы наблюдений азимутов падения швов тектонических брекчий в пределах минерализованной зоны дробления

Математическое ожидание непрерывной случайной величины выражается через плотность распределения  $Mx = \int_{-\infty}^{+\infty} xf(x)dx$ .

Математическое ожидание характеризует среднее значение случайной величины и поэтому его оценки широко используются для количественного описания обычно весьма изменчивых свойств геологических объектов. Они необходимы для решения самых различных задач. Так оценки среднего содержания полезного компонента в руде используются для подсчета запасов месторождений и планирования работы горнодобывающего предприятия, оценки средних значений характеристик прочностных свойств пород участвуют в инженерных расчетах при проектировании различных сооружений, средние размеры обломочных частиц и средний химический и минеральный составы используются в классификациях пород и т. п.

Модой (Mo) случайной величины называют ее наиболее вероятное значение для дискретной величины или значение с максимальной плотностью вероятности для непрерывной величины. Она соответствует вершине (перегибу) на графике функции плотности распределения. Для большинства статистических законов распределения характерна лишь одна мода. Поэтому по наличию на эмпирических графиках распределений (гистограммах) нескольких

модальных значений можно сделать вывод, что изучаемая выборочная совокупность неоднородна.

Медиана ( $Me$ ) случайной величины соответствует значению функции распределения  $F(x)$ , равному 0,5, т. е. такому значению, для которого вероятность ( $P$ ) встречи больших и меньших значений в выборке равны:

$$P(\xi \leq Me) = P(\xi > Me).$$

Медиану легко определить по графику функции распределения, а ее оценку — по кумуляте (см. рис. 7, a).

В симметричных распределениях медиана совпадает с математическим ожиданием и, поскольку ее можно легко определить графическим способом (см. разд. 2.2), часто используется в качестве оценки математического ожидания.

Характеристиками разброса, определяющими степень отклонения значений случайной величины от ее математического ожидания, служат *размах варьирования* (т. е. интервал возможных значений случайной величины) и *центральные моменты* различных порядков.

По выборочным данным размах варьирования оценивается как разница между максимальным и минимальным значениями.

Главной характеристикой разброса случайной величины служит центральный момент второго порядка — *дисперсия*.

Дисперсия дискретной случайной величины

$$\mu_2 = \sigma^2 = \sum_{i=1}^n (x_i - Mx)^2 p_i, \quad (2)$$

а дисперсия непрерывной случайной величины

$$\mu_2 = \sigma^2 = \int_{-\infty}^{+\infty} (x - Mx)^2 f(x) dx.$$

Производными характеристиками от дисперсии являются *стандартное* или среднее квадратическое отклонение  $\sigma = \sqrt{\sigma^2}$  и *коэффициент вариации*  $V = \frac{\sigma}{Mx} \cdot 100\%$ .

Коэффициент вариации — величина безразмерная, поэтому его применяют в тех случаях, когда необходимо сравнить по степени изменчивости свойства, выражения в разных единицах измерения, например, мощность рудного тела и содержание в нем полезного компонента.

Дисперсия и производные от нее характеристики используются в геологии в качестве классификационных признаков для оценки величин погрешностей измерений и анализов и при решении ряда других задач, когда требуется учитывать степень изменчивости изучаемых свойств геологических объектов.

Для характеристики степени асимметрии распределения случайной величины относительно ее математического ожидания используется центральный момент третьего порядка:

$$\mu_3 = \sum_{i=1}^n (x_i - Mx)^3 p_i \quad (3)$$

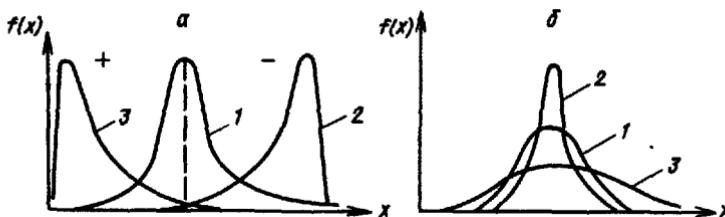


Рис. 9. Графики функции плотности распределения:  
 а—распределение симметричное (1), отрицательно асимметричное (2),  
 положительно асимметричное (3); б—показатель эксцесса равен нулю (1),  
 положителен (2), отрицателен (3)

для дискретной случайной величины и

$$\mu_3 = \int_{-\infty}^{+\infty} (x - Mx)^3 f(x) dx$$

для непрерывной случайной величины.

Для симметричных распределений третий центральный момент равен нулю. В распределениях, у которых мода, медиана и математическое ожидание смешены от середины размаха вариирования в сторону малых значений («хвост» на графике функции  $f(x)$  справа от центра), он имеет знак «плюс», а в случае их смешения в сторону больших значений («хвост» слева)—знак «минус» (рис. 9, а). Поэтому распределения первого типа называют право- или положительно асимметричными, а второго—лево- или отрицательно асимметричными.

Мерой остроты графика функции плотности распределения (эксцесса) служит центральный момент четвертого порядка

$$\mu_4 = \sum_{i=1}^n (x_i - Mx)^4 p_i,$$

$$\text{или } \mu_4 = \int_{-\infty}^{+\infty} (x - Mx)^4 f(x) dx.$$

Оценки моментов третьего и четвертого порядка используются для решения вопроса о соответствии выборочных данных определенному типу распределения. Для этого обычно вычисляют производные от них  $A = \mu_3/\sigma^3$  и  $E = (\mu_4/\sigma^4) - 3$ , которые называются показателями асимметрии и эксцесса.

При оценке характеристик положения и разброса по сгруппированным выборочным данным вероятности ( $p_i$ ) в формулах заменяются частотами попадания значений в каждый интервал группирования ( $n_j$ ), вместо значений случайной величины  $x$  подставляются значения центров интервалов группирования ( $\hat{x}_j$ ), вместо математического ожидания  $Mx$ —его оценка ( $\bar{x}$ ), а операция интегрирования для непрерывных величин заменяется суммированием. Таким образом, формулы для расчета оценок математического ожидания ( $\bar{x}$ ), дисперсии

$(S^2)$ , показатели асимметрии ( $A$ ) и эксцесса ( $E$ ) для сгруппированных и несгруппированных данных принимают следующий вид:

$$\bar{x} = \frac{\sum_{j=1}^k n_j \dot{x}_j}{n} = \frac{\sum_{i=1}^n (x_i)}{n}; \quad (4)$$

$$S^2 = \frac{\sum_{j=1}^k n_j (\dot{x}_j - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}; \quad (5)$$

$$A = \frac{\sum_{j=1}^k n_j (\dot{x}_j - \bar{x})^3}{n S^3} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n S^3}; \quad (6)$$

$$E = \frac{\sum_{j=1}^k n_j (\dot{x}_j - \bar{x})^4}{n S^4} - 3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n S^4} - 3, \quad (7)$$

где  $k$  — количество классов группирования.

Характеристики положения и разброса угловых величин имеют некоторые специфические особенности. В этом нетрудно убедиться на следующем примере. В табл. 5 и на рис. 8 приведены результаты замеров азимутов падения швов тектонических брекчий в пределах минерализованной зоны трещиноватости. Если оценивать математическое ожидание азимутов падения прожилков по формуле (4), то получим  $\bar{x} = 162^\circ$ , что соответствует падению на юго-юго-восток. В то же время, по диаграмме розы наблюдений (см. рис. 8) отчетливо видно, что основная масса прожилков имеет падение в северных румбах.

Таблица 5 Замеры азимутов (в градусах) падения швов тектонических брекчий в пределах минерализованной зоны дробления

Азимут	Частота $n_j$	Частость $(n_j/n) \times 100\%$	Азимут	Частота $n_j$	Частость $(n_j/n) \times 100\%$
0—10	85	12,5	180—190	4	0,6
10—20	72	10,7	190—200	—	—
20—30	67	9,9	200—210	—	—
30—40	42	6,2	210—220	7	1,0
40—50	24	3,5	220—230	—	—
50—60	13	1,9	230—240	8	1,2
60—70	27	4,0	240—250	—	—
70—80	7	1,0	250—260	8	1,2
80—90	13	1,9	260—270	8	1,2
90—100	—	—	270—280	20	2,9
100—110	3	0,4	280—290	—	—
110—120	1	0,1	290—300	21	3,1
120—130	1	0,1	300—310	22	3,4
130—140	—	—	310—320	20	2,9
140—150	2	0,3	320—330	16	2,3
150—160	13	2,9	330—340	43	6,3
160—170	5	0,7	340—350	46	6,7
170—180	1	0,1	350—360	78	11,5

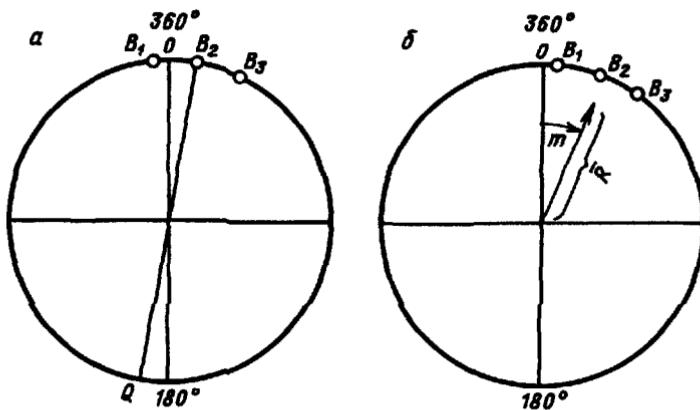


Рис. 10. Изображение замеров азимутов углов падения прожилков в виде точек на окружности:  
а—по первой зоне; б—по второй зоне

Аналогичная ситуация происходит и с оценками мер разброса. Допустим, что по двум зонам трещиноватости, аналогичным рассмотренной выше, выполнено по три замера азимутов падения прожилков. По первой зоне они оказались равными  $355, 10, 25^\circ$ , а по второй— $5, 20, 35^\circ$ . Очевидно, что в обоих случаях замеры не выходят за пределы сектора с углом  $30^\circ$  (рис. 10), т. е. разброс замеров примерно одинаков. Однако по формуле (5) получим, что оценка дисперсии по первой выборке составляет 12 675, а по второй—225.

Эти парадоксы обусловлены тем, что в угловых измерениях значения  $0$  и  $360^\circ$  совпадают, и величина угла при переходе через это направление меняется скачкообразно.

В качестве характеристик положения угловых величин удобно использовать выборочное круговое среднее направление, выборочную круговую медиану и моду.

Если представить замеры угла  $\vartheta_i$  ( $i=1, 2, \dots, n$ ) в виде точек на окружности— $B_i$  (см. рис. 10), то выборочное круговое среднее направление определяется как направление суммы единичных векторов  $OB_1, \dots, OB_n$ . Если всем этим точкам присвоить одинаковую «массу», равную  $\frac{1}{n}$ , то координаты «центра масс» будут определяться формулами

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \vartheta_i \text{ и } \bar{S} = \frac{1}{n} \sum_{i=1}^n \sin \vartheta_i. \quad (8)$$

Длина суммарного вектора  $\overrightarrow{OB}_1, \dots, \overrightarrow{OB}_n$  будет равна  $R=n\bar{R}$ , где  $\bar{R}=\sqrt{\bar{C}^2+\bar{S}^2}$ , а направление этого вектора, т. е. выборочное круговое среднее направление  $m$ , определяется из системы уравнений:

$$\bar{C}=\bar{R} \cos m; \quad \bar{S}=\bar{R} \sin m. \quad (9)$$

Таблица 6. Расчет выборочного кругового среднего направления азимутов падения прожилков во второй зоне

$\theta_i$	$\cos \theta_i$	$\sin \theta_i$	
5	0,996	0,087	
20	0,940	0,342	
35	0,819	0,574	
$\Sigma$	2,755	1,003	
			$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i = 2,755/3 = 0,918;$ $\bar{S} = \frac{1}{n} \sum_{i=1}^n \sin \theta_i = 1,003/3 = 0,334;$ $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} = \sqrt{0,918^2 + 0,334^2} = 0,977;$ $\cos m = \bar{C}/\bar{R} = 0,918/0,977 = 0,940;$ $\sin m = \bar{S}/\bar{R} = 0,334/0,977 = 0,342; m = 20^\circ.$

Таблица 7. Расчет выборочного кругового среднего направления азимутов падения швов тектонических брекчий (исходные данные в табл. 5)

$\theta_j$	$n_j$	$\cos \theta_j$	$n_j \cos \theta_j$	$\sin \theta_j$	$n_j \sin \theta_j$
5	85	0,996	84,68	0,087	7,41
15	72	0,966	69,55	0,259	18,63
25	67	0,906	60,72	0,423	28,32
35	42	0,819	34,40	0,574	24,11
45	24	0,707	16,97	0,707	16,97
55	13	0,574	7,46	0,819	10,65
65	27	0,423	11,41	0,906	24,47
75	7	0,259	1,81	0,966	6,76
85	13	0,087	1,13	0,996	12,95
105	3	-0,259	-0,78	0,966	2,90
115	1	-0,423	-0,42	0,906	0,91
125	1	-0,574	-0,57	0,819	0,82
145	2	-0,819	-1,64	0,574	1,15
155	13	-0,906	-11,78	0,423	5,49
165	5	-0,966	-4,83	0,259	1,29
175	1	-0,996	-1,00	0,087	0,09
185	4	-0,966	-3,98	-0,087	-0,35
215	7	-0,819	-5,73	-0,574	-4,02
235	8	-0,574	-4,59	-0,819	-6,55
255	8	-0,259	-2,07	-0,966	-7,73
265	8	-0,087	-0,70	-0,966	-7,97
275	20	0,087	1,74	-0,996	-19,92
295	21	0,423	8,87	-0,906	-19,03
305	22	0,574	12,62	-0,819	-18,02
315	20	0,707	14,14	-0,707	-14,14
325	16	0,819	13,11	-0,574	-9,18
335	43	0,906	38,97	-0,423	-18,17
345	46	0,966	44,43	-0,259	-11,91
355	78	0,996	77,70	-0,087	-6,80
$\Sigma$	677		461,62		19,13

$$\bar{C} = 461,62/677 = 0,682; \bar{S} = 19,13/677 = 0,028;$$

$$\bar{R} = \sqrt{0,682^2 + 0,028^2} = 0,683; \cos m = \bar{C}/\bar{R} = 0,682/0,683 = 0,999;$$

$$\sin m = \bar{S}/\bar{R} = 0,028/0,683 = 0,041; m \approx 2,5^\circ.$$

Для сгруппированных данных формулы (8) принимают следующий вид:  $\bar{C} = \frac{1}{n} \sum_{j=1}^k n_j \cos \theta_j$ ;  $\bar{S} = \frac{1}{n} \sum_{j=1}^k n_j \sin \theta_j$ , где  $\theta$  — средняя точка  $j$ -го интервала группировки (предполагается, что все интервалы имеют одинаковые длины  $h = \frac{2\pi}{k}$ ,  $k$  — целое число), а  $n_j$  — частота, соответствующая  $j$ -му интервалу.

Расчеты выборочного кругового среднего направления для примеров, изображенных на рис. 8, 10, приведены в табл. 6 и 7.

*Выборочной круговой медианой* Мe называется точка  $B$  на окружности, обладающая двумя свойствами:

— половина точек выборки лежит по одну сторону от диаметра  $BQ$ ;

— большинство точек выборки ближе к  $B$ , чем к  $Q$  (см. рис. 10, a).

При малом количестве замеров эту характеристику можно легко найти по графику распределения точек на окружности. Так, например, очевидно, что для случая, приведенного на рис. 10, a, свойствам выборочной круговой медианы удовлетворяет точка  $B_2$ , соответствующая азимуту падения  $10^\circ$ .

Для угловых измерений, наряду с модой, используют также характеристику, называемую *антимодой*. Она соответствует значению с минимальной частотой.

Для некоторых угловых величин (например, для азимутов падения пород в областях развития линейной складчатости) свойственны распределения с двумя модами, отстоящими друг от друга на  $180^\circ$ .

В качестве выборочной характеристики рассеяния угловых величин удобно использовать *выборочную круговую дисперсию направлений*, которая рассчитывается по формуле  $S_g^2 = 1 - \bar{R}$ , где  $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$ .

## 2.4. ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ, ИСПОЛЬЗУЕМЫЕ В ГЕОЛОГИИ

Для аппроксимации (приближенного описания) эмпирических распределений свойств геологических объектов предлагалось использовать различные законы распределения — нормальный, логарифмически нормальный, биномиальный, Вейбулла, Пуассона, гамма- и бета-распределения и др. Однако многолетний опыт применения статистических методов в геологии показал, что в большинстве случаев можно ограничиться нормальным и логарифмически нормальным для непрерывных величин, биномиальным и Пуассона для дискретных и законом Мизеса для угловых величин.

*Нормальное распределение* непрерывное, полностью определяемое двумя параметрами — математическим ожиданием  $Mx$  и дисперсией  $\sigma^2$ :

$$F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-Mx)^2}{2\sigma^2}} dx.$$

Функция плотности такого распределения имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M_x)^2}{2\sigma^2}}.$$

Нормальное распределение симметрично относительно математического ожидания, следовательно его математическое ожидание, мода и медиана совпадают (рис. 11).

В практике статистического моделирования обычно используют не аналитические выражения нормального закона распределения, а его таблицы, составленные для распределения с математическим ожиданием, равным нулю, и дисперсией, равной 1 (нормированная функция Лапласа). Любое нормальное распределение можно привести к распределению с такими параметрами путем замены значений изучаемой случайной величины ( $X$ ) их отклонениями от математического ожидания, деленными на стандартное отклонение  $\tau = (x - M_x)/\sigma$ .

Решения многих статистических задач основаны на сравнении эмпирических распределений с функцией Лапласа  $\Phi(\tau)$ , поэтому полезно запомнить некоторые ее значения:  $\Phi(-3) \approx 0,0013$ ;  $\Phi(-2) \approx 0,023$ ;  $\Phi(-1) \approx 0,16$ ;  $\Phi(+1) \approx 0,84$ ;  $\Phi(+2) \approx 0,98$ ;  $\Phi(+3) \approx 0,999$ . В условиях нормального закона распределения вероятность значений, отличающихся от математического ожидания больше чем на 3 стандартных отклонения, настолько мала, что попадание их в выборку ограниченного объема можно считать событием практически невозможным, а большинство выборочных значений ( $\sim 95\%$ ) будет находиться в интервале  $M_x \pm 2\sigma$  (см. рис. 11). Таблицы функции Лапласа можно найти в работе Л. Н. Большева и Н. В. Смирнова.

Нормально распределенные случайные величины характеризуют такие свойства геологических объектов, которые зависят от очень большого количества независимых факторов, когда влияние каждого из них равномерное и незначительное.

Нормально распределенными являются случайные технические погрешности измерений и анализов проб. Близки к нормальному распределению содержаний породообразующих минералов и входящих в их состав химических элементов, содержаний некоторых полезных компонентов в рудах, когда они составляют целые проценты, а также распределения характеристик некоторых физико-механических свойств пород и руд — плотности, объемной массы, пористости. Иногда нормальный закон можно использовать для аппроксимации замеров мощностей тел полезных ископаемых.

К нормальному закону в определенных условиях сходятся или путем соответствующих преобразований сводятся некоторые другие распределения.

Логарифмически нормальным (логнормальным) называется закон, при котором нормально распределены логарифмы значений случайной величины. Такое распределение является положительно асимметричным и имеет положительный эксцесс (рис. 12). Математическое ожидание, мода и медиана логнормально распределенной случайной величины не совпадают, причем  $M_0 < M_e < M_x$ .

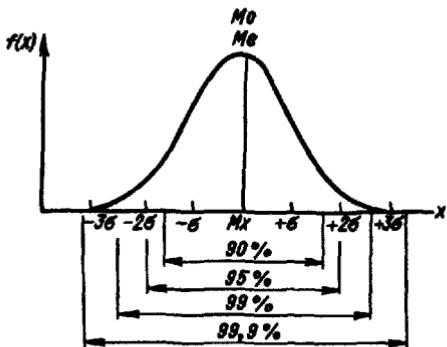


Рис. 11. График функции плотности нормального распределения

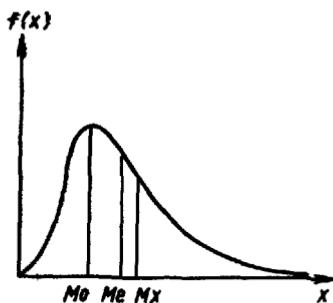


Рис. 12. График функции плотности логнормального распределения

Функция логнормального распределения соответствует функции нормального распределения, где вместо значений самой случайной величины вводятся характеристики, рассчитанные для ее логарифмов:

$$F(\ln x) = \frac{1}{\sigma_{\ln x} \sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(\ln x - M_{\ln x})^2}{2\sigma_{\ln x}^2}} d\ln x.$$

Функция плотности логнормального распределения принимает вид

$$f(\ln x) = \frac{1}{\sigma_{\ln x} \sqrt{2\pi}} e^{-\frac{(\ln x - M_{\ln x})^2}{2\sigma_{\ln x}^2}}.$$

Значение  $Me = e^{M_{\ln x}}$  для логнормально распределенной величины будет соответствовать медиане, а значение  $Mo = e^{M_{\ln x} - \sigma_{\ln x}^2}$  моде.

Появление логнормального распределения, как и нормального, связано с воздействием на исследуемое свойство множества факторов. Однако сила воздействия каждого фактора не одинакова, а пропорциональна изменению силы проявления данного фактора. В некоторых относительно редких случаях эффект пропорциональности можно предсказать исходя из характера процессов формирования изучаемых объектов. Так, например, акад. А. Н. Колмогоров теоретически обосновал, что логарифмически нормальному закону должно соответствовать распределение диаметров частиц при дроблении пород. Это подтверждается результатами изучения гранулометрического состава обломочной составляющей терригенных пород.

Близкие к логарифмически нормальному — положительно асимметричные распределения характерны также для свойств геологических объектов, изменчивость которых ограничена определенным жестко заданным интервалом, а их математическое ожидание смешено от середины этого интервала в сторону малых значений (рис. 13, б, г). Хотя содержание любого элемента в горной породе или руде не может быть меньше 0 и больше 100%, фактический интервал варьирования конкретных химических элементов значительно уже. Однако, несмотря на это, для большинства элементов асимметричность размаха варьиро-

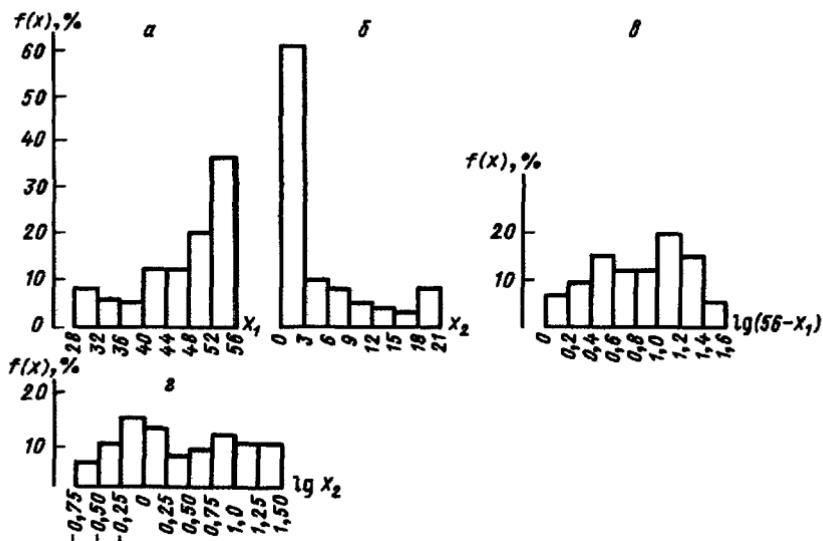


Рис. 13. Гистограммы распределений процентных содержаний химических элементов в известняке:  
 а—содержание  $\text{CaO}$  ( $X_1$ ); б—содержание  $\text{MgO}$  ( $X_2$ ); в—значения  $\lg(56-X_1)$ ;  
 г—значения  $\lg X_2$

вания относительно их математического ожидания очевидна. Например, среднее содержание (кларк) в литосфере золота и висмута составляет всего лишь 0,0000001%. Однако в рудах некоторых месторождений содержание золота по отдельным пробам доходит до 0,3%, а висмута—до 3%. Таким образом, в литосфере в целом размах варьирования содержаний этих элементов в сторону больших значений ( $x_{\max} - M_x$ ) в несколько миллионов раз больше, чем в сторону малых значений ( $M_x - x_{\min}$ ). Симметричные интервалы варьирования содержаний в породах характерны, как правило, лишь для элементов, кларк которых превышает 1%,—O, Si, Ca, Al, Fe, K, Mg, Na.

В отдельных породах и рудах концентрации этих элементов могут возрастать до нескольких десятков процентов. При этом математическое ожидание часто смещается от середины размаха варьирования в сторону больших значений, что приводит к возникновению отрицательно асимметричных распределений, являющихся как бы зеркальным отображением логнормального распределения. Такие *отраженные логнормальные распределения* характерны для содержаний Fe в богатых железных рудах, CaO в известняках (рис. 13, а), K и Na в залежах каменных солей и  $\text{Al}_2\text{O}_3$  в высококачественных бокситах. Отраженное логнормальное распределение легко преобразуется в обычное логнормальное путем вычитания изучаемой случайной величины  $X$  из константы  $C$ . За константу удобно принять содержание данного элемента в минерале-концентрате. При этом распределение значений  $C - X$  будет соответствовать логнормальному закону, а значений  $\eta = \lg(C - X)$ —нормальному. Эмпирический график плотности распределения для такого преобразования приведен на рис. 13, в.

При резко положительно асимметричном распределении изучаемых величин нередко возникает ситуация, когда мода оказывается в интервале значений ниже порога чувствительности анализов или измерений. В этом случае распределение выборочных данных оказывается как бы усеченным или сдвинутым со стороны малых значений на величину порога чувствительности. Такие усеченные логнормальные распределения можно привести к обычному логнормальному, если из каждого выборочного значения  $X$  вычесть некоторую постоянную  $C$ , меньшую минимального выборочного значения. Распределение величины  $\eta = \lg(X - C)$  в этом случае также может соответствовать нормальному закону.

Величину  $\eta$  с распределением, хорошо согласующимся с нормальным законом, можно также получить, если взять логарифмы от значений исследуемой величины  $X$ , к которым прибавлена определенная постоянная  $C$ :  $\eta = \lg(X + C)$ .

Величина константы  $C$  подбирается опытным путем. Такое распределение, названное логнормальным трехпараметрическим, широко используется при оценке запасов месторождений полезных ископаемых южноафриканскими, французскими и канадскими геологами.

Для описания свойств дискретных случайных величин в геологии чаще всего применяются биномиальное распределение и распределение Пуассона.

*Биномиальное распределение* удобно использовать в тех случаях, когда при проведении испытаний наблюдается одно из двух событий—*A* или *B*. Такая ситуация при изучении геологических объектов возникает довольно часто. Например, при разведке рудных тел с прерывистым характером оруденения путем бурения скважин по определенной сети каждая скважина может либо пересечь руды с промышленным содержанием полезного компонента (событие *A*), либо попасть в безрудные участки (событие *B*). Соотношение числа скважин, вскрывших оруденение ( $x$ ), к общему их числу в контуре рудного тела ( $n$ )—так называемый коэффициент рудоносности—широко используется как характеристика степени прерывистости оруденения и имеет важное значение при подсчете запасов полезных ископаемых, так как он выражает долю промышленной руды от общей площади или объема рудного тела.

В примере, приведенном на рис. 14, коэффициент рудоносности составил:

$$K_p = \frac{x}{n} = \frac{9}{28} \approx 0,32.$$

Однако, если сдвинуть сеть скважин в широтном направлении на половину расстояния между ними, то в рудные скопления попадут не 9, а лишь 8 скважин, и коэффициент рудоносности «уменьшится» до 0,29. При других вариантах смещения сети будут получены значения коэффициентов рудоносности как большие, так и меньшие первоначального. Следовательно, значения коэффициентов рудоносности по каждому варианту расположения скважин являются



Рис. 14. Распределение кондиционного оруденения по уступу карьера:

1—площади кондиционных руд по результатам отработки;  
2—скважины детальной разведки ( $\alpha$ —вскрывшие руду,  $\beta$ —вскрывшие безрудные участки)

лишь оценками неизвестного истинного значения доли рудной площади.

Вероятность появления тех или иных значений оценок коэффициента рудоносности зависит от истинного соотношения рудных и безрудных площадей и общего количества скважин и может быть определена по формуле  $P_n(x) = C_n^x p^x (1-p)^{n-x}$ , где  $P_n(x)$ —вероятность того, что при количестве испытаний  $n$  событие  $A$  произойдет  $x$  раз (в нашем примере это выборочное значение коэффициента рудоносности);  $p$ —вероятность события  $A$ , т. е. истинная доля рудной площади;  $C_n^x$ —число сочетаний из  $n$  по  $x$ .

Совокупность вероятностей  $P_n(x)$  при  $x=0, 1, 2, \dots, n$  называется биномиальным распределением. Сумма всех возможных значений  $P_n(x)$  равна 1.

Величина  $C_n^x$  называется биномиальным коэффициентом, а величины  $n$  и  $p$ —параметрами биномиального распределения.

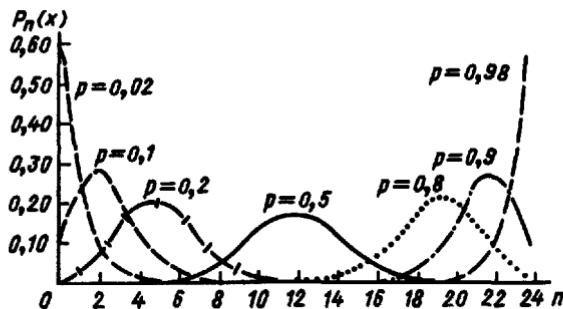


Рис. 15. Графики функций биномиального распределения в зависимости от  $p$  при постоянном  $n=23$

Биномиальные коэффициенты могут быть найдены с помощью так называемого треугольника Паскаля.

Средним значением (математическим ожиданием) биномиального распределения называют величину  $Mx=np$ , а дисперсией — величину  $\sigma^2=np(1-p)$ .

Коэффициент асимметрии для биномиального распределения определяется по формуле

$$A = \frac{1-2p}{\sqrt{np(1-p)}}, \quad (10)$$

а коэффициент эксцесса — по формуле

$$E = \frac{1-6p+6p^2}{np(1-p)}. \quad (11)$$

Из формул (10) и (11) очевидно, что с увеличением  $n$  коэффициенты  $A$  и  $E$  биномиального распределения стремятся к нулю. При этом само распределение приближается к нормальному. При  $p=0,5$  биномиальное распределение становится симметричным (рис. 15).

Операции центрирования и нормирования позволяют получить величину  $\frac{(x-np)}{\sqrt{np(1-p)}}$ , которая приближенно подчиняется нормальному распределению с нулевым средним и дисперсией, равной 1 (нормированная функция Лапласа).

Биномиальный закон распределения применяется при анализе частоты встречаемости ископаемых организмов определенного вида в различных горизонтах осадочных пород, количества различных минералов в шлихах, случаев аварий и производственного травматизма при проведении геологоразведочных работ и т. п.

Если число испытаний велико, а вероятность появления случайного события в каждом испытании очень мала, то для описания вероятностей того, что событие  $A$  в серии из  $n$  испытаний произойдет  $X$  раз, используется распределение Пуассона:  $P_n(x) = \lambda^x e^{-\lambda} / x!$  (где  $\lambda = np$ , т. е. среднее число появления события  $A$  в  $n$  испытаниях).

Для распределения Пуассона среднее число и дисперсия совпадают и равны  $\lambda$ . Оно используется для описания процессов радиоактивного распада химических элементов, вероятности встречи в пробах крупных алмазов и самородков золота, вероятности обнаружения выходов тел полезных ископаемых при проведении поисковых работ и т. п.

Для аппроксимации распределений угловых величин часто привлекают равномерное распределение и распределение Мизеса.

Случайная угловая величина называется равномерно распределенной, если ее плотность распределения вероятностей выражается формулой  $f(\theta) = 1/2\pi$ .

Для этого случая характерно отсутствие концентрации значений около какого-нибудь направления. Выборочные значения имеют максимальный разброс, поскольку угловая дисперсия при этом распределении равна 1.

Это распределение характерно, например, для замеров ориентировки обломков в делювиальных отложениях и эруптивных брекчиях.

Для распределения Мизеса плотность распределения вероятностей выражается формулой

$$f(\theta) = \frac{1}{2\pi J_0(k)} e^{k \cos(\theta - \mu)}$$

при  $|\mu| < \infty$  и  $k > 0$ ,

где

$$J_0(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{k \cos \theta} d\theta = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} \left(\frac{k}{2}\right)^{2r}.$$

Распределение Мизеса, так же как и нормальное распределение, определяется двумя параметрами —  $\mu$  и  $k$ . Причем  $\mu$  — круговое среднее направление случайной угловой величины — схоже с математическим ожиданием  $Mx$ , а параметр  $k$  можно рассматривать как характеристику концентрации распределения около  $\mu$ . Распределение Мизеса при  $k=0$  превращается в равномерное, а при  $k \rightarrow \infty$ , с параметрами  $\mu$  и  $k$  оно асимптотически ведет себя как нормальное с параметрами  $Mx=\mu$  и  $\sigma^2=1/k$ . Таким образом параметр  $1/k$  в распределении Мизеса играет ту же роль, что и дисперсия в случае нормального распределения.

Графики плотности распределения Мизеса с параметрами  $\mu=0$  и  $k=0,5, 1, 2, 4$  изображены на рис. 16.

Аналогично нормальному распределение Мизеса одномодально и симметрично. Его удобно использовать для решения задач проверки

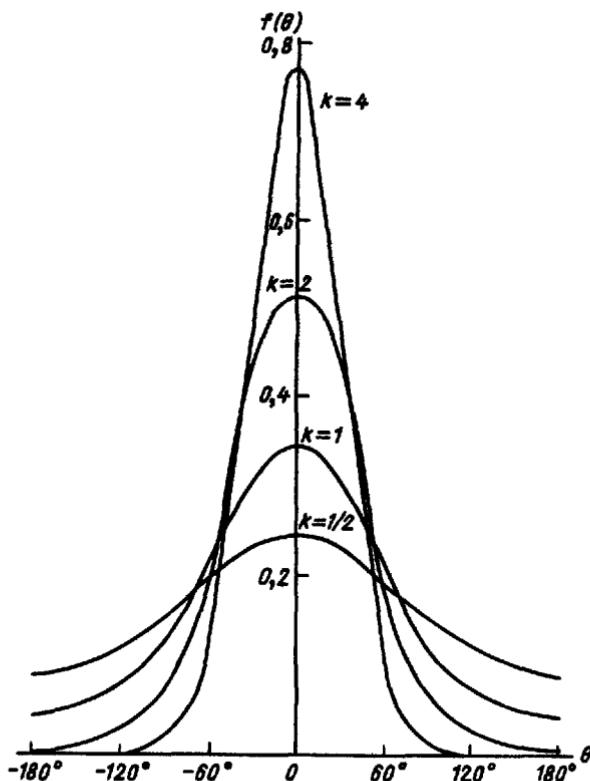


Рис. 16. Графики функций плотности распределения Мизеса при различных значениях  $k$

гипотез и нахождения точечных и интервальных оценок параметров распределений угловых величин.

## 2.5. ПРОВЕРКА ГИПОТЕЗ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ПАРАМЕТРОВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

Большинство статистических методов решения геологических задач основано на использовании свойств тех или иных законов распределения. Однако геолог обычно не может заранее знать какими свойствами будут обладать полученные в результате исследования выборочные совокупности. Поэтому решению конкретных задач предшествует этап сравнения эмпирических распределений с известными теоретическими.

Способы проверки гипотез о соответствии эмпирического распределения определенным законам можно разделить на графические и аналитические.

Графические способы проверки гипотез о типе распределения заключаются в построении эмпирических графиков-гистограмм, кумулят или диаграмм розы наблюдений и сравнении их с теоретическими кривыми функций распределения.

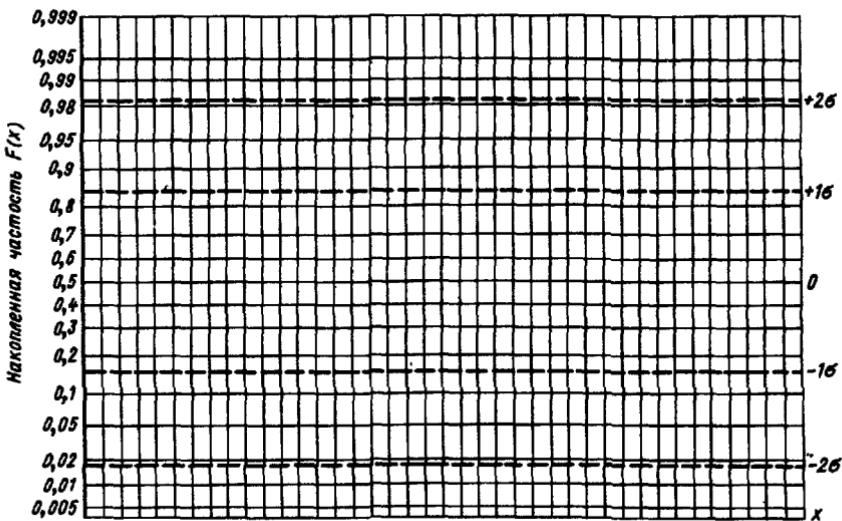


Рис. 17. Вероятностный бланк для проверки гипотезы о нормальном законе распределения

Для графической проверки гипотезы о соответствии выборочных данных нормальному закону удобно использовать специальный вероятностный бланк (рис. 17). Масштаб по оси ординат, где откладываются накопленные частоты, на этом бланке изменен таким образом, что построенная в данных координатах функция  $F(x)$  нормально распределенной случайной величины имеет вид прямой линии. О сходстве выборочного распределения с нормальным судят по близости построенной для него кумуляты к прямой линии. Для проверки гипотезы о соответствии эмпирического распределения логнормальному закону используется аналогичный бланк, отличающийся лишь тем, что по оси абсцисс принят не линейный, а логарифмический масштаб (рис. 18).

Графические способы просты, наглядны и не требуют трудоемких вычислений. Однако это, по существу, качественные способы, так как вопрос о сходстве или отличии эмпирических и теоретических графиков решается субъективно, «на глаз», а это может привести к ошибкам, в особенности при не очень большом объеме выборки. Судить о линейности эмпирических графиков накопленных частот можно лишь в том случае, если они построены не менее чем по 6—7 классам группирования, с количеством замеров в каждом классе не меньше 5—10. Поэтому при объеме выборки меньше 50—60 значений этим способом пользоваться не рекомендуется.

Наиболее надежно гипотеза о соответствии эмпирических распределений нормальному или логарифмически нормальному закону проверяется с помощью критерия Пирсона ( $\chi^2$ ). Этот способ заключается в разделении выборочных данных на  $k$  класс-интервалов и сравнении

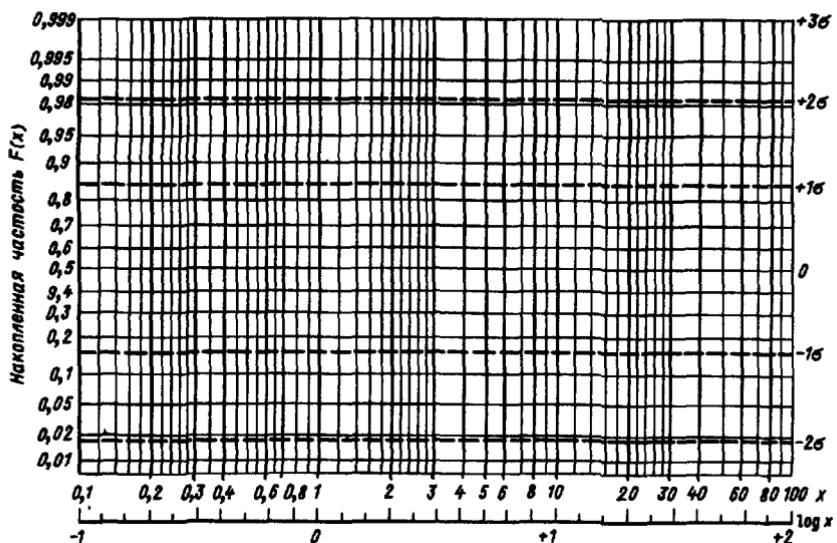


Рис. 18. Вероятностный бланк для проверки гипотезы о логнормальном законе распределения

эмпирических частот по классам ( $n_j$ ) с теоретическими ( $n'_j$ ) для нормального закона распределения.

Предварительно по выборочным данным с помощью формул (4), (5) вычисляют оценки среднего значения  $\bar{x}$  и дисперсии  $S^2$  и производят операцию нормирования и центрирования значений  $x_j$ , соответствующих границам класс-интервалов:  $t_j = (x_j - \bar{x})/S$ .

Это дает возможность воспользоваться для расчета теоретических частот таблицей функции Лапласа (прил. 11). С ее помощью для каждой величины  $t_j$  можно определить значение функции распределения  $\Phi(t_j)$  и теоретические вероятности попадания случайной величины в класс-интервал от  $t_j$  до  $t_{j+1}$ :  $\Delta\Phi(t_j) = \Phi(t_{j+1}) - \Phi(t_j)$ .

Теоретические частоты по класс-интервалам находятся путем перемножения соответствующих значений  $\Delta\Phi(t_j)$  на объем изучаемой выборки:  $\bar{n}_j = \Delta\Phi(t_j)n$ .

В случае соответствия эмпирических данных нормальному закону распределения величина

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - \bar{n}'_j)^2}{\bar{n}'_j}$$

имеет распределение  $\chi^2$  с  $k-3$  степенями свободы, для которого составлены специальные таблицы. Полученное по выборочным данным значение  $\chi^2$  сравнивается с табличным для принятого уровня значимости  $\alpha$  и  $k-3$  степеней свободы.

Если  $\chi^2 < \chi^2_{\alpha, k-3}$ , то гипотеза о соответствии выборочных данных нормальному распределению не отвергается.

Гипотеза о соответствии выборочных данных логнормальному закону проверяется аналогичным образом, но вместо самих значений

изучаемой величины берутся их логарифмы. Для успешного применения этого способа границы класс-интервалов необходимо подбирать таким образом, чтобы теоретические частоты по каждому классу не были бы меньше 5, а количество класс-интервалов должно быть не меньше 6—7. Поэтому способ проверки гипотез о законе распределения с помощью критерия  $\chi^2$  обычно применяется в том случае, когда объем выборки превышает 60 значений.

Для выборок меньшего объема лучше использовать способ, основанный на расчете показателей асимметрии и эксцесса. В случае нормального распределения эти показатели равны нулю (см. разд. 2.3), а их выборочные оценки, вычисляемые по формулам

$$A = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{nS^3} \quad \text{и} \quad E = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{nS^4} - 3,$$

имеют асимптотически нормальное распределение. В случае логнормального распределения теми же свойствами обладают  $A$  и  $E$ , рассчитанные для логарифмов изучаемых величин.

Обычно в качестве критериев соответствия эмпирических распределений нормальному используют отношения выборочных показателей асимметрии и эксцесса к их стандартным отклонениям, которые приближенно оцениваются по формулам  $\sigma_A \approx \sqrt{6/n}$ ;  $\sigma_E \approx \sqrt{24/n}$  (где  $n$  — количество замеров в выборке).

Если эти отношения по абсолютной величине превышают 3, гипотеза о нормальном распределении отвергается. Для нормального распределения вероятность того, что выборочное значение этих отношений будет отличаться от математического ожидания больше чем на 3 стандартных отклонения, очень мала (равна 0,001).

Для проверки данной гипотезы при других доверительных вероятностях (уровнях значимости) можно воспользоваться специальными таблицами распределения показателей асимметрии и эксцесса. Однако распределение оценок асимметрии и эксцесса по выборкам объемом меньше 40—50 значений заметно отличается от нормального, поэтому высказывались сомнения о возможности применения данного метода в случае малых выборок. В то же время в геологической практике, особенно при разведке месторождений полезных ископаемых, часто приходится иметь дело с выборками объемом 10—30 значений.

Характер ошибок при проверке гипотезы о виде распределения по оценкам асимметрии и эксцесса в условиях малых выборок был исследован на числовой модели, составленной по результатам эксплуатационного опробования молибденового штокверка. На одном из уступов карьера исходную густую сеть опробованных буровзрывных скважин разредили таким образом, что было получено 122 варианта по 9—11 скважин, 77 вариантов по 19—22 скважин и 44 варианта по 29—31 скважине.

По совокупности всех опробованных скважин (1117) распределение содержания молибдена (рис. 19) хорошо согласуется с логнормальным законом. Гипотеза о нормальном распределении отвергается.

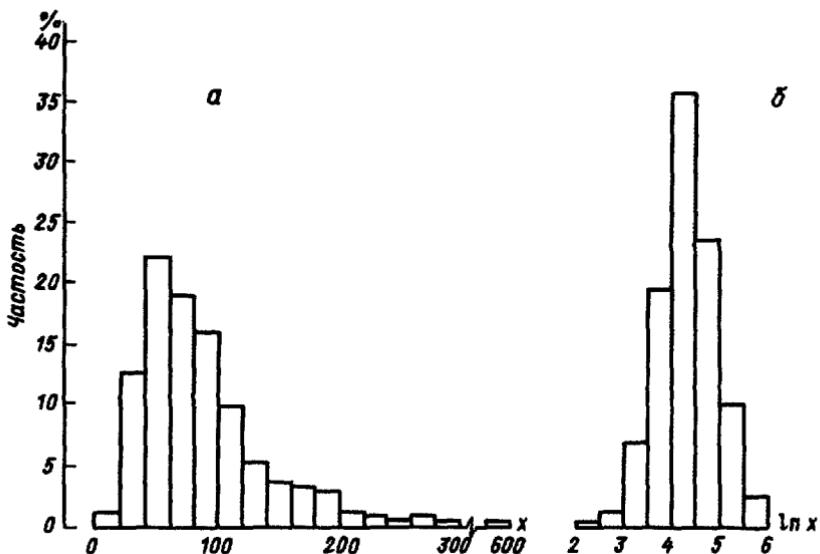


Рис. 19. Гистограммы частотного распределения содержаний молибдена:  
а—содержания молибдена в условных единицах; б—логарифмы содержаний

По выборкам малого объема гипотеза о соответствии нормальному и логнормальному закону проверялась для критических значений  $K$ , равных 3, 2 и 1. При этом возникало четыре возможных варианта решения:

- 1) гипотеза о нормальном законе отвергается, а гипотеза о логнормальном принимается (правильное решение);
- 2) отвергаются обе гипотезы (ошибка I рода);
- 3) обе гипотезы принимаются (ошибка II рода);
- 4) отвергается гипотеза о логнормальном законе, а гипотеза о нормальном принимается (ошибка I и II рода).

По выборкам объемом в 30 значений максимальная вероятность правильного решения достигается при критическом значении для отношений асимметрии и эксцесса к их стандартам ( $K$ ), равном 3. Однако с уменьшением объема выборки при  $K=3$  резко возрастает вероятность ошибки II рода, которая заключается в том, что не отвергается ошибочная гипотеза о нормальном распределении (рис. 20) (табл. 8). По данным эксперимента можно сделать вывод, что проверять гипотезу о виде закона распределения по выборкам объемом в 15—10 и менее значений нецелесообразно, так как вероятность правильного вывода весьма мала, а по выборкам объемом 15—25 значений критическую величину отношений оценок асимметрии и эксцесса к их стандартам лучше принимать равной 2.

Гипотезу о соответствии эмпирических распределений биномциальному закону можно проверить с помощью критерия  $\chi^2$ . Для этого теоретические частоты вычисляются по формуле:  $n_i = NC_m^x p^x (1-p)^{m-x}$  (где  $N$ —число серий испытаний;  $m$ —число испытаний в серии;  $x$ —частота события в серии испытаний;  $p$ —вероятность изучаемого события).

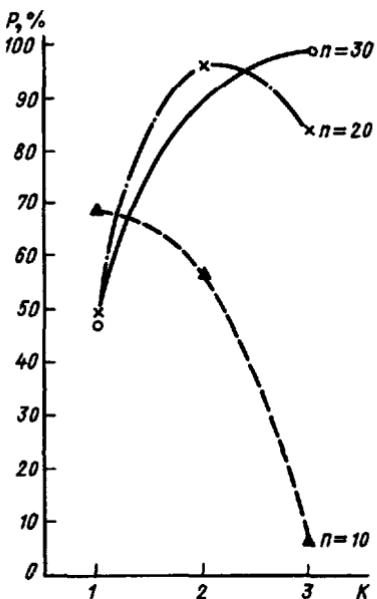


Рис. 20. Вероятности принятия правильного вывода при проверке гипотез о законе распределения содержания молибдена по выборкам малого объема при различных предельных значениях показателей асимметрии и эксцесса ( $K$ )

Таблица 8. Результаты проверки гипотез о законе распределения содержания молибдена по малым выборкам

Объем выборки	Критическое значение $K$	Количество случаев, %			
		правильный вывод	ошибка I рода	ошибка II рода	ошибка I и II рода
30	1	47	53	—	—
	2	91	9	—	—
	3	98	2	—	—
20	1	48	52	—	—
	2	96	1	3	—
	3	84	—	16	—
10	1	68	27	—	5
	2	57	—	43	—
	3	7	—	92	—

Число степеней свободы при этом принимается  $k=N-2$ .

Гипотезу о соответствии распределения закону Пуассона также можно проверить с помощью критерия  $\chi^2$ .

В серии из  $N$  испытаний случайное событие может реализоваться в каждом испытании  $i$  раз. При этом  $i=0, 1, 2, \dots, r$ , где  $r$  — максимальное число наблюдавшихся событий в одном испытании, а  $n_i$  — количество испытаний, когда случайное событие наблюдалось  $i$  раз. По выборочным данным находят оценку средней вероятности случайного события:

$$\lambda \approx \bar{x}_o = \frac{\sum_{i=1}^r n_i i}{N}.$$

Подставив ее в формулу распределения Пуассона

$$P_n(i) = \frac{\lambda^i e^{-\lambda}}{i!},$$

рассчитывают теоретические вероятности  $P_i$  появления ровно  $i$  событий в одном испытании в серии из  $N$  испытаний. Теоретические частоты таких случаев в серии из испытаний находят по формуле  $n'_i = N \cdot P_i$ . Затем по разнице теоретических ( $n'_i$ ) и фактических ( $n_i$ ) частот определяют эмпирическое значение критерия  $\chi^2$ , которое сравнивают с табличным для заданного уровня значимости  $\alpha$  и числа степеней свободы  $K = s - 2$ , где  $s$  — число различных групп выборки, обычно совпадающее с количеством наблюдавшихся различных вариантов реализации случайного события в одном испытании, т. е.  $s = r + 1$ . Исключение составляют случаи, когда теоретические частоты  $n'_i$  для некоторых значений  $i$  очень малы (меньше 5). Обычно это группы наблюдений, где  $i$  близко к  $r$ . Тогда теоретические и фактические частоты для соседних групп наблюдений суммируются, и общее количество групп уменьшается.

**Пример.** На месторождении алмазоносных кимберлитов отобрано 200 проб массой 2 т. В каждой пробе обнаружено от 0 до 4 алмазов размером больше 1 мм. Число проб  $n_i$  с разным количеством  $i$  алмазов класса +1 мм приведено в табл. 9. Требуется проверить гипотезу о соответствии распределения количества алмазов по закону Пуассона.

Таблица 9. Проверка гипотезы о соответствии распределения количества алмазов в пробах по закону Пуассона

$i$	$n_i$	$P_i$	$n'_i$	$n_i - n'_i$	$(n_i - n'_i)^2$	$(n_i - n'_i)^2 / n'_i$
1	2	3	4	5	6	7
0	123	0,580	116	7	49	0,4224
1	52	0,316	63	-11	121	1,9206
2	20	0,086	17	3	9	0,5294
3	3	0,016	3	1	1	0,2500
4	2	0,002	1			

$$\chi^2 = 3,12$$

Средняя вероятность ( $\bar{x}_n$ ) попадания кристалла в пробу равна

$$\frac{\sum_{i=1}^r n_i x_i}{N} = \frac{123 \cdot 0 + 52 \cdot 1 + 20 \cdot 2 + 3 \cdot 3 + 2 \cdot 4}{200} = 0,545.$$

Следовательно, формула закона Пуассона принимает вид

$$P_{200}(i) = \frac{(0,545)^i e^{-0,545}}{i!}.$$

Рассчитанные по этой формуле значения теоретических вероятностей приведены в графе 3 табл. 9. Теоретические частоты получаем путем умножения значений  $P_i$  на объем выборки равный 200 (табл. 9, графа 4). В связи с тем, что частоты при  $i=3$  и  $i=4$  малы, они объединяются. Эмпирическое значение критерия  $\chi^2$  для данной выборки равно 3,12. По таблицам распределения  $\chi^2$  при  $\alpha=0,05$  и  $K=4-2=2$  находим критическое значение:  $\chi^2_{(0,05; 2)}=5,991$ .

Так как  $\chi^2 > \chi^2_{\alpha}$ , то нет оснований отвергать гипотезу о соответствии распределения алмазов класса +1 мм в пробах массой 2 т закону Пуассона.

Проверка гипотезы о равномерном распределении угловой величины  $\vartheta$  при малом объеме выборки осуществляется с помощью критерия равномерности Релея. По выборочным данным вычисляется статистика

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}, \text{ где } \bar{C} = \frac{\sum_{i=1}^n \cos \vartheta_i}{n}, \text{ а } \bar{S} = \frac{\sum_{i=1}^n \sin \vartheta_i}{n},$$

которая сравнивается с ее критическим значением  $\bar{R}_0$  для этого объема выборки  $n$  и принятого уровня значимости  $\alpha$  (прил. 1).

При  $n > 100$  можно воспользоваться тем, что величина  $2n\bar{R}^2$  распределена приближенно по закону  $\chi^2$  — Пуассона с двумя степенями свободы. Критические значения этой величины для различных уровней значимости также приведены в прил. 1.

**Пример.** На месторождении бокситов, отработка которого ведется открытым способом, за 5 лет произошло 19 случаев оползания бортов карьера. Распределение этих случаев по различным месяцам приведено в графе 2 табл. 10. На основании этих данных высказано предположение, что устойчивость бортов карьера снижается в весенний период. Обоснованность этого предположения можно оценить путем проверки гипотезы о равномерном распределении случаев оползания бортов карьера в течение года. Рассматривая год как цикл с периодом  $2\pi$ , определим, что каждому месяцу будет соответствовать интервал на окружности, равный  $2\pi/12$ , т. е.  $30^\circ$ . Середины интервалов, соответствующих каждому месяцу начиная с января, приведены в графе 3 табл. 10. Выполнив необходимые расчеты (табл. 10, графы 4—7), получим, что  $\bar{R}=0,402$  и превышает критическое значение  $\bar{R}_0=0,394$  для  $n=19$  и  $\alpha=0,05$  (прил. 1). Таким образом, гипотеза о равномерном распределении случаев оползания бортов карьера в течение года отвергается. Следовательно, нет оснований отрицать возможность сезонного изменения устойчивости бортов карьера.

Таблица 10. Проверка гипотезы о равномерном распределении случаев оползания бортов карьера в течение года

Месяц	Количество случаев оползания $n_i$	Средняя точка $\vartheta_i$	$\sin \vartheta_i$	$\cos \vartheta_i$	$n_i \sin \vartheta_i$	$n_i \cos \vartheta_i$
1	2	3	4	5	6	7
I	—	15°	—	—	—	—
II	—	45	—	—	—	—
III	3	75	0,966	0,259	2,898	0,777
IV	6	105	0,966	-0,259	5,796	-1,554
V	2	135	0,707	-0,707	1,414	-1,414
VI	1	165	0,259	-0,966	0,259	-0,966
VII	2	195	-0,259	-0,966	-0,518	-1,932
VIII	1	225	-0,707	-0,707	-0,707	-0,707
IX	2	255	-0,966	-0,259	-1,932	-0,518
X	—	285	—	—	—	—
XI	2	315	-0,707	0,707	-1,414	1,414
XII	—	345	—	—	—	—
$\Sigma$					5,796	-4,970

$$\bar{C}^2 = \left( \frac{-4,970}{19} \right)^2 = 0,0684; \quad \bar{S}^2 = \left( \frac{5,796}{19} \right)^2 = 0,0931; \quad \bar{R} = \sqrt{0,0684 + 0,0931} = 0,402.$$

*Проверка гипотезы о соответствии распределения случайной угловой величины распределению Мизеса* осуществляется следующим образом.

Выборочные значения случайной угловой величины группируются в класс-интервалы так, чтобы для большинства классов количество замеров составляло не менее 2—5.

По выборочным данным с помощью формул (8), (9) рассчитываются оценки  $m$  и  $\bar{R}$ . Оценка параметра  $\bar{R}$  используется для нахождения по специальным таблицам (прил. 2) оценки  $k$  параметра концентрации  $k$  распределения Мизеса. Получив оценки  $m$  и  $k$  по таблицам распределения Мизеса (прил. 12) можно найти теоретические вероятности, соответствующие границам класс-интервалов. Таблицы Мизеса составлены для распределения с параметрами  $m = 180^\circ$  и  $k = 0; 0,2; \dots; 10$ . Поэтому значения изучаемой угловой величины необходимо центрировать, т. е. как бы «сдвинуть» на величину  $180^\circ - m$  или  $m - 180^\circ$ , чтобы выборочное угловое среднее направление  $m$  совпало с направлением  $180^\circ$ . Для упрощения пользования таблицей Мизеса выборочные данные целесообразно перегруппировать таким образом, чтобы выборочное значение  $m$  совпало с границей одного из класс-интервалов. Теоретическая вероятность попадания случайной угловой величины в каждый класс-интервал определяется как разница между вероятностями, соответствующими верхней и нижней границам данного класс-интервала:  $\Delta p_i = p_i - p_{i+1}$ , а теоретическая частота рассчитывается путем умножения  $\Delta p_i$  на объем выборки:

$$n'_i = \Delta p_i \cdot n. \quad (12)$$

Проверка гипотезы о соответствии выборочных данных распределению Мизеса производится путем сравнения теоретических и фактических частот по критерию Пирсона  $\chi^2$  при числе степеней свободы  $K = k - 3$ , где  $k$  — число классов группирования.

*Пример.* В табл. 11 приведены замеры азимутов падения кварцевых прожилков по документации канав и рудопроявления золота. Количество замеров невелико, поэтому ширину класс-интервалов при группировании целесообразно принять равной  $20^\circ$ . По структурированным данным с помощью расчетов (табл. 12) находим, что выборочное среднее направление азимутов падения прожилков равно  $173,5^\circ$ . По прил. 2 определяем, что для  $\bar{R} = 0,33$  оценка параметра концентрации  $k$  равна 0,7. В таблице Мизеса значения параметра  $k$  приведены с точностью до 0,1, поэтому вычислять оценку  $k$  с большой точностью нецелесообразно.

Таблица 11. Замеры азимутов (в градусах) падения кварцевых прожилков

№ п/п	Азимут						
1	132	13	330	25	178	37	105
2	302	14	88	26	335	38	130
3	304	15	191	27	110	39	144
4	162	16	198	28	112	40	177
5	130	17	325	29	200	41	42
6	58	18	214	30	257	42	190
7	159	19	211	31	270	43	169
8	144	20	199	32	171	44	41

№ п/п	Азимут						
9	315	21	124	33	141	45	205
10	162	22	84	34	260	46	225
11	318	23	181	35	185	47	270
12	92	24	3	36	15	48	260

Таблица 12. Оценка параметров  $m$  и  $k$  азимутов падения кварцевых прожилков

Азимут, градусы	Частота $n_i$	Средний азимут $\theta_i$	$\sin \theta_i$	$n_i \sin \theta_i$	$\cos \theta_i$	$n_i \cos \theta_i$
1	2	3	4	5	6	7
0—20	2	10	0,174	0,348	0,985	1,970
20—40	—	30	—	—	—	—
40—60	3	50	0,766	2,298	0,643	1,929
60—80	—	70	—	—	—	—
80—100	3	90	1,000	3,000	0	0
100—120	3	110	0,940	2,820	-0,342	-1,026
120—140	4	130	0,766	3,064	-0,643	-2,572
140—160	4	150	0,500	2,000	-0,866	-3,464
160—180	6	170	0,174	1,044	-0,985	-5,910
180—200	7	190	-0,174	-1,218	-0,985	-6,895
200—220	3	210	-0,500	-1,500	-0,866	-2,598
220—240	1	230	-0,766	-0,766	-0,643	-0,643
240—260	3	250	-0,940	-2,820	-0,342	-1,026
260—280	2	270	-1,000	-2,000	0	0
280—300	—	290	—	—	—	—
300—320	4	310	-0,766	-3,064	0,643	2,572
320—340	3	330	-0,500	-1,500	0,866	2,598
340—360	—	350	—	—	—	—
$\Sigma$	48			1,706		-15,065

$$\bar{S} = \frac{1,706}{48} = 0,0355; \quad S^2 = 0,00126;$$

$$\bar{C} = \frac{-15,065}{48} = -0,3139; \quad C^2 = 0,0985;$$

$$\bar{R} = \sqrt{0,00126 + 0,0985} = 0,316;$$

$$\sin m = 0,0355 / 0,316 = 0,112; \quad \cos m = \frac{-0,3139}{0,316} = -0,993;$$

$$m \approx 173,5^\circ; \quad R = \bar{R}n = 15,17; \quad k \approx 0,7.$$

Для упрощения дальнейших расчетов сместим границы класс-интервалов таким образом, чтобы одна из них совпадала со средним круговым направлением, т. е. с  $173,5^\circ$  (табл. 13, графа 1), и определим частоты для новых класс-интервалов (табл. 13, графа 2). Для нахождения теоретических вероятностей попадания замеров азимутов в каждый класс-интервал сдвигем границы этих интервалов на  $180—173,5^\circ$ , чтобы выборочное среднее направление ( $173,5^\circ$ ) совпало со средним направлением теоретического распределения Мизеса.

Таблица 13. Проверка гипотезы о соответствии распределения азимутов (в градусах) падения кварцевых прожилков распределению Мизесса

Азимут $\theta$	$n_i$	Азимут $\theta + (180^\circ - m)$	$p_i$	$n'_i$	$n_i$	$n'_i$	$n_i - n'_i$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	2	3	4	5	6	7	8	9
353,5—13,5	1	0—20	0,025	1,20				
13,5—33,5		20—40	0,027	1,30				
33,5—53,5		40—60	0,032	1,54				
53,5—73,5		60—80	0,039	1,87	5	5,91	-0,91	0,140
73,5—93,5	3	80—100	0,049	2,35				
93,5—113,5		100—120	0,063	3,02	6	5,37	0,63	0,074
113,5—133,5	4	120—140	0,077	3,70				
133,5—153,5		140—160	0,090	4,32	7	8,02	-1,02	0,130
153,5—173,5	5	160—180	0,098	4,70	5	4,70	0,30	0,019
173,5—193,5	6	180—200	0,098	4,70	6	4,70	1,30	0,360
193,5—213,5	5	200—220	0,090	4,32				
213,5—233,5		220—240	0,077	3,70	7	8,02	-1,02	0,130
233,5—253,5	—	240—260	0,063	3,02				
253,5—273,5		260—280	0,049	2,35	5	5,37	-0,37	0,025
273,5—293,5	—	280—300	0,039	1,87				
293,5—313,5		300—320	0,032	1,54				
313,5—333,5	4	320—340	0,027	1,30				
333,5—353,5		340—360	0,025	1,20	7	5,91	1,09	0,201
$\Sigma$	48				48	48	0	1,079

Значения теоретических вероятностей  $p_i$  находят по разности значений распределения Мизесса для верхней и нижней границ класса. Так как последнее симметрично относительно среднего направления, в прил. 12 вероятности приведены только для интервала  $0—180^\circ$ . Теоретические вероятности для углов  $\theta$  больше  $180^\circ$  равны вероятностям для углов  $360^\circ - \theta$ .

В связи с тем, что рассчитанные по формуле (12) теоретические частоты  $n'_i$ , в данном примере для большинства классов, оказались меньше 5 (см. табл. 13, графа 5), ширину класс-интервалов для дальнейших расчетов целесообразно увеличить, просуммировав значения теоретических и фактических частот по соседним классам (см. табл. 13, графа 6 и 7).

Число класс-интервалов после их объединения равно 8, поэтому полученное значение критерия  $\chi^2 = 1,079$  сравниваем с табличными для числа степеней свободы, равного 5. Критическое значение

$\chi^2 = 1,145$  при  $K = 5$  и доверительной вероятности 0,95. Следовательно, гипотеза о соответствии распределения азимутов падения прожилков закону Мизеса не отвергается.

Принятие гипотезы о соответствии изучаемого свойства определенному закону не исключает возможного соответствия этих же выборочных данных другому теоретическому распределению. В геологической практике нередки случаи, когда по выборочным данным не отвергается гипотеза о их соответствии как нормальному, так и логнормальному закону, а распределения дискретных случайных величин могут удовлетворительно аппроксимироваться как биномиальным, так и нормальными законами и т. п. В этих случаях при выборе теоретической модели распределения необходимо учитывать характер решаемой геологической задачи, свойства оценок параметров распределений различного типа по выборкам имеющегося объема и наличие соответствующих статистических таблиц. При этом предпочтение следует отдавать наиболее простым и хорошо изученным распределениям.

## 2.6. ТОЧЕЧНЫЕ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

Большинство геологических объектов отличается сильной изменчивостью свойств, определяемых по единичным замерам. Поэтому в практике геологических исследований часто возникает необходимость в оценке средних значений этих свойств и количественном выражении степени их изменчивости. Эти числовые характеристики используются при решении весьма широкого круга задач в различных областях геологии.

Средние содержания химических элементов в каких-либо геохимических системах — литосфере, совокупности пород определенного состава, интрузивном комплексе и т. п., называемые «кларками», являются важнейшими характеристиками этих систем и широко используются при прогнозировании и поисках месторождений полезных ископаемых.

По среднему содержанию  $\text{SiO}_2$  решается вопрос о принадлежности интрузивной породы к группе кислых, средних и основных.

На средних размерах обломочных частиц основана классификация терригенных пород, а по изменчивости этого параметра можно судить о степени сортированности обломочного материала.

Средние значения параметров тел полезных ископаемых — их мощности, содержания полезных компонентов в руде, объемные массы руды, коэффициенты рудоносности — входят в формулы подсчета запасов при промышленной оценке месторождений.

Для оценки инженерно-геологических свойств грунтов и пород при строительстве промышленных объектов и сооружений используются средние значения и характеристики изменчивости таких параметров, как плотность, коэффициент крепости, сопротивление сжатию и сдвигу, коэффициент внутреннего трения и т. д.

Оценки степени изменчивости свойств геологических объектов — дисперсии, стандартного отклонения и коэффициента вариации используются для определения величин возможных ошибок при расчетах средних значений этих свойств по выборочным данным, а также для определения количества наблюдений, необходимого для вычисления средних значений с наперед заданной точностью.

Статистические оценки могут быть *точечными* и *интервальными*. При точечной оценке неизвестная характеристика случайной величины оценивается некоторым числом, а при интервальной — некоторым интервалом значений. В пределах последнего с заданной вероятностью должно находиться истинное значение оцениваемой величины.

Точечные оценки должны удовлетворять требованиям состоятельности, несмещенности и максимальной эффективности.

Состоятельной называется оценка, сходящаяся по вероятности к оцениваемому параметру с увеличением объема выборки. Таким образом, эта оценка может обладать систематической ошибкой при малом объеме выборки, но величина ошибки асимптотически уменьшается до нуля при увеличении объема выборки до бесконечности.

Несмещенной называется оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки, т. е. не имеющая систематической ошибки.

Максимально-эффективной называется оценка, обладающая минимально возможной дисперсией (минимальной случайной ошибкой) при фиксированном числе наблюдений.

Несмещенная оценка всегда состоятельная, однако состоятельная оценка не всегда будет несмещенной. Так, например, оценка дисперсии нормально распределенной случайной величины, рассчитанная по формуле

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \quad (13)$$

будет состоятельной и несмещенной, в то время, как аналогичная оценка, рассчитанная по формуле

$$S_{cm}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad (14)$$

будет состоятельной, но смещенной.

Для оценки средних значений пользуются:

средним арифметическим

$$\bar{x}_{ap} = \frac{\sum_{i=1}^n x_i}{n}; \quad (15)$$

средним квадратическим

$$\bar{x}_{kv} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}};$$

средним гармоническим

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}};$$

средним геометрическим

$$\bar{x}_{\text{геом}} = \sqrt[n]{x_1 x_2 \dots x_n};$$

средним логарифмическим

$$\bar{x}_{\text{лог}} = 10^{\overline{\lg x}}, \text{ где } \overline{\lg x} = \frac{\sum_{i=1}^n \lg x_i}{n};$$

средними взвешенными оценками

$$\bar{x}_{\text{вз}} = \frac{\sum_{i=1}^n k_i x_i}{\sum_{i=1}^n k_i}, \quad (16)$$

где  $k$  — коэффициент взвешивания.

Для определения степени изменчивости случайных величин также применяются различные оценки:

среднее отклонение

$$\Delta = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n},$$

среднее квадратическое отклонение  $\delta = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$ , дисперсия, коэффициент вариации.

Чаще всего исследователи прибегают к среднему квадратическому и производным от него — дисперсии и коэффициенту вариации (см. разд. 2.3). Это связано с тем, что при возведении отклонений в квадрат усиливается влияние на оценку значений наиболее сильно отличающихся от математического ожидания, т. е. именно таких значений, которые определяют понятие «изменчивость». Кроме того дисперсия, рассчитываемая через средние квадраты отклонений, входит в качестве параметра в аналитические выражения большинства законов распределения.

Оценки, полученные в результате эксперимента по выборкам ограниченного объема, отличаются от математического ожидания оцениваемого параметра и являются случайными величинами. Функции распределения выборочных оценок, а следовательно, и их свойства зависят от закона распределения самой изучаемой случайной величины.

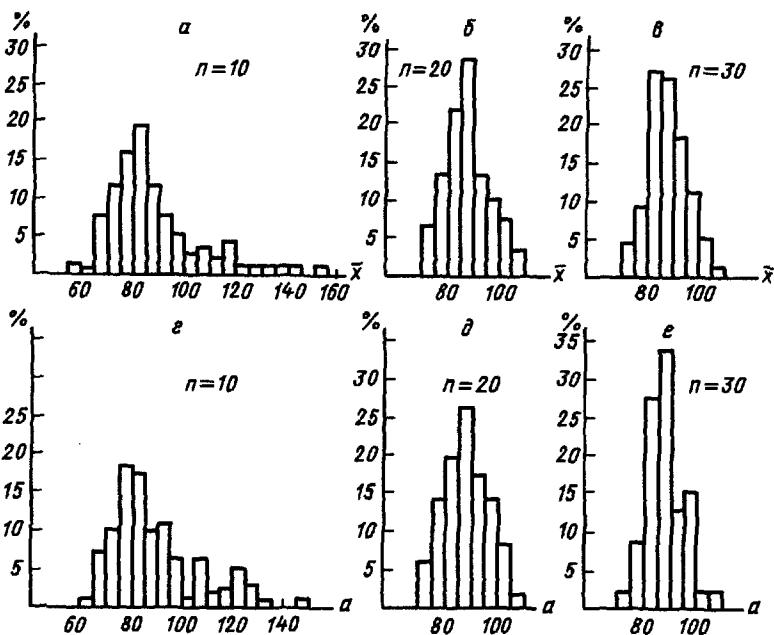


Рис. 21. Гистограммы частотного распределения выборочных оценок среднего содержания молибдена по выборкам объемом 10, 20 и 30 значений:

*a, б, в*—средние арифметические ( $\bar{X}$ ); *г, д, е*—максимально правдоподобные оценки ( $a$ )

Поэтому для выбора наилучшей оценки необходимо проверить гипотезы о соответствии выборочных данных известным статистическим законам распределения.

Наилучшей оценкой математического ожидания *нормально распределенной* случайной величины является среднее арифметическое. Распределение этой оценки в данном случае будет также соответствовать нормальному закону при любом объеме выборки, ее математическое ожидание совпадает с математическим ожиданием изучаемой случайной величины, а дисперсия будет минимальной по сравнению с другими средними. Таким образом, среднее арифметическое является оценкой состоятельной, несмещенной и максимально эффективной.

Оценка среднего квадратического отклонения, рассчитанная по формуле (14), в условиях нормального распределения состоятельная, эффективная, но немного смещена при малом объеме выборки. Для устранения этого смещения объем выборки при расчете необходимо уменьшить на 1, т. е. следует пользоваться формулой (13).

Для логнормально распределенных случайных величин распределения выборочных средних арифметических являются асимптотически нормальными, т. е. приближаются к нормальным лишь при достаточно большом объеме выборок (рис. 21). При малом объеме выборок они становятся явно асимметричными (см. рис. 21, *a, б*). Это связано

с тем, что на величину среднего арифметического оказывают сильное влияние единичные большие, но редко встречающиеся значения.

В табл. 14 приведены результаты опробования по 20 буровзрывным скважинам, равномерно расположенным в пределах одного уступа карьера, вскрывающего молибденовое месторождение. Среднее содержание молибдена по этим скважинам составляет 100 усл. ед. и существенно превышает среднее содержание, определенное по всем 1117 скважинам этого уступа, равное 89,5 усл. ед. В то же время, если исключить из выборки одну скважину с содержанием 370 усл. ед., то оценка среднего содержания молибдена по оставшимся 19 скважинам окажется равной 86 усл. ед., т. е. будет ниже «истинного» среднего.

Эффективность среднего арифметического при оценивании математического ожидания логнормально распределенной случайной величины снижается с увеличением дисперсии изучаемой случайной величины.

Для оценки математического ожидания логнормальных величин по выборкам малого объема была предложена максимально правдоподобная оценка  $a$ , которую можно вычислить с помощью следующего выражения:

Таблица 14. Расчет статистических характеристик распределения содержания молибдена по 20 скважинам

№ скв.	$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$\ln x_i$	$\ln x_i - \bar{\ln x}$	$(\ln x_i - \bar{\ln x})^2$
1	122	22	484	4,8040	-0,3666	0,1344
2	50	-50	2500	3,9120	-0,5254	0,2760
3	78	-22	484	4,3567	-0,0807	0,0065
4	370	270	72900	5,9135	1,4761	2,1789
5	97	-3	9	4,5747	0,1373	0,0189
6	175	75	5625	5,1648	0,7274	0,5291
7	50	-50	2500	3,9120	-0,5254	0,2760
8	65	-35	1225	4,1744	-0,2630	0,0692
9	47	-53	2809	3,8502	-0,5872	0,3448
10	110	10	100	4,7005	0,2631	0,0692
11	120	20	400	4,7875	0,3501	0,1226
12	82	-18	324	4,4067	-0,0307	0,0009
13	85	-15	225	4,4426	0,0052	0
14	59	-41	1681	4,0775	-0,3599	0,1295
15	145	45	2025	4,9767	0,5393	0,2908
16	101	1	1	4,6151	0,1777	0,0316
17	89	-11	121	4,4886	0,0512	0,0026
18	77	-23	529	4,3438	-0,0936	0,0088
19	52	-48	2304	3,9512	-0,4862	0,2364
20	22	-73	5329	3,2958	-1,1416	1,3032

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2001}{20} = 100; S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{101575}{19} = 5346;$$

$$\bar{\ln x} = \frac{88,7483}{20} = 4,4374; S_{\ln x}^2 = \frac{6,0294}{19} = 0,3173;$$

$$a = e^{4,4374} \Psi_{20} \left( \frac{0,3173}{2} \right) = 84,56 \cdot 1,1631 \approx 98;$$

$$b^2 = e^{2 \cdot 4,4374} \left[ \Psi_{20}(2 \cdot 0,3173) - \Psi_{20}\left(\frac{18}{19} \cdot 0,3173\right) \right] = 2467.$$

$$a = e^{\overline{\ln x}} \Psi_n(t), \quad (17)$$

где  $t = \frac{1}{2} S_{\ln x}^2$ ,

$$\text{а } \Psi_n(t) \approx e^t \left\{ 1 - \frac{t(t+1)}{n} + \frac{t^2(3t^3 + 22t + 21)}{6n^2} \right\}.$$

Таблицы функции  $\Psi_n(t)$  приведены в прил. 3.

Для десятичных логарифмов формула (17) принимает вид  $a = 10^{\overline{\ln x}} \Psi_n(t)$ , где  $t = 2,65 S_{\ln x}^2$ .

Однако при не очень больших значениях дисперсии применение этой оценки не приводит к существенному уточнению. Из приведенного в табл. 14 примера видно, что максимально правдоподобная оценка  $a = 98$  усл. ед. немного ближе к истинному среднему содержанию молибдена, чем среднее арифметическое  $\bar{x} = 100$ , но разница между этими оценками невелика.

Величина расхождения между оценками  $a$  и  $\bar{x}$  быстро уменьшается с увеличением объема выборки (см. рис. 21,  $a$ — $e$ ), и при  $n \geq 100$  они в большинстве случаев практически совпадают.

Следует также учитывать, что в случае отличия распределения случайной величины от логнормального максимально правдоподобная оценка может оказаться смещенной и несостоительной, в то время как среднее арифметическое сохраняет эти свойства и в условиях логнормального распределения. В то же время решить вопрос о соответствии эмпирических распределений тому или иному закону по малым выборкам затруднительно. Поэтому пользоваться максимально правдоподобной оценкой целесообразно лишь тогда, когда имеются какие-либо априорные сведения о соответствии распределения изучаемой случайной величины логнормальному закону.

Оценка дисперсии, рассчитанная по формуле (5) в условиях логнормального закона, достаточно эффективна лишь при очень большом объеме выборки. Поэтому во всех случаях, когда гипотеза о логнормальном законе не отвергается, а объем выборки не превышает 100, целесообразно пользоваться максимально правдоподобной оценкой дисперсии, рассчитываемой по формуле

$$b^2 = e^{\overline{2 \ln x}} \left\{ \Psi_n(2S_{\ln x}^2) - \Psi_n\left(\frac{n-2}{n-1} S_{\ln x}^2\right) \right\}.$$

По малым выборкам эта оценка обычно существенно отличается от оценки  $S^2$ , которая весьма чувствительна к отдельным выдающимся значениям. В примере, приведенном в табл. 14, оценки  $S^2$  и  $b^2$  отличаются друг от друга более чем в два раза. При этом оценка  $b^2$  гораздо ближе к оценкам дисперсии  $S^2$  и  $b^2$  по выборке большого объема ( $S^2 = 3392$ ,  $b^2 = 3346$ ), которые практически совпадают и поэтому могут быть признаны близкими к «истинному» значению дисперсии содержания молибдена по данному уступу карьера.

Для снижения влияния редко встречающихся (аномальных) значений на величину выборочных оценок математического ожидания иногда используют процедуру *винзоризации*. При этом выборочные значения изучаемого признака ранжируются, т. е. записываются в порядке возрастания:  $x_1 \leq x_2 \leq x_3 \dots \leq x_{n-1} \leq x_n$ . Несколько значений в начале ряда:  $x_1, x_2 \dots x_g$  заменяется на  $x_{g+1}$ , т. е. первым значением в ранжированием ряду, следующим за этой группой. Столько же значений в конце ряда, т. е.  $x_{n-g}, x_{n-g+1}, \dots, x_n$ , заменяется на  $x_{n-g-1}$ . Число значений, подлежащих замене, определяется по гистограмме. К редко встречающимся обычно относят значения с частотой менее 5—3. Однако процедура винзоризации допустима лишь для симметричных распределений. В условиях асимметричного распределения (например, логнормального) эта процедура приводит к получению смещенных оценок. По винзоризованным выборкам нельзя оценивать характеристики разброса — дисперсию, коэффициент вариации и т. п.

Необходимо помнить, что рассмотренные выше способы вычисления точечных оценок исходят из условий однородности, случайности и независимости выборочных данных, которые в практике геологоразведочных работ выполняются далеко не всегда. Реальные геологические объекты обладают размерностью и их свойства определенным образом изменяются в пространстве. Геолога обычно интересуют средние значения изучаемого свойства в каком-либо объеме или на определенной площади. Поэтому в геологической практике часто применяются так называемые взвешенные оценки (16), когда каждому замеру изучаемого свойства приписывается определенная зона влияния, границы которой проводятся на половине расстояния между точками или линиями замеров. Площади, объемы или массы этих зон влияния принимаются в качестве значений весового коэффициента  $k$  в формуле (16).

Так, например, когда оценивается среднее содержание полезного компонента в каком-либо рудном теле, то под этим понимается среднее количество полезного компонента в единице объема или массы данного рудного тела. В то же время определение содержания полезного компонента часто производится по пробам переменной длины, расположенным на разном расстоянии друг от друга. При этом фактические зоны влияния каждой пробы оказываются неодинаковыми. Если между размером зоны влияния и содержанием полезного компонента существует зависимость, например, содержание снижается в местах увеличения мощности рудного тела (длины пробы), то среднее арифметическое содержание, рассчитанное по формуле (15), будет смещенной оценкой среднего содержания в объеме всего рудного тела. Несмешенная оценка в данном случае рассчитывается по формуле

$$\bar{C}_{\text{нн}} = \frac{\sum_{i=1}^n l_i S_i C_i}{\sum_{i=1}^n l_i S_i} = \frac{\sum_{i=1}^n V_i C_i}{\sum_{i=1}^n V_i},$$

где  $C_i$  — содержание полезного компонента по пробе длиной  $l_i$ ;  $S_i$  и  $V_i$  — соответственно площадь и объем зоны влияния данной пробы.

Если сеть наблюдений равномерна, т. е. площади зон влияния всех проб равны, то взвешивание производится только на их длину:

$$\bar{C}_{\text{вз}} = \frac{\sum_{i=1}^n l_i C_i}{\sum_{i=1}^n l_i}.$$

Состоятельной, несмещенной и максимально эффективной оценкой параметра  $p$ , т. е. вероятности случайного события  $A$  при биномиальном распределении, является величина  $\hat{p} = \frac{x}{n}$ , т. е. частота этого события.

Для параметра  $\lambda$  распределения Пуассона этими же свойствами будет обладать оценка  $\lambda \approx \bar{x}_n = \sum_{i=1}^n n_i i / N$ , т. е. средняя частота появления случайного события  $A$  в выборках объемом  $N$ .

Наилучшими точечными оценками параметров  $\mu$  и  $k$  распределения Мизеса будут выборочное среднее круговое направление  $m$ , вычисляемое по формулам (8), (9), и оценка  $\hat{k} = A^{-1}(\bar{R})$ , которую можно определить с помощью специальных таблиц (см. прил. 2). Приближенную оценку параметра концентрации  $k$  при  $\bar{R} < 0,45$  можно вычислить по формуле  $\hat{k} \approx \frac{1}{6} \bar{R} (12 + 6\bar{R}^2 + 5\bar{R}^4)$ , а при значениях  $\bar{R}$  близких к 1 по формуле  $\frac{1}{k} \approx 2(1 - \bar{R}) - (1 - \bar{R})^2 - (1 - \bar{R})^3$ .

Точечная оценка не содержит информации о точности полученного результата. Чем меньше выборка и чем сильнее изменчивость признака, тем большей может оказаться ошибка. Поэтому в условиях малых выборок всегда желательно знать интервал значений признака, в который с заданной вероятностью попадает его неизвестное истинное среднее значение.

Границы этого доверительного интервала определяются величиной возможного отклонения  $\lambda$  выборочной оценки параметра  $\hat{\theta}$  от его истинного значения  $\theta$ :

$$P(\hat{\theta} - \lambda \leq \theta \leq \hat{\theta} + \lambda) = 1 - \alpha.$$

Вероятность  $P = 1 - \alpha$  накрытия доверительным интервалом истинного значения параметра называется доверительной вероятностью. Величина  $\alpha$  определяет вероятность того, что истинное значение окажется за пределами этого интервала.

Для получения интервальной оценки необходимо располагать данными о функции распределения выборочных точечных оценок, которая зависит от вида функции распределения изучаемой случайной величины.

Если изучаемая генеральная совокупность распределена нормально с параметрами  $Mx$  и  $\sigma^2$ , то выборочные средние арифметические

оценки  $\bar{x}$ , полученные по выборкам объемом  $n > 60$  значений, также будут распределены нормально с математическим ожиданием  $Mx$  и дисперсией  $\sigma_{\bar{x}}^2$ , прямо пропорциональной  $\sigma^2$  и обратно пропорциональной объему выборки  $n$ :  $\sigma_{\bar{x}}^2 = \sigma^2/n$ .

Следовательно, величина  $\lambda = (\bar{x} - Mx)/\sigma_{\bar{x}} = (\bar{x} - Mx)\sqrt{n}/\sigma$  будет распределена нормально с параметрами  $M\lambda = 0$  и  $\sigma_{\lambda}^2 = 1$ . Поэтому вероятность того, что единичная выборочная оценка  $\bar{x}$  будет отличаться от математического ожидания  $Mx$  не более чем на величину  $\lambda$  может быть найдена по таблицам интегральной функции Лапласа (см. прил. 11).

В геологической практике дисперсия  $\sigma^2$  изучаемой случайной величины обычно неизвестна, и вместо нее используется ее выборочная оценка  $S^2$ . Поэтому задача нахождения доверительных интервалов имеет лишь приближенные решения.

При объеме выборки меньше 60 значений величина  $\lambda = (\bar{x} - Mx)\sqrt{n-1}/S$  распределена по закону Стьюдента с  $f = n-1$  степенями свободы. Доверительный интервал для  $Mx$  в этом случае можно записать как

$$P\left\{-t_{f,1-\alpha} \leq \frac{(\bar{x} - Mx)\sqrt{n-1}}{S} \leq t_{f,1-\alpha}\right\} = 1 - \alpha,$$

или

$$P\left(\bar{x} - t_{f,1-\alpha} \frac{S}{\sqrt{n-1}} \leq Mx \leq \bar{x} + t_{f,1-\alpha} \frac{S}{\sqrt{n-1}}\right) = 1 - \alpha, \quad (18)$$

где  $t_{f,1-\alpha}$  — значения критерия Стьюдента для объема выборки  $n$  и доверительной вероятности  $1 - \alpha$  (см. прил. 13).

*Пример.* Вернемся к задаче определения петрографического типа породы из горизонта неогеновых лав, рассмотренной в разд. 2.2. Более корректно вопрос о названии данной породы можно решить путем расчета интервальной оценки среднего содержания. Расчеты оценок асимметрии и эксцесса показывают, что гипотеза о соответствии эмпирических данных нормальному закону распределения не отвергается. Это позволяет для расчета интервальной оценки среднего содержания  $\text{SiO}_2$  в лавах воспользоваться формулой (18). При доверительной вероятности 0,96 и числе наблюдений 30 значение критерия Стьюдента  $t_{29;0,05} = 2,05$ . Находим границы доверительного интервала, в %:

$$\bar{x} \pm t_{29;0,05} \frac{s}{\sqrt{n-1}} = 65,5 \pm \frac{2,05 \cdot 4,4}{\sqrt{29}} = 65,6 \pm 1,7.$$

Таким образом, с вероятностью 0,95 истинное среднее содержание  $\text{SiO}_2$  в породах данного горизонта находится в интервале от 63,9 до 67,3%, т. е. не выходит за пределы, установленные для дацита: 63,0—68,5%.

Построение интервальной оценки дисперсии нормально распределенной случайной величины основано на том, что распределение величины  $(n-1)S^2/\sigma^2$  соответствует распределению  $\chi^2$  Пирсона с  $n-1$  степенями свободы.

Поэтому

$$P[c_1 < (n-1)S^2/\sigma^2 < c_2] = 1 - \alpha,$$

или

$$P[(n-1)S^2/c_2 < \sigma^2 < (n-1)S^2/c_1] = 1 - \alpha,$$

где  $c_1 = \chi^2_{n, 1-\alpha/2}$  и  $c_2 = \chi^2_{n, \alpha/2}$  — значения распределения  $\chi^2$ , соответствующие вероятностям  $1-\alpha/2$  и  $\alpha/2$ , которые можно получить из соответствующей таблицы.

В приведенном выше примере точечная оценка дисперсии содержания  $\text{SiO}_2$  в лавах составила 19,2. Приняв доверительную вероятность равной 90 с помощью таблиц критерия Пирсона находим  $c_1 = \chi^2_{29, 0,95} = 17,708$ ;  $c_2 = \chi^2_{29, 0,05} = 42,557$ .

Таким образом, с вероятностью 0,90 истинная дисперсия содержания  $\text{SiO}_2$  в лавах находится в интервале

$$\text{от } \frac{(30-1) \cdot 19,2}{42,557} = 13,1\% \text{ до } \frac{(30-1) \cdot 19,2}{17,708} = 31,4\%.$$

Простых и точных методов построения доверительных интервалов для оценок среднего при логнормальном распределении изучаемой совокупности не существует.

При достаточно большом объеме выборки ( $n > 60$ ) для построения доверительного интервала максимально правдоподобной оценки  $a$  можно воспользоваться формулой

$$P(a - \lambda_a \leq Mx \leq a + \lambda_a) \approx 1 - \alpha,$$

где

$$\lambda_a = Z_{1-\alpha/2} \frac{a}{\sqrt{n}} \sqrt{S_{\ln x}^2 + 0,5 S_{\ln x}^4},$$

а  $Z_{1-\alpha/2}$  — значение функции Лапласа для доверительной вероятности  $1-\alpha/2$  (см. прил. 11). Вероятность  $\alpha$  в данном случае делится на 2 в связи с тем, что вероятность непопадания истинного среднего в доверительный интервал складывается из вероятностей двух событий:  $Mx < a - \lambda_a$  и  $Mx > a + \lambda_a$ .

В случае малых выборок, когда распределение оценок  $a$  является асимметричным, лучше пользоваться упрощенной формулой, дающей несимметричный доверительный интервал. При этом приближенная оценка среднего рассчитывается как

$$a = \text{antiln}(\overline{\ln x} + 0,5 S_{\ln x}^2),$$

или

$$a = \text{antilg}(\overline{\lg x} + 1,1513 S_{\lg x}^2).$$

Нижняя и верхняя границы доверительного интервала определяются по формуле

$$a \pm \lambda = \text{antiln}(\overline{\ln x} \pm t_{\alpha, n-1} \frac{\sqrt{S_{\ln x}^3}}{\sqrt{n}}),$$

где  $t_{\alpha, n-1}$  — значение критерия Стьюдента (см. прил. 13).

*Пример.* Найдем доверительный интервал для оценки среднего содержания молибдена в штокверке по 20 скважинам (см. табл. 14), приняв доверительную вероятность  $P$  равной 0,95.

По первому способу получим:

$$\lambda_a = Z_{1-\alpha/2} \frac{a}{\sqrt{n}} \sqrt{S_{\ln x}^2 + 0,5 S_{\ln x}^4} = 1,96 \frac{98}{20} \sqrt{0,3173 + 0,5 \cdot 0,3173^2} \approx 26.$$

Таким образом, среднее содержание молибдена с вероятностью 0,95 находится в интервале  $a + \lambda_a = 98 \pm 26$ , т. е.  $72 \leq Mx \leq 124$  усл. ед.

По второму способу:

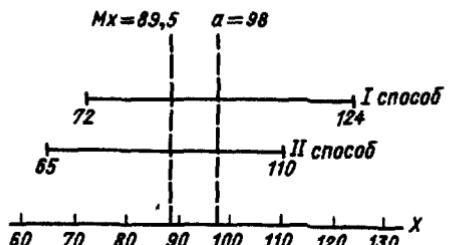


Рис. 22. Доверительные интервалы оценок среднего содержания молибдена по 20 скважинам по I и II способу

$$a = e^{\ln \bar{x} + 0.5 S_{\ln x}} = e^{4.4374 + 0.5 \cdot 0.3173} = 99;$$

$$a \pm \lambda = \text{antiln} \left[ \ln \bar{x} \pm t_{a, n-1} \frac{\sqrt{S_{\ln x}^2}}{\sqrt{n}} \right] = \text{antiln} \left[ 4.4374 \pm \frac{2.09 \cdot \sqrt{0.3173}}{4.472} \right] =$$

$$= \text{antiln} (4.4374 \pm 0.2633).$$

Следовательно,  $65 \leq Mx \leq 110$ .

В приведенном примере «истинное» среднее содержание молибдена составляет 89,5 усл. ед. Поэтому можно сделать вывод, что в данном случае второй способ дает лучший результат, так как доверительный интервал оказывается симметричным относительно «истинного» среднего (рис. 22), в то время как по первому способу он симметричен относительно оценки  $a$ , которая по данной выборке существенно отклоняется от истинного среднего.

Для получения интервальных оценок параметров *биномиального распределения* при достаточно большом количестве испытаний ( $n > 100$ ) в случае, когда оцениваемая вероятность случайного события  $P$  не очень близка к 0 или 1, можно воспользоваться нормальным приближением биномиального распределения:

$$P = (\hat{p} - \lambda_{\hat{p}} \leq P \leq \hat{p} + \lambda_{\hat{p}}) \approx 1 - \alpha,$$

где  $\lambda_{\hat{p}} = Z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$ ;  $\hat{p} = x/n$  — частота случайного события, определенная по  $n$  испытаниям, а  $Z_{1-\alpha/2}$  — значение функции Лапласа для заданной доверительной вероятности  $P = 1 - \alpha$ .

По малым выборкам оценки  $P$  распределены асимметрично, за исключением случаев, когда  $P \approx 0,5$  (см. рис. 15). Поэтому более универсальным является способ нахождения границ доверительного интервала с помощью формулы

$$\hat{p} \pm \lambda_{\hat{p}} = \frac{n}{Z_{1-\alpha/2}^2 + n} \left[ \hat{p} + \frac{Z_{1-\alpha/2}}{2n} \pm Z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \left( \frac{Z_{1-\alpha/2}}{2n} \right)^2} \right].$$

*Пример.* Построим доверительный интервал для оценки коэффициента рудоносности, приведенной в разд. 2.4 (см. рис. 14), приняв доверительную вероятность равной 0,95. Точечная оценка коэффициента рудоносности по 28 скважинам составила

$$\hat{K}_p = \hat{p} = \frac{x}{n} = \frac{9}{28} \approx 0,32.$$

По первому способу

$$\lambda_{\hat{p}} = Z_{1-0,10/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96 \sqrt{\frac{0,32(1-0,32)}{28}} = 0,17.$$

Следовательно, неизвестный истинный коэффициент рудоносности с вероятностью 0,95 находится в интервале от 0,15 до 0,49.

По второму способу

$$\begin{aligned}\hat{p} \pm \lambda &= \frac{28}{1,96^2 + 28} \left[ 0,32 + \frac{1,96^2}{2,28} \pm 1,96 \sqrt{\frac{0,32(1-0,32)}{28} + \left(\frac{1,96}{2 \cdot 28}\right)^2} \right] = \\ &= 0,879 (0,389 \pm 0,186); \\ &0,18 \leq K_p \leq 0,51.\end{aligned}$$

Оба способа дают довольно близкие результаты. Это объясняется тем, что величина коэффициента рудоносности в данном примере не слишком сильно отличается от 0,5.

Доверительный интервал для параметра  $\mu$  распределения Мизеса можно записать в следующем виде:

$$P(m - \lambda_m \leq \mu \leq m + \lambda_m) \approx 1 - \alpha,$$

где  $m$  — выборочное круговое среднее направление. Величину  $\lambda_m$ , зависящую от объема выборки  $n$  и выборочной результирующей длины  $\bar{R}$ , можно найти с помощью номограммы (см. прил. 4).

Интервальная оценка для параметра концентрации  $k$  определяется аналогичным образом:

$$P(\hat{k} - \lambda_k \leq k \leq \hat{k} + \lambda_k) \approx 1 - \alpha.$$

Верхняя и нижняя граница интервала находится по номограммам, изображенными в прил. 5.

*Пример.* Построим интервальные оценки параметров  $\mu$  и  $k$  для азимутов падения кварцевых прожилков, приведенных ранее в табл. 12. По 48 замерам были получены следующие оценки (см. табл. 12):  $m \approx 173,5^\circ$ ,  $\bar{R} = 0,316$ ,  $k \approx 0,7$ . Приняв доверительную вероятность равной 0,95, по номограмме (см. прил. 4, б) находим  $\lambda_m \approx 34^\circ$ . Следовательно, оцениваемое круговое среднее направление  $\mu$  азимутов падения (в градусах) кварцевых прожилков с вероятностью 0,95 находится в интервале:

$$173,5 - 34 = 139,5 \leq \mu \leq 173,5 + 34 = 207,5.$$

Границы доверительного интервала для  $k$  определяются по номограмме, изображенной на прил. 5:  $0,2 \leq k \leq 1,0$ .

Способы построения доверительных интервалов не только позволяют найти величины возможных ошибок при оценке средних значений изучаемых свойств по выборочным данным, но могут также использоваться для решения обратной задачи, т. е. для нахождения объема выборки, обеспечивающей получение оценок с заданной точностью. Задача определения оптимального числа выборочных данных при исследовании геологических объектов возникает постоянно и является весьма важной, особенно при разведке месторождений, когда каждое выборочное наблюдение требует проходки специальной горной выработки или бурения скважины.

Для решения обратной задачи в случае нормального распределения исследуемой случайной величины формула (18) принимает вид

$$n = \left( \frac{t_{1-\alpha, n-1} S}{\lambda} \right)^2. \quad (19)$$

Обычно удобнее пользоваться не абсолютной ошибкой  $\lambda$ , а относительной —  $\lambda_0 = \frac{\lambda}{\bar{x}} \cdot 100\%$ , поэтому в формуле (19) стандартное отклонение  $S$  заменяется коэффициентом вариации  $V$ :

$$n = \left( \frac{t_{1-\alpha, n-1} V}{\lambda_0} \right)^2.$$

При малом объеме выборки ( $n < 60$ ) величина  $t$  зависит от  $n$ , поэтому для решения обратной задачи используется способ последовательного приближения. Первоначально в формулу подставляется значение распределения Стьюдента  $t$  для  $n = \infty$ .

Если полученное  $n_1$  окажется меньше 60, в формулу подставляется значение  $t$  для полученного  $n_1$ , и эта операция повторяется до тех пор, пока полученное в результате очередного расчета значение  $n$ , не совпадет с величиной  $n_{i-1}$ , принятой для определения  $t$  при ее расчете.

*Пример* В приведенном выше примере расчет интервальной оценки среднего содержания  $\text{SiO}_2$  в эфузивных породах по 30 пробам показал, что с вероятностью 0,95 возможная ошибка составляет  $\pm 1,6\% \text{ SiO}_2$ . Определим число проб, необходимое для того, чтобы с той же вероятностью ошибка в определении среднего содержания не превысила 1%  $\text{SiO}_2$ . Воспользовавшись оценкой стандарта  $S$ , получим

$$n_1 = \left( \frac{t_{0.05, n=\infty} S}{\lambda} \right)^2 = \left( \frac{1.96 \cdot 4.4}{1} \right)^2 \approx 74$$

В таблице критерия Стьюдента (см. прил 13) находим, что  $t_{0.05, n=74-1} \approx 1.99$ . Следовательно,  $n_2 = \left( \frac{1.99 \cdot 4.4}{1} \right)^2 \approx 77$

Значения  $t_{0.05, n=73}$  и  $t_{0.05, n=77}$  практически совпадают, поэтому в дальнейших расчетах нет необходимости

Решая задачи на основе интервальных статистических оценок надо помнить о том, что методы их расчета предполагают строгое соблюдение условия независимости выборочных данных. В то же время в изменении свойств геологических объектов в пространстве часто наблюдаются закономерности, и между результатами наблюдений в соседних пунктах при регулярном их расположении возникают определенные зависимости. В этом случае величины возможных ошибок, рассчитанные по приведенным выше формулам, оказываются завышенными. То же происходит и с оценками необходимого объема выборки при решении обратной задачи. Величина этого завышения зависит от характера и степени проявления закономерных изменений. Поэтому прежде чем использовать интервальные статистические оценки для решения каких-либо практических задач, необходимо убедиться в отсутствии закономерных изменений изучаемого свойства в пространстве. Способы проверки гипотезы об отсутствии закономерной изменчивости подробно рассмотрены в гл. 5.

## 2.7. СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГЕОЛОГИЧЕСКИХ ГИПОТЕЗ

Решение многих геологических задач основано на принципе аналогии, когда для объяснения особенностей строения слабо изученных объектов используют закономерности, установленные при изучении аналогичных объектов. Для правильного выбора объекта-аналога необходимо оценить степень его сходства с исследуемым объектом.

В других случаях (например, при интерпретации многих геофизических данных) возникает необходимость оценить степень различия геологических объектов по тем или иным физическим свойствам.

Для объективного решения вопроса о сходстве или различии геологических объектов используются статистические методы проверки гипотез о равенстве числовых характеристик их свойств. В геологической практике чаще всего эти методы применяются для суждения:

- о равенстве средних значений изучаемого признака, полученных разными методами для одного и того же объекта или одним методом для различных объектов;

- о равенстве дисперсий двух случайных величин по выборочным данным;

- об однородности изучаемого объекта.

Статистическая проверка гипотез производится с помощью критериев согласия.

*Критерием согласия* называется значение некоторой функции  $K=f(X_1, X_2, \dots, X_n)$ , где  $X_1, X_2, \dots, X_n$  — случайные величины, характеризующие проверяемую гипотезу. Функция выбирается таким образом, чтобы в случае правильности проверяемой гипотезы ее значения представляли бы собой случайную величину с заранее известным распределением.

Проверяемая гипотеза принимается, если значение  $K$ , вычисленное через выборочные значения величин  $X_1, X_2, \dots, X_n$ , окажется меньше или больше (в зависимости от формулировки гипотезы) теоретического значения  $K$  для аналогичных условий и заданной вероятности  $\alpha$ , которое берется по известному распределению. Вероятность  $\alpha$  при этом соответствует уровню вероятности практически невозможного события и называется *уровнем значимости*.

Соответственно вероятность  $(1-\alpha)$ , определяющая область, в пределах которой правильность принятого решения будет практически достоверным событием, называется *доверительной*.

Ошибка, заключенная в непринятии гипотезы, в действительности являющейся справедливой, называется *ошибкой первого рода*, а принятие ложной гипотезы — *ошибкой второго рода*.

Если вероятность ошибки второго рода обозначить через  $\beta$ , то  $(1-\beta)$ , т. е. вероятность отсутствия такой ошибки, будет величиной, называемой *мощностью данного критерия* относительно конкурирующей гипотезы.

Увеличение доверительной вероятности (уменьшение уровня значимости) снижает вероятность ошибки первого рода, но увеличивает вероятность ошибки второго рода.

Область применения определенных критериев согласия обычно ограничивается некоторыми условиями, а их мощность зависит от характера конкурирующей (альтернативной) гипотезы и объема выборки.

Для решения задач на основе статистической проверки гипотез геолог должен выполнить следующие операции:

— четко сформулировать проверяемую ( $H_0$ ) и альтернативную ( $H_1$ ) гипотезу исходя из существа поставленной геологической задачи;

— выбрать наиболее мощный при данном объеме выборки критерий, условия применения которого не противоречат свойствам изучаемых случайных величин;

— оценить последствия ошибки первого и второго рода в условиях решаемой геологической задачи и выбрать уровень значимости исходя из требования минимизации ущерба в результате неправильного решения;

— рассчитать эмпирическое значение критерия согласия  $K$  по выборочным данным, сравнить его с теоретическим значением  $K$  для принятого уровня значимости и принять решение относительно гипотезы  $H_0$ ;

— интерпретировать полученный результат применительно к поставленной геологической задаче.

При формулировке проверяемой гипотезы  $H_0$  трудностей обычно не возникает, однако вопрос о том, какую гипотезу принять в качестве альтернативной не всегда решается однозначно, так как для одной и той же гипотезы  $H_0$  может существовать несколько альтернативных гипотез  $H_1$ . Например, при расчете интервальных оценок гипотеза  $H_0$  заключается в том, что неизвестное математическое ожидание  $Mx$  находится в определенном интервале значений, т. е.

$$H_0: \bar{x} - \lambda \leq Mx \leq \bar{x} + \lambda.$$

В то же время в качестве альтернативных могут выступать разные гипотезы:

— математическое ожидание меньше нижней границы доверительного интервала, т. е.

$$H_1^1: \bar{x} - \lambda > Mx;$$

— математическое ожидание больше верхней границы доверительного интервала, т. е.

$$H_1^2: \bar{x} + \lambda < Mx;$$

— математическое ожидание больше верхней или меньше нижней границы доверительного интервала, т. е.

$$H_1^3: \bar{x} - \lambda > Mx < \bar{x} + \lambda.$$

В рассмотренных выше примерах в качестве альтернативной принималась гипотеза  $H_1^3$ . Однако при подсчете запасов месторождений часто целесообразнее пользоваться альтернативной гипотезой  $H_1^1$ , так как вопрос возможности промышленного использования определенных объемов руды решается путем сравнения полученных

данных оценок среднего содержания полезного компонента с минимальным промышленным содержанием.

Неправильная формулировка альтернативной гипотезы может вызвать ошибки при пользовании статистическими таблицами, поскольку существуют таблицы для критериев двух типов — односторонних и двусторонних. В таблицах односторонних критериев приводятся доверительные вероятности или уровни значимости, соответствующие простым альтернативным событиям типа  $H_1^1$  или  $H_1^2$ . Таблицы двусторонних критериев построены для сложных альтернатив типа  $H_1^3$ , когда учитывается вероятность сразу двух событий.

Интегральная функция Лапласа (см. прил. 11) относится к таблицам первого типа, поэтому при нахождении по ней вероятностного критерия  $Z$  для построения двустороннего доверительного интервала, т. е. при альтернативе  $H_1^3$ , уровень значимости необходимо уменьшать в два раза. Таблицы второго типа строятся только для симметрично распределенных критериев. Они более компактны и удобны для построения доверительных интервалов и проверки гипотез при альтернативах типа  $H_1^3$ . Примером таблиц этого типа является таблица двустороннего  $t$ -критерия Стьюдента (см. прил. 13). Используя эту таблицу для построения односторонних доверительных интервалов или проверки гипотез при альтернативах типа  $H_1^1$  или  $H_1^2$  значения функции принимаются для уровня значимости  $2\alpha$ .

Статистические критерии согласия разделяются на параметрические и непараметрические. Параметрические критерии выводятся из свойств тех или иных статистических законов распределения и могут использоваться лишь в том случае, если распределение выборочных данных согласуется с этим законом. Непараметрические критерии могут применяться даже в том случае, если закон распределения изучаемых величин неизвестен или их распределения не соответствуют никакому из известных законов. Непараметрические критерии обычно обладают несколько меньшей мощностью по сравнению с параметрическими аналогами, но область их применения значительно шире. Как было отмечено в разд. 2.4, фактические распределения свойств геологических объектов часто отклоняются от теоретических, поэтому геологи проявляют большой интерес к непараметрическим критериям.

Выбор уровня значимости при статистической проверке гипотез является весьма важным, но отнюдь не всегда простым вопросом. Он решается исключительно исходя из особенностей геологической задачи на основе анализа возможных последствий от ошибок первого и второго рода. Для правильного выбора уровня значимости геологу необходимо четко представлять себе конечную цель проводимых исследований, а иногда даже выполнять укрупненные технико-экономические расчеты для оценки возможного ущерба за счет принятия неправильного решения. В случае затруднения с выбором уровня значимости гипотезу целесообразно проверить при разных его значениях.

При интерпретации полученных результатов необходимо следить за тем, чтобы вывод по геологической задаче строго логически соответствовал проверяемой гипотезе  $H_0$ .

## 2.8. ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ СРЕДНИХ (МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ)

Необходимость сравнения средних значений изучаемых свойств геологических объектов возникает при решении широкого круга задач во всех отраслях геологических наук. Так, например, по мнению многих петрологов, средний химический состав лав вулканов и интрузивных пород отражает в общих чертах особенности состава породивших их глубинных магматических очагов. Путем сравнения различных эфузивных и интрузивных пород по среднему содержанию в них химических элементов можно судить о комагматичности (т. е. генетическом родстве) эфузивных и интрузивных образований, о принадлежности интрузивных образований к определенному магматическому комплексу или двух вулканических построек к одному глубинному магматическому очагу.

Известно, что метаморфические породы характеризуются устойчивыми парагенетическими ассоциациями с небольшим (2—4) числом породообразующих минералов. Различия в наборе и процентных соотношениях этих минералов отражают различия в химическом составе исходных пород, претерпевших метаморфизм. Статистические методы проверки гипотезы о равенстве средних содержаний породообразующих минералов используются для стратиграфического расчленения метаморфических комплексов и корреляции их разрезов при детальном геологическом картировании.

В палеонтологии статистические методы проверки гипотезы о равенстве средних способствуют объективному разделению семейств ископаемых организмов на виды. Для выделения нового вида необходимо доказать, что данная группа ископаемых организмов существенно отличается по среднему значению какого-либо морфологического признака, например, по степени сферичности или углу между линиями замкового шва и краем вентрального синуса.

В процессе разведки месторождения о надежности выбранного способа отбора проб обычно судят по контрольным пробам, которые отбираются другим, более надежным способом, но, как правило, более трудоемким и дорогим. Проверка гипотезы о равенстве средних содержаний полезного компонента, рассчитанных по рядовым и контрольным пробам, позволяет объективно решить вопрос о наличии или отсутствии систематических ошибок в результатах рядового опробования. Число подобных примеров можно было бы увеличить. Общим во всех перечисленных случаях является невозможность уверенного решения задач такого типа путем визуального сравнения средних значений свойств, так как эти свойства характеризуются большой изменчивостью, а объем выборок часто бывает невелик. Как правило, выборочные оценки средних обладают значительными дисперсиями и могут заметно различаться даже для совершенно аналогичных объектов.

Для решения перечисленных задач используются параметрические и непараметрические критерии согласия.

Наиболее часто в геологической практике употребляется параметрический критерий Стьюдента  $t$ . Его применение основано на том, что если из нормально распределенной совокупности отобраны выборки  $X_1, X_2, \dots, X_k$  объемом в  $n_1$  значений и выборки  $Y_1, Y_2, \dots, Y_k$  объемом в  $n_2$  значений, то величина

$$t = |\bar{x} - \bar{y}| / \sqrt{S_1^2/n_1 + S_2^2/n_2}, \quad (20)$$

где  $\bar{x}, \bar{y}$  — выборочные оценки среднего, а  $S_1^2, S_2^2$  — выборочные оценки дисперсии, подчиняется закону распределения Стьюдента с  $n_1 + n_2 - 2$  степенями свободы. Проверка гипотезы о равенстве двух выборочных средних заключается в подстановке в формулу (20) оценок  $\bar{x}$  и  $S_1^2$  по первой и  $\bar{y}$  и  $S_2^2$  по второй выборке и сравнении полученного значения критерия  $t$  с табличным для данного числа степеней свободы и заданной доверительной вероятности. Если расчетное значение критерия превышает табличное, то гипотеза о равенстве выборочных средних отвергается.

В случае соответствия выборочных данных логнормальной модели для проверки гипотезы о равенстве средних рекомендуется использовать критерий Родионова. Д. А. Родионовым было установлено, что величина

$$Z = \left\{ |\overline{\lg x} - \overline{\lg y}| + 1,153 |S_{\lg x}^2 - S_{\lg y}^2| \right\} / \sqrt{S_{\lg x}^2/n_1 + S_{\lg y}^2/n_2 + 2,65 (S_{\lg x}^4/n_1 - 1 + S_{\lg y}^4/n_2 - 1)}$$

распределена асимптотически нормально с математическим ожиданием 0 и дисперсией 1. Поэтому при проверке гипотезы о равенстве средних с помощью этого критерия теоретическое значение величины  $Z$  находим по таблице значений интегральной функции Лапласа (см. прил. 11).

В некоторых геологических задачах, например при проверке гипотезы о комагматичности пород, гипотезу о равенстве математических ожиданий можно заменить гипотезой о равенстве центров распределения, т. е. медиан. В этом случае можно воспользоваться критерием Стьюдента даже при логнормальном распределении изучаемых величин, использовав оценки средних значений и дисперсий логарифмов:

$$t = |\overline{\lg x} - \overline{\lg y}| / \sqrt{S_{\lg x}^2/n_1 + S_{\lg y}^2/n_2}.$$

**Пример.** В районе широко развиты интрузии гранитов средне- и позднепалеозойского возраста. Со среднепалеозойскими гранитами связаны месторождения редкometалльных пегматитов. Среднепалеозойские граниты отличаются от нерудоносных верхнепалеозойских повышенным содержанием  $\text{Na}_2\text{O}$  и  $\text{TiO}_2$  и пониженным содержанием  $\text{K}_2\text{O}$  (табл. 15). Распределение содержаний  $\text{Na}_2\text{O}$  и  $\text{K}_2\text{O}$  в гранитах обоих комплексов соответствует нормальному закону, а содержание  $\text{TiO}_2$  — логнормальному. Эти различия можно использовать для определения возраста интрузий и оценки перспектив их рудоносности в тех случаях, когда эту задачу нельзя решить по возрастным взаимоотношениям гранитов и осадочных пород

Таблица 15. Исходные данные оценки рудоносности гранитов неизвестного возраста (в %)

Возраст гранитов	Число проб	Na <sub>2</sub> O		K <sub>2</sub> O		TiO <sub>2</sub>	
		$\bar{x}$	$S^2$	$\bar{x}$	$S^2$	$\lg \bar{x}$	$S_{\lg x}^2$
Средний палеозой	100	3,90	1,21	4,51	1,42	-0,886	0,268
Поздний палеозой	100	3,46	1,52	5,02	1,65	-1,426	0,321
Неизвестен	30	3,38	1,83	4,83	1,88	-1,352	0,225

В нижней строке табл. 15 приведены числовые характеристики содержания оксидов по одной из интрузий неизвестного возраста. Ее сопоставление с интрузиями определенного возраста по содержанию Na<sub>2</sub>O и K<sub>2</sub>O можно провести с помощью критерия Стьюдента, а по содержанию TiO<sub>2</sub> с помощью критерия Родионова.

В данной задаче проверяемая гипотеза  $H_0$  состоит в том, что  $\bar{x}$  и  $\bar{y}$  равны, т. е.  $H_0: \bar{x} = \bar{y}$ , при альтернативе  $H_1: \bar{x} \neq \bar{y}$ , т. е.  $\bar{x} > \bar{y}$  или  $\bar{x} < \bar{y}$ . Поэтому уровень значимости  $\alpha$  принимается для двусторонних критерии  $t$  и  $Z_{1-\alpha/2}$ .

При сравнении интрузии неизвестного возраста с рудоносными среднепалеозойскими гранитами первого рода будет состоять в том, что фактически перспективная интрузия того же возраста будет признана нерудоносной и поисковые работы на этом участке будут прекращены, что может привести к пропуску месторождения. Ошибка второго рода произойдет, если фактически нерудоносная интрузия другого возраста будет отнесена к рудоносному комплексу, поисковые работы будут продолжены, но не дадут положительного результата. Экономический ущерб от этой ошибки в данном случае заключается в непроизводительных затратах на поисковые работы в пределах оцениваемой локальной площади. Определить ущерб от ошибки первого рода сложнее, но он может быть гораздо значительнее, так как пропуск месторождения обычно можно восполнить только путем изучения гораздо больших по площади территорий в другом районе. Поэтому при проверке данной гипотезы уровень значимости, т. е. вероятность ошибки первого рода, целесообразно принять довольно низким — 0,05, несмотря на то, что при этом увеличивается вероятность ошибки второго рода.

Сравнивая ту же интрузию с заведомо нерудоносными верхнепалеозойскими гранитами, наоборот следует стремиться снизить вероятность ошибки второго рода. Поэтому уровень значимости целесообразно увеличить до 0,1.

Сравнение интрузии неизвестного возраста со среднепалеозойскими гранитами:

$$\text{по } \text{Na}_2\text{O} \left( t = |3,90 - 3,38| / \sqrt{\frac{121}{100} + \frac{138}{30}} = 2,16 \right);$$

$$\text{по } \text{K}_2\text{O} \left( t = |4,51 - 4,83| / \sqrt{\frac{1,42}{100} - \frac{1,88}{30}} = 1,15 \right);$$

$$\text{по } \text{TiO}_2 \left( Z = \{ |-0,886 + 1,352| + 1,153 |0,268 - 0,225|^2 \} / \sqrt{\frac{0,268}{100} + \frac{0,225}{30} + 2,65 \left( \frac{0,0718}{99} + \frac{0,0506}{29} \right)} = 3,99 \right).$$

Для уровня значимости  $\alpha=0,05$  критическое значение критерия Стьюдента при числе степеней свободы  $n_1+n_2-2=128$  равно 1,98.

Таким образом, гипотеза об идентичности средних содержаний  $\text{Na}_2\text{O}$  отвергается, в то время как различие по содержанию  $\text{K}_2\text{O}$  можно признать несущественным. Критическое значение  $Z$  для тех же условий равно 1,96, поэтому гипотеза о равенстве содержаний также отвергается. Последняя гипотеза отвергается даже в том случае если уровень значимости снизить до 0,001 ( $Z_{0,9995} \approx 3,3$ ).

Сравнение гранитов неизвестного возраста с верхнепалеозойскими гранитами:

$$\begin{aligned} \text{по } \text{Na}_2\text{O} & \left( t = |3,46 - 3,38| \sqrt{\frac{1,52}{100} + \frac{1,38}{30}} = 0,37 \right); \\ \text{по } \text{K}_2\text{O} & \left( t = |5,02 - 4,83| \sqrt{\frac{1,65}{100} + \frac{1,88}{30}} = 0,75 \right); \\ \text{по } \text{TiO}_2 & \left( Z = \{ |-1,426 + 1,352| + 1,153 |0,321 - 0,225| \} / \right. \\ & \left. \sqrt{\frac{0,321}{100} + \frac{0,225}{30} + 2,65 \left( \frac{0,1030}{99} + \frac{0,0506}{29} \right)} = 1,37 \right). \end{aligned}$$

Критическое значение критерия Стьюдента ( $t$ ) и критерия  $Z$  для уровня значимости  $\alpha=0,1$  равны:  $t_{0,1;128} \approx 1,66$ ;  $Z_{0,95} \approx 1,645$ . Поэтому гипотеза об идентичности химического состава изучаемой интрузии с верхнепалеозойскими гранитами принимается для всех трех компонентов.

По совокупности проверенных гипотез можно достаточно уверенно дать отрицательную оценку перспектив рудоносности в пределах данной интрузии.

Непараметрические критерии ( $X$ -критерий Ван-дер-Вардена, Вилкоксона) используются обычно при малом объеме выборок или в тех случаях, когда средние значения рассчитаны по полукачественным данным — например по результатам полукачественного спектрального анализа.

Проверка гипотезы о равенстве средних, определенных по двум выборкам ( $A$  и  $B$ ) с помощью  $X$ -критерия Ван-дер-Вардена, начинается с того, что все значения по обеим выборкам ранжируются, т. е. записываются в один ряд в порядке возрастания.  $X$ -критерий представляет собой величину

$$X = \sum_1^h \psi \left( \frac{i}{n+1} \right),$$

где  $n$  — общее число значений по двум выборкам;  $h$  — число наблюдений в выборке  $B$ ;  $i$  — порядковый номер каждого значения выборки  $B$  в общем ряду;  $\psi(\dots)$  — функция, обратная функции нормального распределения.

При  $n > 20$  величина  $X$  распределена асимптотически нормально с математическим ожиданием 0 и дисперсией  $\sigma_X^2$ . Процедура проверки гипотезы сводится к расчету всех значений аргумента  $i/(n+1)$ , нахождению по таблицам обратной функции нормального распределения значений функции  $\psi$  для этих аргументов, суммированию значений функции  $\psi$  и сравнению полученного значения критерия  $X$  с табличным для заданного уровня значимости, общего числа

наблюдений  $n$  и разницы между объемами выборок  $A$  и  $B$ . Если расчетное значение  $X$  по абсолютной величине больше табличного, гипотеза о равенстве выборочных средних отвергается.

Для функции  $\psi(\dots)$  существуют специальные таблицы. Однако ее значения можно найти и с помощью обычных таблиц интегральной функции нормального распределения с параметрами 0,1 (см. прил. 11), используя ее в обратном порядке. При этом значения аргумента  $i/(n+1)$  приравниваются к вероятностям  $p$ , а величина  $\psi(i/(n+1))$  находится по значениям  $Z$ , соответствующим этим вероятностям.

Для  $i/(n+1)$  значения  $\psi(i/(n+1))$  будут отрицательными, а для  $i/(n+1) > 0,5$  — положительными.

Если систематических расхождений между выборками  $A$  и  $B$  нет, то в ранжированном ряду значения каждой выборки будут располагаться симметрично относительно середины этого ряда, соответствующей  $i=n/2$  и  $i/(n+1)=0,5$ , число отрицательных и положительных значений  $\psi(i/(n+1))$  для каждой выборки будет примерно равным, а их алгебраические суммы, т. е. значения  $X$ -критерия, близки к нулю.

**Пример.** Для снижения затрат на разведку на одном из участков россыпного месторождения золота часть шурфов (примерно каждый второй) была заменена скважинами ударно-канатного бурения. Необходимо убедиться в том, что результаты опробования скважин не имеют систематической ошибки. Так как число скважин и шурfov на опытном участке невелико — 13 и 10, для сопоставления вычисленных по ним содержаний золота можно использовать непараметрический критерий Ван-дер-Вардена. Результаты опробования шурfov и скважин приведены в табл. 16. По этим данным составлен общий вариационный ряд (табл. 17).

Таблица 16. Результаты опробования разведочных выработок на россыпном месторождении золота

Выработка $A$ (скважины)		Выработка $A$ (скважины)		Выработка $B$ (шурфы)		Выработка $B$ (шурфы)	
№ п/п	Содержание Au, м <sup>2</sup> /м <sup>3</sup>	№ п/п	Содержание Au, м <sup>2</sup> /м <sup>3</sup>	№ п/п	Содержание Au, мг/м <sup>3</sup>	№ п/п	Содержание Au, мг/м <sup>3</sup>
1	322	7	192	1	431	6	221
2	250	8	375	2	397	7	548
3	225	9	381	3	462	8	478
4	315	10	538	4	457	9	299
5	399	11	198	5	251	10	541
6	348	12	317				
		13	293				

Проверяемая гипотеза  $H_0$  в данной задаче заключается в том, что систематических расхождений в определении содержаний золота по шурфам и скважинам нет (т. е.  $H_0: X=0, A=B$ ), при альтернативе  $H_1: X\neq 0, A\neq B$  (т. е.  $A>B$  или  $A<B$ ). Поэтому для нахождения критического значения  $X$  нужно воспользоваться таблицами двустороннего критерия Ван-дер-Вардена (см. прил. 6).

Ошибка первого рода в данной задаче заключается в том, что правильная гипотеза об отсутствии систематического расхождения между результатами опробования шурфов и скважин будет отвергнута, а это не позволит снизить затраты на дальнейшую разведку за счет использования более дешевых скважин. Ошибка второго рода, т. е. принятие неверной гипотезы  $H_0$ , состоит в признании несущественным (случайным) расхождения между данными по шурфам и скважинам, в то время как на самом деле оно носит систематический

Таблица 17. Проверка гипотезы о равенстве средних содержаний золота по критерию Ван дер Вардена

№ п/п	Содержание	Выборка	$t/(n+1)$	$\psi t/(n+1)$
1	192	А		
2	198	А		
3	221	Б		
4	225	А	0,125	-1,15
5	250	А		
6	251	Б	0,250	-0,67
7	293	А		
8	299	Б	0,333	-0,43
9	315	А		
10	317	А		
11	322	А		
12	348	А		
13	375	А		
14	381	А		
15	397	Б	0,625	0,32
16	399	А		
17	431	Б	0,708	0,55
18	457	Б	0,750	0,67
19	462	Б	0,792	0,81
20	478	Б	0,833	0,97
21	538	А		
22	541	Б	0,917	1,39
23	548	Б	0,958	1,73

$$X=4,19$$

характер ошибки второго рода в данной ситуации может привести к неправильной геолого-экономической оценке месторождения, что нанесет гораздо больший экономический ущерб по сравнению с дополнительными затратами за счет ошибки первого рода. Поэтому для уменьшения риска ошибки второго рода уровень значимости при проверке гипотезы  $H_0$  целесообразно принять не слишком высоким, например 0,1.

Допустимое значение  $X$ -критерия Ван-дер-Вардена для уровня значимости 0,1, числа наблюдений 23 и разности между объемами сравниваемых выборок 3, равно 3,12 (см. прил. 6).

Таким образом, гипотеза о равенстве средних значений содержания золота по скважинам и шурфам отвергается. Для выяснения причин систематического занижения содержания золота по скважинам использовать их для разведки данного месторождения не рекомендуется.

Непараметрический критерий Вилкоксона ( $W$ ) также основан на процедуре ранжирования и представляет собой сумму рангов  $R_i$  членов меньшей выборки в общем ранжированном ряду из обеих выборок:

$$W = \sum_{i=1}^{n_1} R_i, \quad n_1 \leq n_2.$$

Если гипотеза о равенстве средних по совокупностям  $A$  и  $B$  верна, т. е.  $H_0: \bar{x}_1 = \bar{x}_2$ , математическое ожидание статистики Вилкоксона ( $MW$ ) и величины возможных отклонений от нее выборочных оценок ( $W$ ) зависят только от объемов выборок  $n_1$  и  $n_2$ .

Для случаев, когда  $n_1$  и  $n_2 < 25$  значения удвоенного математического ожидания критерия Вилкоксона ( $2MW$ ) и его нижнего критического

значения  $W_1$  для заданного уровня значимости  $\alpha$  приведены в специальных таблицах (см. прил. 7). Верхнее критическое значение критерия  $W_2$  определяется из уравнения  $W_2 = 2M\bar{W} - W_1$ . Уровень значимости для  $W_1$  в этих таблицах дан для альтернативы  $H_1: \bar{x}_1 \neq \bar{x}_2, \bar{x}_1 < \bar{x}_2$ . Поэтому при альтернативе  $H_1: \bar{x}_1 \neq \bar{x}_2, \bar{x}_1 < \bar{x}_2$  или  $\bar{x}_1 > \bar{x}_2$  уровень значимости для нахождения  $W_1$  необходимо уменьшить в два раза.

Для приведенного выше примера сравнения содержаний золота по шурфам и скважинам (см. табл. 17) значение критерия Вилкоксона равно 151. Для уровня значимости  $\alpha/2 = 0,05$  и  $n_1 = 10, n_2 = 13$  с помощью прил. 7 находим:  $W_1 = 92, 2M\bar{W} = 240, W_2 = 148$ . Таким образом, эмпирическое значение критерия  $W$  превышает его верхнее критическое значение, следовательно, с вероятностью 0,9 гипотеза о равенстве средних содержаний золота по шурфам и скважинам отвергается.

Если  $n_1$  или  $n_2 > 25$ , критические значения критерия Вилкоксона можно определить по следующим приближенным формулам:

$$W_1 \approx 0,5 [n_1(n_1 + n_2 + 1) - 1] - Z_{1-\alpha/2} \sqrt{\frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)};$$

$$W_2 \approx n_1(n_1 + n_2 + 1) - W_1,$$

где  $Z_{1-\alpha/2}$  — значения функции нормального распределения с параметрами 0,1.

При наличии в объединенной выборке совпадающих значений им дается одинаковый средний ранг, равный среднему арифметическому из всех рангов, приходящихся на данную группу повторяющихся значений, а формула принимает следующий вид:

$$W_1 \approx \frac{n_1(n_1 + n_2 + 1) - 1}{2} - Z_{1-\alpha/2} \times$$

$$\times \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \left[ 1 - \frac{\sum_{i=1}^k (t_i^3 - t_i)}{(n_1 + n_2 + 1)(n_1 + n_2)(n_1 + n_2 - 1)} \right]^2},$$

где  $k$  — число групп из повторяющихся значений, принадлежащих разным выборкам;  $t_i$  — число совпавших значений в группе с номером  $i$  ( $i = 1, 2, 3, \dots, k$ ).

Группы повторяющихся значений, состоящие полностью из значений выборки  $A$  или  $B$ , можно не учитывать при введении поправки.

При сравнении двух случайных угловых величин обычно проверяется гипотеза о равенстве круговых средних направлений  $\mu$ . В случае соответствия обеих выборочных распределений закону Мизеса гипотезу  $H_0: \mu_1 = \mu_2$  при альтернативе  $H_1: \mu_1 \neq \mu_2$  можно проверить с помощью параметрического критерия Ватсона — Вильямса. Для этого по приведенным выше формулам (8, 9) по выборкам рассчитываются суммарные величины  $R_1$  и  $R_2$ , результирующие длины  $\bar{R}_1$  и  $\bar{R}_2$ , выборочные круговые средние направления  $m_1$  и  $m_2$ , а также общая векторная величина  $R = \sqrt{R_1^2 + R_2^2 + 2R_1 R_2 \cos(m_1 - m_2)}$ .

С помощью этих параметров вычисляется средняя общая результирующая длина  $R = R/(n_1 + n_2)$  и статистика Ватсона — Вильямса  $R' =$

$(R_1 + R_2)/(n_1 + n_2)$ . По значению  $\bar{R}$  с помощью таблиц функции  $k = A^{-1}(\bar{R})$  (см. прил. 2) определяется общая оценка параметра концентрации  $k$ . Предполагается, что параметры концентрации  $k_1$  и  $k_2$  сравниваемых угловых величин равны.

Если  $0 < \bar{R} < 0,7$ , а  $n_1$  отличается от  $n_2$  не более чем в два раза, критическое значение статистики Ватсона—Вильямса  $\bar{R}'_{kp}$  для уровня значимости  $\alpha=0,05$  можно определить с помощью номограмм, приведенных в прил. 8. Если  $\bar{R}' > \bar{R}'_{kp}$ , гипотеза о равенстве средних круговых направлений отвергается.

При  $\bar{R} > 0,7$  для проверки данной гипотезы можно использовать критерий Фишера. В случае  $0,7 < \bar{R} < 0,98$  эмпирическое значение  $F$ -критерия рассчитывается по формуле

$$F = \left(1 + \frac{3}{8k}\right) \frac{(n_1 + n_2 - 2)(R_1 + R_2 - R)}{n_1 + n_2 - R_1 - R_2},$$

а если  $\bar{R} > 0,98$ , по более простой формуле

$$F = \frac{(n_1 + n_2 + 2)(R_1 + R_2 - R)}{n_1 + n_2 - R_1 - R_2}.$$

Гипотеза  $H_0$  отвергается, если рассчитанное значение  $F$  превышает табличное значение критерия Фишера для заданного уровня значимости  $\alpha$  при степенях свободы  $f_1 = 1$  и  $f_2 = n_1 + n_2 - 2$ .

Пример. Наrudопроявлении золота были замерены азимуты падения кварцевых прожилков (см. табл. 12) и неминерализованных трещин. Для обеих выборок гипотеза о соответствии распределений закону Мизеса не отвергается. Параметры распределения азимутов падения кварцевых прожилков  $R_1 = 15,17$ ,  $\bar{R} = 0,316$ ,  $m_1 = 173,5$ ,  $n_1 = 48$ ,  $k_1 = 0,7$  и неминерализованных трещин  $R_2 = 16,47$ ,  $\bar{R}_2 = 0,328$ ,  $m_2 = 166$ ,  $n_2 = 50$ ,  $R_2 \approx 0,7$ .

По формулам (8, 9) рассчитаем параметры  $R_1$ ,  $\bar{R}$ ,  $k$  и  $\bar{R}'$ :

$$R = \sqrt{15,17^2 + 16,41^2 + 2 \cdot 15,17 \cdot 16,41 \cdot \cos(173,5^\circ - 166^\circ)} = 31,51^\circ;$$

$$\bar{R} = \frac{31,51}{98} = 0,3215; \quad \bar{R}' = \frac{15,17 + 16,41}{98} = 0,3222;$$

$$k = A^{-1}(0,3215) \approx 0,7.$$

Так как  $\bar{R} < 0,4$ , а  $n_1 \approx n_2$  для нахождения критического значения критерия Ватсона—Вильямса  $\bar{R}'_{kp}$  воспользуемся номограммой (см. прил. 8). Для  $\bar{R}$  равного 0,32  $\bar{R}'_{kp} \approx 0,35$ , следовательно, гипотеза о равенстве круговых средних направлений азимутов падения кварцевых прожилков и неминерализованных трещин не отвергается с доверительной вероятностью 0,95.

Непараметрическим аналогом критерия Ватсона—Вильямса является ранговый критерий равномерных меток Вилера—Ватсона—Ходжеса, применение которого не ограничивается условием соответствия выборочных данных какому-либо определенному закону распределения.

Построение этого критерия основано на объединении двух сравниваемых выборок объемом  $n_1$  (меньшая выборка) и  $n_2$  в общую выборку объемом  $n = n_1 + n_2$  и ранжировании всех замеров в порядке возрастания угловых величин. Фактические замеры заменяются величинами  $\beta = 360 R_i/n$ , где  $R_i$  — ранг, т. е. порядковый номер замера в общем ранжированном ряду. Эта операция соответствует размеще-

нию замеров по двум выборкам на окружности  $2\pi$  в порядке их возрастания на одном расстоянии друг от друга, равном  $2\pi/n$ . Ранговый критерий равномерных меток представляет собой величину  $R^* = 2(n-1) R_1^2 / n_1 n_2$ , где

$$R_1^2 = C_1^2 + S_1^2 = \left( \sum_{i=1}^{n_1} \cos \beta_i \right)^2 + \left( \sum_{i=1}^{n_1} \sin \beta_i \right)^2,$$

т. е. квадрат длины результирующего вектора для точек на окружности, соответствующих значениям меньшей по объему выборки.

Для малых выборок ( $n \leq 20$ ) рассчитаны таблицы критических значений статистики  $R^*$ . При  $n > 20$ , если гипотеза  $H_0$  о равенстве круговых средних верна, статистика  $R^*$  распределена приближенно как  $\chi^2$  с двумя степенями свободы. Поэтому ее критические значения для заданного уровня значимости можно найти с помощью прил. 2.

*Пример.* В пределах рудного поля оси линейных складок и основные рудоконтролирующие разрывные нарушения ориентированы в северо-восточном направлении (табл. 18). Проверим гипотезу о равенстве круговых средних их азимутов простирации. Для определения рангов общей выборки (табл. 18, графы 3, 5) используются замеры по обеим выборкам, а расчет значений  $\beta_i$  (табл. 18, графа 6) и статистик  $R_1^2$  производится только по замерам азимутов разрывных нарушений, составляющим выборку меньшего объема. Значение критерия  $\chi^2$  для уровня значимости  $\alpha = 0,1$  и числа степеней свободы  $f = 2$  равно 4,605 и значительно превышает рассчитанную величину статистики. При уменьшении уровня значимости величина  $\chi^2$  возрастает, поэтому отвергать гипотезу о равенстве круговых средних направлений осей складок и азимутов простирации разрывных нарушений нет оснований.

Таблица 18. Проверка гипотезы о равенстве азимутов простирации разрывных нарушений и осей складок

№ п/п	Складки		Разрывные нарушения					
	азимут	ранг	азимут	ранг	$\beta_i$	$\sin \beta_i$	$\cos \beta_i$	
1	2	3	4	5	6	7	8	
1	42	23	23	11	146,67	0,549	-0,836	
2	40	20	26	13	173,33	0,116	-0,993	
3	20	8	16	5	66,67	0,918	0,396	
4	24	12	62	27	360,00	0	1	
5	28	14	38	19	253,33	-0,958	-0,287	
6	14	4	18	6	80,00	0,985	0,174	
7	32	17	12	2	26,67	0,449	0,894	
8	42	24	30	16	213,33	-0,549	-0,836	
9	56	26	10	1	13,33	0,231	0,973	
10	22	10	12	3	34,00	0,559	0,829	
11	40	21	30	15	200,00	-0,342	-0,940	
12	20	9						
13	40	22						
14	32	18						
15	19	7						
16	46	25						
						$\Sigma 1,958$	0,374	

$$R_1^2 = \left( \sum_{i=1}^{n_1} \cos \beta_i \right)^2 + \left( \sum_{i=1}^{n_1} \sin \beta_i \right)^2 = 3,834 + 0,140 = 3,974$$

$$R^* = \frac{2(n-1)R_1^2}{n_1 n_2} = \frac{2,26 \cdot 3,974}{11 \cdot 16} = 1,17$$

$$\chi^2_{\alpha=0,1; f=2} = 4,605$$

Для проверки гипотезы о равенстве вероятностей  $p_1$  и  $p_2$  двух биномиально распределенных дискретных случайных величин можно воспользоваться следующим отношением:

$$Z_{\text{наб}} = |\hat{p}_1 - \hat{p}_2| / \sqrt{\left( \frac{x_1 + x_2}{n_1 + n_2} \right) \left( 1 - \frac{x_1 + x_2}{n_1 + n_2} \right) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad (21)$$

где  $\hat{p}_1$  и  $\hat{p}_2$  — оценки вероятностей случайного события  $A$ , равные  $x_1/n_1$  и  $x_2/n_2$ ;  $x_1$  и  $x_2$  — частоты, появления события  $A$ ;  $n_1$  и  $n_2$  — число испытаний в сравниваемых экспериментах.

Если гипотеза  $H_0: p_1 = p_2$  верна, величина  $Z_{\text{наб}}$  распределена асимптотически нормально с параметрами 0 и 1, что дает возможность использовать таблицу функции Лапласа (см. прил. 11) для нахождения критического значения  $Z_{\text{кр}}$  при проверке данной гипотезы с заданным уровнем значимости  $\alpha$ . Если альтернативная гипотеза заключается в том, что  $p_1 \neq p_2$ , значение  $Z_{\text{кр}}$  по таблицам функции Лапласа находится для вероятности  $P=1-\alpha/2$ , а при альтернативах  $H_1: p_1 > p_2$  или  $p_2 > p_1$  для вероятности  $p=1-\alpha$ .

*Пример.* Месторождение молибдена отрабатывается карьером. По уступу карьера, где пробурена густая сеть буровзрывных скважин, установлено, что коэффициент рудоносности, т. е. отношение скважин, вскрывших кондиционную

руду ( $x_1$ ) к их общему количеству ( $n_1$ ), составляет:  $K_{\text{руд}} = x_1/n_1 = \hat{p}_1 = \frac{358}{1117} = 0,32$ .

Нижележащий уступ вскрыт редкой сетью разведочных скважин, по которым установлено, что  $K_{\text{руд}} = x_2/n_2 = \hat{p}_2 = \frac{10}{25} = 0,40$ . Высказано предположение, что

коэффициент рудоносности, а следовательно и отношение объема руды к объему пустых пород с глубиной увеличивается. Это предположение можно подтвердить путем проверки гипотезы о равенстве коэффициентов рудоносности ( $H_0: p_1 = p_2$ ) при альтернативе  $H_1: p_1 < p_2$ . По формуле (21) получим:

$$Z_{\text{наб}} = |0,32 - 0,40| / \sqrt{\left( \frac{358+10}{1117+25} \right) \left( 1 - \frac{358+10}{1117+25} \right) \left( \frac{1}{1117} + \frac{1}{25} \right)} =$$

$$= \frac{0,08}{0,0945} \approx 0,85.$$

Значение  $Z_{\text{кр}}$  для уровня значимости  $\alpha=0,05$ , т. е. для доверительной вероятности  $1-\alpha=0,95$  равно 1,65 (см. прил. 11).

По той же таблице находим, что значению  $Z=0,85$  соответствует вероятность 0,8023. Следовательно, гипотезу о равенстве коэффициентов рудоносности можно отвергнуть лишь в том случае, если уровень значимости  $\alpha$ , т. е. вероятность ошибки первого рода, будет увеличен до 0,2.

Расчеты показывают, что различие в оценках коэффициентов рудоносности по двум уступам не очень существенно и может быть обусловлено малым количеством скважин на нижнем уступе, а не фактическим увеличением рудонасыщенности с глубиной.

## 2.9. ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ ДИСПЕРСИЙ

Сравнение геологических объектов по степени изменчивости, которая оценивается по величине дисперсии или коэффициента вариации тех или иных свойств, необходимо для обоснованного применения принципа аналогии при их изучении. Так, например, дисперсия мощности рудных тел характеризует сложность их строения.

Различие в дисперсиях свойств аналогичных по составу геологических объектов может указывать и на различие в истории их формирования. Так различие дисперсий содержаний основных породообразующих минералов в двух схожих по составу комплексах магматических пород может указывать на то, что комплекс, для которого характерна большая степень рассеяния содержаний, формировался в течение более длительного периода и в нем сильнее проявились процессы дифференциации.

Различные горные породы, сходные по средним значениям физических свойств — магнитной восприимчивости, электропроводимости и т. п., часто отличаются по степени изменчивости этих свойств. Поэтому путем проверки гипотез о равенстве (различии) дисперсий можно проводить литологическое расчленение разрезов по данным геофизического каротажа скважин при бескерновом бурении, а также интерпретировать результаты геофизических съемок при составлении геологических карт.

На сравнении дисперсий основаны также методы определения величин случайных погрешностей различных способов опробования и анализов. Если количественные данные о свойствах геологического объекта получены различными способами, то более надежным следует признать тот способ, который дает меньший разброс значений изучаемого свойства, т. е. характеризуется меньшей дисперсией.

Для проверки гипотезы о равенстве дисперсий  $\sigma_1^2$  и  $\sigma_2^2$  обычно используется критерий Фишера F. Р. Фишером было установлено, что в случае равенства дисперсий двух нормально распределенных случайных величин, величина  $F = S_1^2/S_2^2$  при  $S_1^2 > S_2^2$  распределена по закону Фишера с  $n_1 - 1$  и  $n_2 - 1$  степенями свободы, где  $n_1$  — количество членов в выборке, по которой получена большая оценка дисперсии  $S_1^2$ , а  $n_2$  — объем второй выборки. Процедура проверки гипотезы сводится к нахождению эмпирического значения F-критерия и сравнению его с табличным значением для принятой доверительной вероятности  $(1-\alpha)$  и степенях свободы  $f_1 = n_1 - 1$  и  $f_2 = n_2 - 1$ . Если вычисленное значение критерия Фишера превышает табличное, то гипотеза о равенстве двух дисперсий отвергается.

В приложении приведены критические значения F-критерия для уровня значимости  $\alpha$  и альтернативной гипотезы  $H_1: \sigma_1^2 > \sigma_2^2$ . При сложной альтернативе  $H_1: \sigma_1^2 \neq \sigma_2^2$ , т. е.  $\sigma_1^2 > \sigma_2^2$  или  $\sigma_1^2 < \sigma_2^2$ , критическое значение критерия Фишера находят для уровня значимости  $\alpha/2$ .

В условиях асимметричных распределений критерий Фишера обладает малой мощностью. В случае логнормального распределения сравниваемых совокупностей при использовании этого критерия необходимо пользоваться максимально правдоподобными оценками

дисперсий или проверять гипотезу о равенстве дисперсий логарифмов значений исследуемого признака.

*Пример.* Воспользуемся статистическими характеристиками, приведенными в табл. 15, и проверим гипотезу о равенстве дисперсий содержаний  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$  и  $\text{TiO}_2$  в гранитах неизвестного возраста и гранитах средне- и верхнепалеозойского комплексов при альтернативе  $H_1: \sigma_1^2 \neq \sigma_2^2$  и уровне значимости  $\alpha/2 = 0,05$ .

Сравнение гранитов неизвестного возраста со среднепалеозойскими гранитами:

$$\text{по } \text{Na}_2\text{O} \left( F = \frac{1,38}{1,21} = 1,14, f_1 = 29, f_2 = 99, F_{kp} \approx 1,60 \right);$$

$$\text{по } \text{K}_2\text{O} \left( F = \frac{1,88}{1,42} = 1,32, f_1 = 29, f_2 = 99, F_{kp} \approx 1,60 \right);$$

$$\text{по } \text{TiO}_2 \left( F = \frac{0,268}{0,225} = 1,19, f_1 = 99, f_2 = 29, F_{kp} \approx 1,80 \right).$$

Сравнение гранитов неизвестного возраста с верхнепалеозойскими гранитами:

$$\text{по } \text{Na}_2\text{O} \left( F = \frac{1,52}{1,38} = 1,10, f_1 = 99, f_2 = 29, F_{kp} \approx 1,80 \right);$$

$$\text{по } \text{K}_2\text{O} \left( F = \frac{1,88}{1,65} = 1,11, f_1 = 29, f_2 = 99, F_{kp} \approx 1,60 \right);$$

$$\text{по } \text{TiO}_2 \left( F = \frac{0,321}{0,225} = 1,43, f_1 = 99, f_2 = 29, F_{kp} \approx 1,80 \right).$$

Во всех случаях рассчитанные значения критерия Фишера оказались меньше критических, следовательно рассматриваемые граниты существенно не отличаются по степени изменчивости содержаний данных химических элементов как от среднепалеозойских, так и от верхнепалеозойских гранитов. Поэтому характеристики изменчивости в данном случае нельзя использовать в качестве классификационного признака.

В процессе решения геологических задач иногда приходится сравнивать по степени изменчивости разные признаки, например мощность рудного тела и содержание в нем полезного компонента. В этом случае проверяется гипотеза о равенстве их коэффициентов вариации  $V_1$  и  $V_2$ . Для этого используется величина

$$F = \frac{V_1^2}{1 + V_1^2} \left( \frac{n_1}{n_1 + 1} \right) / \frac{V_2^2}{1 + V_2^2} \left( \frac{n_2}{n_2 + 1} \right),$$

которая также распределена по закону Фишера с  $n_1 - 1$  и  $n_2 - 1$  степенями свободы, при условии, что в знаменателе стоит меньшая из величин  $\frac{V^2}{1 + V^2} \left( \frac{n}{n + 1} \right)$ .

Непараметрическим аналогом критерия Фишера является критерий Сиджела—Тьюки, по процедуре вычисления во многом сходный с критерием Вилкоксона. Он применим для распределений любого вида и не чувствителен к аномальным значениям, поэтому весьма удобен для решения геологических задач, особенно по выборкам малого объема.

Критерий Сиджела—Тьюки построен исходя из предположения о равенстве центров распределения сравниваемых совокупностей. Поэтому в случае несоблюдения этого условия исходные данные по каждой выборке необходимо центрировать относительно их медиан, т. е. сравнивать не сами значения изучаемых параметров, а их отклонения от медиан.

Значения сравниваемых выборочных совокупностей объединяются в общую выборку и записываются в виде вариационного ряда в порядке их возрастания:  $x_1 < x_2 < x_3 < \dots < x_{N-1}$ , где  $N = n_1 + n_2$  — объем общей выборки;  $n_1$  — объем меньшей выборки. Члены вариационного ряда, в свою очередь, ранжируются следующим образом: ранг 1 приписывается наименьшему члену ряда  $x_1$ ; ранг 2 — наибольшему, т. е.  $x_N$ ; ранг 3 — значению  $x_2$ ; ранг 4 — значению  $x_{N-1}$  и т. д. Если  $N$  нечетно, то медианному значению ранг не присваивается. При таком ранжировании значениям выборки с меньшей дисперсией будут присваиваться преимущественно большие ранги, а значениям выборки с большей дисперсией — наоборот, малые. В случае равенства дисперсий значения из разных выборок будут чередоваться в ранжированном ряду случайно, и сумма рангов, относящихся к членам меньше по объему выборки, будет обладать всеми свойствами рассмотренного выше критерия Вилкоксона ( $W$ ). Дальнейшая проверка гипотезы о равенстве дисперсий сводится к нахождению критических значений  $W_1$  и  $W_2$  по описанной выше процедуре и сравнению с ними рассчитанного значения  $W$ .

*Пример.* По одному из участков молибден–вольфрамового месторождения для контроля бороздовых проб (выборка А) отобрано 16 валовых проб большой массы (выборка Б). Проверим гипотезу о равенстве средних значений и дисперсий содержаний молибдена по пробам, отобранным разным способом, при альтернативах  $H_1$ :  $\bar{x} \neq \bar{x}_B$ ,  $H'_1$ :  $\sigma_A^2 \neq \sigma_B^2$  и уровне значимости  $\alpha = 0,1$ .

Запишем приведенные в табл. 19 результаты опробования в виде ранжированного ряда (табл. 20, графа 2) и определим ранги  $R_1$  для расчета критерия Вилкоксона  $W$  (см. табл. 20, графа 4). С помощью прил. 6 находим критические значения критерия Вилкоксона для  $n_1 = 16$ ,  $n_2 = 20$  и  $\alpha/2 = 0,05$ :  $W_1 = 243$ ,  $W_2 = 349$ . Расчетное значение критерия  $W$ , определенное по сумме рангов значений из меньшей выборки Б, равно 250, т. е. находится в интервале между нижним и верхним критическими значениями критерия Вилкоксона. Следовательно, гипотеза о равенстве средних содержаний молибдена по пробам разной массы не отвергается.

Ввиду равенства средних значений при проверке гипотезы о равенстве дисперсий нет необходимости в центрировании данных. Выполнив ранжирование по способу Сиджела—Тьюки (см. табл. 20, графа 5), получим значение критерия  $W'$  для проверки гипотезы о равенстве дисперсий. Оно равно 354 и превышает верхнее критическое значение критерия  $W_2$ . Следовательно гипотеза о равенстве дисперсий отвергается. Дисперсия содержаний молибдена по бороздовым пробам существенно выше, чем по валовым.

Результаты проверки гипотез позволяют сделать следующие выводы:

- различие в средних содержаниях молибдена по пробам разной массы несущественно, следовательно опробование малыми пробами не приведет к систематическим ошибкам в оценке среднего качества руд;

- содержания по бороздовым пробам отличаются большим разбросом, следовательно для оценки среднего содержания молибдена с заданной точностью их требуется значительно больше, чем валовых.

Таблица 19. Содержание (в %) молибдена по бороздовым (выборкам А) и валовым пробам (выборка Б)

Выборка А				Выборка Б			
№ п/п	Содержание Mo	№ п/п	Содержание Mo	№ п/п	Содержание Mo	№ п/п	Содержание Mo
1	0,065	11	0,082	1	0,056	11	0,046
2	0,075	12	0,061	2	0,078	12	0,056
3	0,110	13	0,022	3	0,067	13	0,088
4	0,140	14	0,070	4	0,120	14	0,104
5	0,182	15	0,140	5	0,038	15	0,054
6	0,010	16	0,060	6	0,062	16	0,064
7	0,192	17	0,025	7	0,072		
8	0,310	18	0,230	8	0,167		
9	0,037	19	0,180	9	0,068		
10	0,160	20	0,105	10	0,059		

Таблица 20. Проверка гипотезы о равенстве средних и дисперсий содержаний (в %) молибдена по бороздовым и валовым пробам

№ п/п	Содержание Mo	Выборка	Ранг по критерию Вилкоксона ( $R_1$ )	Ранг по критерию Сиджела — Тьюки ( $R_2$ )
1	0,010	А	1	1
2	0,022	А	2	3
3	0,025	А	3	5
4	0,037	А	4	7
5	0,038	Б	5	9
6	0,046	Б	6	11
7	0,054	Б	7	13
8	0,056	Б	8,5	15
9	0,056	Б	8,5	17
10	0,059	Б	10	19
11	0,060	А	11	21
12	0,061	А	12	23
13	0,062	Б	13	25
14	0,064	Б	14	27
15	0,065	А	15	29
16	0,067	Б	16	31
17	0,068	Б	17	33
18	0,070	А	18	35
19	0,072	Б	19	36
20	0,075	А	20	34
21	0,078	Б	21	32
22	0,082	А	22	30
23	0,088	Б	23	28
24	0,104	Б	24	26
25	0,105	А	25	24
26	0,110	А	26	22
27	0,120	Б	27	20
28	0,140	А	28,5	18
29	0,140	А	28,5	16
30	0,160	А	30	14
31	0,167	Б	31	12
32	0,180	А	32	10

№ п/п	Содержание Mo	Выборка	Ранг по критерию Вилкоксона ( $R_1$ )	Ранг по критерию Сиджела—Тьюки ( $R_2$ )
33	0,182	A	33	8
34	0,192	A	34	6
35	0,230	A	35	4
36	0,310	A	36	2

$n_1=16$ ;  $n_2=20$ ;  $\alpha/2=0,05$ ;  
 $\hat{W}=\sum R_1$  (по выборке Б)=250;  $W'=\sum R_2$  (по выборке Б)=354;  $W_1=243$ ;  
 $2MW=592$ ;  $W_2=2MW-W_1=349$ .

## 2.10. АНАЛИЗ ОДНОРОДНОСТИ ВЫБОРОЧНЫХ ГЕОЛОГИЧЕСКИХ СОВОКУПНОСТЕЙ

При использовании одномерных статистических моделей для описания свойств геологических объектов предполагается, что данный объект однороден в отношении изучаемого свойства. Обычно вопрос об однородности решается исходя из принятой геологической модели. Исследуемый объект считается статистически однородным, если он однороден по геологическому строению. Однако на ранних стадиях изучения трудно однозначно решить вопрос о геологической однородности на основе только качественной геологической информации. В этих случаях можно использовать обратный прием—получать суждение о геологической однородности объекта путем проверки гипотезы о его статистической однородности, используя количественные данные о характере изменчивости его свойств.

Задачи, основанные на проверке гипотезы о статистической однородности геологических объектов, можно разделить на три типа:

- выделение аномальных значений;
- разделение неоднородных выборочных совокупностей;
- оценка степени влияния различных факторов на характер изменчивости свойств геологических объектов.

Выявление локальных неоднородностей (аномалий) в строении геологических объектов имеет исключительно важное практическое значение при проведении поисковых работ, так как они часто используются в качестве признаков, указывающих на наличие повышенных концентраций полезных ископаемых.

Наличие же в выборочных совокупностях резко выдающихся значений, обусловленных локальными причинами и не характерными для данного геологического объекта, в целом, снижает точность вычисления точечных и интервальных оценок средних параметров и затрудняет решение рассмотренных выше задач на основе проверки гипотез о равенстве средних и дисперсий.

Для выделения аномальных значений совокупность результатов наблюдений рассматривается как выборка из двух различных генеральных совокупностей—«фоновой» и «аномальной». При этом аномаль-

ные значения присутствуют в выборке в очень небольшом количестве или совсем отсутствуют.

В случаях нормального распределения фоновой генеральной совокупности эта задача решается с помощью параметрических критериев Смирнова и Фергюсона.

Н. В. Смирновым было установлено, что если максимальный по значению член выборочной совокупности не является аномальным, то величина  $t = (x_{\max} - \bar{x}) / S_{cm}^2$  имеет распределение, названное его именем. В данной формуле  $x_{\max}$  — максимальный член выборки;  $\bar{x}$  — среднее арифметическое;  $S_{cm}^2$  — смещенная оценка дисперсии, которая рассчитывается через несмещенную оценку дисперсии  $S^2$  по формуле  $S_{cm}^2 = S^2 \left( \frac{n-1}{n} \right)$ .

Если рассчитанное значение критерия больше допустимого, определенного по таблицам распределения Смирнова для заданной доверительной вероятности и  $n$  степеней свободы, то максимальное значение выборки следует считать аномальным.

Критерий Фергюсона основан на том, что если выборочная совокупность не содержит аномальных значений, то оценка коэффициента асимметрии  $A$  будет распределена асимптотически нормально с математическим ожиданием 0 и дисперсией  $\sigma_A^2$ .

Если рассчитанное значение коэффициента асимметрии превышает табличное для заданной доверительной вероятности и  $n$  степеней свободы, то максимальное значение выборки следует признать аномальным. Если распределение фоновой совокупности отличается от нормального, то «аномальными» будут признаваться все редко встречающиеся большие значения, принадлежащие исследуемой генеральной совокупности. Это ограничивает область применения обоих критериев. Они могут применяться только в том случае, если заранее известно, что распределение фоновой совокупности является нормальным.

В практике геохимических исследований за аномальные значения часто принимают маловероятные значения, по абсолютной величине превышающие  $3\sigma \pm \bar{x}$  или  $\bar{x} \pm 2\sigma$  (т. е. отличающиеся от среднего на утроенное или удвоенное значение стандартного отклонения). Однако этот способ нельзя признать корректным, так как он не гарантирует от ошибок как первого, так и второго рода, причем вероятность этих ошибок оценить нельзя.

Таким образом, задача выявления аномальных значений не имеет универсального решения статистическими методами. Аномальное значение должно определяться опытным путем на основе анализа геологических причин изменения значений изучаемых свойств. Статистические характеристики при этом будут иметь вспомогательное значение.

Если количество наблюдений, принадлежащих разным геологическим совокупностям в неоднородной выборке велико, то возникает необходимость и возможность ее разделения на несколько однородных совокупностей.



Рис. 23. Классификация амфиболитов по содержанию роговой обманки. По Б. М. Роненсону и др.:

*а* — экспериментальный график распределения содержаний роговой обманки; *б* — графики плотности вероятностей для совокупностей *A*, *B* и *C*; *в* — графики интегральной функции распределения для совокупностей *A*, *B* и *C*

К задачам такого типа относятся:

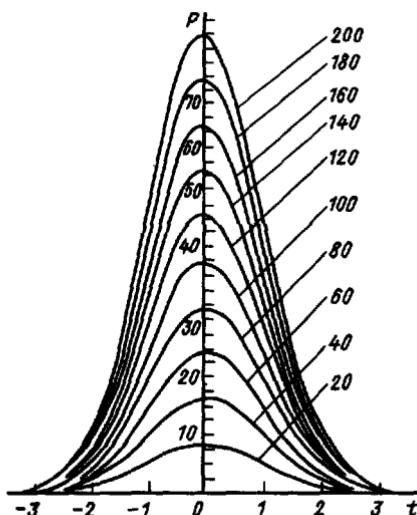
- расчленение разреза осадочных пород по микрофауне;
- расчленение немых толщ по петрографическому, минеральному или химическому составу;
- расчленение древней коры выветривания по комплексу геохимических признаков;
- выделение маркирующих горизонтов;
- определение рационального комплекса геофизических методов для геологического картирования;
- таксономическая классификация ископаемых организмов по морфологическим признакам;
- классификация метаморфических и интрузивных горных пород по химическому составу;
- выделение однородных подсчетных блоков при подсчете запасов месторождений.

Решение этих задач включает операции по проверке гипотезы о неоднородности исходной выборочной совокупности, нахождению границ входящих в нее однородных совокупностей и оценке параметров каждой из этих совокупностей. При этом для разделения исследуемого множества наблюдений или объектов на однородные совокупности используются как формальные математические методы, так и экспертный метод, когда геолог на основании опыта и представлений о структуре изучаемого сложного геологического объекта первоначально выделяет элементарные условно однородные совокупности, а окончательные границы однородных совокупностей определяются после объединения элементарных совокупностей, различия между которыми признаются статистически незначимыми.

Простейшие способы разделения неоднородных выборочных совокупностей основаны на анализе графиков эмпирических кривых распределения. На неоднородность выборки может указывать наличие на графике нескольких максимумов (рис. 23). Неоднородные выборочные совокупности можно разделить с помощью специальных палеток.

Палетки (рис. 24) представляют собой набор кривых функций плотности распределения нормированной и центрированной случайной величины, умноженной на объем выборки. Для пользования палетками необходимо, чтобы вертикальный масштаб исследуемой эмпирической кривой распределения соответствовал масштабу палетки. В первую

Рис. 24 Палетка для подбора эталонных кривых плотности вероятностей



очередь подбираются кривые, наилучшим образом аппроксимирующие крайние участки эмпирической кривой. Объем каждой однородной выборки, входящей в смешанную совокупность, определяется по индексу кривой, положение нуля палетки указывает на математическое ожидание однородной выборки, а отношение горизонтальных масштабов палетки и эмпирической кривой соответствует ее стандарту. По полученным числовым характеристикам однородных совокупностей можно построить графики интегральных функций распределений (см. рис. 23), с помощью которых удобно оценивать надежность выделения одних однородных объектов на фоне других. Надежность (вероятность) разделения будет определяться по расстоянию  $P_{ab}$  между интегральными функциями распределения вдоль линии, проходящей через точку пересечения  $x_0$  кривых плотности вероятностей (см. рис. 23, в). При расчленении комплексов горных пород с помощью геофизических методов этот способ, с одной стороны, позволяет решить вопрос о том, какие породы в районе выделяются наиболее надежно, т. е. дает возможность выбрать маркирующие горизонты, а с другой — позволяет выбрать наиболее эффективный комплекс геофизических методов для картирования каждого типа пород.

Разделение полимодальных распределений можно выполнять также путем анализа геологических неоднородностей изучаемого объекта.

*Пример.* На гистограмме частотного распределения содержания  $\text{Fe}_2\text{O}_3$  по пробам метровой длины из месторождений никеля района Моя (о. Куба) отчетливо выделяются два модальных значения (рис. 25, а), что, по-видимому, обусловлено неоднородностью коры выветривания ультрамафитов, с которой связано оруденение. В кондиционной по содержанию никеля части коры выветривания визуально по текстурным особенностям и минеральному составу выделяется сверху вниз пять зон: железистых стяжений, бесструктурных охр, зрелых структурных охр, первичных структурных охр, выщелоченных и обогащенных материнских пород. Количество проб по каждой из этих зон

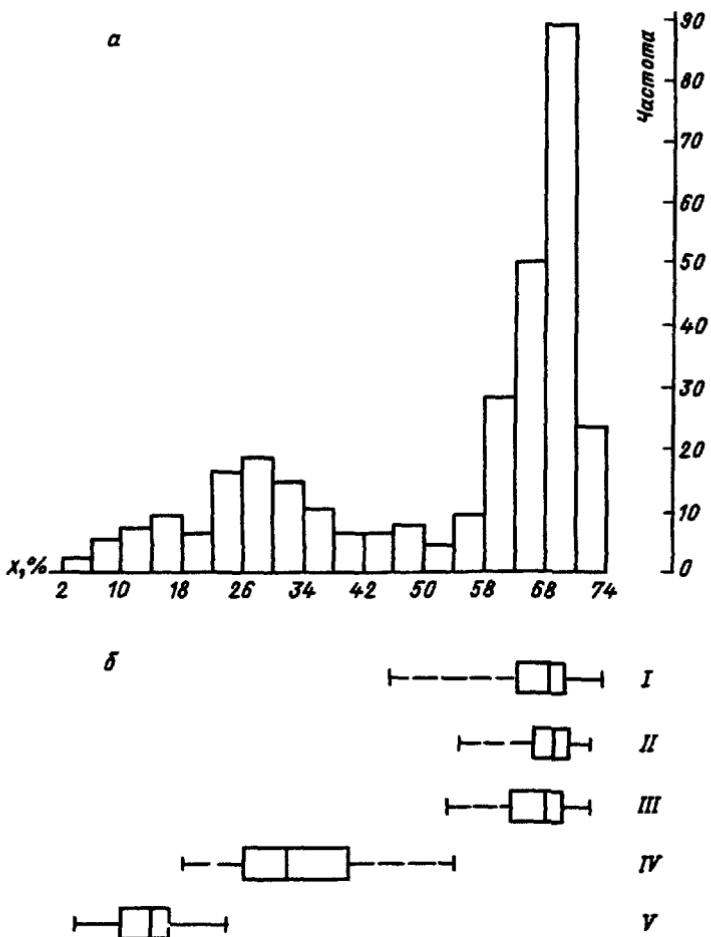


Рис. 25. Распределение содержания  $\text{Fe}_2\text{O}_3$  (в %) в рудах месторождений района Moa:  
 а—гистограмма распределения содержания  $\text{Fe}_2\text{O}_3$ ; б—схематические диаграммы распределения  $\text{Fe}_2\text{O}_3$  по разным зонам коры выветривания

невелико, поэтому распределения содержаний  $\text{Fe}_2\text{O}_3$  по ним целесообразнее представить в виде числовых диаграмм «ящик с усами» (см. рис. 25, б), а не гистограмм. Сравнение диаграмм позволяет сделать вывод, что по содержанию  $\text{Fe}_2\text{O}_3$  кора выветривания отчетливо разделяется на две части. Верхняя часть, включающая первые три зоны, характеризуется резко повышенным содержанием железа и может рассматриваться как залежь комплексных железонikelевых руд, а нижняя, объединяющая зоны первичных структурных охр и обогащенных материальных пород, представляет промышленный интерес только как источник никеля и кобальта. При этом две нижние зоны существенно различаются между собой по содержанию  $\text{Fe}_2\text{O}_3$ .

В тех случаях, когда визуальное сравнение гистограмм или числовых диаграмм, характеризующих однородные с геологических позиций участки, не позволяет однозначно решить вопрос об их сходстве или различии, эта задача решается путем последовательной проверки гипотез о равенстве средних

и дисперсий по выборкам из разных участков. Участки, для которых гипотеза о равенстве параметров распределений не отвергается, объединяются. Эта процедура продолжается до тех пор, пока все элементарные участки не объединятся в совокупности, различие между которыми признается значимым для принятой доверительной вероятности.

## 2.11. ОДНОФАКТОРНЫЙ И ДВУХФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ В ГЕОЛОГИИ

Свойства геологических объектов, как и любой другой сложной природной системы, обычно зависят от ряда факторов, обуславливающих их изменчивость. Выявление этих факторов и оценка степени их влияния на изменчивость (неоднородность) свойств изучаемых объектов осуществляется с помощью дисперсионного анализа.

Этот статистический метод основан на следующем принципе: если на случайную величину действуют взаимонезависимые факторы  $A, B, \dots, D$ , то общую дисперсию этой случайной величины  $\sigma^2$  можно рассматривать как сумму дисперсий  $\sigma^2 = \sigma_A^2 + \sigma_B^2 + \dots + \sigma_D^2$ .

По количеству оцениваемых факторов дисперсионный анализ подразделяется на одно-, двух- и многофакторный.

Каждый фактор представляет собой переменную дискретную или непрерывную величину, которая разделяется на определенное количество постоянных интервалов (уровней). Если количество замеров изучаемой случайной величины на всех уровнях по всем факторам одинаково, дисперсионный анализ принято называть *равномерным*, а если разное — *неравномерным*.

Суждение о влиянии определенного фактора или комбинации факторов на изменчивость изучаемой случайной величины основано на группировке ее замеров по факторам и их уровням и проверке гипотезы о равенстве дисперсий, обусловленных данными факторами, с остаточной (случайной) дисперсией, вызванной неучтенными факторами. Если гипотеза отвергается, то делается вывод о том, что данный фактор или взаимодействие факторов оказывают существенное влияние на изменение изучаемого свойства геологического объекта.

С помощью дисперсионного анализа решается широкий круг геологических задач:

- проверяются гипотезы о влиянии литологических, петрофизических, геохимических, структурных и других факторов на локализацию оруденения;
- выявляются элементы зональности различных геологических объектов;
- определяется влияние способа отбора проб на их достоверность и представительность;
- оценивается влияние ландшафтных условий на интенсивность проявления различных поисковых признаков;
- решается вопрос о влиянии гипергенных процессов на качество руд;
- выявляются факторы, определяющие прочностные свойства грунтов и пород и т. д.

Таблица 21. Однофакторный дисперсионный анализ

Номер изменения	Уровень фактора			
	$A_1$	$A_2$	...	$A_k$
1	$x_{11}$	$x_{12}$	...	$x_{1k}$
2	$x_{21}$	$x_{22}$	...	$x_{2k}$
.	.	.	...	.
.	.	.	...	.
$n$	$x_{n1}$	$x_{n2}$	...	$x_{nk}$
Групповые средние	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_k$

При равномерном однофакторном дисперсионном анализе случайной величины  $x$  относительно фактора  $A$ , имеющего  $k$  уровней при количестве замеров на каждом уровне равном  $n$ , результаты наблюдений обозначаются как  $x_{ij}$ , где  $i$ —номер наблюдения ( $i=1, 2, \dots, n$ ), а  $j$ —номер уровня фактора ( $j=1, 2, \dots, k$ ) и записываются в виде табл. 21.

По этим данным рассчитываются следующие статистики:

1) общая сумма квадратов отклонений наблюдаемых значений признака от общей средней  $\bar{x}$ :

$$C_{\text{общ}} = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x})^2;$$

2) факторная сумма квадратов отклонений групповых средних от общей средней, характеризующая рассеяние между группами:

$$C_{\text{факт}} = n \sum_{j=1}^k (\bar{x}_j - \bar{x})^2;$$

3) остаточная сумма квадратов отклонений наблюдаемых значений от своей групповой средней, характеризующая рассеяние внутри групп:

$$C_{\text{ост}} = \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2 + \dots + \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2;$$

4) общая, факторная и остаточная дисперсии:

$$S_{\text{общ}}^2 = C_{\text{общ}} / k(n-1); \quad S_{\text{факт}}^2 = C_{\text{факт}} / (k-1); \quad S_{\text{ост}}^2 = C_{\text{ост}} / (k(n-1));$$

5) значение критерия Фишера:

$$F = S_{\text{факт}}^2 / S_{\text{ост}}^2.$$

Значение критерия Фишера сравнивается с критическим для заданного уровня значимости  $\alpha$  и числа степеней свободы  $k-1$  и  $k(n-1)$ .

При неравномерном однофакторном дисперсионном анализе, когда количество наблюдений на уровне  $A_1$  равно  $n_1$ , на уровне  $A_2-n_2$ , ..., на уровне  $A_k-n_k$ , а общее их число равно  $N=\sum_{j=1}^k n_j$ , факторная и остаточная дисперсии находятся по формулам

Таблица 22. Проверка гипотезы о влиянии предрудного метасоматического изменения гранитов на содержание свинца в руде

№ замеса $i$	Уровни факторов							
	$A_1$			$A_2$			$A_3$	
	$x_{i1}$	$(x_{i1} - \bar{x})$	$(x_{i1} - \bar{x})^2$	$x_{i2}$	$(x_{i2} - \bar{x})$	$(x_{i2} - \bar{x})^2$	$x_{i3}$	$(x_{i3} - \bar{x})$
1	1,17	0,56	0,12	2,28	0,13	0,08	1,80	0,01
2	1,52	0,16	0,00	2,46	0,29	0,22	2,38	0,21
3	1,90	0,00	0,15	0,88	1,08	1,23	2,62	0,49
4	1,76	0,03	0,06	2,03	0,01	0,00	2,91	0,98
5	1,54	0,14	0,00	1,22	0,49	0,59	1,60	0,10
6	0,63	1,66	0,77	2,29	0,14	0,09	2,83	0,83
7	2,30	0,14	0,62	1,80	0,01	0,04	2,13	0,04
8	1,32	0,36	0,04	1,79	0,02	0,04	2,06	0,02
9	0,94	0,96	0,32	1,61	0,10	0,14	2,23	0,10
10	1,15	0,59	0,13	2,30	0,14	0,10	3,06	1,30
11	0,75	1,37	0,58	2,60	0,46	0,37	1,86	0,00
12	2,49	0,32	0,96	1,76	0,03	0,05	1,92	0,00
13	2,14	0,05	0,40	2,14	0,05	0,02	2,16	0,06
14	1,62	0,09	0,01	2,73	0,66	0,55	2,27	0,12
15	1,40	0,27	0,01					
	22,63	6,70	4,17	27,89	3,61	3,52	31,83	4,26
								2,50

$$\bar{x}_1 = 1,51$$

$$\bar{x}_2 = 1,99$$

$$\bar{x}_3 = 2,27$$

$$\bar{x} = \frac{82,35}{43} = 1,92;$$

$$S_{\text{общ}}^2 = \frac{6,70 + 3,61 + 4,26}{42} = \frac{14,57}{42} = 0,347;$$

$$S_{\text{факт}}^2 = \frac{15(1,51 - 1,92)^2 + 14(1,99 - 1,92)^2 + 14(2,27 - 1,92)^2}{2} = \\ = \frac{2,52 + 0,07 + 1,72}{2} = \frac{4,31}{2} = 2,155.$$

$$S_{\text{окт}}^2 = \frac{4,17 + 3,52 + 2,50}{40} = \frac{10,19}{40} = 0,255;$$

$$\frac{S_{\text{факт}}^2}{S_{\text{окт}}^2} = \frac{2,155}{0,255} = 8,45.$$

При  $f_1 = k - 1 = 2$ ,  $f_2 = N - k = 40$  и  $\alpha = 0,05$   $F_{\text{кр}} = 3,24$ , а при  $\alpha = 0,01$   $F_{\text{кр}} = 5,18$ .

$$S_{\text{факт}}^2 = \frac{1}{k-1} \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2;$$

$$S_{\text{окт}}^2 = \frac{1}{N-k} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2.$$

Остальные операции выполняются так же, как при равномерном анализе.

*Пример.* При изучении гидротермального свинцово-цинкового месторождения в гранитах высказано предположение, что на интенсивность процесса

рудоотложения влияла степень предрудного метасоматического изменения пород. Для проверки этой гипотезы результаты опробования на свинец по 43 разведочным пересечениям были разделены на три группы: в слабо измененных (уровень  $A_1$ ), в средне измененных (уровень  $A_2$ ) и сильно измененных (уровень  $A_3$ ) гранитах. Расчеты по схеме однофакторного дисперсионного анализа приведены в табл. 22.

Таким образом с достаточно высокой доверительной вероятностью гипотеза об отсутствии влияния степени метасоматического изменения гранитов на содержание свинца в руде отвергается, и предрудный метасоматоз должен рассматриваться как один из ведущих рудоконтролирующих факторов.

Вычислительные операции при однофакторном дисперсионном анализе можно упростить, используя равенство  $C_{\text{ост}} = C_{\text{общ}} - C_{\text{факт}}$ .

При двухфакторном дисперсионном анализе сумма квадратов отклонений от общего среднего разделяется на компоненты, отвечающие двум предполагаемым факторам изменчивости —  $A$  и  $B$ . Если по фактору  $A$  выделяется  $p$  уровней, а по фактору  $B$  —  $q$  уровней, то общее количество групп будет равно  $m = pq$ , а исходные данные можно записать в виде табл. 23.

Если для каждого сочетания факторов  $A_iB_j$  произведено по  $n$  наблюдений (двуухфакторный дисперсионный анализ с повторением), то в каждую клетку табл. 23 помещается  $n$  значений, а единичное наблюдение обозначается как  $x_{ijk}$ , где  $k=1, 2, \dots, n$ . Оценки средних значений по группам ( $\bar{x}_{ij}$ ), по факторам ( $\bar{x}_{i..}$  и  $\bar{x}_{.j}$ ) и общее среднее ( $\bar{x}$ ) в этом случае рассчитываются по формулам

Таблица 23. Двуухфакторный дисперсионный анализ

A	Уровни фактора B						Среднее
	$B_1$	$B_2$	$\dots$	$B_j$	$\dots$	$B_q$	
$A_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1j}$	$\dots$	$x_{1q}$	$\bar{x}_1$
$A_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2j}$	$\dots$	$x_{2q}$	$\bar{x}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_i$	$x_{i1}$	$x_{i2}$	$\dots$	$x_{ij}$	$\dots$	$x_{iq}$	$\bar{x}_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_p$	$x_{p1}$	$x_{p2}$	$\dots$	$x_{pj}$	$\dots$	$x_{pq}$	$\bar{x}_p$
Среднее	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\dots$	$\bar{x}_{.j}$	$\dots$	$\bar{x}_{.q}$	$\bar{x}$

$$\bar{x}_{ij.} = \frac{1}{n} \sum_{k=1}^n x_{ijk};$$

$$\bar{x}_{i..} = \frac{1}{qn} \sum_{j=1}^q \sum_{k=1}^n x_{ijk} = \frac{1}{q} \sum_{j=1}^q \bar{x}_{ij};$$

$$\bar{x}_{ij} = \frac{1}{pn} \sum_{i=1}^p \sum_{k=1}^n x_{ijk} = \frac{1}{p} \sum_{i=1}^p \bar{x}_{ij}.$$

Общая схема вычисления дисперсий при двухфакторном анализе приведена в табл. 24.

Проверка гипотезы о влиянии на изменчивость изучаемого свойства каждого фактора в отдельности и их совместного влияния производится по критерию Фишера:

$$F_A = S_1^2 / S_4^2; \quad F_B = S_2^2 / S_4^2; \quad F_{AB} = S_3^2 / S_4^2.$$

Полученные значения  $F$ -критерия сравниваются с критическим для заданного уровня значимости и числа степеней свободы, приведенного в табл. 24.

Таблица 24. Схема вычисления дисперсий при двухфакторном дисперсионном анализе

Вид дисперсии	Сумма квадратов отклонений	Число степеней свободы	Дисперсия
Факторная по фактору $A$	$C_1 = nq \sum_{i=1}^p (\bar{x}_{i..} - \bar{x})^2$	$p-1$	$S_1^2 = \frac{C_1}{p-1}$
Факторная по фактору $B$	$C_2 = np \sum_{j=1}^q (\bar{x}_{..j} - \bar{x})^2$	$q-1$	$S_2^2 = \frac{C_2}{q-1}$
Смешанная по факторам $AB$	$C_3 = n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij..} - \bar{x}_{i..} - \bar{x}_{..j} + \bar{x})^2$	$(p-1)(q-1)$	$S_3^2 = \frac{C_3}{(p-1)(q-1)}$
Остаточная	$C_4 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij..})^2$	$pq(n-1)$	$S_4^2 = \frac{C_4}{pq(n-1)}$
Общая	$C = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x})^2$	$npq-1$	$S^2 = \frac{C}{npq-1}$

При расчете  $F$ -критерия в данном случае в знаменателе всегда берется остаточная дисперсия. Поэтому его значение иногда может получиться меньше 1.

Приведенные схемы дисперсионного анализа основаны на свойствах нормального закона распределения и предположении о равенстве дисперсий на разных уровнях одного и того же фактора. Однако  $F$ -критерий по выборкам достаточно большого объема устойчив и для совокупностей умеренно отклоняющихся от нормальных. Умеренное различие в дисперсиях так же не является препятствием для его применения при условии приблизительного равенства объемов выборок по группам. Если возможность применения  $F$ -критерия все же вызывает сомнения можно воспользоваться непараметрическими критериями.

Однофакторный непараметрический дисперсионный анализ с применением критерия Краскала — Уоллиса включает следующие операции:

— ранжирование всех наблюдений по возрастанию от 1 до  $N$ , где  $N = \sum_{j=1}^k n_j$  — объем всей выборки;

— нахождение сумм рангов  $R_1, R_2, \dots, R_k$  для каждого уровня анализируемого фактора;

— вычисление критерия Краскала — Уоллиса по формуле

$$H = \frac{12}{N(N+1)} \left( \frac{R_1^2}{n_1} + \dots + \frac{R_k^2}{n_k} \right) - 3(N+1);$$

— сравнение полученного значения  $H$  с его критическим значением ( $H_c$ ) для принятого уровня значимости  $\alpha$ .

При достаточно большом объеме выборки, когда количество наблюдений по каждому уровню превышает 5, значение  $H_c$  определяется по таблицам распределения  $\chi^2$  для числа степеней свободы  $f=k-1$ , где  $k$  — количество уровней исследуемого фактора.

Для малых выборок критические значения критерия Краскала — Уоллиса определяются по специальным таблицам (прил. 9).

Если рассчитанное значение критерия  $H$  превышает  $H_c$ , то гипотеза об отсутствии влияния анализируемого фактора на изменчивость изучаемого свойства отвергается.

При двухфакторном непараметрическом анализе Фридмана исходные данные записываются в виде табл. 24.

Для проверки влияния фактора  $A$  значения в строках таблицы ранжируются, т. е. заменяются цифрами от 1 до  $q$ . По каждому столбцу вычисляется сумма рангов —  $R_j (j=1, \dots, p)$ , рассчитывается статистика  $S = (R_1^2 + \dots + R_q^2) - (R_1 + \dots + R_q)^2/q$  и критериальная статистика  $\chi_A^2 = 12S/pq(q+1)$ .

Для количества уровней по факторам  $A$  или  $B$  больше 4, критическое значение статистики  $\chi_A^2$  определяется по таблицам критерия  $\chi^2$  для количества степеней свободы  $f=q-1$ , а при малом количестве уровней используются специальные таблицы (прил. 10).

Для проверки гипотезы о влиянии фактора  $B$  строки и столбцы в табл. 24 меняются местами.

Использованный в данной главе математический аппарат описан в специальных пособиях по теории вероятностей и математической статистике. Приемы первичного анализа результатов массовых наблюдений подробно рассмотрены в монографии Дж. Тьюки. Много полезных рекомендаций по вычислительным методам статистики можно почерпнуть из справочника Дж. Полларда.

Применению статистических методов при геохимических исследованиях посвящены работы В. Н. Бондаренко, Д. А. Родионова. Многочисленные примеры использования одномерных статистических моделей в различных областях геологии приведены в монографиях Р. Миллера и Дж. Кана, У. Крамбейна и Ф. Грейбилла, Дж. Дэвиса.

Статистическому анализу угловых величин посвящена монография  
К. Мардса.

### Контрольные вопросы

1. Почему при изучении свойств геологических объектов можно применять методы математической статистики?
2. Какие требования предъявляются к выборочной геологической совокупности при статистическом моделировании?
3. Что такое «стебель с листьями» и «ящик с усами»?
4. Какие свойства геологических объектов можно описать непрерывными случайными величинами, а какие — дискретными?
5. В чем специфика случайных угловых величин?
6. Какие геологические задачи можно решать с помощью гистограмм и кумулят?
7. Что такое круговое среднее направление, круговая мода и круговая медиана?
8. Какие статистические законы распределения используются в геологии?
9. Как проверяются гипотезы о типе статистического распределения?
10. Для чего рассчитываются точечные и интервальные оценки свойств геологических объектов?
  11. Какими свойствами должны обладать точечные оценки?
  12. Что такое «максимально правдоподобная оценка»?
  13. Как строятся доверительные интервалы оценок средних значений в условиях нормального, логнормального, биномиального закона и распределений Пуассона и Мизеса?
  14. Какие типы геологических гипотез можно проверять статистическими методами?
  15. Чем различаются параметрические критерии согласия от непараметрических?
  16. Чем различаются ошибки первого и второго рода при статистической проверке гипотез?
  17. Какие геологические задачи решаются путем проверки гипотез о равенстве средних?
  18. Для чего используются критерии Стьюдента, Ван-дер-Вардена, Родионова, Вилкоксона, Ватеона — Вильямса, Вилера — Ватсона — Ходжеса?
  19. Какие геологические задачи решаются путем проверки гипотез о равенстве дисперсий?
  20. Для чего используются критерии Фишера, Сиджела — Тьюки?
  21. Какие геологические задачи решаются с помощью критериев Смирнова и Фергюсона?
  22. Какие геологические задачи решаются с помощью однофакторного и двухфакторного дисперсионного анализа?

## ГЛАВА 3

### ДВУМЕРНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

---

#### 3.1. СУЩНОСТЬ И УСЛОВИЯ ПРИМЕНЕНИЯ ДВУМЕРНЫХ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Моделирование геологических образований и процессов как сложных природных систем часто вызывает необходимость совместного рассмотрения нескольких их свойств с целью выяснения общей структуры изучаемого объекта. Так, например, при изучении полезных ископаемых по керну скважин одновременно определяют мощность залежи, содержание в ней полезных компонентов,

породообразующих элементов, значения эффективной пористости или различные другие свойства вмещающих пород и руд.

В одних случаях изучаемые свойства геологических объектов проявляются независимо друг от друга, а в других между ними могут быть выявлены более или менее отчетливые взаимосвязи.

Например, в редкometалльных пегматитах тантал и ниобий входят в состав только одного минерала — tantalит-колумбита ( $\text{Fe}, \text{Mn}$ )  $(\text{Nb}, \text{Ta})_2\text{O}_6$ . Между содержаниями этих элементов в рудах всегда наблюдается прямая зависимость: чем больше тантала, тем больше ниobia, а между их содержаниями в мономинеральных фракциях — обратная. Это объясняется тем, что в рудах содержания обоих элементов прямо пропорциональны концентрациям рудного минерала, а в минерале тантал и ниобий изоморфно замещают друг друга в кристаллической решетке.

В других случаях для объяснения природы наблюдаемых зависимостей необходимо проследить длинную цепочку взаимозависимых процессов и явлений. Так, в результате обработки многолетних статистических данных о случаях тяжелого производственного травматизма на угольных шахтах было установлено, что их частота определенным образом связана с фазами луны. Эта на первый взгляд весьма странная связь объясняется влиянием положения луны на приливные силы, которые проявляются не только в гидросфере, но и в литосфере, и часто играют роль «спускового крючка» для таких явлений как горный удар, выбросы газа и т. п.

Часто связь между различными свойствами геологических объектов вообще не поддается объяснению с генетических или причинно-следственных позиций, так как наблюдаемые зависимости могут быть не связаны с геологическими процессами, а обусловлены методикой изучения геологических объектов или другими причинами. Так, например, существует зависимость между частотой обнаружения коренных рудных выходов и типом современных ландшафтов опиcкованных площадей. Большинство коренных рудных выходов концентрируется в ландшафтных зонах эрозионного типа, а в аккумулятивных ландшафтах они обнаруживаются гораздо реже. Очевидно, что никаких причинно-следственных связей между процессами рудообразования и более поздними процессами формирования современных ландшафтов нет, а выявленная закономерность обусловлена тем, что в ландшафтных зонах аккумулятивного типа резко снижается эффективность поисков.

Изучение взаимозависимостей между значениями свойств геологических образований способствует более глубокому пониманию особенностей геологических процессов и выявлению факторов, влияющих на эффективность методов исследования геологических объектов. В ряде случаев оно позволяет получить количественные оценки некоторых свойств по значениям других, легко определяемых свойств. Так как изучаемые взаимозависимости имеют статистический характер и практически всегда отличаются от функциональных, для их изучения и описания используются двумерные и многомерные статистические модели.

В двумерной статистической модели объект исследования рассматривается как двумерная статистическая совокупность, а ее основной

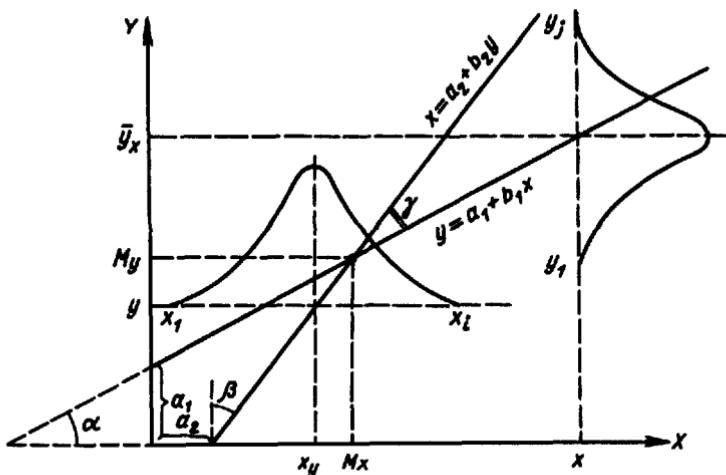


Рис. 26. Параметры двумерной случайной величины  $XY$ .  $M_x$  и  $M_y$ —математические ожидания величин  $X$  и  $Y$ ;  $\bar{Y}$ —центр условного распределения  $Y$  для  $X=x$ ;  $\bar{X}$ —центр условного распределения  $X$  для  $Y=y$ ;  $b_1=\operatorname{tg} \alpha$ ;  $b_2=\operatorname{tg} \beta$

характеристикой является двумерная функция распределения случайных величин  $X$  и  $Y$ . Между двумя случайными величинами проявляются стохастические (вероятностные) связи, когда заданному значению случайной величины  $X=x$  соответствует не какое-либо значение величины  $Y$ , а набор ее значений  $y_1, y_2, \dots, y_n$ , каждому из них свойственна определенная вероятность  $p_1, p_2, \dots, p_n$ . Функция распределения величины  $Y$ , соответствующая значению  $X=x$ , характеризуется математическим ожиданием  $\bar{y}_x$  и дисперсией  $\sigma_y^2$ .

Распределения величины  $\bar{Y}$ , соответствующие выбранным значениям величины  $X$ , называются *условными распределениями*, а дисперсии  $\sigma_y^2$ —*условными дисперсиями*. Геометрическое место точек, соответствующих центрам условных распределений  $\bar{y}$ , называется *линией регрессии*, а уравнение этой линии—*уравнением регрессии*. Аналогично каждому значению случайной величины  $Y=y$  соответствует некоторая функция распределения величины  $X$  с математическим ожиданием  $\bar{x}_y$  и дисперсией  $\sigma_x^2$ .

Системе из двух случайных величин всегда будет соответствовать две линии регрессии:  $y_x=f(x)$ —регрессия  $Y$  по  $X$  и  $x_y=f(y)$ —регрессия  $X$  по  $Y$  (рис. 26). Если линии регрессии прямые, то регрессия двух величин называется *линейной*. В более сложных случаях линии регрессии соответствуют кривым линиям, а регрессия случайных величин называется *нелинейной*.

В прямоугольной системе координат линии регрессии могут быть заданы аналитически. Для линейной регрессии будем иметь следующую пару уравнений:

$$y=a_1+b_1x \text{ (регрессия } Y \text{ на } X); \\ x=a_2+b_2y \text{ (регрессия } X \text{ на } Y).$$

Уравнения нелинейной регрессии зависят от вида кривой. Например, для параболической регрессии

$$y=a_1+b_1x+c_1x^2; \quad x=a_2+b_2y+c_2y^2.$$

Регрессия может быть однозначно описана, если известны вид уравнения и значения коэффициентов  $a$ ,  $b$ ,  $c$  и т. д. В системе двух уравнений линейной регрессии коэффициенты  $a_1$  и  $a_2$  характеризуют положения начальных точек линий регрессии. При  $a_1$  и  $a_2$  равных нулю, линии проходят через начало координат. Степень зависимости случайных величин определяется коэффициентами  $b_1$  и  $b_2$ , которые называются коэффициентами линейной регрессии. Они представляют собой тангенсы наклона прямой регрессии  $y=a_1+b_1x$  к оси абсцисс (угол  $\alpha$ ) и прямой регрессии  $x=a_2+b_2y$  к оси ординат (угол  $\beta$ ), если величины  $X$  и  $Y$  центрированы и нормированы по стандарту, т. е. соответствуют  $(x-\bar{x})/\sigma_x$  и  $(y-\bar{y})/\sigma_y$ .

В общем случае прямые регрессии пересекаются в точке, координаты которой равны математическим ожиданиям величин  $X$  и  $Y$ , а угол  $\gamma$  между ними изменяется от  $0$  до  $90^\circ$ . Чем меньше угол  $\gamma$ , тем сильнее связь между величинами. В частном случае одна из линий регрессии может быть параллельна одной из осей координат, а соответствующий ей коэффициент регрессии равен нулю, в то время как другая останется не параллельной другой оси, а соответствующий ей коэффициент будет отличаться от нуля. Следовательно, степень связи между величинами зависит от того, какая величина взята в качестве аргумента, а для полной характеристики связи всегда необходимо знать оба коэффициента регрессии. Если угол  $\gamma=0$  и обе линии регрессии сливаются в одну прямую, то  $b_{y/x}=\frac{1}{b_{x/y}}$ , а связь между величинами становится функциональной. Если же линии регрессии параллельны осям координат и взаимно перпендикулярны, то зависимости между случайными величинами вызываются тем, что среди действующих на них факторов имеются как общие, так и факторы, влияющие только на величину  $X$  и только на величину  $Y$ . Статистическими методами можно установить наличие стохастических связей и оценить их силу, но нельзя объяснить появление этих связей причинно-следственными отношениями.

Задача совместного исследования двух (и более) признаков сводится к выявлению их стохастической сопряженности. При наличии такой сопряженности можно обосновать прогноз тех пределов, в которых с наперед заданной надежностью содержится искомая случайная величина, если сопряженная с ней величина принимает определенное значение. Для этого необходимо знать функции распределения величины  $Y$  или хотя бы некоторые их числовые характеристики — генеральные математические ожидания, дисперсии величин  $X$  и  $Y$ , а также центры и дисперсии условных распределений этих величин.

Основными числовыми характеристиками двумерного распределения случайных величин являются показатели их связи: ковариация, или корреляционный момент (момент связи), коэффициент корреляции и корреляционное отношение.

Ковариация, или корреляционный момент, представляет собой математическое ожидание произведения отклонений двух случайных величин от их математических ожиданий

$$\text{cov}(x, y) = M[(x - Mx)(y - My)] = M[(\bar{x}_y - M\bar{x})(\bar{y}_x - M_y)].$$

Коэффициент корреляции представляет собой ковариацию, нормированную по стандартам  $\rho = \text{cov}(x, y)/\sigma_x \sigma_y$ . Пределами изменения коэффициента корреляции являются  $\rho = -1$  и  $\rho = +1$ , причем значение  $+1$  соответствует функциональной связи величин, а  $\rho = 0$  полному отсутствию линейной связи. Знак коэффициента (+) или (-) указывает на характер связи (прямая или обратная).

Если оба уравнения регрессии линейные, вида  $y = a_1 + b_1 x$  и  $x = a_2 + b_2 y$ , то коэффициент корреляции  $\rho = \sqrt{b_1 b_2}$ .

Корреляционным отношением называется отношение дисперсий (стандартов) центров условных распределений к общей дисперсии (стандарту) величины. Таких отношений в двумерном распределении, так же как и уравнений регрессии, очевидно, может быть два:

$$\eta_{y/x} = \sigma_{\bar{y}_x}/\sigma_y; \quad \eta_{x/y} = \sigma_{\bar{x}_y}/\sigma_x.$$

В случае линейности обоих уравнений регрессии значения  $\eta_{y/x}$  и  $\eta_{x/y}$  совпадают, т. е.  $\eta_{y/x} = \eta_{x/y}$ .

Величины корреляционных отношений изменяются в пределах от 0 до 1. Значение  $\eta = 0$  свидетельствует о независимости величин, образующих двумерное распределение.

Двумерная статистическая модель использует математический аппарат теории вероятностей и статистики случайных величин, поэтому изложенные в разд. 2.1 требования к массовости, однородности и случайности выборочной совокупности для данной модели также являются обязательными. Нарушение этих условий снижает эффективность корреляционного и регрессионного анализа и может привести к ошибочным выводам. В то же время условие независимости для большинства задач, решаемых с помощью этой модели, не является обязательным.

Следует помнить, что числовые показатели силы корреляционной связи, характер этой связи и вид уравнений регрессии могут существенно меняться в зависимости от методики изучения свойств геологических объектов и в первую очередь от геометрии проб или от способа измерения.

### 3.2. ПРОСТЕЙШИЕ ПРЕОБРАЗОВАНИЯ ДВУМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Исследования двумерных случайных величин, так же как и одномерных, целесообразно начинать с предварительного анализа их свойств с помощью простейших графических преобразований.

Двумерную случайную величину  $X, Y$  можно наглядно изобразить в виде корреляционного поля точек, когда каждая пара значений изображается в виде точки с координатами  $x_i$  и  $y_i$  (рис. 27). Если предположить, что одно из свойств ( $Y$ ) зависит от другого ( $X$ ), то по горизонтальной оси следует откладывать  $X$  (аргумент), а по вертикальной —  $Y$  (функция). В случае отсутствия каких-либо

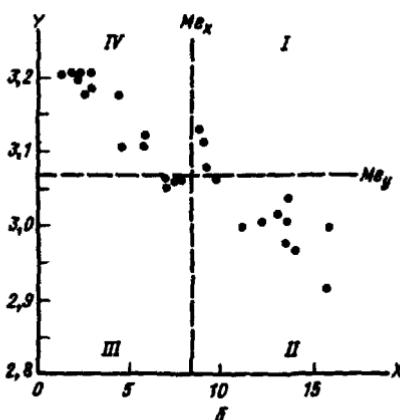
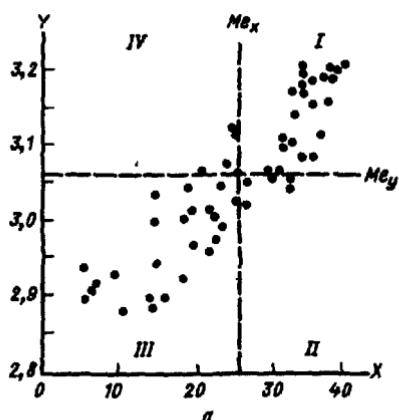


Рис. 27. Корреляционные поля точек соотношений объемной массы  $Y$  (в  $\text{t}/\text{м}^3$ ) и содержаний  $X$  (в %)  $\text{P}_2\text{O}_5$  (а)  $\text{Al}_2\text{O}_3$  (б) для апатит-нефелиновых руд месторождения Коашва

априорных сведений о причинно-следственных связях между изучаемыми свойствами оси выбираются произвольно. Масштаб и начало отсчета по каждой оси выбираются по размаху варьирования каждого признака таким образом, чтобы поле графика было квадратным или прямоугольным с отношением сторон  $X:Y$  не более чем 2:1. Если значения признака изменяются в интервале, удаленном от 0, как в случае объемной массы руды на рис. 27, начало координат по соответствующей оси целесообразно сместить на величину немногого меньшую минимального значения.

С помощью корреляционного поля точек можно получить весьма ценные сведения о свойствах изучаемой двумерной случайной величины.

Форма и ориентировка корреляционного поля точек позволяют судить о наличии корреляционной связи, о ее характере (прямая или обратная) и виде (линейная или нелинейная). Если связь между изучаемыми свойствами существует, то корреляционное поле точек имеет форму вытянутого эллипса, длинная ось которого наклонена относительно осей координат (см. рис. 27). По направлению наклона определяется характер связи: положительная (прямая) — см. рис. 27, а, или отрицательная (обратная) — см. рис. 27, б. Когда связь отсутствует, корреляционное поле точек имеет изометрическую форму (рис. 28, а) или представляет собой эллипс, длинная ось которого параллельна одной из осей координат. Наличие перегибов оси корреляционного поля точек указывает на нелинейный вид связи.

Корреляционное поле точек позволяет также проверить однородность выборочной совокупности. Если на графике имеются точки, значительно удаленные от основного поля, это указывает на возможное засорение выборки аномальными значениями, не характерными для данного объекта. Причины появления таких значений необходимо выяснить. Нередко они возникают в результате ошибок при замерах или анализах, а также за счет арифметических ошибок и опечаток

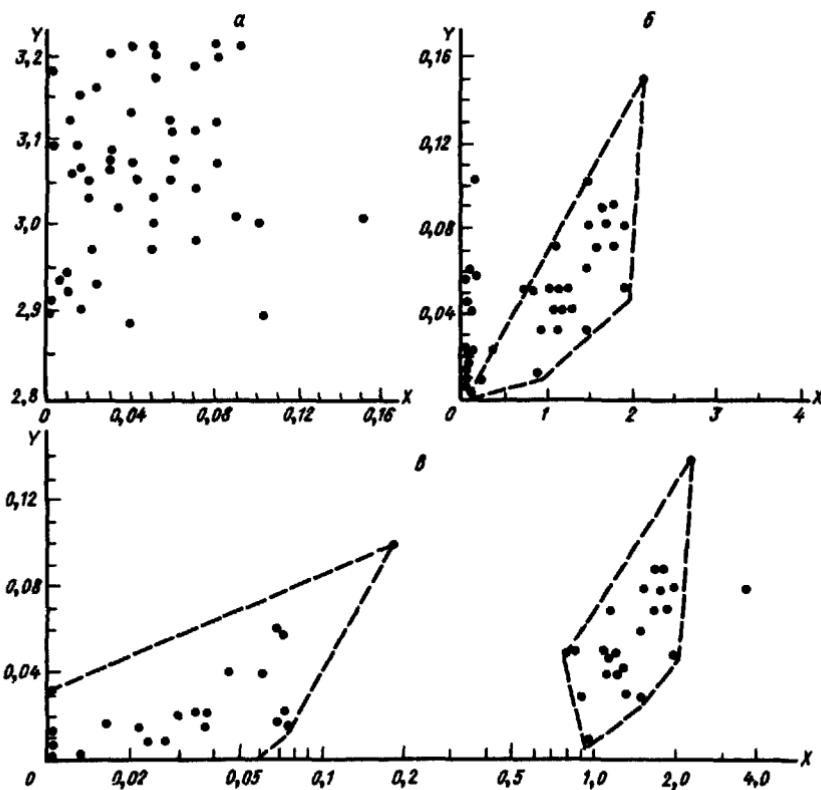


Рис. 28. Корреляционные поля точек для физических свойств апатит-нефелиновых руд месторождения Коашва:

*a*—влажность  $X$  (в %)—объемная масса  $Y$  (в  $\text{т}/\text{м}^3$ ); *b*—пористость  $X$  (в %)—влажность  $Y$  (в %); *c*—пористость (в %) ( $X$  в логарифмическом масштабе)—влажность  $Y$  (в %)

в процессе предварительной обработки результатов наблюдений. Аномальные значения резко снижают точность оценок показателей силы связи и коэффициентов уравнений регрессии, поэтому их целесообразно исключать из выборки.

В геологической практике приходится сталкиваться также со случаем, когда корреляционное поле точек разделяется на два, реже три самостоятельных поля. Это является признаком неоднородности исследуемой совокупности. Так, на графике зависимости пористость—влажность (см. рис. 28, б), кроме аномальной пробы с резко выдающимся значением пористости (3,66%), наблюдаются пробы с очень низкой пористостью, но довольно высокой влажностью, которые также не попадают в основное поле точек. Для более детального рассмотрения этой группы точек вместо обычного линейного масштаба по оси  $X$  целесообразно принять логарифмический масштаб (см. рис. 28, в). Это позволяет убедиться в том, что пробы данной выборки явно неоднородны по значениям пористости. На

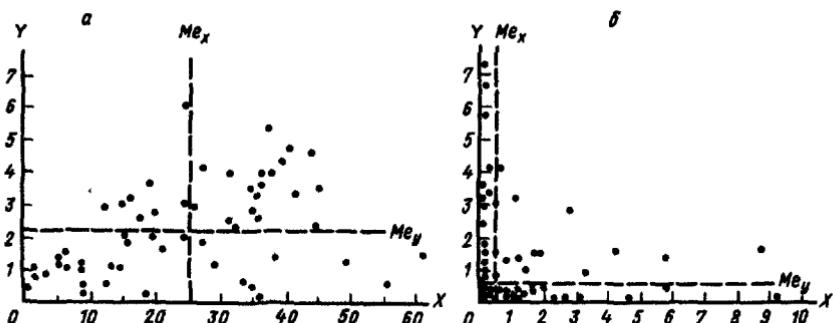


Рис. 29. Корреляционные поля точек соотношений процентных содержаний минералов в рудах месторождения Коашва:

*a*—нефелин ( $X$ )—титанит ( $Y$ ); *б*—полевой шпат ( $X$ )—титано-магнетит ( $Y$ )

графике отчетливо выделяются две группы проб, причем для обеих групп характерна прямая корреляционная зависимость между пористостью и влажностью.

Корреляционное поле точек позволяет судить и о виде статистического распределения случайной двумерной величины. Для этого необходимо на графике провести линии, соответствующие медианам значений  $X$  и  $Y$ . Положение медиан легко определить перемещая по графику прозрачную линейку параллельно осям координат до такого положения, когда число точек над и под линейкой окажется одинаковым. Если медианы пересекаются в центре графика, и место

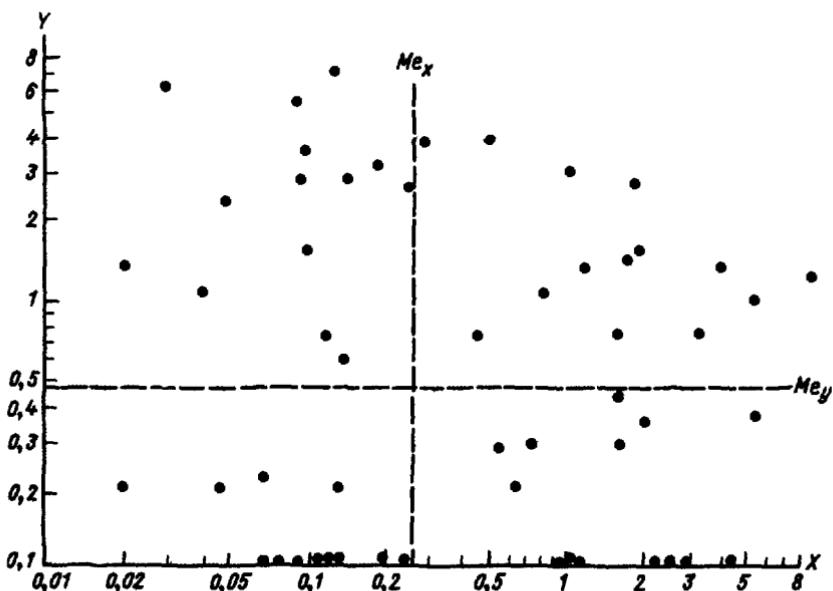


Рис. 30. Корреляционное поле точек процентных содержаний полевого шпата ( $X$ ) и титано-магнетита ( $Y$ ) в рудах месторождения Коашва в логарифмическом масштабе

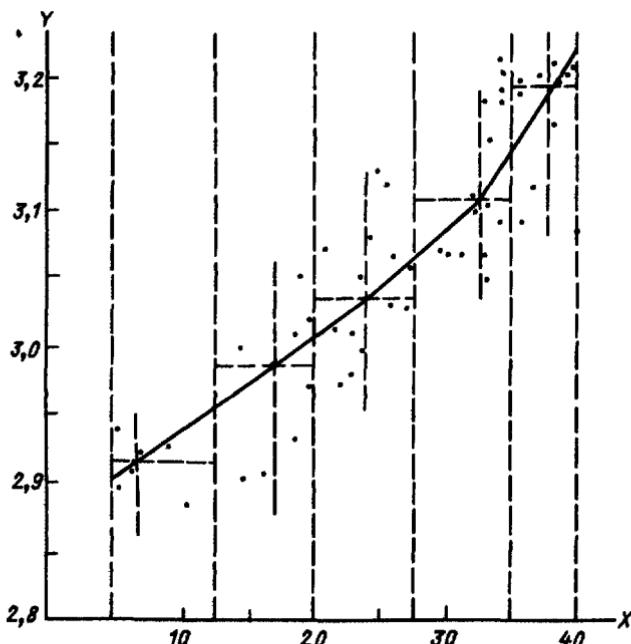


Рис. 31. Корреляционное поле точек и медианы условных распределений соотношения содержания  $P_2O_5$  ( $X$ , %)—объемная масса ( $Y$ ,  $t/m^3$ ) в рудах месторождения Коашва

их пересечения совпадает с максимальным количеством точек на единицу площади, то распределение изучаемой двумерной случайной величины является симметричным и, возможно, близко к двумерному нормальному (см. рис. 27). Если большинство точек и место пересечения медиан смещены от центра графика к началу координат, то двумерная случайная величина распределена асимметрично и, возможно, мы имеем дело с двумерным логнормальным распределением (рис. 29, б). Смещение к началу координат одной из медиан указывает на то, что асимметричное распределение характерно для одного из рассматриваемых свойств. В двух последних случаях целесообразно построить графики корреляционных полей точек, используя логарифмический (рис. 30) и полулогарифмический масштабы.

Для обобщения информации о свойствах распределения двумерной случайной величины корреляционное поле точек можно разделить на интервалы по одной из осей (обычно по оси  $X$ ) и построить по точкам, попадающим в каждый интервал, медианы условных распределений или схематические диаграммы «ящик с усами» (см. разд. 2.2). Медианы условных распределений легко найти изложенным выше способом с помощью перемещения линейки (рис. 31). По характеру изменения положения центров условных распределений можно более надежно сделать вывод о наличии корреляционной связи и судить о ее линейности. По схематическим диаграммам (рис. 32) можно также проследить характер изменения размаха

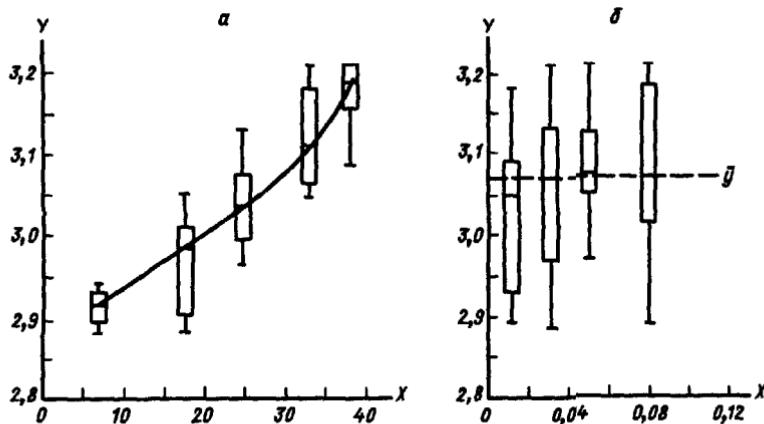


Рис. 32. Схематические диаграммы условных распределений для свойств руды месторождения Коашва:  
а—содержание  $P_2O_5$  ( $X, \%$ )—объемная масса ( $Y, t/m^3$ ); б—влажность ( $X, \%$ )—объемная масса ( $Y, t/m^3$ )

варьирования значений  $Y$  от центров условных распределений при разных значениях  $X$ . Так, например, на рис. 32 отчетливо видно, что центры условных распределений объемной массы ( $Y$ ) при увеличении содержания  $P_2O_5$  ( $X$ ) закономерно смещаются вверх по координате  $Y$ . При этом амплитуда смещения медиан для начальных и конечных интервалов явно превышает размах варьирования величины  $Y$  в пределах интервалов. Это указывает на явное наличие корреляционной связи. Можно отметить, что размах варьирования значений объемной массы при низких и высоких содержаниях  $P_2O_5$  меньше, чем в пробах со средним содержанием. Соединив центры условных распределений можно заметить, что зависимость объемной массы от содержания  $P_2O_5$  (см. рис. 32, а) не очень сильно, но все-таки отличается от линейной. В то же время по схематическим диаграммам условных распределений для соотношения влажность—объемная масса (см. рис. 32, б) прослеживается, что связи между этими свойствами руды нет, так как центры условных распределений для трех класс-интервалов находятся практически на одном уровне, соответствующем среднему значению объемной массы, а медиана для первого класс-интервала отклоняется от этой линии на величину, примерно на порядок меньшую по сравнению с размахом варьирования значений объемной массы в этом интервале значений влажности.

### 3.3. ПРОВЕРКА ГИПОТЕЗ О НАЛИЧИИ КОРРЕЛЯЦИОННОЙ СВЯЗИ

Выявление корреляционных связей между различными свойствами геологических объектов способствует решению весьма широкого круга задач. Наиболее часто корреляционный анализ используется при

изучении геологических процессов, разработке поисковых критериев и факторов рудоконтроля, а также при выборе рациональных комплексов методов исследований при геологическом картировании, поисках и разведке месторождений.

Изучение геологических процессов может идти по-разному. В одних случаях первоначально создается гипотетическая модель процесса, на основе которой высказываются предположения о характере зависимостей между отдельными свойствами геологических образований, сформированных этим процессом. Затем выборочным методом производится определение этих свойств и статистическая обработка полученных наблюдений. Если гипотеза о наличии и характере корреляционной связи подтверждается, это служит косвенным подтверждением правомерности исходной геологической модели. В других случаях статистический анализ результатов наблюдений предшествует теоретическим заключениям, а выявленные корреляционные связи учитываются при разработке детерминированных моделей, описывающих зависимости между геологическими явлениями и изучаемыми физическими, химическими, биологическими и другими факторами.

К задачам такого типа относятся:

— выявление парагенетических ассоциаций минералов и химических элементов в горных породах и рудах. Так, например, наличие корреляционных связей между петрогенными и редкими элементами способствует оценке роли процессов дифференциации магмы и асимиляции ею вмещающих пород;

— изучение поведения химических элементов в процессе гипергенного и метасоматического изменения пород. Нарушение корреляционных связей, свойственных неизмененным породам и появление новых связей позволяет судить о подвижности химических элементов, устойчивости тех или иных минералов и общем характере геохимических процессов;

— палеогеографический анализ условий формирования осадочных пород. Так, например, выявление зависимостей между гранулярным и минеральным составом и степенью окатанности обломочного материала позволяет определять области сноса и границы бассейна осадконакопления при формировании определенных стратиграфических горизонтов;

— выяснение источников рудного вещества путем изучения корреляционной связи между концентрацией рудных элементов во вмещающих породах и рудах.

Для целенаправленного ведения геологоразведочных работ очень важно знать факторы, влияющие на размещение в недрах полезных ископаемых,— химический и минеральный составы вмещающих пород, их физико-механические свойства (эффективная пористость, трещиноватость и т. п.), элементы залеганиярудовмещающих структур и т. д. Проверка гипотез о наличии корреляционной связи между этими свойствами и концентрацией полезного компонента позволяет оценить роль каждого из них и наметить участки, наиболее благоприятные для локализации оруденения. Выявление таких связей между составом пород и руд и их физическими свойствами—

магнитностью, электропроводностью, плотностью, естественной радиоактивностью и т. п., позволяет выбрать наиболее информативные геофизические методы для геологического картирования и поисков полезных ископаемых.

Иногда практический интерес представляют сведения не о наличии, а об отсутствии корреляционной связи. Установлено, что для неизменных горных пород различного состава характерна прямая корреляция между плотностью и магнитной восприимчивостью. В метасоматически измененных породах эта связь не наблюдается. На этом эффекте основан оригинальный способ картирования метасоматитов.

Проверка гипотезы о наличии корреляционной связи обычно основана на том, что для двумерной нормально распределенной случайной величины  $XY$  при отсутствии корреляции между  $X$  и  $Y$  коэффициент корреляции и корреляционное отношение равны нулю. Поэтому процедура проверки заключается в расчете выборочных оценок этих характеристик и оценке значимости их отличия от нуля.

Выборочная оценка коэффициента корреляции может быть рассчитана по формуле

$$r = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / nS_x S_y,$$

где  $\bar{x}$  и  $\bar{y}$  — выборочные оценки средних значений случайных величин  $X$  и  $Y$ ;  $S_x$  и  $S_y$  — выборочные оценки их стандартов;  $n$  — количество сравниваемых пар значений.

При расчетах вручную удобнее пользоваться формулой

$$r = \left[ \sum_{i=1}^n x_i y_i - \frac{1}{n} \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right) \right] / \sqrt{\left[ \sum_{i=1}^n x_i^2 - \frac{1}{n} \times \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ \sum_{i=1}^n y_i^2 - \frac{1}{n} \times \left( \sum_{i=1}^n y_i \right)^2 \right]}.$$

Когда математическое ожидание выборочного коэффициента корреляции равно нулю, величина  $t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$  имеет распределение Стьюдента с  $n-2$  степенями свободы. Если рассчитанное по этой формуле значение величины  $t$  превышает табличное значение критерия Стьюдента (см. прил. 13) для принятой доверительной вероятности и числа степеней свободы  $n-2$ , гипотеза об отсутствии корреляционной связи отвергается.

При выборе доверительной вероятности (уровня значимости) следует помнить, что в данном случае проверяется гипотеза об отсутствии корреляционной связи ( $H_0: r=0$ ), а гипотеза о ее наличии ( $H_1: r \neq 0$ ) выступает в качестве альтернативной. Поэтому увеличение доверительной вероятности (снижение уровня значимости) приводит к увеличению вероятности ошибки второго рода, которая будет заключаться в принятии гипотезы об отсутствии связи, в то время как она существует. Другими словами снижение уровня значимости приводит к пропуску слабо проявленных корреляционных связей, что при решении многих задач нежелательно.

Приближенная оценка коэффициента корреляции может быть получена графическим способом с помощью корреляционного поля точек. Поле точек разделяется на четыре квадранта линиями, соответствующими медианам величин  $X$  и  $Y$  (см. рис. 27). Для оценки коэффициента корреляции используется формула

$$r = (n_1 - n_2) / (n_1 + n_2),$$

где  $n_1$  — число точек в квадрантах I и III, а  $n_2$  — в квадрантах II и IV.

Если корреляционная связь отсутствует, то количество точек во всех квадрантах будет примерно одинаковым, а величина  $r$  — близка к нулю. В случае прямой корреляционной связи большинство точек попадет в квадранты III и I и  $r$  будет величиной положительной, а при обратной связи точки сконцентрируются преимущественно в квадрантах II и IV, а  $r$  будет величиной отрицательной. Так, например, приближенная оценка коэффициента корреляции для соотношения объемная масса — содержание  $P_2O_5$  (см. рис. 27, a):

$$r = \frac{(23 + 24) - (3 + 5)}{(23 + 24) + (3 + 5)} \approx 0,71;$$

а для соотношения объемная масса — содержание  $Al_2O_3$ :

$$r = \frac{(4 + 3) - (10 + 10)}{(4 + 3) + (10 + 10)} \approx -0,48.$$

Как все моментные характеристики второго и более высокого порядка, коэффициент корреляции чувствителен к виду функций эмпирических распределений величин, входящих в двумерную систему. Поэтому, если эти распределения заметно отличаются от нормальных, проверка гипотез о связи по коэффициенту корреляции требует предварительной нормализации эмпирических распределений с помощью логарифмического или каких-либо других преобразований.

Таким образом, при проверке гипотез о стохастической связи случайных величин по коэффициенту корреляции необходимо учитывать функцию их эмпирических распределений.

Если не удается проверить гипотезу о соответствии эмпирического распределения определенному закону из-за малого количества данных (или распределения существенно отличаются от нормального закона и не поддаются нормализации), то для проверки гипотезы о наличии корреляционной связи можно использовать ранговый коэффициент корреляции Спирмена. Его расчет основан на замене выборочных значений исследуемых случайных величин их рангами в порядке возрастания. При этом предполагается, что если между значениями случайных величин нет корреляционной зависимости, то ранги этих величин тоже будут независимыми. Выражение для расчета рангового коэффициента корреляции имеет вид

$$r = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)},$$

где  $d_i$  — разность рангов сопряженных значений изучаемых величин  $x_i$  и  $y_i$ ;  $n$  — количество пар в выборке.

Если ранги значений  $X$  и  $Y$  являются независимыми случайными величинами, то выборочная оценка  $r$  распределена нормально с математическим ожиданием 0 и дисперсией  $1/(n-1)$ .

Для проверки значимости рангового коэффициента корреляции можно использовать величину  $r_{kp} = Z(P)/\sqrt{n-1}$ , где  $Z(P)$  — значение обратной функции нормального распределения при доверительной вероятности  $P$ .

Если расчетное значение коэффициента Спирмена ( $r$ ) больше критического ( $r_{kp}$ ), то гипотеза о независимости исследуемых величин отвергается.

*Пример.* При проведении гидрогеологических изысканий по одному из профилей буровых скважин были выполнены опытные электроразведочные работы. Для оценки эффективности этого метода необходимо знать, существует ли зависимость между электрическим сопротивлением пород ( $\rho_k$ ) и относительной мощностью горизонта гравийно-галечных отложений ( $m_r$ ), к которым приурочены основные водоносные горизонты. В профиле пробурено всего 12 скважин, поэтому для проверки гипотезы о наличии корреляционной связи между значениями  $\rho_k$  и  $m_r$  можно воспользоваться ранговым коэффициентом корреляции (табл. 25).

$$\text{Ранговый коэффициент корреляции } r = 1 - \frac{6 \cdot 104}{12(144-1)} = 1 - 0,39 = 0,61.$$

Величина обратной функции нормального распределения для доверительной вероятности 0,95 равна 1,64, а критическое значение рангового коэффициента корреляции при этой доверительной вероятности и объеме выборки в 12 значений будет  $r_{kp} = \frac{1,64}{\sqrt{12-1}} = 0,49$ .

Расчетное значение рангового коэффициента корреляции больше критического, что свидетельствует о статистически значимой корреляционной связи между значениями относительной мощности и электрическим сопротивлением горизонта гравийно-галечных отложений. Следовательно, данный геофизический метод можно рекомендовать для широкого использования при проведении гидрогеологических изысканий в данном районе.

Таблица 25. Результаты вычисления рангового коэффициента корреляции для значений  $m_r$  и  $\rho_k$

№ скв.	$m_r$		$\rho_k$		$d_i$	$d_i^2$
	значение, %	ранг	значение, Ом·м	ранг		
1	67	9	253	10	-1	1
2	80	12	115	7	5	25
3	40	5	126	8	-3	9
4	24	2	82	6	-4	16
5	25	3	66	5	-2	4
6	38	4	25	1	3	9
7	18	1	44	3	-2	4
8	72	10	180	9	1	1
9	44	6	32	2	4	16
10	51	8	319	11	-3	9
11	76	11	421	12	-1	1
12	50	7	51	4	3	9

$$\sum d_i^2 = 104$$

При наличии нескольких пар с одинаковыми значениями вычисление коэффициента корреляции несколько усложняется:

$$r = 1 - \left[ 6 \left( \sum_{i=1}^n d_i^2 + T_x^2 + T_y^2 \right) / n(n^2 - 1) \right],$$

где  $T_x^2$  и  $T_y^2$  — поправки на повторяющиеся содержания первого и второго элемента соответственно, вычисляемые из выражения

$$T^2 = \frac{1}{12} \sum_{i=1}^m (l_i^3 - l_i),$$

где  $l_i$  — количество данных с повторяющимися содержаниями элемента;  $m$  — количество таких групп с повторяющимися содержаниями элемента.

Коэффициент корреляции может служить надежной характеристикой стохастической связи лишь в условиях линейности обоих уравнений регрессии. В равной мере это относится и к ранговому коэффициенту корреляции Спирмена. Однако в геологической практике зависимость между свойствами изучаемых объектов часто отличается от линейной. Так, например, в рудах одного из свинцовых месторождений присутствует золото, которое рассматривается как сопутствующий полезный компонент. Линейная прямая корреляционная связь между концентрациями золота и свинца в рудах проявляется только при содержаниях свинца ниже 1,5%, для богатых руд она практически отсутствует, а руды среднего качества характеризуются обратной корреляционной связью (рис. 33). Это объясняется тем, что в бедных вкрапленных рудах галенит первой генерации тесно ассоциирует с золотоносным пиритом, а высокие концентрации свинца в богатых рудах связаны с наличием более поздних незолотоносных кварц-карбонат-галенитовых прожилков.

Нелинейный характер зависимости между исследуемыми величинами может возникнуть как результат специфических условий эксперимента. Так, например, оконтуривание рудных тел по заданной минимальной мощности и минимальному содержанию полезного компонента в пробе приводит к появлению сложных нелинейных зависимостей между мощностями рудных тел и содержанием в них полезных компонентов (рис. 34).

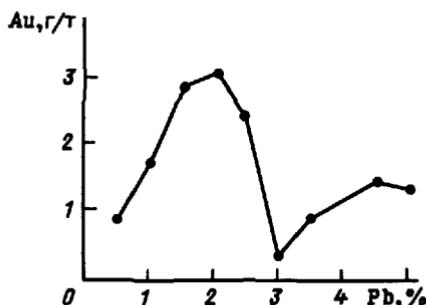


Рис. 33. Зависимость между содержанием золота и свинца в рудах свинцового месторождения



Рис. 34. Линия регрессии для мощности рудного тела  $m$  (в м) и содержания полезного компонента  $C$  (в усл. ед.)

О характере связи судят по виду эмпирических линий регрессии. Если они заметно отличаются от прямой, гипотезу о наличии корреляционной связи следует проверить с помощью отношения.

Для вычисления оценок корреляционных отношений выборочные данные группируются в классы по значениям одного из исследуемых свойств. По каждому классу рассчитываются групповые средние  $\bar{y}_i$  или  $\bar{x}_i$  и оценки стандартных отклонений групповых средних  $S_{\bar{y}_i}$  и  $S_{\bar{x}_i}$  по формулам

$$S_{\bar{y}_i} = \sqrt{\frac{1}{N} \sum_{i=1}^m n_i (\bar{y}_i - \bar{y})^2}; \quad S_{\bar{x}_i} = \sqrt{\frac{1}{N} \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2}, \quad (22)$$

где  $n_i$  — число наблюдений в  $i$ -ой группе;  $m$  — число групп;  $N$  — общее число наблюдений.

Выборочные значения корреляционных отношений определяются

$$\eta_{y/x} = S_{\bar{y}_i}/S_y; \quad \eta_{x/y} = S_{\bar{x}_i}/S_x,$$

где  $S_y$  и  $S_x$  — оценки общего стандартного отклонения исследуемых случайных величин.

Статистическая значимость различия корреляционного отношения от нуля проверяется с помощью критерия

$$\theta_y = \frac{\eta_{y/x}^2 (N-m-2)}{(1-\eta_{y/x}^2)(m-2)} \sqrt{\frac{(m-2)(N-m-4)}{2(N-4)}}. \quad (23)$$

При равенстве истинного корреляционного отношения нулю величина  $\theta_y$ , распределена нормально с математическим ожиданием 0 и дисперсией 1, что позволяет определять критические значения  $\theta_y$  для заданных доверительных вероятностей по таблицам нормального распределения. Если расчетное значение  $\theta_y$  превышает критическое, гипотеза об отсутствии корреляционной связи отвергается. Аналогично проверяется гипотеза о наличии корреляционной связи по  $\eta_{x/y}$ .

*Пример.* Проверим гипотезу о наличии корреляционной связи между содержаниями золота и свинца в рудах упомянутого выше свинцового месторождения. С помощью коэффициента корреляции устанавливается отсутствие связи (табл. 26). Однако, учитывая нелинейный характер графика линии регрессии (см. рис. 33), гипотезу о наличии корреляционной связи следует проверить повторно по корреляционному отношению (табл. 27). С этой целью по формуле (22) рассчитываем стандартное отклонение групповых средних:

$$S_{\bar{y}_i} = \sqrt{\frac{1}{60} \cdot 33,479} = 0,747.$$

Затем, используя данные табл. 26, рассчитываем стандартное отклонение величины  $y$  по формуле

$$S_y = \sqrt{\frac{1}{N-1} \left[ \sum_i^N y_i^2 - \frac{(\sum y_i)^2}{N} \right]} = \sqrt{\frac{1}{59} \left( 250,98 - \frac{107,11^2}{60} \right)} = 1,006.$$

Оценка корреляционного отношения  $\eta_{y/x}$  для содержания золота по содержанию свинца составит

$$\eta_{y/x} = S_{\bar{y}_i}/S_y = \frac{0,747}{1,006} = 0,743.$$

Таблица 26. Расчет коэффициента корреляции между содержаниями золота  $Y$  (в г/т) и свинца  $X$  (в %)

$t$	$x_t$	$y_t$	$x_t y_t$	$x_t^2$	$y_t^2$
1	2,05	3,76	7,71	4,20	14,14
2	5,03	2,09	10,51	25,30	4,37
3	0,80	1,98	1,58	0,64	3,92
4	0,31	0,20	0,06	0,10	0,04
5	0,77	3,10	2,39	0,59	9,61
6	4,01	1,67	6,70	16,01	2,79
7	1,19	2,59	3,08	1,42	6,71
8	1,26	1,70	2,14	1,59	2,89
9	0,68	0,23	0,16	0,46	0,05
10	0,91	1,21	1,10	0,83	1,46
11	4,33	0,91	3,94	18,75	0,83
12	2,38	1,68	4,00	5,66	2,82
13	0,98	2,44	2,39	0,96	5,95
14	0,42	0,50	0,21	0,18	0,25
15	1,71	1,21	2,07	2,92	1,46
16	3,51	1,15	4,04	12,32	1,32
17	1,11	2,30	2,55	1,23	5,29
18	2,10	3,48	7,31	4,41	12,11
19	1,21	0,61	0,74	1,46	0,37
20	2,92	0,40	1,17	8,53	0,16
21	0,74	0,27	0,20	0,55	0,07
22	1,53	2,57	3,93	2,34	6,60
23	3,70	0,90	3,33	13,69	0,81
24	2,71	1,69	4,58	7,34	2,86
25	0,79	0,78	0,62	0,62	0,61
26	1,90	4,32	8,21	3,61	18,66
27	1,51	2,30	3,47	2,28	5,29
28	0,21	1,22	0,26	0,04	1,49
29	4,81	1,05	5,05	23,14	1,10
30	1,38	2,09	2,88	1,90	4,37
31	3,96	2,54	10,06	15,68	6,45
32	1,96	1,58	3,10	3,84	2,50
33	0,52	0,82	0,43	0,27	0,67
34	2,95	0,20	0,59	8,70	0,04
35	1,10	1,44	1,58	1,21	2,07
36	0,93	3,15	2,93	0,86	9,92
37	1,78	1,21	2,15	3,17	1,46
38	5,16	0,87	4,49	26,62	0,76
39	0,37	1,15	0,42	0,14	1,32
40	0,44	0,91	0,40	0,19	0,83
41	2,21	4,25	9,39	4,88	18,06
42	4,67	2,03	9,48	21,81	4,12
43	1,44	4,31	6,21	2,07	18,58
44	3,13	0,25	0,78	9,80	0,06
45	1,35	0,39	0,53	1,82	0,15
46	0,81	1,35	1,09	0,66	1,82
47	1,32	3,51	4,63	1,74	12,32
48	0,99	1,62	1,60	0,98	2,62
49	2,41	3,98	9,59	5,81	15,84
50	1,03	0,35	0,36	1,06	0,12
51	1,55	2,80	4,34	2,40	7,84
52	3,39	0,41	1,39	11,49	0,17
53	1,23	1,58	1,94	1,51	2,50
54	1,48	4,22	6,24	2,19	17,81
55	4,03	1,19	4,80	16,24	1,42

<i>i</i>	<i>x<sub>i</sub></i>	<i>y<sub>i</sub></i>	<i>x<sub>i</sub>y<sub>i</sub></i>	<i>x<sub>i</sub><sup>2</sup></i>	<i>y<sub>i</sub><sup>2</sup></i>
56	0,64	2,31	1,48	0,41	5,34
57	1,15	1,75	2,01	1,32	3,06
58	1,52	3,22	4,89	2,31	10,37
59	3,40	1,21	4,11	11,56	1,46
60	0,67	0,98	0,66	0,45	0,96
$\sum$	114,55	107,11	198,05	324,26	250,98
	$198,05 - \frac{1}{60} \cdot 114,55 \cdot 107,11$			$-6,44$	
	$r = \frac{-6,44}{\sqrt{(324,26 - \frac{1}{60} \cdot 114,55^2)(250,98 - \frac{1}{60} \cdot 107,11^2)}}$			$79,44$	$= -0,081;$
	$t = \frac{0,081 \sqrt{58}}{\sqrt{1 - 0,0066}} = 0,62.$				

Таблица 27. Расчет корреляционного отношения между содержаниями золота (*Y*) и свинца (*X*)

№ п/п	Классы группирования по <i>X</i>	$\bar{y}_i$	<i>n<sub>i</sub></i>	$\bar{y}_i - \bar{y}$	$(\bar{y}_i - \bar{y})^2$	$(\bar{y}_i - \bar{y})^2 n_i$
1	0,20 – 0,75	0,80	10	-0,99	0,980	9,800
2	0,76 – 1,25	1,75	15	-0,04	0,002	0,030
3	1,26 – 1,75	2,90	11	1,11	1,232	1,367
4	1,75 – 2,25	3,10	6	1,31	1,716	10,296
5	2,26 – 2,75	2,45	3	0,66	0,436	1,308
6	2,76 – 3,25	0,28	3	-1,51	2,280	6,840
7	3,26 – 3,75	0,92	4	-0,87	0,757	3,028
8	3,76 – 4,25	1,80	3	0,01	0	0
9	4,26 – 4,75	1,47	2	-0,32	0,102	0,204
10	4,76 – 5,25	1,34	3	-0,45	0,202	0,606

$$(\bar{y}_i - \bar{y}) n_i = 33,479$$

Для проверки статистической значимости отличия корреляционного отношения от нуля по формуле (23) рассчитываем значение величины  $\theta$ ,

$$\theta = \frac{0,743^2 (60 - 10 - 2)}{(1 - 0,743^2)(10 - 2)} \sqrt{\frac{(10 - 2)(60 - 10 - 4)}{2(60 - 4)}} = 13,69.$$

Поскольку критическое значение величины  $\theta$  при доверительной вероятности 0,95 составляет 1,96, гипотеза об отсутствии корреляционной связи между исследуемыми величинами отвергается. Следовательно, между этими величинами существует статистически значимая нелинейная корреляционная связь.

### 3.4. ИСПОЛЬЗОВАНИЕ КОРРЕЛЯЦИОННЫХ СВЯЗЕЙ ДЛЯ ПРЕДСКАЗАНИЯ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

Если для двух величин на основании представительной выборки доказано наличие корреляционной связи, определен ее вид и подобрано описывающее его уравнение, то создается возможность прогноза значений одной из случайных величин по значениям другой. Подобные

задачи часто возникают в геологической практике. Широко распространен случай, когда эмпирическое распределение одной из случайных величин не противоречит нормальному закону (или может быть нормализовано с помощью какого-либо преобразования), а значения другой случайной величины могут выбираться произвольно. Для изучения связей между такими величинами используются методы регрессионного анализа, позволяющие установить влияние произвольно выбранных значений одной величины (например,  $X$ ) на значения другой, нормально распределенной случайной величины (например,  $Y$ ). В отличие от корреляционного анализа в этом случае анализируются только регрессии  $Y$  и  $X$ , но не наоборот.

Регрессионный анализ применяется для:

— оценки содержания сопутствующих компонентов по содержанию основных компонентов в рудах. Например, кадмий в полиметаллических рудах обычно входит в состав сфалерита, а рений в молибденсодержащих медно-порфировых рудах — в состав молибдениита. Непосредственное определение редких и рассеянных элементов в рудах затруднительно ввиду их низких концентраций, сопоставимых с точностью анализов. Надежные количественные оценки содержаний этих элементов получаются только при анализе мономинеральных фракций, отбор которых весьма трудоемок. Поэтому по ограниченному количеству мономинеральных проб рассчитываются характеристики корреляционной зависимости между содержанием основного и попутного компонента, которые в дальнейшем используются для определения среднего содержания попутного компонента в каждом подсчетном блоке;

— определения объемной массы руд. На месторождениях железа, свинца, хрома, барита и других полезных ископаемых, где руды характеризуются высоким содержанием полезного минерала, значительно отличающегося по удельной массе от вмещающихся пород, обычно наблюдается отчетливая зависимость между объемной массой руды и содержанием полезного компонента. В этом случае объемная масса по каждому подсчетному блоку также может определяться корреляционным методом;

— интерпретации результатов геофизических методов опробования. Геофизические методы опробования основаны на замере физических свойств руды (естественного или вызванного излучения, магнитности и т. п.), изменения которых пропорциональны концентрации полезного компонента. Кроме концентрации полезного компонента на изменение данного свойства оказывают влияние и другие факторы, не всегда поддающиеся точному учету. Поэтому на практике для интерпретации результатов геофизических методов опробования часто применяются «кривые соответствия», которые представляют собой эмпирические линии регрессии, рассчитанные по замерам физических свойств руд и результатам анализов проб, отобранных в месте замера;

— уточнения оценок параметров рудных тел по результатам отработки. Опыт эксплуатации месторождений свидетельствует о том, что средние содержания полезных компонентов в блоках богатых руд по данным разведки оказываются обычно завышенными, а в бед-

ных — заниженными. В процессе эксплуатации по отработанным блокам могут быть рассчитаны уравнения регрессии истинных средних содержаний и содержаний, определенных по данным разведки. Эти уравнения можно использовать для уточнения оценок средних содержаний в оставшихся блоках.

Решение задач данного типа основано на построении эмпирических линий регрессии или расчете их аналитических выражений — уравнений регрессии. Для правильного решения таких задач необходимо не только оценить силу корреляционной связи, но и выявить ее характер. Поэтому приближенный способ проверки гипотезы о линейности связи по виду эмпирической линии регрессии в данном случае обычно дополняется аналитическими расчетами.

Аналитический способ проверки гипотезы о линейности связи основан на том, что при ее наличии коэффициент корреляции и корреляционное отношение совпадают по абсолютной величине.

Подходящим критерием для проверки данной гипотезы является критерий Фишера

$$F = (\eta_{y/x}^2 - r^2)(N-m)^2 / (1-\eta_{y/x}^2)(m-2),$$

где  $\eta_{y/x}$  — корреляционное отношение признака  $Y$  по классам группирования  $X$ ;  $m$  — число классов группирования;  $N$  — количество пар значений  $XY$ .

Полученные значения  $F$  сравниваются с табличными значениями  $F_{kp}$  для заданного уровня значимости  $\alpha$  при  $f_1=(m-2)$  и  $f_2=(N-m)$  степенях свободы. Корреляция должна считаться нелинейной, если  $F > F_{kp}$ .

*Пример.* Проверим гипотезу о линейности корреляционной связи между содержаниями золота и свинца в рудах свинцового месторождения (исходные данные см. в табл. 26, 27) с помощью критерия Фишера

$$F = \frac{(\eta_{y/x}^2 - r^2)(N-m)}{(1-\eta_{y/x}^2)(m-2)} = \frac{(0,743^2 - 0,081^2)(60-10)}{(1-0,743^2)(10-12)} = 10,65.$$

Табличное значение критерия Фишера для доверительной вероятности 0,95 при 8 и 50 степенях свободы равно 2,13. Следовательно, гипотеза о линейности корреляционной связи между содержаниями данных компонентов в руде отвергается.

Аналитическим выражением линейной связи являются уравнения регрессии вида:  $y = a_1 + b_1 x$ ;  $x = a_2 + b_2 y$ .

Оценки коэффициентов линейной регрессии определяются по формулам:

$$b_1 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (x_i - \bar{x})^2;$$

$$b_2 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (y_i - \bar{y})^2;$$

$$a_1 = \bar{y} - b_1 \bar{x}; \quad a_2 = \bar{x} - b_2 \bar{y}.$$

Нелинейная связь может быть выражена аналитически с помощью уравнения параболы  $y = a + bx + cx^2$ .

Коэффициенты этого уравнения находятся путем группирования одного параметра по классам значений другого и решения системы уравнений

$$an + b \sum_{i=1}^k n_i x + c \sum_{i=1}^k n_i x^2 = \sum_{i=1}^k n_i \bar{y}_i;$$

$$a \sum_{i=1}^k n_i x + b \sum_{i=1}^k n_i x^2 + c \sum_{i=1}^k n_i x^3 = \sum_{i=1}^k n_i x \bar{y}_i;$$

$$a \sum_{i=1}^k n_i x^2 + b \sum_{i=1}^k n_i x^3 + c \sum_{i=1}^k n_i x^4 = \sum_{i=1}^k n_i x^2 \bar{y}_i,$$

где  $n_i$  — количество проб в классах группирования;  $\bar{y}_i$  — частное среднее значение величины  $Y$  в данном классе.

Иногда нелинейность корреляционной зависимости бывает вызвана асимметричностью статистических распределений одного или обоих исследуемых параметров. В этом случае линия регрессии становится близкой к прямой, если при построении воспользоваться полулогарифмическим или логарифмическим масштабом. Уравнения регрессии для таких параметров будут иметь вид:  $y = a + b \lg x$ ;  $\lg y = a + bx$ ;  $\lg y = a + b \lg x$ .

Другой причиной возникновения нелинейной связи может явиться неоднородность изучаемого объекта. Так, в рассмотренном ранее примере нелинейность связи между содержаниями золота и свинца в рудах свинцового месторождения вызвана наличием различных типов руд. В этом случае нелинейность связи можно устраниТЬ, разделив выборочные значения на несколько совокупностей, однородных с генетическими позициями. Так, выделив из данных, приведенных в табл. 26, только пробы, характеризующие вкрапленные руды (табл. 28), можно убедиться, что связь между содержаниями золота и свинца для этого типа руд близка к линейной (рис. 35) и выражается уравнением  $y = 1,69x$ .

Таблица 28. Расчет уравнения регрессии для содержаний золота и свинца во вкрапленных рудах

$i$	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	0,80	1,98	-0,15	0,37	0,137	-0,097	0,022
2	0,31	0,20	-0,64	-1,41	1,988	0,902	0,410
3	0,77	3,10	-0,18	1,49	2,220	-0,268	0,032
4	1,19	2,59	0,24	0,98	0,960	0,235	0,058
5	1,26	1,70	0,31	0,09	0,008	0,028	0,096
6	0,68	0,23	-0,27	-1,38	1,904	0,373	0,073
7	0,91	1,21	-0,04	-0,40	0,160	0,016	0,002
8	0,98	2,44	0,03	0,83	0,689	0,025	0,001
9	0,42	0,50	-0,53	-1,11	1,232	0,588	0,028
10	1,71	1,21	0,76	-0,40	0,160	-0,304	0,058
11	1,11	2,30	0,16	0,69	0,476	0,110	0,026
12	0,74	0,27	-0,21	-1,34	1,796	0,281	0,044
13	0,79	0,78	-0,16	-0,83	0,689	0,133	0,026
14	1,51	2,30	0,56	0,69	0,476	0,386	0,166
15	0,21	1,22	-0,74	-0,39	0,152	0,289	0,548
16	1,38	2,09	0,43	0,48	0,230	0,206	0,185
17	0,52	0,82	-0,43	-0,79	0,624	0,340	0,185
18	1,10	1,44	0,15	-0,17	0,029	-0,026	0,024
19	0,37	1,15	-0,58	-0,46	0,212	0,267	0,336
20	0,44	0,91	-0,51	-0,70	0,490	0,357	0,260
21	1,44	4,31	0,49	2,70	7,290	1,323	0,240
22	1,35	0,39	0,40	-1,22	1,488	-0,488	0,160
23	0,81	1,35	-0,14	-0,26	0,068	0,036	0,020

<i>i</i>	<i>x<sub>i</sub></i>	<i>y<sub>i</sub></i>	<i>x<sub>i</sub> - x̄</i>	<i>y<sub>i</sub> - ȳ</i>	( <i>y<sub>i</sub> - ȳ</i> ) <sup>2</sup>	( <i>x<sub>i</sub> - x̄</i> )( <i>y<sub>i</sub> - ȳ</i> )	( <i>x<sub>i</sub> - x̄</i> ) <sup>2</sup>
24	0,99	1,62	0,04	0,01	0	0	0,002
25	1,03	0,35	0,08	-1,26	1,588	-0,100	0,006
26	1,55	2,80	0,60	1,19	1,416	0,714	0,360
27	1,23	1,58	0,28	-0,03	0,001	-0,008	0,078
28	1,48	4,22	0,53	2,61	4,666	1,383	0,281
29	0,64	2,31	-0,31	0,70	0,490	-0,217	0,096
30	0,67	0,98	-0,28	-0,63	0,397	0,176	0,078
<b>Σ</b>	<b>28,39</b>	<b>48,35</b>			<b>32,036</b>	<b>6,660</b>	<b>3,900</b>

$$\bar{x} = 0,95; \quad y = 1,61; \quad b_1 = \frac{6,600}{3,900} = 1,69;$$

$$a_1 = 1,61 - 1,69 \cdot 0,95 \approx 0; \quad y = 1,69x.$$

Для выбора вида преобразования величины *Y* или *X* с целью приближения линии регрессии к прямой удобно пользоваться так называемой «лестницей преобразований»:

$$\begin{aligned}
 & Y^3(X)^3 \\
 & Y^2(X^2) \\
 & Y(X) \\
 & \sqrt{Y}(\sqrt{X}) \\
 & \lg Y(\lg X) \\
 & -\frac{1}{\sqrt{Y}} \left( -\frac{1}{\sqrt{X}} \right) \\
 & -\frac{1}{Y} \left( -\frac{1}{X} \right) \\
 & -\frac{1}{Y^2} \left( -\frac{1}{X^2} \right) \\
 & -\frac{1}{Y^3} \left( -\frac{1}{X^3} \right).
 \end{aligned}$$

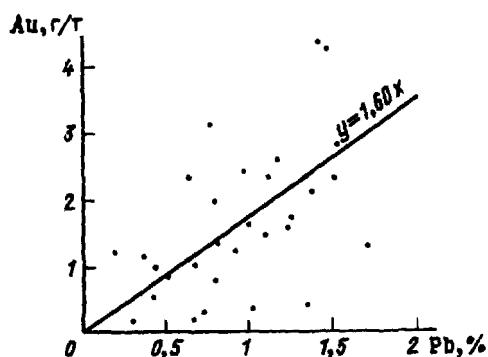
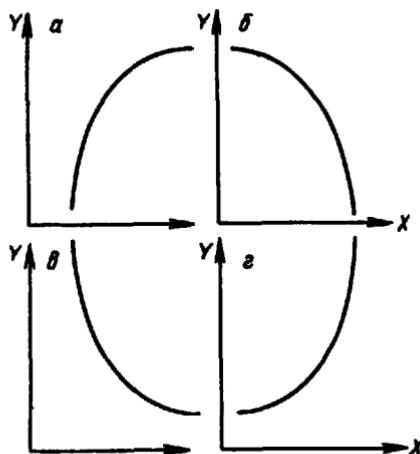


Рис. 35. Линия регрессии для содержания золота в зависимости от содержания свинца для вкрашенных руд свинцового месторождения

Рис. 36. Направление движения по лестнице преобразований:  
 а—по направлению  $Y^2$ ,  $Y^3$  или  $\lg x$ ,  
 $-1/x$  и т. д.; б—по направлению  $Y^2$ ,  
 $Y^3$  или  $x^2$ ,  $X^3$ , в—по направлению  
 $\lg x$ ,  $-1/x$  и т. д. или  $\lg y$ ,  $-1/y$  и т. д.;  
 г—по направлению  $X^2X^3$  или  $\lg y$ ,  
 $-1/y$  и т. д.



Выбор наиболее подходящего преобразования осуществляется путем последовательного перебора вариантов преобразований, расположенных вниз или вверх по «лестнице» от значений  $X$  или  $Y$ . Направление движения по лестнице преобразований определяется по характеру кривизны графика линии регрессии  $Y$  по  $X$  (рис. 36).

Вопросы применения корреляционного и регрессионного анализов в геологии и геохимии рассмотрены в работах Р. Миллера и Дж. Кана, Дж. Дэвиса, В. Н. Бондаренко. Интересные примеры использования двумерной статистической модели для решения некоторых задач разведки месторождений приведены в работах Д. Криге. Приемы построения и анализа графиков зависимостей изложены в монографии Дж. Тьюки.

#### *Контрольные вопросы*

1. Свойства каких геологических объектов можно рассматривать как двумерную статистическую совокупность?
2. Каков характер связи между свойствами геологических объектов?
3. Что можно узнать о двумерной статистической совокупности с помощью корреляционного поля точек, схематических диаграмм условных распределений и эмпирических кривых регрессий?
4. Какие геологические задачи решаются путем проверки гипотезы о наличии корреляционной связи?
5. Как оценивается сила корреляционной связи?
6. Для чего используется корреляционное отношение и ранговый коэффициент корреляции?
7. При каких условиях коэффициент корреляции может служить надежной характеристикой силы корреляционной связи?
8. Для чего в геологии применяется регрессионный анализ?
9. Какими уравнениями можно описать характер корреляционной связи свойств геологических объектов?
10. Как можно проверить гипотезу о линейном характере корреляционной связи?
11. Что может служить причиной и нелинейной зависимости свойств геологических объектов?
12. Как можно устранить нелинейность корреляционной связи?

Любое геологическое явление может быть охарактеризовано множеством признаков, поддающихся наблюдению и измерению. Геологические объекты должны рассматриваться как системы, зависящие от большого числа факторов и требующие для своего описания многомерного признакового пространства. Так, например, магматические породы сходного минерального и химического составов могут обладать некоторыми петрохимическими особенностями, определяющими их специфическую рудоносность. Эти особенности не поддаются выявлению с первого взгляда, однако они могут быть установлены путем целенаправленной статистической обработки результатов химических анализов пород. При решении подобных задач необходимо совместное рассмотрение комплекса изучаемых признаков, т. е. создание многомерной статистической модели.

#### 4.1. СУЩНОСТЬ И УСЛОВИЯ ПРИМЕНЕНИЯ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

В качестве математической модели значений комплекса признаков рассматривается многомерная случайная величина, которая часто называется случайным вектором. Многомерные модели подразумевают вероятность нормального статистического распределения рассматриваемых случайных величин или хотя бы возможности их нормализации. Однако статистические критерии для большинства процедур многомерного анализа разработаны при очень сильных ограничениях или основываются на логических соображениях. Некоторые многомерные модели и методы (например, метод главных компонент и многие методы распознавания образов) вообще не имеют статистического обоснования, а критерий значимости для них еще не созданы.

Вследствие сложных стохастических взаимосвязей между изучаемыми признаками (переменными) часто не удается принять правильного решения относительно каждой из них. В таких случаях очень эффективно всестороннее исследование системы с выделением наиболее важных факторов, объединяющих влияние нескольких переменных.

Многомерные методы статистических исследований сложны как с теоретических, так и с методологических позиций.

В большинстве многомерных геологических задач приходится иметь дело со сложными сочетаниями действующих факторов, которые не удается выделить в чистом виде и изучить изолированно. Тем не менее многомерные методы являются весьма перспективными и многообещающими средствами геологических исследований, поскольку они позволяют геологу одновременно работать с большим числом переменных, чем он может осознать сам. Совместное изучение

комплексов взаимосвязанных переменных (признаков) способствует выявлению дополнительной, часто весьма существенной, информации об изменчивости свойств изучаемых объектов и обеспечивает возможность прогнозирования их неизвестных свойств.

Для работы с многомерными математическими моделями необходимо знание основ линейной алгебры, поскольку записи исходных данных и математические действия над ними производятся в матричной форме. Общие сведения о матрицах наблюдений и результатов вычислений можно найти в работе Дж. Дэвиса.

## 4.2. МНОГОМЕРНЫЙ КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Многомерный корреляционный анализ применяется для выявления зависимостей между наблюденными значениями различных геологических характеристик и разделения множества признаков по характеру их внутренних связей.

Статистические свойства случайных величин с  $n$ -мерным нормальным распределением задаются их ковариационными и корреляционными матрицами, которые могут быть вычислены по исходным матрицам. С этой целью исходная матрица  $[A]$  порядка  $m \times n$  умножается на ее транспонированный аналог  $[A]^T$ , расположенный слева\*, причем между матрицами  $[A]^T$  и  $[A]$  вводится промежуточный множитель  $\left( [E] - \frac{[J]}{m} \right)$ , где  $[E]$  — единичная матрица порядка  $n \times n$ ;  $[J]$  — матрица порядка  $n \times n$ , состоящая только из единиц;  $m$  — число строк исходной матрицы. Полученная матрица  $[L]$  имеет порядок  $n \times n$ :  $[A]^T \left( [E] - \frac{[J]}{m} \right) [A] = [L]$ .

Путем умножения матрицы  $[L]$  на величину  $\frac{1}{(m-1)}$  получается ковариационная матрица  $[C]$ , диагональные элементы которой являются дисперсиями, а недиагональные — ковариациями:

$$C = \frac{1}{m-1} \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 & \dots & \dots & \Sigma x_1 x_m \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \Sigma x_m x_1 & \Sigma x_m x_2 & \dots & \dots & \Sigma x_m^2 \end{bmatrix}.$$

Если вычислить ковариационную матрицу для стандартизованных переменных (т. е. умножить ее на величину  $1/S_i S_j$ ) получится корреляционная матрица  $[R]$ , в которой по диагонали расположены единицы, а недиагональные элементы представляют собой парные коэффициенты корреляции между соответствующими признаками:

\* Матрица  $[S] = [A]^T \cdot [A]$  называется матрицей сумм квадратов и смешанных произведений.

$$[R] = \begin{bmatrix} 1 & r_{x_1 x_2} & \dots & \dots & r_{x_1 x_m} \\ \dots & 1 & \dots & \dots & \dots \\ \dots & \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & 1 & \dots \\ r_{x_m x_1} & r_{x_m x_2} & \dots & \dots & 1 \end{bmatrix}.$$

Для изучения взаимосвязей нескольких случайных величин вычисляют соответствующие парные и частные корреляционные коэффициенты, а для оценки зависимости одной случайной величины от других случайных величин — множественный коэффициент корреляции, выражющий меру линейности такой зависимости (или выборочное корреляционное отношение, выражющее меру нелинейности связи).

*Частный коэффициент корреляции* оценивает меру линейной зависимости между двумя случайными величинами  $x_i$  и  $x_j$  при условии, что влияние всех остальных случайных величин устранено. (В отличие от частного коэффициента, парный коэффициент корреляции двух случайных величин служит мерой их линейной зависимости). Оценка частного коэффициента обозначается как  $r_{ij.q}$ , где  $q$  — набор индексов 1, 2, 3, ...,  $m$  без  $i$  и  $j$ . В данном случае этот коэффициент корреляции оценивает линейную связь двух признаков, указанных слева от точек в индексе при  $r$ , а влияние всех остальных признаков, номера которых расположены справа от точки, устранено. Расчет частного коэффициента производится по формуле

$$r_{ij.q} = \frac{r_{ij} - r_{iq} r_{jq}}{\sqrt{(1 - r_{iq}^2)(1 - r_{jq}^2)}}.$$

Например, для корреляций двух величин при исключении третьей

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}.$$

В матричной записи  $r_{ij.q} = -\frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}$ , где  $C_{ij}$  — алгебраическое дополнение к соответствующему элементу  $r_{ij}$  в определителе корреляционной матрицы.

Множественный коэффициент корреляции позволяет оценить меру линейной зависимости одной случайной величины  $x_i$  от совокупности других случайных  $x_{m-1}$  величин. Выборочный множественный коэффициент корреляции оценивается как  $R_{i.k} = \sqrt{1 - \frac{1}{C^{ii}}}$  (где  $k$  — набор  $m-1$  индексов;  $C^{ii}$  — диагональный элемент матрицы  $[R]^{-1}$ , обратной корреляционной матрице  $[R]$ ) или  $R = \sqrt{1 - \frac{|L|}{\beta_{||} |L'|}}$  (где  $|L|$  — определитель ковариационной матрицы;  $\beta_{||}$  — первый член ковариационной матрицы;  $|L'|$  — определитель ковариационной матрицы без первого столбца и первой строки).

Методы многомерного корреляционного анализа используются в геологии для изучения зависимостей между случайными величинами, чаще всего между признаками, зависящими от совокупного влияния факторов неясной физической природы. Если выборки значений изучаемых признаков согласуются с  $m$ -мерным нормальным распределением, достоверности парных, частных и множественных коэффициентов корреляции могут быть оценены путем проверки различных статистических гипотез с определением вероятных доверительных областей. Когда о функциональных видах распределений случайных величин ничего не известно (или они характеризуются различными функциями статистических распределений), могут использоваться ранговые коэффициенты парной, частной и множественной корреляций, а также коэффициенты конкордации (согласованности) ранжированных рядов, не имеющие своего параметрического аналога. Методов проверки статистической значимости ранговых коэффициентов частной, множественной корреляции и конкордации пока что не разработано. Помимо оценок частных и множественных коэффициентов корреляции, корреляционный анализ включает в себя различные статистические методы выделения и группирования ассоциаций химических элементов.

#### 4.3. СТАТИСТИЧЕСКИЕ МЕТОДЫ ВЫДЕЛЕНИЯ АССОЦИАЦИЙ ХИМИЧЕСКИХ ЭЛЕМЕНТОВ

Методы выделения парагенетических химических элементов сводятся к разделению исходных совокупностей на несколько классов по степени сходства, составляющих их объектов. При этом предполагается, что степень сходства объектов, объединяемых в один класс, должна быть существенно больше степени сходства между объектами из разных классов.

Основой классификации исходных совокупностей химических элементов служат корреляционные матрицы их содержаний в изучаемых объемах (минералах, породах или рудах), а мерой сходства — оценки парных коэффициентов их корреляции. Выделение парагенетических ассоциаций элементов сводится к выявлению внутри корреляционной матрицы  $[R]$  однородных блоков (подматриц типа  $[R_i]$ ), соответствующих таким совокупностям элементов, содержания которых связаны линейной положительной зависимостью.

Строгое математическое решение задачи классификации исходных совокупностей по корреляционным характеристикам практически невозможно, поскольку коэффициенты парной корреляции элементов имеют стохастическую природу, в связи с чем во многих конкретных случаях возникают возможности многовариантного группирования объектов.

В геологической практике предложены различные методы исследования структур корреляционных матриц, среди которых Б. И. Смирнов выделяет шесть групп:

- 1) анализ корреляционной матрицы с позиций теории графов;
- 2) иерархическое группирование (клuster-анализ) с помощью дендрограмм;

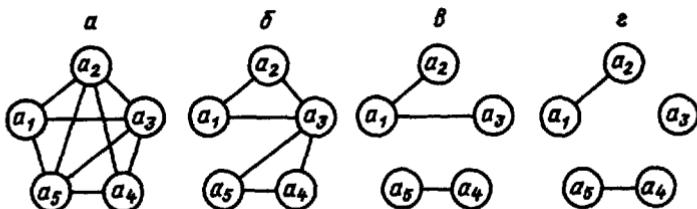


Рис. 37. Графы при изменяющемся пороговом значении:  
 $a - r=0,2$ ;  $b - r=0,3$ ;  $v - r=0,4$ ;  $z - r=0,5$

- 3) методы, опирающиеся на понятия компактности;
- 4) метод корреляционных профилей; 5) коническая корреляция;
- 6) методы, использующие идеи и приемы факторного анализа.

Ниже рассматриваются возможности использования методов первой и второй групп, а в разделе «метод главных компонент» — методы шестой группы. Для знакомства с остальными методами можно рекомендовать работы С. Миллера, Дж. Кана и Б. Н. Смирнова.

*Анализ корреляционной матрицы с позиций теории графов* отличается наиболее простыми приемами классификации. Графом  $G(A)$  называется геометрическая схема, включающая две и более точки конечного множества  $A = \{a_1, \dots, a_k, \dots, a_l, \dots, a_p\}$ , соединенные линиями, если между ними установлены линейные корреляционные связи. Каждая из точек (кружков) обозначает один из химических элементов. Элементы множества  $A$  называют вершинами, а линии их соединяющие — ребрами. Вершины, соединяемые ребрами, называются смежными, а не соединяемые — изолированными. Если все вершины графа смежны, причем реализованы все возможные для них соединения, граф называется полным (рис. 37).

Две вершины  $a_k$  и  $a_l$  ( $k \neq l$ ) называются связными, если существует цепь, концы которой находятся в вершинах  $a_k$  и  $a_l$ . Граф называется связным, если любая пара его вершин связана. На рис. 37, б, в показано два частичных подграфа (компоненты связности), обладающих свойством связности  $G(A') = \{a_1, a_2, a_3\}$  и  $G(A'') = \{a_4, a_5\}$ . Граф называется взвешенным, если каждому его ребру поставлено в связь некоторое число.

При решении многих геологических задач результаты вычисления корреляционных матриц часто изображаются в виде графов без дополнительной математической обработки данных. С этой целью весь диапазон значений коэффициентов парной корреляции условно разбивается на несколько интервалов (например, высокие — более 0,75, средние — 0,5—0,75, низкие — 0,3—0,5 и незначимые — менее 0,3), а длины ребер между смежными вершинами выбираются обратно пропорционально значениям соответствующих парных коэффициентов. Для высоких значений они показываются толстыми линиями, для средних — тонкими сплошными, а для низких — прерывистыми. Несмотря на возможность субъективных решений, иллюстрация многомерных линейных связей элементов с помощью подобных графов весьма наглядна, а число возможных вариантов их группирования

заметно снижается с повышением пороговых значений коэффициентов парной корреляции (см. рис. 37, а—г).

Для получения более строгих решений описание графов выполняется на основе матриц смежности, по алгоритмам нахождения максимально полных подграфов (максимально обособленных групп), методами «ветвящихся связей» или «максимального корреляционного пути», описанными в монографии Б. Н. Смирнова.

#### 4.4. КЛАСТЕР-АНАЛИЗ (ДЕНДРОГРАММЫ И ДЕНДРОГРАФЫ)

Методы иерархической группировки (численной таксономии) исходных множеств получили название *клластер-анализа*, т. е. анализа групп. Первоначально они применялись в биологии и палеонтологии, а в настоящее время широко используются в различных отраслях геолого-минералогических наук, в частности, при классификации парагенетических ассоциаций элементов земной коры.

Задача кластер-анализа сводится к разбиению множества элементов корреляционной матрицы признаков [ $R$ ] на группы таким образом, чтобы в них объединялись объекты с наивысшими значениями характеристик сходства, а разобщенные группы оставались бы при этом максимально изолированными по данному признаку. В качестве меры сходства могут использоваться непосредственно парные коэффициенты корреляции,  $m$ -мерное евклидово расстояние или другие дистанционные коэффициенты.

Первый шаг анализа групп методом попарного объединения элементов матрицы состоит в выявлении наивысших коэффициентов корреляции между отдельными парами, которые объединяются и принимаются за центры групп. Число таких центров может изменяться от одного до трех (редко более).

Далее матрица вычисляется снова, причем сгруппированные элементы считаются за один элемент, а коэффициенты их корреляции с другими группами вычисляются заново с помощью различных методов осреднения. По результатам вычисления составляется новая матрица, меньшей размерности, в которой изменяются лишь значения коэффициентов, связанные с членами объединенных групп. Сокращенная и пересчитанная матрица вновь подвергается сокращению, путем выявления и объединения пар с максимальными значениями признаков сходства, и последующим осреднением новых групповых коэффициентов. Операция последовательного сокращения и пересчета коэффициентов матрицы повторяется до тех пор, пока значения групповых коэффициентов сходства не достигнут порогового значения или размерность матрицы не станет минимальной. В первом случае процедура последовательного группирования прекращается на том ее шаге, которому соответствует принятое пороговое значение признака.

Для целей корреляционного анализа парагенетических ассоциаций элементов рекомендуется использовать дистанционный коэффициент  $d_t$ , который может быть рассчитан, как арккосинус коэффициента

парной корреляции и обладает свойствами метрики. С этой целью исходная матрица  $[R]$  трансформируется с помощью тригонометрических таблиц в матрицу дистанционных коэффициентов  $[D_T]$ , которая и подвергается процедуре последовательного группирования.

Результаты кластер-анализа изображаются в виде древовидного графа—дендограммы, в которой по оси абсцисс располагаются символические обозначения объектов исследования (векторов матрицы), а по оси ординат—минимальные значения дистанционных коэффициентов, соответствующих каждому шагу классифицирующей процедуры. Таким образом ось ординат используется для масштабного представления иерархических уровней группирования.

Наглядность и содержательная ценность древовидных графов существенно повышаются, если в них отражена информация не только о тесноте внутригрупповых связей, но и о межгрупповых расстояниях

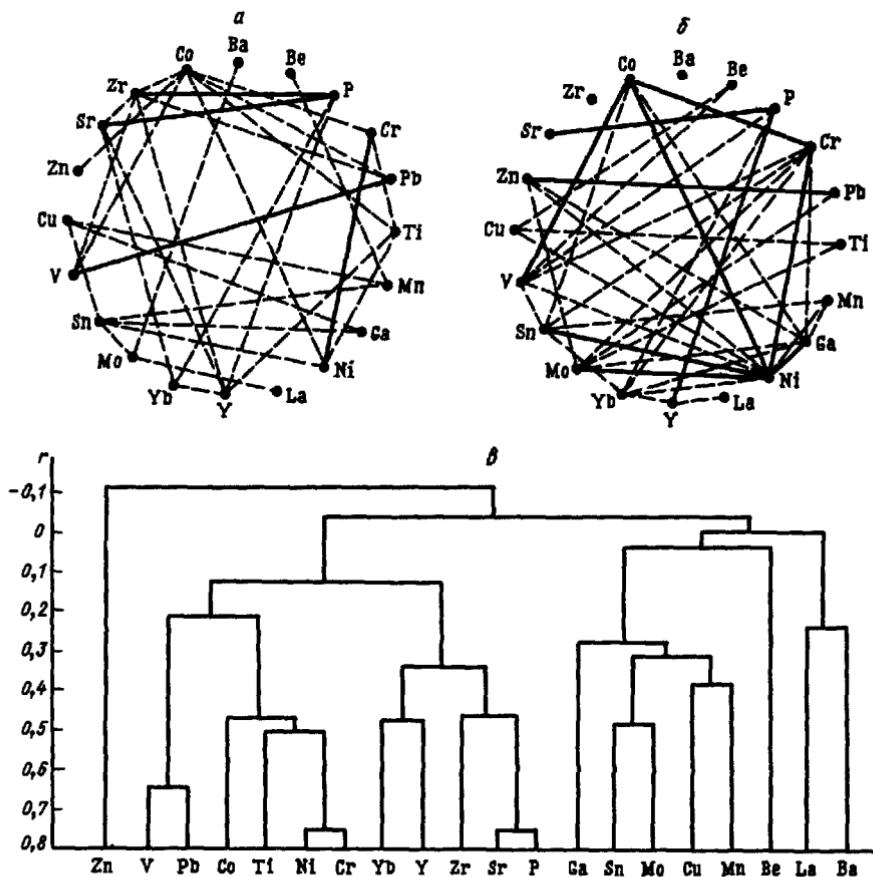


Рис. 38. Характеристики корреляционных связей между содержаниями химикатов  
а—граф по безрудным альбититам; б—граф по рудным альбититам; в—дендограмма

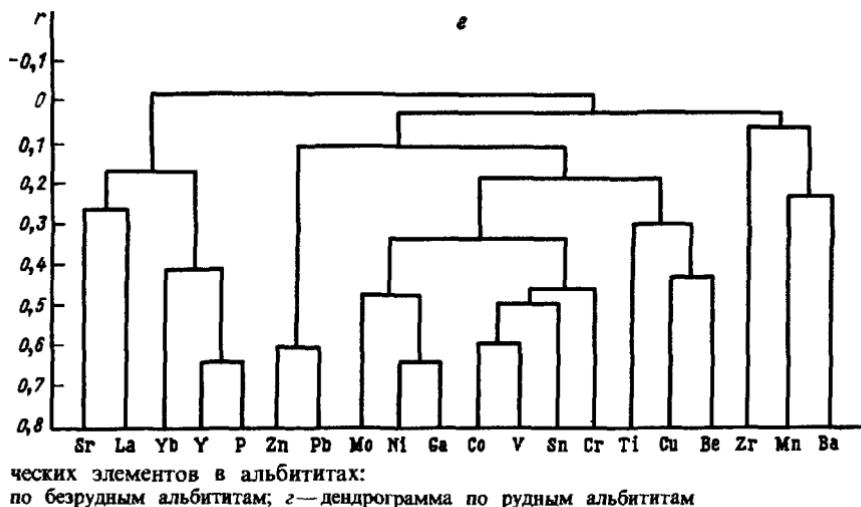
h. С этой целью по матрицам  $[D_T]$ ,  $[D'_T]$ ,  $[D''_T]$  и другим последовательно вычисляются значения  $h$ ,  $h''$  и т. д., которые затем выносятся на ось абсцисс. Такой дендривидный граф, учитывающий не только внутригрупповые расстояния, но и средние расстояния между группами, называется *дендрографом*.

**Пример.** Рудные тела редкometалльного месторождения приурочены к зонам натровых метасоматитов (альбититов). В результате детального изучения минерального состава метасоматитов было установлено, что на месторождении развиты альбититы двух типов. Причем редкometалльное орудение характерно лишь для одного из них. По химическому составу рудные и безрудные альбититы весьма близки, поэтому различить их по содержанию отдельных химических элементов не удается. Однако некоторые различия в минеральном составе проявляются в особенностях корреляционных связей между элементами. Наглядно эти различия отражаются на графах (рис. 38, а, б) и дендрограммах (рис. 38, в, г). В качестве меры близости элементов при построении дендрограмм в данном случае используются непосредственно парные коэффициенты корреляции, рассчитанные по 50 пробам, отобранным из рудных и безрудных альбититов, тип которых однозначно определен минералогическими исследованиями. Предельное значение коэффициента корреляции для доверительной вероятности 0,95 при объеме выборок в 50 проб равен 0,28. Поэтому для целей классификации целесообразно сравнивать лишь те группы, для которых коэффициенты корреляции, отражающие тесноту внутригрупповой связи, превышают эту величину.

Для обеих дендрограмм характерна группа, объединяющая фосфор и редкоземельные элементы, что, по-видимому, обусловлено присутствием в альбититах обоих типов апатита, в составе которого отмечены повышенные концентрации этих элементов.

Основной отличительной особенностью безрудных альбититов является тесная ассоциация сидерофильных элементов ( $Ni—Cr—Ti—Co$ ), которая в рудных альбититах распадается.

Для рудных альбититов характерна ассоциация халькофильных элементов ( $Pb—Zn$ ), в то время как в безрудных альбититах корреляционная связь между этими элементами отрицательная. Таким образом кластер-анализ позволяет оперативно и достаточно надежно определить тип альбититов по результатам спектральных анализов, не прибегая к детальному изучению шлифов.



#### 4.5. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И ЕЕ ИСПОЛЬЗОВАНИЕ ДЛЯ ПРЕДСКАЗАНИЯ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

В отличие от двумерной регрессии в методах множественной регрессии зависимая переменная  $Y$  рассматривается как функция не одной, а нескольких независимых переменных  $X_1, X_2, \dots, X_m$ .

Уравнение множественной регрессии зависимой переменной  $Y$  относительно  $m$  независимых переменных  $X_{1,2,\dots,m}$  записывается как линеаризированная функция вида

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m = \beta_0 + \sum_{i=1}^m \beta_i x_i, \quad (24)$$

где  $\beta_0, \beta_1, \beta_2, \dots, \beta_m$  — коэффициенты регрессионной модели. Оно наилучшим способом (в смысле наименьших квадратов) описывает тенденцию расположения наблюденных точек в  $m$ -мерном пространстве и позволяет оценить совместное влияние всех изучаемых параметров на зависимую переменную.

Множественная регрессия строится на основе учета всех возможных взаимодействий между переменными и их сочетаниями. В ее задачи входит оценка общего вклада всех переменных ( $R^2$ ) в изменчивость  $Y$ , а также определение относительного влияния каждой из них с помощью коэффициентов  $\beta_i$ . Таким образом, множественный регрессионный анализ сводится к вычислению значений коэффициентов регрессионной модели  $\beta_0, \beta_1, \beta_2, \dots, \beta_m$  по совокупности  $n$  наблюдений над переменными  $X_1, X_2, \dots, X_m$  и  $Y$ , оценке влияния каждой переменной и их общего вклада в оценку зависимой переменной  $Y$ .

В матричной форме уравнение (24) записывается как:

$$[\Sigma Y] = [\Sigma X] [\beta],$$

где  $[\Sigma Y]$  — вектор-столбец, состоящий из сумм квадратов и смешанных произведений переменной  $Y$  с переменными  $X_1, X_2, \dots, X_m$ ;  $[\Sigma X]$  — матрица сумм квадратов и смешанных произведений  $X_1, X_2, \dots, X_m$ ;  $[\beta]$  — вектор-столбец неизвестных коэффициентов регрессии. Коэффициенты регрессии  $\beta_i$  рассчитываются как *частные коэффициенты регрессии*, характеризующие изменения данной независимой переменной при условии, что влияние всех остальных переменных устранено. Приведенное уравнение может быть решено путем обращения матрицы  $[\Sigma X]$ , однако в процессе обращения матрицы возникают вычислительные трудности, связанные с резким возрастанием числа цифр в суммах квадратов, что приводит к потерям значащих цифр при их округлении. Поэтому для решения уравнений обычно используются корреляционные матрицы зависимой и независимой переменных  $[R]$ , со стандартизованными частными коэффициентами

регressии типа  $B_k = b_k \frac{S_k}{S_y}$ , где  $S_k$  — оценка стандартного отклонения переменной  $X_m$ ;  $S_y$  — оценка стандартного отклонения переменной  $Y$ . В матричной форме уравнение записывается как  $[R] [B] = [r_{xy}]$ , где  $[r_{xy}]$  — вектор коэффициентов корреляции между переменными

$Y$  и  $X_{1,2,\dots,m}$ , а его решение с обращением матрицы  $[R]^{-1}$ , как

$$[B] = [R]^{-1} [r_{xy}].$$

Рассчитанные коэффициенты  $B$  переводятся в  $\beta$  по формуле  $b_k = B_k (S_y / S_k)$ , а постоянный член  $b_0$  рассчитывается по формуле  $b_0 = \bar{y} - b_1 x_1 - b_2 x_2 - \dots - b_m x_m$ . Оценка общего вклада всех независимых переменных в оценку определяется значением квадрата множественного коэффициента корреляции  $R^2$ . Значение  $R = 1 - 1/C^{\text{ii}}$  (где  $C^{\text{ii}}$  — диагональный элемент матрицы  $[R]^{-1}$ , обратной корреляционной матрице  $[R]$ ).

Для сравнительной оценки вклада каждой зависимой переменной коэффициент  $R^2$  сначала рассчитывается для пары  $y$  и  $x_k$  с максимальным коэффициентом корреляции, а затем последовательно с тремя и более переменными (до  $m$  переменных).

Модели множественной регрессии используются для предсказаний значений зависимой переменной (например, содержаний ценного элемента, объемной массы руды и глубины формирования минерала и др.) по набору независимых переменных (например, содержаний пордообразующих элементов, объемных масс тяжелых минералов в рудах, содержаний элементов-индикаторов в минералах и др.).

*Пример.* Рассмотрим результаты статистического изучения особенностей распределения содержаний рубидия, цезия, лития и калия в микроклинах крупного пегматитового тела. Многими исследователями было отмечено, что калиевые полевые шпаты, взятые из различных частей пегматитовых жил, характеризуются разным содержанием щелочей, причем наблюдается закономерное уменьшение содержаний редких щелочей в калиевых полевых шпатах от центра жилы к зальбандам и с глубиной. Однако эти выводы основываются на ограниченных аналитических данных и носят скорее качественный, чем количественный характер. Пегматитовое тело в виде крутопадающей секущей жилы залегает среди переслаивающихся кварц-биотитовых и андалузит-биотитовых сланцев среднепротерозойского возраста. Оно обладает сложной неправильной формой (характерна изменчивость элементов залегания как по простирианию, так и по падению: наличие раздупов, пережимов, изгибов и апофиз), отличается крупными размерами и довольно ясно выраженным симметрично-зональным строением. Призальбандовую часть тела слагает эндоконтактовая мусковит-кварцевая оторочка шириной от 10 до 40 см. Эта оторочка сменяется зоной мелкозернистого существенно альбитового пегматита мощностью от 1 до 33 м. Центральная часть жилы сложена неравномерно-зернистым средне- и крупнозернистым, часто блоковым, неравномерно альбитизированным сподумен-микроклин-кварцевым пегматитом директивной структуры, в котором встречаются гнезда и полосы существенно альбитового состава и участки, сложенные кварц-сподуменовым агрегатом.

Из описанного пегматитового тела была отобрана серия проб микроклина (56 образцов), для которых количественным спектральным методом были определены содержания рубидия, цезия, лития и калия.

Для выяснения зависимости между расположением микроклина в пределах пегматитового тела и такими его характеристиками, как содержание редких щелочей и калия, величина отношения калия к рубидию и рубидия к цезию, интенсивность окраски был проведен многомерный корреляционный анализ взаимосвязи перечисленных выше параметров с координатами, определяющими положение пробы в пределах тела. В качестве координат были приняты:  $l$  — относительное расстояние места взятия пробы от контакта;  $h$  — относительная высота места взятия. Содержания элементов ( $C_{\text{Rb}}$ ,  $C_{\text{Cs}}$ ,  $C_{\text{Li}}$ ,  $C_{\text{K}}$ ) оценивались в тысячных долях процента; относительные расстояния от контакта ( $l$ ) — в процентах от мощности жилы в данном сечении; относительные высоты ( $h$ ) отсчитывались от условного нулевого уровня, совпадающего с нижней границей обнаженной части жилы; интенсивность окраски (оцениваемая субъективно)

была обозначена следующими индексами: 1 — светло-серая; 2 — серая; 3 — темно-серая; 4 — очень темно-серая.

Для того, чтобы исключить возможное влияние краевых зон, все варианты задач по определению множественной корреляции решались дважды: а) для всего набора проб ( $n=56$ ) и б) только для проб из центральной части жилы, т. е. тех, у которых  $l < 25\%$  ( $n=34$ ). Для учета законов распределения случайных величин (содержания редких щелочей в микроклинах распределены логнормально, а калия — нормально) все вычисления производились дважды: для случайных величин и их логарифмов. Но так как проверка гипотезы о равенстве коэффициентов корреляции с помощью преобразования Фишера показала, что различия между оценками этих коэффициентов для случайных величин и их логарифмов несущественны (при 5%-ном уровне значимости), результаты вычислений для логарифмов не приводятся.

Таблица 29. Статистические характеристики коррелируемых параметров микроклинов из пегматитовой жилы

Признак	Средне-арифметическое $\bar{x}$	Средне-квадратичное отклонение $\sigma_x$	Средне-арифметическое логарифмов $\lg x$	Среднеквадратичное отклонение логарифмов $S_{\lg x}$	Средне-геометрическое $\bar{x}$
56 проб					
Относительная высота $h$	164,50	114,45	1,9854	0,572	96,70
Относительное расстояние от контакта $l$	33,20	18,25	1,4054	0,376	25,43
Содержание (в $n \cdot 10^{-3}\%$ ):					
калия ( $C_K$ )	11 429,46	1849,26	4,0140	0,138	10 330,0
рубидия ( $C_{Rb}$ )	891,62	462,28	2,9107	0,184	814,10
цезия ( $C_{Cs}$ )	56,50	54,63	1,6405	0,284	43,70
лития ( $C_{Li}$ )	45,86	67,58	1,3523	0,697	22,51
Величина отношения $K/Rb$	13,18	5,40	1,0848	0,171	12,15
$Rb/Cs$	21,51	11,51	1,2682	0,251	18,55
Цветовой индекс	2,34	0,71	0,3470	0,144	2,22
34 пробы					
Относительная высота $h$	163,03	117,89	1,9730	0,574	91,83
Содержание (в $n \cdot 10^{-3}\%$ ):					
калия ( $C_K$ )	10 145,00	959,84	4,0062	0,041	10 140,0
рубидия ( $C_{Rb}$ )	946,06	456,74	2,9386	0,176	868,20
цезия ( $C_{Cs}$ )	58,09	43,28	1,6845	0,248	48,37
лития ( $C_{Li}$ )	52,04	75,19	1,4661	0,657	29,25
Величина отношения $K/Rb$	12,66	4,91	1,0669	0,186	11,66
$Rb/Cs$	19,81	8,66	1,2507	0,214	17,81
Цветовой индекс	2,44	0,74	0,3649	0,148	2,32

Таблица 30. Парные коэффициенты корреляции параметров микроклинов из крупного пегматитового тела (выше диагонали — для выборки  $n=56$ ; ниже — для выборки  $n=34$ )

Параметры	$h$	$C_{Rb}$	$C_{Cs}$	$C_{Li}$	$C_K$
$h$		0,503	0,485	(-0,125)	-0,312
$C_{Rb}$	0,539		0,442	(0,117)	-0,481
$C_{Cs}$	0,570	0,586		(-0,078)	-0,681
$C_{Li}$	(-0,237)*	(-0,084)	(-0,111)		-0,416
$C_K$	(-0,287)	(-0,193)	(-0,329)	(0,193)	

\* В скобках указаны коэффициенты корреляции незначимо отличные от нуля.

Таблица 31. Множественные корреляционные коэффициенты параметров для микроклинов по выборке проб ( $n$ )

Признаки	$n=34$			
	6*	5	4	3
Относительная высота $h$		0,6360	0,6282	0,6236
Содержание:				
рубидия ( $C_{Rb}$ )	0,4826	0,6716	0,6687	0,6367
цезия ( $C_{Cs}$ )	0,4843	0,7022	0,6798	0,6586
лития ( $C_{Li}$ )	(0,1180)**	(0,3591)	(0,3237)	
калия ( $C_K$ )	(0,3535)	(0,3846)		
Относительное расстояние от контакта $l$	(0,2669)			

\* Число признаков, участвующих в расчете.

\*\* В скобках указаны коэффициенты корреляции, не значимо отличные от нуля.

В результате вычислений были получены основные статистические характеристики коррелируемых параметров (табл. 29), их парные и множественные коэффициенты корреляции (табл. 30, 31) и составлены в ряде случаев уравнения множественной регрессии.

Анализируя полученные данные, можно сделать некоторые выводы.

1. Сравнение статистических характеристик проб только из центральной части жилы ( $n=34$ ) с характеристиками всего набора проб ( $n=56$ ) показывает (см. табл. 29), что исключение из расчета приконтактовых проб (у которых  $l > 25\%$ ) вызывает увеличение средних содержаний рубидия, цезия, лития, уменьшение содержаний калия и величины отношений калия к рубидию и рубидия к цезию. Следовательно, микроклин из краевых частей жилы содержит более низкие концентрации редких щелочей, в нем несколько больше калия, выше отношения калия к рубидию и рубидия к цезию.

2. Результаты определения величины парных коэффициентов корреляции (см. табл. 30) показывают, что при обоих вариантах расчета наблюдается значимая линейная связь между параметром  $h$  и содержаниями рубидия и цезия, а также между содержаниями рубидия и цезия, тогда как между содержаниями цезия и лития, рубидия и лития она отсутствует.

3. Результаты определения оценок коэффициентов множественной корреляции (см. табл. 31) показывают, что в исследованном комплексе признаков существует значимая совокупная взаимосвязь между величиной  $h$  и содержаниями редких щелочей, главным образом рубидия и цезия.

Наличие достаточно сильной множественной линейной связи между этими величинами позволяет построить уравнение множественной регрессии типа

$$h = 43,08 + 0,08 C_{Rb} + 1,01 C_{Cs} - 0,27 C_{Li};$$

$$h = 25,48 + 0,08 C_{Rb} + 1,05 C_{Cs}.$$

Эти уравнения могут быть использованы для примерной оценки величины эрозионного среза и прогноза жильного тела на глубину.

#### 4.6. ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ В ГЕОЛОГИИ

Многие прогнозные и интерпретационные задачи решаются в практической геологии путем сопоставления комплекса признаков изучаемого объекта с комплексом тех же признаков эталонного объекта. Совокупность подобных методов, основанных на принципе аналогии, получила название методов распознавания образов.

Проблема распознавания образов геологических объектов тесно связана с проблемой их классификации по совокупности известных признаков. С позиций многомерного математического анализа

реальному геологическому объекту ставится в соответствие набор действительных чисел  $x_1, x_2, \dots, x_m$ , которые выражают значения измеренных геологических, геохимических или геофизических его признаков. Каждая совокупность таких чисел интерпретируется как вектор или точка в многомерном пространстве, а множествам объектов одного класса в пространстве признаков соответствуют некоторые множества точек.

Области многомерного пространства, в которых сосредоточены точки одного класса, называются собственными областями классов. Если объекты тождественны друг другу, их собственные области совпадают. Собственные области объектов, резко различающихся друг от друга, не пересекаются в пространстве, а для объектов, близких по некоторым признакам, характерно большее или меньшее пересечение их собственных областей.

Разделение непересекающихся собственных областей двух классов возможно с помощью гиперповерхностей (в трехмерном пространстве — плоскостей, в двухмерном — линий), а для разделения пересекающихся областей необходимо отыскать *решающую функцию*, минимизирующую возможность принятия ошибочного решения.

Статистический подход к классификации объектов основан на байесовской теории принятия решений. В терминах этой теории «состояние природы»  $A$  может быть охарактеризовано вероятностью  $P(A_i)$  и условной вероятностью  $P(X/A_i)$ . Под условной вероятностью подразумевается вероятность того, что совокупность признаков объекта примет значение  $X$ , когда состояние природы будет  $A_i$  (т. е. объект будет относиться к классу  $A_i$ ). Если удается оценить априорную  $P(A_i)$  и условную  $P(X/A_i)$  вероятность, можно предсказать то состояние природы, которое будет свойственно объекту в зависимости от того, какое значение  $X$  примут характеристизующие его признаки:

$$P(A_i/X) = \frac{P(A_i) P(X/A_i)}{\sum_{i=1}^m P(A_i) P(X/A_i)},$$

где  $m$  — число возможных состояний природы (т. е. классов).

Для совокупностей многомерных статистических признаков геологических объектов заранее не известно пересекаются ли их собственные области или нет, в связи с чем допускается возможность их пересечения. В качестве меры пересечения  $\Phi$  используется произведение условных вероятностей  $\Phi = P(X/A_i) P(X/A_j)$ , где  $P(X/A_i)$  — вероятность попадания точки  $X$  в область пересечения при условии, что она относится к классу  $A_i$ ;  $P(X/A_j)$  — то же при условии, что она относится к классу  $A_j$ .

Правило, в соответствии с которым производится разграничение объектов на классы, называется *решающим правилом*. Для его реализации в общем случае подыскивается система разделяющих функций  $g_1(X), g_2(X), \dots, g_m(X)$ . Если  $g_i(X) > g_j(X)$ , точка  $X$  принадлежит к  $A_i$  и наоборот. Если  $g_i(X) = g_j(X)$ , то области  $A_i$  и  $A_j$  соприкасаются. Если при подстановке признаков  $X_m$  окажется, что для всех  $i \neq j$  соблюдается неравенство  $g_i(X) > g_j(X)$ , то можно утверждать принадлежность объекта к классу  $i$ .

Для нормального распределения признака  $i$   $g_i(X) = -(\bar{x} - \bar{x}_i)^2 / 2\sigma^2$ . Задача обучения распознаванию образов сводится к подбору математи-

ческой модели алгоритма аппроксимации и построению в пространстве признаков некоторой поверхности (в общем случае — гиперплоскости), разделяющей множества точек, соответствующих различным классам объектов, или к отысканию решающей функции.

Модели распознавания образов геологических объектов весьма разнообразны. При решении конкретных геологических задач их выбор зависит от природы геологических объектов, числа, полноты описания эталонных объектов, типов и информативности их признаков. В зависимости от типов исходных признаков выделяют две группы моделей: дискретные и непрерывные.

Дискретные модели применяются в тех случаях, когда измеряемые признаки рассматриваются как независимые или частично зависимые детерминированные величины.

Непрерывные модели используются для распознавания образов таких объектов, измеряемые признаки которых могут рассматриваться как случайные величины и поддаются статистическому описанию многомерными функциями плотностей вероятности.

В качестве критерии оптимальности распознавания используются решающие правила, определяющие пороговые значения решающих функций. Они могут определяться статистическими, логическими или эвристическими\* методами.

Для целей эффективного использования математических методов распознавания образов необходимо последовательно решить несколько взаимосвязанных задач: выбрать эталонные объекты, признаки и методы распознавания; построить решающее правило и провести классификацию контрольных объектов. Лишь после получения удовлетворительных результатов классификации контрольных объектов можно приступать к распознаванию образов новых объектов.

При выборе эталонных объектов (формировании выборки обучения) следует иметь в виду, что использование статистических алгоритмов распознавания требует, чтобы число объектов в выборке обучения было бы в несколько раз больше числа признаков. Чем больше неизвестных параметров (например, средних значений и дисперсий признаков), которые требуют определения в процессе обучения, тем больше должна быть выборка обучения. Поэтому на практике обычно возникает проблема сокращения числа признаков и повышения их информативности. При использовании алгоритмов, в которых измеренные признаки рассматриваются как детерминированные величины, число объектов в выборке обучения может быть соизмеримым или даже меньше числа измеряемых признаков.

При выборке признаков, характеризующих особенности оцениваемых геологических объектов (формировании признакового пространства объектов), решающее значение имеют знания и аналитические способности исследователей, от эрудиции которых зависит правильная отбраковка заведомо неинформативных и малонинформативных признаков и пре-

---

\* Эвристическими называют такие методы, когда пороговые значения решающих функций оцениваются эмпирически или с позиций «здравого смысла».

дельно возможное ограничение размерности признакового пространства.

Выбор метода распознавания основывается на априорных значениях специалистов-геологов и результатах предварительного анализа выборок объектов по их признаковым описаниям. В «односвязных» объектах классы распознаваемых объектов характеризуются многими особенностями, члены одного и того же класса располагаются близко, а члены различных классов — далеко друг от друга. В таких случаях применяются простые и быстродействующие линейные методы распознавания образов (разделение гиперплоскостью или кусочно-линейной гиперповерхностью). Если же классы состоят из подклассов, один из которых находится в окружении точек другого класса («многосвязные» объекты), для их распознавания используются более сложные и емкие во времени нелинейные методы, например, полиномиальные разделяющие функции.

При использовании любых алгоритмов распознавания следует стремиться к построениям решающих функций как можно более простых видов, поскольку они легче поддаются реализации и обеспечивают более устойчивые решения, особенно при малых выборках обучения.

#### **4.7. СТАТИСТИЧЕСКАЯ ОЦЕНКА ПЕРСПЕКТИВ РУДОНОСНОСТИ ТЕРРИТОРИЙ**

Для прогнозирования перспектив рудоносности территорий, вероятных масштабов оруденения и ожидаемых геологопромышленных типов месторождений используются совокупности (наборы) геологических, геофизических и минералого-geoхимических признаков, для которых изначально установлены пространственные или временные связи с прогнозируемыми геологическими объектами.

В простейших случаях удается различить сравниваемые объекты по суммам кодируемых признаков. В таких схемах различающая способность (информативность) каждого из признаков изучается и оценивается раздельно с помощью условных балльных оценок, коэффициентов парной линейной или ранговой корреляции и другими способами. Однако оценивая информативности каждого признака изолированно, нельзя получить характеристики их суммарной эффективности, поскольку месторождения полезных ископаемых представляют собой сложные геологические системы, сформированные под совокупным влиянием многих рудообразующих факторов. Для оценок их важнейших пространственно-временных связей, необходимо выявление суммарного эффекта доминирующих рудообразующих факторов. Используя более совершенные схемы распознавания перспективных площадей или крупномасштабных рудных объектов с применением человеко-машинных методов предварительного преобразования признаков, принятия оптимальных решений и оценки их надежности можно заметно повысить информативность используемых исходных данных. При этом, однако, вопросы выбора объектов и территорий прогноза, элементарных участков прогнозирования, совокупностей исходных признаков и систем их описания основываются, прежде всего на существующем опыте человека-машинного прогнозирования и геологической зрудации исполнителей.

Обычно за элементарные участки прогноза принимаются квадратные и прямоугольные ячейки, размеры которых определяются средними размерами прогнозируемых объектов, целями и масштабами проводимых исследований. Как правило, площади элементарных ячеек измеряются на картах любых масштабов квадратными сантиметрами или их долями. Для описания элементарных участков выбираются признаки, наиболее тесно связанные с проявлениями процесса оруденения и масштабами рудных образований. Это могут быть наборы детерминированных признаков, характеризующих литологический, гранулометрический, петрографический или минеральный состав полезных ископаемых и вмещающих их пород, их стратиграфическое положение или формационную принадлежность, пространственные связи с тектоническими или вулкано-тектоническими структурами линейного или центрального типов, метасоматитами или гидротермальными соответствующих масштабов проявления или стохастических характеристик различных свойств полезных ископаемых и вмещающих их пород.

Полученный массив описаний разбивается на несколько совокупностей: 1) эталонная выборка перспективных объектов (месторождений, рудных полей и др.); 2) эталонная выборка неперспективных территорий; 3) выборка объектов экзамена, включающая как перспективные, так и неперспективные объекты; 4) массив описания диагностируемых территорий (если необходимых сведений нет, выборки «2», а иногда и «1» могут отсутствовать. В таких случаях, возможна процедура распознавания образов только в самообучающемся режиме «автоматической классификации»).

Дальнейшая обработка данных зависит от видов описываемых признаков, качества и объема эталонных выборок.

При большом числе дискретных детерминированных неметризованных признаков и ограниченном числе эталонных выборок исходных данных, что характерно для подавляющего большинства задач подобного рода, массив исходных векторных переменных преобразуется в альтернативную форму (есть — 1, нет — 0). С целью минимизации размерности признакового пространства определяется информативность каждого признака, отбрасываются наименее информативные из них и выбирается оптимальный алгоритм распознавания.

В случаях отсутствия нулевых вероятностей признаков (невырожденные распределения) возможно применение алгоритмов распознавания типа ШОУ-1, ШОУ-2, Образ-3 и других с использованием статистического дискретного или эвристического решающего правила.

При наличии выборочных частот вероятностей признаков равных нулю (вырожденные распределения) применяются детерминистские алгоритмы распознавания образов типа «Тупиковых тестов», Кора-1, Кора-2, Кора-3, «Перебор», «Гиперпласт» и других с использованием логических и эвристических решающих правил.

Детерминистский алгоритм распознавания «Гиперпласт» использован в человеко-машинной информационно-прогнозирующей системе «Регион-ОС», принятой к внедрению в геологоразведочной службе нашей страны.

## **4.8. ОЦЕНКА ИНФОРМАТИВНОСТИ ГЕОЛОГИЧЕСКИХ ПРИЗНАКОВ**

При решении задач распознавания образов часто возникает проблема сокращения размерности признакового пространства, путем исключения наименее информативных признаков. Сокращение числа признаков не только упрощает используемые алгоритмы распознавания, но и способствует улучшению качества классификации объектов, удаляя случайные помехи («шумы»), искажающие и затушевывающие полезную информацию. Оценка информативности признаков полезна, поскольку она способствует выявлению рудоформирующих факторов и их ранжированию по степени значимости.

Строгих математических методов оценок оптимальных мер относительной информативности признаков пока что не разработано. С этой целью в геологической практике используются различные эвристические, детерминистские и статистические подходы, обеспечивающие формирование признакового пространства меньшей размерности по отношению к исходному. Формирование пространства меньшей размерности достигается либо путем отбора из  $m$  исходных признаков некоторого (меньшего) числа наиболее информативных признаков, либо конструированием новой системы признаков с определенными свойствами, либо одновременно двумя упомянутыми методами. Сложность решения этой задачи связана с тем, что оценка индивидуальной информативности любого признака в отдельности от других признаков часто не обеспечивает достоверных результатов, вследствие наличия статистических связей между ними. Даже относительно независимые признаки, весьма информативные по результатам индивидуальных оценок, могут обеспечить в совокупности с другими признаками лишь весьма незначительную дополнительную информацию и, наоборот, относительно малоинформационные признаки могут значительно увеличивать суммарную информативность системы признаков.

К числу эвристических методов оценки информативности и ранжирования признаков относятся сбор и обработка сведений об изучаемых объектах, получаемых в результате опроса экспертов — специалистов. Поскольку мнения специалистов об информативности конкретных признаков, как правило, расходятся — результаты опроса подвергаются систематизации и математической обработке методами ранговой корреляции или дифференциального ранжирования. Использование информации, обобщающей коллективный опыт специалистов, повышает достоверность исходных данных.

Детерминистские и статистические методы оценки информативности признаков основаны на их ранжировании по данным выборки обучения, т. е. по фактическим данным, характеризующим изучаемые классы объектов.

Ранжирование признаков по выборке обучения можно проводить в предположении их статистической независимости либо в предположении о наличии взаимозависимости между ними.

При детерминированном подходе информативность независимого признака определяется тем, насколько сильно отличаются его значения

в разных распознаваемых классах. Так, например, при использовании алгоритмов «Тупиковые тесты» информационный вес признака  $r$  определяется числом  $R_r = M_r/M$  (где  $M$  — общее число всех тупиковых тестов,  $M_r$  — число тех тупиковых тестов, в которые входит столбец с номером  $r$ ).

Статистический подход к оценке информативности независимых признаков опирается на разность или отношение вероятностей признака в разных классах. Так, например, информативность признака оценивается по нормированной разности средних его арифметических значений в выборках обоих классов:  $J = (\bar{x}_1 - \bar{x}_2)^2 / S^2$  (где  $S^2$  — средневзвешенное стандартное отклонение по обеим выборкам).

Для целей ранжирования условно независимых признаков по их информативности могут быть использованы и оценки парных коэффициентов корреляции их значений в выборках обоих классов.

В практических задачах при ограниченном числе признаков требуется оценка информативности не каждого признака, а их сочетаний. Для сочетаний независимых признаков их информативность может быть определена суммированием информативностей отдельных признаков, выраженных в долях единицы. Однако при решении большинства геологических задач условие взаимозависимости признаков не соблюдается, вследствие такой способ неправомерен.

Для оценок информативности и ранжирования взаимозависимых признаков разработаны специфические методы выделения их наиболее информативных сочетаний, однако многие из них малоэффективны, из-за необходимости перебора огромного числа комбинаций. В практике решения задач по распознаванию образов для оценок информативности признаков и их ранжирования используются методы дискриминантного анализа и метод главных компонент.

#### 4.9. ЛИНЕЙНЫЕ ДИСКРИМИНАНТНЫЕ ФУНКЦИИ

Дискриминантный анализ является мощным статистическим средством разделения (дискриминации) многомерных нормально распределенных совокупностей на группы таким образом, чтобы была достигнута максимальная однородность внутри групп и минимальная между ними.

Задачи дискриминации отличаются от близких к ним задач классификации тем, что число групп (в простейшем случае две) задается заранее и зависит от априорных сведений о соотношениях между пробами, в то время как классификация внутренне замкнута, а число кластеров (групп) не может быть заранее определено.

Линейная дискриминантная функция преобразует множество измерений, входящих в выборку в единственное дискриминантное число (дискриминантную метку). Последним определяется граница разделения совокупностей.

Для нахождения линейной дискриминантной функции строится уравнение регрессии, в котором в качестве зависимых переменных выступают разности между многомерными средними двух анализируемых групп  $\bar{H}_j = \bar{U}_j - \bar{V}_j$ . С этой целью решается матричное уравнение

вида  $[C][a] = [H]$  (где  $[C]$  — ковариационная матрица объединенной выборки размерности  $m \times m$ ;  $[a]$  — вектор-столбец коэффициентов дискриминантной функции;  $[H]$  — вектор-столбец разностей между средними значениями признаков групп). Уравнение решается с помощью операции обращения и умножения матриц  $[a] = [C]^{-1}[H]$ , после чего по вычисленным коэффициентам  $a_1, a_2, \dots, a_p, \dots, a_k$  строится дискриминантная функция

$$DX_1, X_2, \dots, X_p, \dots, X_k = \sum_{p=1}^k a_p X_p. \quad (25)$$

Дискриминантный индекс  $D_0$ , определяющий принадлежность испытуемого объекта к той или иной группе, рассчитывается по формуле

$$D_0 = \frac{1}{2} \sum_{p=1}^k a_p (\bar{u}_p + \bar{v}_p). \quad (26)$$

Он соответствует середине «обобщенного расстояния» между центрами обеих групп в многомерном пространстве изучаемых признаков.

Процедура построения линейной дискриминантной функции сводится к следующему. Обозначим через  $u_{ij}$  результат измерения признака с номером  $i$  в пробе с номером  $j$ , взятой из первой совокупности. В результате мы получим матрицу  $U$  порядка  $k \times n$  результатов наблюдений над этой совокупностью

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n_1} \\ u_{21} & u_{22} & u_{23} & \cdots & u_{2n_1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{k1} & u_{k2} & u_{k3} & \cdots & u_{kn_1} \end{bmatrix}.$$

Обозначим через  $v_{ij}$  результат измерения признаков с номером  $i$  в пробе с номером  $j$ , взятой из второй совокупности, и получим матрицу  $[V]$  порядка  $k \times n_2$

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n_2} \\ v_{21} & v_{22} & \cdots & v_{2n_2} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ v_{k1} & v_{k2} & \cdots & v_{kn_2} \end{bmatrix}.$$

Используя эти данные, вычисляем элементы выборочной ковариационной матрицы  $[B]$ , которую можно представить в матричной записи

$$B = \frac{1}{n_1 + n_2 - 2} (Su + Sv),$$

где  $S_u$  и  $S_v$  — матрицы сумм центрированных квадратов и смешанных произведений, вычисленные по элементам матриц  $[U]$  и  $[V]$ . Далее находим обращенную матрицу  $(C_{ij})$ , вычисляем коэффициенты  $a_1, a_2, \dots, a_k$  и строим дискриминантную функцию

$$D(x) = \sum_{j=1}^p a_j x_j.$$

После расчета порогового значения (дискриминантного индекса)  $D_0$  по формуле (26) оцениваются значения центров обеих исходных групп  $D_{\bar{u}}$  и  $D_{\bar{v}}$ , путем подстановки начала значений  $\bar{u}_j$ , а затем  $\bar{v}_j$  в уравнение дискриминантной функции (25). «Обобщенное расстояние» между двумя многомерными средними  $D^2$  называют расстоянием (или критерием) Махalanобиса. Оно вычисляется путем вычитания  $D_{\bar{u}}$  из  $D_{\bar{v}}$ , что эквивалентно подстановке вектора разности между средними двух групп  $\bar{u}_j$  и  $\bar{v}_j$  в дискриминантную функцию

$$D^2 = \sum_{i=1}^p a_i (\bar{u}_i - \bar{v}_i).$$

Проверка значимости разделения объектов имеет смысл только при соблюдении ряда условий: случайности наблюдений признаков в каждой группе; равной вероятности того, что неизвестное наблюдение принадлежит к любой из групп; нормального распределения признаков и одинакового порядка ковариационных матриц различных групп.

Чем больше  $D^2$ , тем более уверенное разделение можно провести между объектами. Уровень ошибочной классификации определяется как вероятность попадания наблюдения, принадлежащего первому объекту, в область, определяющую второй объект, и наоборот. Численное выражение ошибок классификации определяется из соотношения  $P = 1 - \Phi(D^2)$  (где  $\Phi(D^2)$  — функция нормального распределения).

Величину  $D^2$  можно использовать в дискриминантном анализе при последовательном наборе признаков и минимизации их числа в классификации. Для этой цели нужно расположить признаки по величине их влияния на  $D(x)$ . Первым ставится признак, для которого величина  $D^2$  наибольшая из всех остальных, взятых в отдельности. Вторым ставится признак, который по величине больше всех других, исключая первый, и т. д. Признаки, характеризующиеся отрицательными значениями обобщенного расстояния, отбрасываются.

Дискриминантный анализ позволяет не только надежно решать классификационные задачи, но и определять информативность используемых для классификации признаков и выбирать из первоначального набора признаков оптимальную комбинацию, т. е. подбирать рациональный комплекс поисковых или разведочных исследований.

Отнесение наблюдаемого объекта  $r$  к одной из заданных совокупностей  $R_i$  по значениям дискриминантной функции подразумевает, что классифицируемый объект  $r$  принадлежит к одной из обучающих совокупностей  $R_i$ . На практике же могут встретиться ситуации, когда  $r$  принадлежит классу, не вошедшему в число этих совокупностей. Тогда применение дискриминантного анализа может привести к неправильным представлениям и выводам.

Методы линейного дискриминантного анализа широко используются при решении самых разнообразных геологических задач — для разделения фаций осадочных и магматических пород, перспективных отложений или территорий от неперспективных, оgneупорных и неогнеупорных глин и других геологических объектов.

В качестве примера разделения рассмотрим классификацию кимберлитовых пород одного из месторождений Сибири по петрохимическим данным.

На месторождении выделены три типа кимберлитов — афировые туфы северо-восточной части месторождения, кластопорфировые брекчии центральной части месторождения и афировые туфы юго-запада. Если брекчии резко отличаются от туфов, то туфы двух участков очень близки между собой по внешнему облику и петрографическим особенностям. В результате статистической обработки для каждой пары кимберлитовых пород были получены линейные дискриминантные функции. Так, при сравнении афировых туфов двух участков была получена линейная дискриминантная функция следующего вида:  $45,157\text{MnO} - 1,236\text{FeO} + 1,642\text{CaO} - 33,489\text{N}_2\text{O} + 21,809\text{K}_2\text{O} + 43,213\text{P}_2\text{O}_5 - 1,486\text{S}_{\text{сульф}} - 1,040\text{CO}_2 - 61,386 = D(x)$ .

Если  $D(x) > 0$ , то кимберлитовая порода относится к афировым кимберлитам северо-востока, при  $D(x) < 0$  — к афировым туфам юго-запада.

Из 16 компонент, определенных в каждой пробе, только десять оказались информативными. Ошибка классификации не превышает 0,1%. Для оценки ошибки классификации был проведен экзамен на других массивах (табл. 32).

Таблица 32. Теоретическая и эмпирическая ошибки неправильной классификации для кимберлитовых пород одного месторождения

Афировые туфы юго-запада	Кластопорфировые брекчии	Тип кимберлитов
0,1% / 0,00% (18)	0,3% / 0,00% (42)	Афировые туфы (северо-восточная часть)
	2,0% / 10,00% (44)	Афировые туфы (юго-западная часть)
	5% / 8,0% (52)	Все туфы

Примечание В числителе — теоретическая ошибка классификации, в знаменателе — эмпирическая. В скобках — число проэкзаменованных проб

Анализ данных, приведенных в табл. 32, показывает, что полученные дискриминантные функции дают возможность достаточно надежно определить тип кимберлитовых пород по их химическому составу.

#### 4.10. МЕТОД ГЛАВНЫХ КОМПОНЕНТ

С увеличением размерности признакового пространства возрастают трудности изучения геологических объектов и возникает проблема замены многочисленных наблюдаемых признаков меньшим их числом, без существенной потери полезной информации. Одним из наиболее распространенных методов решения этой задачи является метод главных компонент.

Основой метода главных компонент является линейное преобразование  $m$  исходных переменных (признаков) в  $m$  новых переменных, где каждая новая переменная представляет собой линейное сочетание

исходных. В процессе преобразования векторы наблюдаемых переменных заменяются новыми векторами (главными компонентами), которые вносят резко различные вклады в суммарную дисперсию многомерных признаков. Сокращение пространства признаков достигается путем отбора нескольких наиболее информативных компонент, обеспечивающих основную долю суммарной дисперсии, что приводит к заметному уменьшению их общего числа за счет наименее информативных компонент, отражающих малые доли суммарной дисперсии.

Главные компоненты — это собственные векторы ковариационных матриц исходных признаков. Число собственных векторов ковариационной матрицы определяется числом изучаемых признаков, т. е. равно числу ее столбцов (или строк). Каждый собственный вектор (главная компонента) характеризуется собственным значением и координатами.

*Собственные значения* ковариационной матрицы ( $\lambda_j$ ) — это длины ее собственных векторов, т. е. их дисперсии. Суммы собственных значений ковариационной матрицы равны ее следу, т. е. сумме ее диагональных элементов.

*Координаты* собственного вектора ковариационной матрицы ( $\omega_{ij}$ ) — это числовые коэффициенты, характеризующие его положение в  $m$  мерном признаковом пространстве. Число точечных координат каждого собственного вектора ( $\omega_{ij}$ ) —  $\omega_1, \omega_2, \dots, \omega_m$  определяется размерностью пространства, а их численные значения — это коэффициенты линейных уравнений данного собственного вектора.

Собственные значения ковариационной матрицы находятся как характеристические корни полиномиальных уравнений путем их решения. Однако осуществить это для больших значений  $m$  очень сложно. Поэтому в вычислительной практике их определяют методами матричных преобразований (путем последовательных приближений к собственным значениям), которые могут быть реализованы только с помощью ЭВМ. Методы отыскания координат собственных векторов симметричных матриц также сложны и требуют применения ЭВМ.

Поскольку ковариационные матрицы исходных признаков симметричны, их собственные векторы всегда ортогональны, а составляющие их переменные взаимонезависимы, т. е. не коррелированы между собой.

В методе главных компонент координаты собственных векторов рассматриваются как *нагрузки* соответствующих переменных на тот или иной фактор. Они используются для расчета матриц нового множества совокупностей путем проектирования векторов исходных данных (признаков  $x_1, x_2, \dots, x_m$ ) на оси собственных векторов ( $\gamma_1, \gamma_2, \dots, \gamma_m$ ):

$$\gamma_j = \sum_{i=1}^m \omega_{ji} x_i, \quad (27)$$

где  $\omega_{ji}$  — нагрузки  $j$ -й компоненты в  $i$ -й переменной признака. С помощью формулы (27) исходная матрица наблюдений признаков размерности  $n \times m$  пересчитывается в матрицу новых переменных

(той же размерности), учитывающих собственные значения каждой из компонент. Если статистические (корреляционные) связи между наблюдаемыми признаками многомерного пространства проявляются достаточно отчетливо, то разложение исходной матрицы наблюдений на  $m$  новых компонент приводит к заметному возрастанию контрастности распределения дисперсий по новым компонентам, в сравнении с исходными векторами. Как правило, дисперсия одной из главных компонент достигает половины и более от суммарной дисперсии признаков, а в совокупности с дисперсиями еще одной-двух последующих компонент, их общий вклад в суммарную дисперсию превышает 90%.

Таким образом, без существенной потери информации об изменчивости наблюденных признаков можно заметно сократить размерность пространства наблюденных признаков (до  $p \leq m$ ), ограничившись данными по двум-трем наиболее информативным главным компонентам. Это позволяет считать, что вместо исходной матрицы размерностью  $n \times m$ , для целей геологического анализа может использоваться матрица главных компонент размерностью  $n \times p$  (где  $p$ , как правило не превышает 2—3). Поскольку новые переменные в этой матрице представлены некоррелированными величинами, метод главных компонент может рассматриваться как мощное средство определения истинного числа линейно независимых векторов, содержащихся в исходной матрице.

Реализация метода главных компонент сводится к проведению следующих вычислительных процедур.

1. По исходной матрице наблюдений  $[X]$  размерностью  $n \times m$  формируется ее ковариационная матрица  $[C]$  размерностью  $m \times m$ . С этой целью матрица  $[X]$  умножается слева на ее транспонированный аналог  $[X]^T$ .

2. Для ковариационной матрицы  $[C]$  вычисляются ее собственные значения и собственные векторы. В результате получается матричное уравнение  $[C] = [W]^T [\Lambda] [W]$ , где  $[W]$  — матрица нагрузок, т. е. координат собственных векторов ( $\omega_{ij}$ ) ковариационной матрицы  $[C]$  размерностью  $m \times n$ ;  $[\Lambda]$  — диагональная матрица размерностью  $m \times m$ , главная диагональ которой представлена собственными значениями ковариационной матрицы ( $\lambda_j$ ). В этой матрице столбцы значений  $\lambda_j$  упорядочены (ранжированы) по убыванию собственных значений векторов.

3. По суммарной величине вкладов первых главных компонент, с учетом целей и задач исследования, решается вопрос о минимизации пространства признаков и составляется матрица главных компонент  $[F]$  размерностью  $n \times p$ , где  $p \leq m$  (в большинстве случаев принимается равным 2, реже 3).

Рассмотрим пример использования метода главных компонент при интерпретации геохимических данных. На одном из месторождений редкometалльных пегматитов было установлено, что первичные ореолы рассеяния многих элементов отличаются сложным строением и нечеткими границами. Характер распределения концентраций олова, бериллия и редких щелочей в коренных породах иллюстрируется на рис. 39.

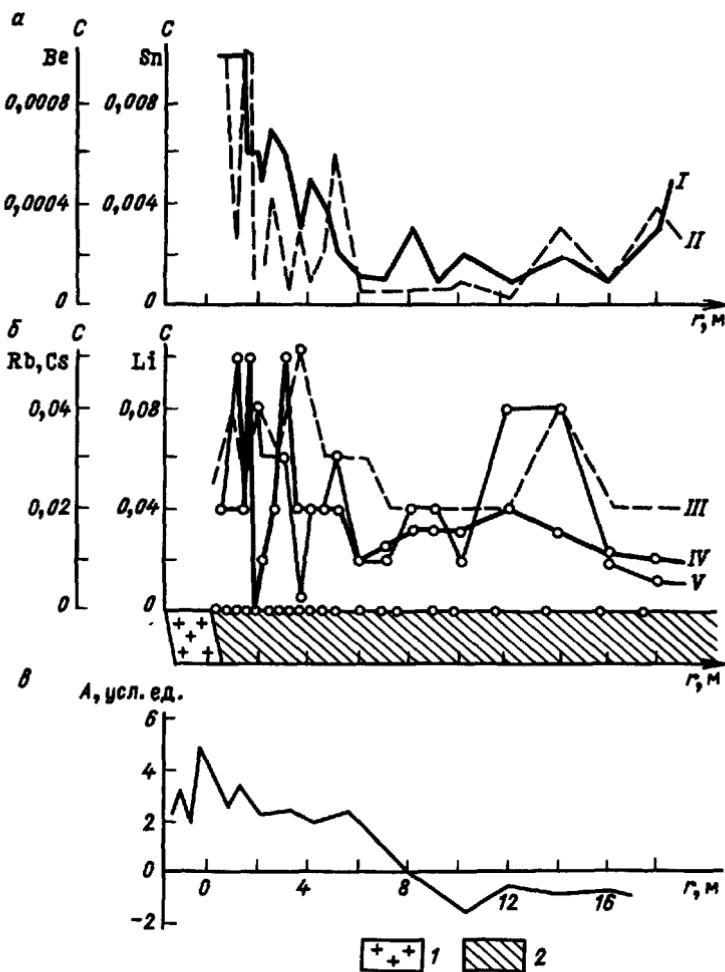


Рис. 39. Характер распределения элементов-индикаторов редкometалльного оруденения в первичных ореолах рассеяния:  
 а—распределение олова (I), бериллия (II); б—распределение рубидия (III), лантана (IV), цезия (V); в—первая главная компонента.  
 1—пегматит, 2—вмещающие биотитовые сланцы

Для выявления геохимических аномалий был проведен корреляционный и компонентный анализ для первой главной компоненты А (табл. 33).

Результаты изучения корреляционных связей дают возможность предполагать значительное влияние фактора удаленности от рудного тела на содержание всех указанных химических элементов в пробах вмещающих пород. Компонентный анализ позволил выявить пять главных компонент, из которых существенной является только первая. Ее дисперсия составляет 67% от суммарной дисперсии. Первая главная компонента отрицательно связана с удаленностью

Таблица 33. Результаты корреляционного и компонентного анализов геохимических данных по месторождению редкometалльных пегматитов

Признак	Коэффициент корреляции					Факторный вес
	Li	Rb	Cs	Sn	Be	
Расстояние от пегматитовой жилы	-0,70	-0,80	-0,67	-0,70	-0,48	-0,81
Содержание:						
лития Li	1,00	0,70	0,60	0,55	0,65	0,74
рубидия Rb		1,00	0,73	0,61	0,45	0,79
цезия Cs			1,00	0,41	0,41	0,62
олова Sn				1,00	0,49	0,59
бериллия Be					1,00	0,48
Дисперсия факторов, %						67,00

Примечание. Коэффициенты корреляции и факторные веса отличаются от нуля с надежностью 0,95 при их абсолютном значении.

от пегматитовой жилы и положительно с концентрациями исследуемых элементов. Эта компонента отражает общий характер размещения элементов-индикаторов редкometалльного оруденения в первичных ореолах рассеяния и может быть с успехом использована для их выделения.

Метод главных компонент используется в качестве основы *факторного анализа* многомерных совокупностей. В практике используются два метода факторного анализа: *R*- и *Q*-метод. Сущность *R*-метода, основанного на использовании не ковариационной, а корреляционной матрицы, близка к задаче метода главных компонент, но в отличие от него факторный анализ считается статистическим методом. Обычное допущение состоит в том, что связи между *m* переменными ставятся в зависимости от корреляционных связей каждой из переменных с *p* взаимно некоррелированными факторами при *p* < *m*. Поэтому дисперсия для *m* переменных вычисляется как сумма дисперсии *p*, так называемых «общих факторов» и независимой от них суммарной добавки от *m* случайных переменных — «фактора специфичности».

Вычисление нормализованных собственных значений и собственных векторов корреляционной матрицы имеет свои специфические особенности. Процедура стандартизации позволяет считать собственные векторы корреляционной матрицы факторами, а их нагрузки, полученные умножением каждой компоненты нормализованного собственного вектора на квадратный корень из соответствующего собственного значения, факторными нагрузками. Суммы квадратов факторных нагрузок по одному фактору равны их собственным значениям. Суммы квадратов факторных нагрузок по каждой переменной (признаки) называются «общностями», которые обеспечивают одинаковые суммы вкладов переменных в факторы. Поскольку значения общностей зависят от числа сохраненных факторов *p*, этот вопрос приобретает в факторном анализе принципиальное значение (хотя и не имеет строгого однозначного решения).

Для более содержательной интерпретации результаты факторного анализа подвергаются процедуре «вращения» вокруг центра координат, для выявления наиболее контрастных сочетаний факторных нагрузок.

Использование метода главных компонент для целей факторного анализа рассмотрим на примере Нерюндинского железорудного месторождения — крупнейшего месторождения Ангаро-Катского района Иркутской области. В его геологическом строении принимают туфогенные отложения корвучанской свиты, песчано-аргиллитовые породы буркуглинской свиты и аргиллиты катской свиты. На месторождении широко развиты интрузии траппов. В структуре месторождения главную роль играет крупный разлом широтного профиля, пересекающий вулканические трубки, картируемые по наличию витров и литокластических туфов, прослеживающихся до глубин 800 м. Рудные тела месторождения, представленные метасоматическими залежами брекчиевидно-вкрашенных руд и магнетитовыми жилами, залегают в породах, существенно измененных пневмогидротермальными растворами и преобразованных в скарны и скарноподобные или скарнированные породы.

Значительная часть руд месторождения подвергнута окислению, что осложняет изучение качества руд. Качество руд месторождения в значительной степени определяется многочисленными природными факторами (конкретная геолого-структурная позиция месторождения, интенсивность развития дорудных и пострудных метасоматических процессов, геохимическая обстановка и др.). Чтобы установить какие факторы сыграли решающую роль в формировании железорудного месторождения, мы обратились к методу главных компонент, позволяющему разложить ковариационную или корреляционную матрицу и сопоставить эти составляющие с реальной обстановкой и существующими генетическими представлениями.

Исходным материалом для решения поставленной задачи явились результаты химических анализов групповых проб, проанализированных на 12 элементов и оксидов.

По общепринятой схеме факторного анализа были получены парные корреляционные коэффициенты, главные компоненты, факторные нагрузки, дисперсии (табл. 34) и значения главных компонент во всех точках наблюдений.

Первая компонента, учитывающая 42% от суммарного воздействия на изученные признаки всей совокупности, определяет общий характер оруденения. Положительные значимые факторные нагрузки, являющиеся коэффициентами корреляции между параметрами и главной компонентой, характеризуют привнос железа и марганца. В то же время отрицательные значения факторных нагрузок свидетельствуют о выносе ряда оксидов.

Вторая и третья компоненты раскрывают роль кальциево-магниевого и калиевого метасоматоза, предшествующего рудообразованию.

Построенная на разрезе тренд-поверхность первой главной компоненты отражает особенности размещения железного оруденения (рис. 40). Положительные значения главной компоненты характеризуют

Таблица 34. Матрица парных коэффициентов корреляции и факторных нагрузок

Признаки	$Fe_{\text{окн}}$	$Fe_2O_3$	FeO	$SiO_2$	$Al_2O_3$	CaO	MgO	MnO	$TiO_2$	P	S	K <sub>2</sub> O	Факторные нагрузки трех компонент		
													1	2	3
$Fe_{\text{окн}}$	1,00	0,63	0,80	-0,75	-0,60	-0,80	-0,06	0,33	-0,58	-0,15	0,30	-0,08	0,90	0,02	0,05
$Fe_2O_3$		1,00	0,64	-0,72	-0,58	-0,71	-0,11	0,29	-0,57	-0,12	0,11	-0,16	0,77	0,00	0,13
FeO			1,00	-0,57	-0,45	-0,74	0,13	0,26	-0,40	-0,27	0,06	0,16	0,67	0,20	-0,16
$SiO_2$				1,00	0,92	0,28	0,46	-0,32	0,77	0,17	-0,02	0,14	-0,78	0,11	0,00
$Al_2O_3$					1,00	0,17	0,37	-0,45	0,68	0,06	-0,14	0,18	-0,59	0,09	0,00
CaO						1,00	-0,46	0,01	0,14	-0,02	0,02	-0,08	-0,42	-0,37	0,10
MgO							1,00	0,48	0,52	0,25	-0,50	0,09	-0,04	0,67	0,03
MnO								1,00	-0,46	-0,28	0,14	0,16	0,22	-0,23	-0,09
$TiO_2$									1,00	0,37	-0,14	0,07	-0,51	0,22	0,01
P										1,00	-0,08	-0,29	-0,06	0,04	0,23
S											1,00	0,02	0,03	-0,05	0,02
K <sub>2</sub> O												1,00	0,00	0,04	-0,31
Дисперсия факторов, %													42,00	18,70	9,30

Примечание. Коэффициенты корреляции и факторные нагрузки отличаются от нуля с надежностью 0,95 при их абсолютных значениях.

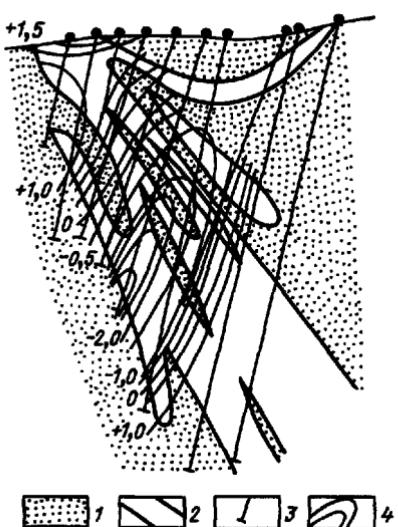


Рис. 40. Тренд поверхности первой главной компоненты третьего порядка, отражающий размещение железного оруденения по разрезу Нерюндинского месторождения:

1 — вмещающие породы; 2 — рудное тело;  
3 — буровые скважины; 4 — изолинии значений первой главной компоненты, усл. ед.

участки разреза с рудой достаточно высокого качества. Отрицательные значения главной компоненты характеризуют участки разреза с бедной рудой и частой перемежаемостью бедных рудных прослоев и пустых пород. Повышение значений главной компоненты свидетельствует об улучшении качества руд. Так, на рис. 40 четко прослеживается тенденция увеличения значений первой главной компоненты ниже скв. 21. Пробуренные буровые скв. 88 и 92, данные которых в расчетах не участвовали, подтвердили наличие руд с высоким содержанием железа.

График первых двух главных компонент (рис. 41) позволяет достаточно надежно разделять руды Нерюндинского месторождения по сортам.

*Q*-метод факторного анализа предназначен для исследования соотношений не между переменными, а между наблюдениями. Цели *Q*-метода факторного анализа сводятся к размещению последовательностей наблюдений в разумном порядке, т. е. к установлению связей между ними. Таким образом, цель *Q*-метода факторного анализа, по существу, совпадает с целью кластер-анализа, однако для его выполнения требуется значительно больше машинного времени.

Сущность *R*- и *Q*-факторного анализов изложена в работе Дж. Дэвиса, а с методикой их проведения можно ознакомиться в работах У. Крамбейна и Ф. Грейбила, Д. А. Родионова.

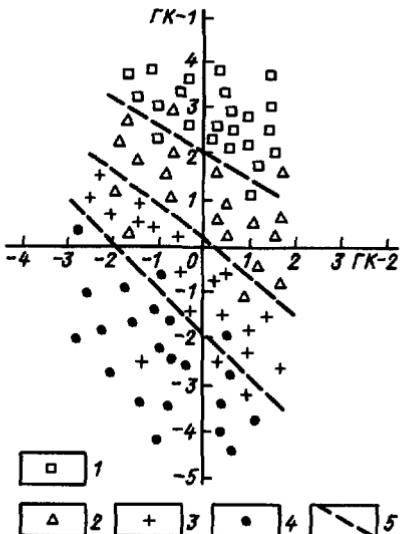


Рис. 41. Распределение типов железных руд по содержанию железа в зависимости от значений первой и второй главных компонент.

Содержание железа, в %: 1 — более 30; 2 — 25—30; 3 — 20—25; 4 — менее 20; 5 — граничные полосы различным типам руд

#### **4.11. ОБЛАСТЬ ПРИМЕНЕНИЯ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ В ГЕОЛОГИИ**

Возможности применения многомерных статистических моделей для изучения взаимозависимостей комплексов самых различных геологических признаков практически не ограничены для любой отрасли геологии. В палеонтологии они используются для статистического описания морфологических признаков ископаемых форм организмов и сопоставления их групп с литолого-фацальными разрезами осадочных пород, с целью оценки достоверности их стратиграфического положения (или установления групп руководящих ископаемых). Корреляционные методы парагенетического анализа химических элементов и минералов находят широкое применение в геохимии и минералогии. Различные методы многомерного описания самых различных физических свойств, химического и минерального состава осадочных и магматических пород используются в линтологии и петрографии для разделения их по фацальным или формационным признакам или для оценок их перспектив на выявление самых различных полезных ископаемых. С каждым годом все шире используются методы «распознавания образов» рудоносных территорий или месторождений полезных ископаемых, основанные на статистических описаниях сочетаний благоприятных элементов геологического строения, влияющих на концентрации полезных ископаемых. В настоящее время алгоритмы «распознавания образов», использующие самые различные статистические, логические и эвристические многомерные модели, реализуются в человеко-машинных информационно-прогнозирующих системах, нашедших широкое применение в геологоразведочной отрасли.

Многомерные статистические описания связей геологических переменных с последующими оценками степени их взаимозависимостей, используются в геологической практике с целью идентификации (отождествления), дискриминации (разделения), классификации (группирования) изучаемых объектов или в поисках наиболее информативных комбинаций признаков для решения прогнозных задач.

Задачи идентификации геологических объектов, например, оценка коллекторских свойств или газоносности пород по совокупности скважинно-геофизических характеристик, обычно выполняются с помощью моделей множественной регрессии.

В целях дискриминации геологических объектов на два заранее заданных класса, например, разделение кимберлитовых пород на алмазосный и неалмазосный типы, по данным их силикатных анализов, может быть использована модель линейной дискриминантной функции.

Классификация геологических объектов, например, иерархическое группирование парагенетических ассоциаций элементов метасоматически измененных пород или руд по данным их полных химических анализов, производится с помощью кластер-анализа, других методов многомерного корреляционного анализа или  $Q$ -метода факторного анализа.

Конечной целью большинства многомерных статистических методов является предсказание (прогнозирование) тех или иных свойств изучаемых геологических объектов.

Прогнозирование свойств геологических объектов, чаще всего выявление перспектив их рудоносности или оценка вероятных масштабов оруденения проводится с помощью алгоритмов «распознавания образов».

В зависимости от характера исходных данных и целей геологических исследований для составления этих алгоритмов используются самые различные многомерные модели. При этом, как правило, возникает проблема поиска наиболее информативных сочетаний признаков и сокращения размерности их пространства, что достигается с помощью метода главных компонент, *R*-метода факторного анализа или других логических и эвристических методов.

Возможности использования многомерных статистических моделей для целей решения геологических задач изучены в настоящее время далеко не полностью и несомненно имеют большое будущее.

#### Контрольные вопросы

1. Какие геологические условия благоприятствуют применению многомерных статистических моделей?
2. В чем различаются понятия парного и частного коэффициентов корреляции?
3. Какие зависимости можно оценить с помощью множественного коэффициента корреляции?
4. Какие статистические методы могут использоваться для выделения взаимосвязанных ассоциаций химических элементов?
5. В чем преимущество и недостатки методики графов?
6. Каковы принципы и области применения кластер-анализа?
7. В чем различие понятия «дendrogramma» и «дайдрограф»?
8. Для решения каких геологических задач эффективно использование множественных регрессионных моделей?
9. На чем основаны «методы распознавания образов» и каковы возможности их использования для решения геологических задач?
10. Какие многомерные статистические модели применяются для оценки параметров рудоносности территорий?
11. В чем заключаются основные сложности оценки информативности геологических признаков?
12. Для решения каких геологических задач могут быть использованы модели линейных дискриминантных функций?
13. В чем смысл использования метода главных компонент?
14. Какие виды факторного анализа основаны на положениях метода главных компонент?
15. Какие геологические задачи решаются с применением факторного анализа?
16. К чему сводятся задачи идентификации, дискриминации и классификации изучаемых геологических объектов и какие математические модели пригодны для их решения?

## ГЛАВА 5

### МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННЫХ ПЕРЕМЕННЫХ

---

При изучении строения земной коры геолога интересуют не только средние характеристики изменчивости и взаимосвязи наблюдаемых значений свойств горных пород, минеральных ассоциаций или полезных ископаемых, но также закономерности их пространственных

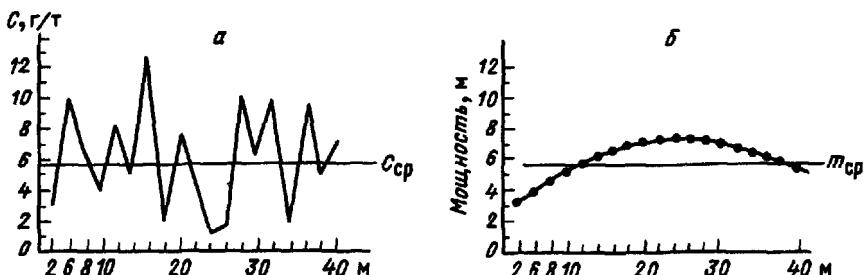


Рис. 42. Графики изменения параметров по золоторудной жиле:  
а—содержание золота; б—мощность жилы

изменений в исследуемых объемах недр. Статистические модели для этих целей непригодны, поскольку любой статистический показатель отражает лишь средний уровень изменчивости изучаемого свойства, независимо от пространственного размещения точек наблюдений, в то время, как закономерности их пространственного размещения могут оказаться принципиально различными, что иллюстрируется схематическими графиками на рис. 42.

К тому же, статистические характеристики обеспечивают объективные оценки уровня наблюдаемой изменчивости признака только в тех случаях, когда выборочные данные представляют собой совокупность независимых случайных величин. Для оценок реальных погрешностей изменчивости корреляционно связанных исходных данных требуется введение поправок за связь.

Для целей математического моделирования закономерностей пространственного размещения изучаемых свойств геологических образований их признаки рассматриваются не как случайные величины, а как пространственные переменные, обладающие рядом специфических характеристик: мерностью, областями существования и воздействия (определения).

Их совокупности образуют поля пространственных переменных, в пределах которых положение каждой переменной определяется координатами пространства.

Геометрические и аналитические методы моделирования геологических, геохимических, геофизических и других полей пространственных переменных способствуют объективному выделению и количественному описанию тенденций, наблюдаемых в изменении свойств исследуемых объектов, а в ряде случаев позволяют выявлять новые, ранее неизвестные закономерности. Для целей моделирования при этом используются результаты геологического картирования, геохимических и шлиховых съемок, геофизических наблюдений, геологоразведочных работ и т. д.

Пространственные закономерности изменения гранулометрического и минерального составов терригенных отложений позволяют установить направление сноса обломочного материала и восстановить палеогеографическую обстановку периода формирования толщ осадоч-

ных пород. По изменениям концентрации различных минералов выявляется зональность интрузивных массивов и месторождений полезных ископаемых. Пространственные закономерности изменения геофизических полей широко используются при геологическом ровании и поисках полезных ископаемых. Выявление закономерностей в изменении параметров рудных тел на ранних стадиях изучения позволяет более обоснованно оценивать не затронутые геологической разведкой фланги и глубокие горизонты месторождений, а также способствует правильному выбору размеров сети наблюдений при продолжении геологоразведочных работ. Математическое моделирование геохимических и геофизических полей позволяет более надежно выявлять аномалии, перспективные на обнаружение месторождений полезных ископаемых.

С помощью моделирования дискретных геологических полей устанавливаются закономерности расположения месторождений ирудопроявлений определенного генетического типа относительно интрузий, тектонических нарушений или других элементов геологического строения.

## 5.1. ГЕОЛОГИЧЕСКИЕ ОБЪЕКТЫ, КАК ПОЛЯ ПРОСТРАНСТВЕННЫХ ПЕРЕМЕННЫХ

Полем пространственной переменной называется область пространства, каждой точке которого поставлено в соответствие некоторое значение изучаемой переменной. В качестве геологического поля может рассматриваться область пространства, при этом каждому элементу последнего соответствует определенное значение изучаемого геологического признака.

В зависимости от природы моделируемых признаков различают геофизические, геохимические, минералогические, морфометрические и другие геологические поля, которые по размерности изучаемого пространства подразделяются на одномерные, двумерные, трехмерные и многомерные.

*Непрерывные и дискретные геологические пространственные переменные.* По характеру распространения (областям существования) в земной коре геологические пространственные переменные разделяются на непрерывные и дискретные.

Непрерывные пространственные переменные выражают свойства горных пород, минеральных ассоциаций или полезных ископаемых, проявленные в любой точке поля, т. е. на всей площади (во всем объеме) исследуемого блока земной коры или геологического тела. К числу этих переменных относятся концентрации химических элементов в горных породах, их физические свойства, мощность изучаемых геологических тел и многие другие свойства пород и руд.

К числу дискретных пространственных переменных относятся пространственно ограниченные геологические образования, области существования (размеры) которых пренебрежимо малы по сравнению с исследуемыми площадями или объемами недр. Они представлены

геологическими телами специфического состава (например, отдельными разновидностями пород), месторождениями полезных ископаемых, вкрапленниками отдельных минералов или минеральных агрегатов в породах и др. карти-

*Скалярные и векторные поля.* По признакам мерности пространственных переменных различают скалярные и векторные геологические поля. Большинство обычно изучаемых геологических переменных относится к скалярным величинам, для задания которых достаточно знать их модуль и знак. Совокупности этих переменных образуют скалярные геологические поля.

Реже в геологической практике используются векторные пространственные переменные, для задания которых в каждой точке пространства необходимо знать не только модуль, но и направление переменной. Векторные случайные поля могут моделироваться как векторы, ориентированные в реальном дву- или трехмерном пространстве (например, магнитные поля) или как комплексы различных скалярных переменных (например, по содержанию нескольких химических элементов в каждой точке). Многие скалярные поля могут быть преобразованы в векторные, если изучать не исходные величины, а их производные, т. е. градиенты геологических полей.

## 5.2. ЭЛЕМЕНТЫ НЕОДНОРОДНОСТИ, ИЗМЕНЧИВОСТЬ И АНИЗОТРОПИЯ ГЕОЛОГИЧЕСКИХ ПОЛЕЙ

Изучение геологических полей сводится, в конечном счете, к выделению и описанию слагающих их неоднородностей по тем или иным признакам. При сборе исходных данных для выявления и описания неоднородностей возможны два наиболее типичных случая:

1) областями наблюдений могут служить реально существующие неоднородности, т. е. ограниченные по размерам разнородные природные образования, четко обособленные от окружающей среды. В этом сравнительно редком случае, моделируются дискретные геологические поля;

2) значительно чаще из-за отсутствия природных границ реальных неоднородностей геологических образований, в их пределах приходится выделять серию искусственных областей наблюдений, например, путем выборки (или замера) части объема (площади) породы при ее опробовании. Результаты наблюдений по таким искусственно создаваемым областям служат основой моделирования непрерывных геологических полей, а сами области наблюдений называются областями воздействия эксперимента. В зависимости от условий эксперимента, целей и задач исследований, размеры и формы областей воздействия могут сознательно изменяться.

Очевидно, что в первом случае моделируются реальные неоднородности геологических образований, а закономерности их пространственного размещения выявляются практически однозначно, в то время как во втором случае моделируются результаты наблюдений по областям воздействия эксперимента, формы и размеры которых

зависят от намерений (или возможностей) исследователей. Поэтому при моделировании непрерывных геологических полей, выявляемые закономерности далеко не полностью отражают реально существующие (истинные) неоднородности геологического строения изучаемых территорий, участков или блоков.

Если размеры областей наблюдений существенно превышают размеры истинных элементов неоднородности изучаемых геологических массивов, то по результатам моделирования создается представление о примерной однородности его геологического поля. При размерах областей наблюдений превышающих размеры истинных элементов неоднородности только на один-два порядка или сопоставимых с ними, результаты наблюдений приобретают случайный характер, в результате чего в моделируемых геологических полях возникает неконтролируемый случайный эффект. Если же размеры областей наблюдений заведомо меньше истинных элементов неоднородности, результаты наблюдений приобретают регулярный характер, а плавные изменения параметров геологического поля отражают влияние структурных особенностей изучаемого массива или крупных его участков.

Неоднородности первого рода придают результатам наблюдений случайный характер, формируя случайную составляющую изменчивости изучаемого признака. Неоднородности второго рода определяют закономерную составляющую пространственной изменчивости признака, выявление которой — основная цель моделирования непрерывных геологических полей, так как только ее значения могут быть воспроизведены и не должны заметно изменяться при небольших смещениях сети наблюдений.

При случайной изменчивости значения признака, наблюдаемые в различных (в том числе в смежных) пунктах, не зависят друг от друга. Они не зависят также от расстояния между пунктами наблюдений и имеют по всем направлениям характер случайных неупорядоченных колебаний. Положительные и отрицательные знаки приращения величины признака по какому-либо направлению часто сменяются в каждом пункте и весьма редко сохраняют постоянство более чем в трех-пяти соседних пунктах.

Для неслучайной (закономерной) изменчивости характерно, что значения признака во всех точках наблюдений функционально связаны между собой. Определяющим свойством неслучайной изменчивости является наличие плавных колебаний значений признака и постоянство знака его приращения на протяжении полупериода каждого колебания.

При моделировании непрерывных геологических полей наблюдаемые признаки содержат в себе как закономерную, так и случайную составляющую. Однако их соотношения зависят в каждом конкретном случае как от природных свойств изучаемых геологических образований, так и от геометрических особенностей принятых областей воздействия эксперимента, а также от густоты сети наблюдений и ряда других факторов.

Моделируемые непрерывные геологические поля часто обладают анизотропными свойствами. Они проявляются в том, что характеристики изменчивости поля зависят от той части пространства и того

направления, по которым они определяются. Признаками анизотропии непрерывных геологических полей могут служить отчетливо вытянутые контуры элементов неоднородности различных размеров, «ступени», проявленные горизонталями модулей градиентов признака.

### **5.3. ГЕОЛОГИЧЕСКИЕ, МЕТОДИЧЕСКИЕ И ТЕХНИЧЕСКИЕ ФАКТОРЫ, ОБУСЛОВЛИВАЮЩИЕ ПОЯВЛЕНИЕ ЗАКОНОМЕРНОЙ И СЛУЧАЙНОЙ СОСТАВЛЯЮЩИХ НАБЛЮДАЕМОЙ ИЗМЕНЧИВОСТИ**

При моделировании геологических полей возможны практически любые соотношения закономерной и случайной составляющих наблюдаемой изменчивости изучаемого признака. Их конкретные значения и долевые соотношения между ними зависят от сочетания ряда перечисленных выше геологических, методических и технических факторов.

Простое геологическое строение изучаемых образований способствует появлению и преобразованию закономерной составляющей изменчивости моделируемых полей. Строение принято считать простым, когда структурные элементы моделируемого объекта несопоставимо малы (неоднородности структур и микротекстур минеральных агрегатов) либо несопоставимо велики (неоднородности крупных геологических тел или блоков земной коры) по сравнению с технически возможными линейными размерами областей воздействия эксперимента. В первом случае (например, при моделировании свойств массивов равномернозернистых магматических или осадочных пород) пространственная изменчивость многих признаков практически отсутствует, а во втором (например, при моделировании степени изменений пород, в зонах контактового метаморфизма крупных plutонов) преобладает закономерная изменчивость признака, которая отражает плавное и постепенное уменьшение степени изменения вмещающих пород по мере удаления от контакта с интрузивом.

Сложное геологическое строение изучаемых образований, структурные элементы которых сопоставимы по размерам с линейными размерами объемов воздействия эксперимента приводит к появлению случайной составляющей изменчивости моделируемых признаков. Такое строение характерно для зон брекчирования и мигматизации, пегматитовых тел, а также для многих природных скоплений полезных ископаемых эпигенетического происхождения. В общем случае доля случайной изменчивости наблюдаемого признака возрастает с усложнением геологического строения объекта, что проявляется в увеличении числа его разномасштабных структурных уровней.

Важнейшими методическими факторами, определяющими соотношения случайной и закономерной составляющих изменчивости признаков геологических полей, являются размеры областей воздействия эксперимента и густота сети наблюдений. Чем крупнее размеры областей воздействия эксперимента и меньше расстояния между смежными наблюдениями, тем больше (при прочих равных условиях) доли наблюдаемой закономерной изменчивости признака. Однако

стущение сети наблюдений до размеров меньших по сравнению с линейными размерами областей воздействия эксперимента практически нецелесообразно. Из сказанного ранее очевидно, что решающее влияние на проявления наблюдаемых закономерностей и случайной изменчивости признака оказывают соотношения размеров истинных элементов неоднородности геологических объектов с размерами выбранных областей влияния эксперимента. Поэтому, используя априорные представления о вероятных уровнях строения и размерах истинных неоднородностей моделируемого объекта, возможно прогнозирование ожидаемых соотношений между случайной и закономерной составляющими изменчивости признака для выбранных размеров и форм областей влияния эксперимента.

#### 5.4. ФОН, АНОМАЛИИ И ПОВЕРХНОСТЬ ТРЕНДА

Наиболее распространенной моделью непрерывного скалярного геологического поля является модель аддитивного случайного поля, когда на плоскости с координатами  $x$  и  $y$  задаются значения непрерывной скалярной переменной  $\hat{u} = f(x, y)$ , значения которой используются для описания аддитивного скалярного поля  $\hat{u} = f(x, y) + \varepsilon$ , где  $f(x, y) = \hat{u}$  — функция координат;  $\varepsilon$  — случайная переменная.

В задачу моделирования поля входит оценка функции  $f(x, y)$  в известных предположениях относительно  $\varepsilon$  и описание случайной части  $\varepsilon$  при некоторых предположениях относительно  $f(x, y)$ . Главной задачей изучения пространственных закономерностей является описание неслучайной (закономерной) компоненты поля, отражающей уровень его значений, характерный для отдельных частей изучаемой территории.

*Неслучайная компонента, характеризующая основную часть моделируемого геологического поля, называется его фоном.* Фоновая часть поля выявляет область относительно повышенных или пониженных значений изучаемого признака и несет в себе полезную геологическую информацию о природе изучаемого геологического объекта. Для выделения фона необходима генерализация основных свойств поля с подавлением более или менее существенных частных отклонений. В каждом конкретном случае отклонения от фона рассматриваются как аномальные.

Методы выделения фоновой части геологического поля с разделением неслучайной и случайной составляющих изучаемых признаков по эмпирическим данным получили название *анализа поверхностей тренда*.

В геологической практике для целей тренд-анализа используют два разных методических подхода: 1) склонение исходных данных скользящими статистическими окнами; 2) аппроксимация полей единой функцией пространственных координат (ортогональными полиномами и др.). Методы скользящих средних более универсальны и обеспечивают лучшие оценки средних параметров пространственно ограниченных участков геологических полей по сравнению с методами

полиномиального тренд-анализа исходных данных, которые используются преимущественно для выявления региональных геологических закономерностей.

Относительный характер закономерной и случайной составляющих наблюдаемой изменчивости признаков оказывает заметное влияние на результаты тренд-анализа геологических полей. В связи с этим, в зависимости от масштабов, целей, задач и условий исследований, под их фонами могут подразумеваться поверхности тренда различной степени плавности, а под аномалиями — любые отклонения от фона, превышающие заданный условный уровень.

## 5.5. ГЕОМЕТРИЧЕСКИЕ МЕТОДЫ ВЫЯВЛЕНИЯ ЗАКОНОМЕРНЫХ СОСТАВЛЯЮЩИХ ПРИЗНАКОВ

Геометрические методы, предложенные П. К. Соболевским для описания закономерностей пространственных изменений признаков, как и статистические методы не обеспечивали объективной оценки их изменчивости. Если статистические методы не учитывали влияния плавных, закономерных изменений, то геометрические методы игнорировали влияния многочисленных случайных отклонений по отдельным пунктам наблюдений.

Разработанные П. К. Соболевским методы геометризации земных недр включают комплекс графоаналитических операций с численными характеристиками геологических объектов. Исследуемую часть пространства недр П. К. Соболевский рассматривал как «геохимическое поле», в пределах которого совокупность наблюдаемых форм, свойств и процессов может рассматриваться под углом зрения единства ее геологического генезиса.

Геометрическая модель П. К. Соболевского основывается на представлении о функциональных связях наблюдаемых геологических параметров с пространственными координатами. Им сформулированы принципы геометрического моделирования геологических полей, согласно которым:

- значения изучаемого признака в точках наблюдений могут рассматриваться, как функция пространственных координат  $a = f(x, y, z)$ ;
- функция  $a = f(x, y, z)$  должна удовлетворять условиям однозначности, непрерывности и плавности.

Функция называется *однозначной*, если она пересекается с нормально восстановленной из точки плоскости или линии проецирования сечения поля один раз.

Функция называется *непрерывной*, если незначительному изменению координат пространства соответствует столь же незначительное изменение свойства.

Функция называется *плавной*, если она дифференцируема в каждой точке пространства.

Геометрическое моделирование числовых значений геологических параметров производится графически с помощью изолиний. На планах или разрезах изолинии образуют топографические поверхности,

которые рассматриваются как топофункции. П. К. Соболевским была показана возможность математических действий с топофункциями и разработаны оригинальные математические операции с ними, включающие любые арифметические и алгебраические действия. Среди них при геометрическом моделировании полей наиболее широко распространены сложение (вычитание), умножение (деление) и дифференцирование (интегрирование) аппроксимирующих топофункций.

Сложение (вычитание) функций используется для оценок мощностей геологических тел, построения аддитивных геохимических ореолов различных элементов и др.

Умножение (деление) функций применяется при расчетах комплексных геохимических и геофизических полей и при оценках продуктивности скоплений полезных ископаемых.

Дифференцирование функций используется для определения скоростей изменения признаков, т. е. для расчета его градиентов, а интегрирование — при расчетах средних значений параметров.

Методы геометрического моделирования широко применяются в геологической практике, так как они весьма наглядны и способствуют пространственному восприятию изучаемых закономерностей. Однако в них отражаются особенности только неслучайной составляющей изменчивости признаков и полностью игнорируются проявления многочисленных случайных отклонений. Поэтому методы геометрического моделирования не позволяют оценивать достоверность установленных пространственных закономерностей и вычисляемых средних значений.

## 5.6. СПОСОБЫ СГЛАЖИВАНИЯ СЛУЧАЙНЫХ ПОЛЕЙ

Совокупности многих наблюденных признаков часто не удовлетворяют требованиям непрерывности и плавности. В таких случаях П. К. Соболевский рекомендует сглаживать их скользящими статистическими окнами, размеры которых должны подбираться эмпирически, в зависимости от характера исходных данных. Позже способ скользящего сглаживания геологоразведочных данных был использован П. Л. Каллистовым для разделения их наблюдаемой изменчивости на две составляющие — случайную и закономерную.

Примером почти чистой случайной изменчивости могут служить беспорядочные колебания содержаний золота по данным опробования одного из жильных месторождений (см. рис. 42, а).

Неслучайная изменчивость свойств реальных геологических тел полезных ископаемых обычно настолько сложна, что выразить ее аналитически как функцию координат пространства не удается. По достаточно большому числу измеренных значений она может быть выражена геометрически в виде профилей или топографических поверхностей.

В качестве наиболее простых обобщающих характеристик неслучайной координированной изменчивости П. Л. Каллистов предложил использовать:

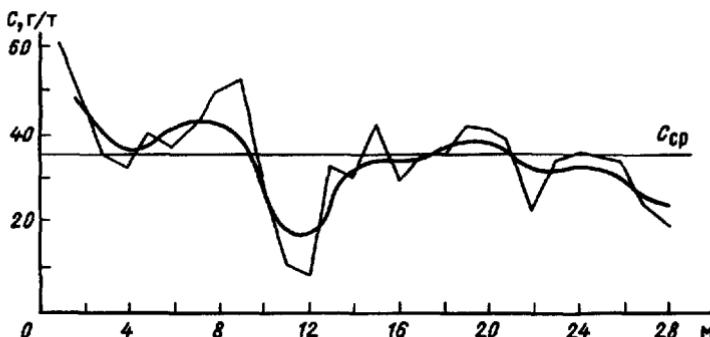


Рис. 43. Изменение содержания золота по штреку. По П. Л. Каллистову

- 1) длины полупериодов колебаний значений изучаемого признака (расстояния между смежными минимумами или максимумами);
- 2) амплитуды колебаний и градиенты изменения значений признака (т. е. скорости изменения его значений).

Обычно обе составляющие изменчивости признака наблюдаются в природных геологических телах совместно. Для разделения общей изменчивости признака на ее случайную и закономерную составляющие П. Л. Каллистовым было предложено сглаживать эмпирические данные по разведочному профилю способом «скользящего окна». В результате такого сглаживания закономерная изменчивость выявляется в виде плавной кривой, которая может быть описана функцией синусоидального типа, а для характеристики случайной изменчивости используется коэффициент вариации, вычисленный через отклонения каждого частного значения от скользящей средней, т. е. уровня неслучайной изменчивости (рис. 43).

Для решения аналогичной задачи не по профилям, а по площади, по точкам среднеарифметических значений признака, рассчитанным для центров разведочных ячеек путем двух- или трехкратного сглаживания, проводятся изолинии (рис. 44), а дисперсия случайной составляющей рассчитывается через отклонения фактических значений от соответствующих изолиний. Способы сглаживания исходных данных методом скользящего окна представляют собой не строго математический, а лишь описательный прием. При сглаживании ряда в него всегда вносится некоторая ложная взаимозависимость между отдельными членами, обусловленная техникой процесса сглаживания. Характер скользящей средней зависит от числа проб в скользящем окне, а также от числа последовательных сглаживаний. Дисперсия случайной составляющей, вычисление при последовательном сглаживании от скользящих средних, зависит не только от свойства исследуемого ряда, но также от способа сглаживания. Поэтому без учета условий эксперимента они не могут использоваться в качестве объективных критериев оценки доли случайной составляющей.

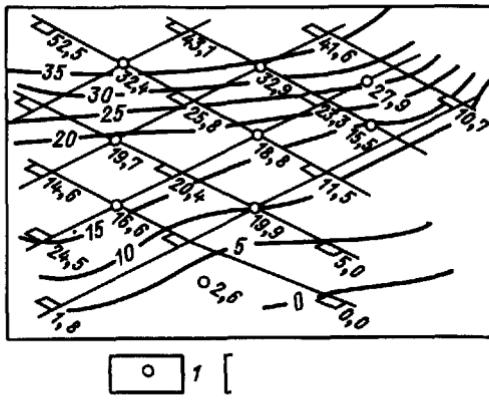


Рис. 44. Изменение мощности (в м) тела полезного ископаемого по площади. По А. А. Петрову:  
1—скважины; 2—шурфы

Общая формула простейшего (невзвешенного) способа сглаживания исходных данных может быть записана как

$$\Phi(x, y) = \hat{u} = \frac{1}{n} \sum_{i=1}^n u_i, \quad (28)$$

где  $x$ ;  $y$  — координаты центра скользящего окна;  $n$  — число точек в сглаживающем окне.

Поскольку при увеличении числа точек в скользящем окне получаемая скользящая средняя все больше маскирует и сдвигает наблюденные точки максимальных и минимальных значений признака, в практике широко используются способы взвешенного сглаживания исходных данных, которые обеспечивают лучшую аппроксимацию эмпирических кривых или поверхностей. В общем виде формула взвешенного сглаживания записывается как

$$\Phi(x, y) = u = g \sum K(\xi, \eta) u(\xi, \eta),$$

где  $x, y$  — координаты центра площадки трансформации;  $K(\xi, \eta)$  — коэффициенты, зависящие от координат  $\xi, \eta$  точки задания поля в пределах площадки трансформации;  $U(\xi, \eta)$  — наблюденные значения признаков в соответствующих точках площадки трансформации;  $\vartheta$  — нормировочный множитель.

Так, например, В. Ф. Мягков рекомендует формулу сглаживания по пяти последовательным точкам ряда:  $\hat{u}_j = 0,0625(u_{i-2} + 4u_{i-1} + 6u_i + 4u_{i+1} + u_{i+2})$ .

В зарубежной геологической практике широко распространены сглаживающие преобразования Вулхауса, Спенсера, Шеппарда и др. В формуле Вулхауса использованы результаты 15 последовательных наблюдений, по семи наблюдений в каждую сторону от величины  $u_0$ , которую необходимо сгладить:

$$u_0^1 = \frac{1}{125} [25u_0 + 24(u_1 + u_{-1}) + 21(u_2 + u_{-2}) + 7(u_3 + u_{-3}) + \\ + 3(u_4 + u_{-4}) - 2(u_6 + u_{-6}) - 3(u_7 + u_{-7})].$$

В формуле Спенсера используется по 10 значений с каждой стороны:

$$u_0^1 = \frac{1}{350} [60u_0 + 57(u_1 + u_{-1}) + 47(u_2 + u_{-2}) + 33(u_3 + u_{-3}) + \\ + 18(u_4 + u_{-4}) + 6(u_5 + u_{-5}) - 2(u_6 + u_{-6}) - 5(u_7 + u_{-7}) - \\ - 5(u_8 + u_{-8}) - 3(u_9 + u_{-9}) - (u_{10} + u_{-10})].$$

При комплексном изучении геологических полей, когда одновременно фиксируется несколько свойств, скрытые закономерности могут быть выявлены с помощью способа «скользящей корреляции». Он основан на расчете коэффициентов корреляции между значениями исследуемых свойств в пределах площадок трансформации. Карта коэффициентов корреляции между содержаниями определенных минералов в обломочных толщах позволяет наметить направление поступления обломочного материала, а графики или поля изменения коэффициентов корреляции между различными физическими свойствами помогают интерпретировать результаты геофизических работ при геологическом картировании.

*Пример.* Мolibденовое оруднение скарнового и грейзенового типов приурочено к лейкократовым гранитам, над которыми отмечается незначительное повышение удельного электрического сопротивления ( $\rho_t$ ) и понижение уровня гравитационного поля ( $\Delta g$ ). Однако выделение рудовмещающего интрузива по какому-либо признаку оказывается затруднительным (рис. 45). Между значениями  $\rho_t$  и  $\Delta g$  были вычислены коэффициенты корреляции по участкам профиля протяженностью 500 м. По графику изменения коэффициентов корреляции рудоносный интрузив четко выделяется пониженными значениями  $r$ .

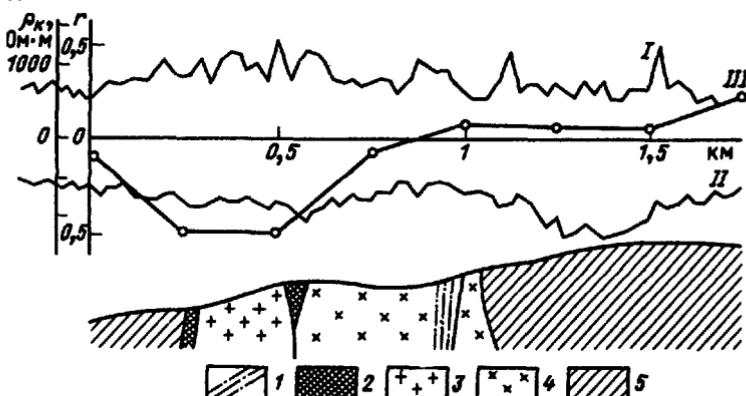


Рис. 45. График геофизических наблюдений и коэффициента корреляции над рудопроявлением Чемуртай. По Л. Л. Ляхову и Г. А. Соловьеву:

I—удельное электрическое сопротивление  $\rho_t$ ; II—уровень гравитационного поля  $\Delta g$ ; III—коэффициент корреляции. 1—зона разлома; 2—скарны; 3—лейкократовые граниты; 4—микродиориты и габбродиориты; 5—песчаники и известняки

Способы, основанные на сглаживании или преобразовании исходных данных, просты и наглядны, но обладают существенными недостатками:

— они не дают объективных количественных критериев для оценки значимости выявленных закономерностей. Вопрос о наличии закономерностей решается по виду сглаженных поверхностей или графиков. При этом «закономерности» могут быть отмечены даже в заведомо случайных данных, так как процесс сглаживания обусловливает корреляцию между соседними значениями;

— результаты сглаживания существенно изменяются в зависимости от вида преобразования и размера площадки трансформации. Выбрать оптимальный способ преобразования можно либо путем перебора множества вариантов, либо с помощью привлечения дополнительной информации о характере изменчивости изучаемого свойства;

— любое преобразование обладает селективными свойствами только по отношению к закономерностям, близким по размеру с площадкой трансформации. Поэтому для выделения в наблюдаемой изменчивости закономерностей разного масштаба необходимо использовать различные варианты преобразований.

## 5.7. СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОВЕРКИ ГИПОТЕЗ О НАЛИЧИИ ТРЕНДА

Для решения вопроса о наличии каких-либо пространственных (или временных) закономерностей наблюденной изменчивости свойств геологического объекта могут быть использованы два наиболее простых и экспрессивных способа — смены знаков и числа скачков.

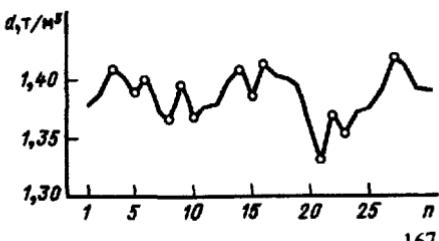
Решение задач этого типа основано на сравнении свойств наблюдаемых полей или упорядоченных последовательностей со свойствами или последовательностями, в которых закономерная составляющая (тренд) заведомо отсутствует.

Точкой смены знака в упорядоченной последовательности называется такой элемент последовательности, в котором знак приращения изменяется на противоположный — убывание на возрастание или наоборот (рис. 46).

Число точек смены знака в случайных последовательностях зависит только от общего количества элементов последовательности  $N$ . Если

Рис. 46. Изменение объемной массы  $d$  угля по простирианию пласта в пределах шахтного поля.

Кружком обозначены точки смены знака.  $n$  — расстояние между скважинами



$N > 10$ , статистическое распределение числа точек смены знака близко к нормальному с математическим ожиданием

$$M(t) = \frac{2N-4}{3} \quad (29)$$

и дисперсией

$$\sigma^2(t) = \frac{16N-29}{90}.$$

Это дает возможность использовать при статистической проверке гипотезы о наличии тренда таблицу функции нормального распределения. Проверка гипотезы основана на сравнении фактического значения числа точек смены знака  $t$ , полученного по исследуемому графику, с теоретическим его значением  $M(t)$ , рассчитанным по формуле (29). Поскольку в таблицах приводятся значения нормированной функции нормального распределения, разницу между фактическим и теоретическим числом точек смены следует разделить на  $\sqrt{\sigma^2(t)}$  для получения вероятностного критерия

$$Z = \frac{t - M(t)}{\sqrt{\sigma^2(t)}}.$$

В неслучайной последовательности значения  $t$  и  $M(t)$  не должны существенно отличаться. Следовательно, вероятность больших по абсолютной величине значений критерия  $Z$  будет мала. Так как значения критерия  $Z$  распределены нормально с параметрами 0 и 1, то по его величине с помощью таблиц нормированного нормального распределения можно оценить вероятность полученного по исследуемому ряду отклонения фактического числа точек смены знака от теоретического при условии, что ряд случаен. Если эта вероятность мала (например, меньше 0,05), то гипотеза о случайному характере исследуемого ряда отвергается и считается, что он имеет тренд.

*Пример.* Проверим гипотезу о наличии тренда в изменении объемной массы угля по продольному профилю из 30 буровых скважин, ориентированному по простиранию угольного пласта (см. рис. 46). Математическое ожидание числа точек смены знака для  $N=30$  будет

$$M(t) = \frac{2 \cdot 30 - 4}{3} \approx 18,7,$$

а дисперсия  $\sigma^2(t) = \frac{16 \cdot 30 - 29}{90} \approx 5,0$ . Фактическое число точек смены знака по профилю 13, следовательно, значение критерия  $Z$  составит

$$Z = \frac{13 - 18,7}{\sqrt{5}} \approx -2,55,$$

что соответствует вероятности 0,0054. Вероятность полученного отклонения фактического числа точек смены знака от теоретического для случайного ряда настолько мала, что гипотезу о случайному характере изменения объемной массы угля по простиранию угольного пласта следует отвергнуть.

Способ проверки гипотезы о наличии тренда по количеству скачков используется в случае, когда упорядоченная последователь-

ность состоит из двух типов элементов, которые условно можно обозначить знаками плюс (+) и минус (-). Скачком называется интервал последовательности, в пределах которого наблюдаются элементы только одного типа. Например, последовательность  $+ + - + - - - + -$  может быть разделена на шесть скачков  $(++) (-) (++) (- - -) (+) (-)$ .

Последовательности такого типа могут быть получены путем разделения всех значений исследуемого свойства на две группы по их отношению к медианному значению. Все значения больше медианного обозначаются знаком (+), а меньше — знаком (-). Число скачков в случайных последовательностях зависит от количества элементов первого ( $n_1$ ) и второго ( $n_2$ ) типа. Статистическое распределение количества скачков ( $u$ ) в случайных последовательностях асимптотически нормальное с математическим ожиданием

$$M(u) = \frac{2n_1 n_2}{n_1 + n_2} + 1,$$

и дисперсией

$$\sigma^2(u) = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}.$$

Как и в предыдущем способе, фактическое значение числа скачков ( $u$ ) сравнивается с теоретическим по исследуемому ряду  $M(u)$  по критерию

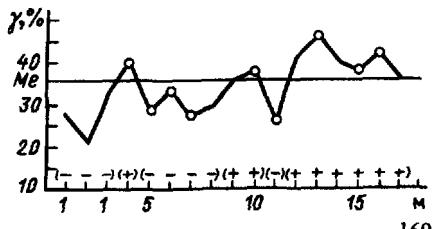
$$Z = \frac{u - M(u)}{\sqrt{\sigma^2(u)}}.$$

С помощью таблиц нормального распределения определяется вероятность полученного значения  $Z$  в случайной последовательности. Если эта вероятность мала, гипотеза о том, что исследуемая последовательность является случайной, отвергается и считается, что она имеет тренд.

*Пример.* Проверим гипотезу о случайном характере изменения общей пористости четвертичных игнимбритов андезитового состава по профилю вкрест простириания потока вулкана (рис. 47). По способу точек смены знака эта гипотеза не отвергается, так как фактическое значение величины  $t$  точно совпадает с математическим ожиданием для случайного ряда значений

$$M(t) = \frac{2 \cdot 17 - 4}{3} = 10.$$

Проведя на профиле линию, соответствующую медианному значению ( $M_e = 36\%$ ), разделим наблюдаемые значения на два класса. При этом количество



значений, меньших медианного —  $n_1 = 9$ , больших медианного —  $n_2 = 8$ , а количество  $u = 6$ .

Вычислим математическое ожидание и дисперсию количества скачков для случайного ряда при  $n_1 = 9$  и  $n_2 = 8$ :

$$M(u) = \frac{2 \cdot 9 \cdot 8}{9 + 8} + 1 \approx 9,5;$$

$$\sigma^2(u) = \frac{2 \cdot 9 \cdot 8 (2 \cdot 9 \cdot 8 - 9 - 8)}{(9 + 8)^2 (9 + 8 - 1)} \approx 4.$$

Значение критерия  $Z$  при этом составит

$$\frac{6 - 9,5}{4} = -1,75,$$

что соответствует вероятности 0,04. Таким образом, при доверительной вероятности 0,95 гипотезу о случайному изменении пористости исследуемых пород по данному направлению следует отвергнуть. Тренд в данном случае устанавливается визуально по тенденции к увеличению пористости по направлению вулканического потока.

Каждый из рассмотренных способов применим для выяснения закономерностей определенного типа. Способом «смены знака» лучше улавливаются локальные закономерности, проявленные плавными изменениями исследуемых свойств, в то время, как способом «скаков» отчетливее устанавливаются общие тенденции, присущие всему исследуемому ряду. Поэтому для принятия гипотезы о наличии тренда достаточно, чтобы она подтверждалась хотя бы одним из этих способов.

Оба способа можно использовать и для проверки гипотезы о характере изменения свойств геологических образований в двумерном пространстве, т. е. по площади. Для этого необходимо построить графики изменения свойств по различным направлениям.

## 5.8. АППРОКСИМАЦИЯ ПОВЕРХНОСТЕЙ ТРЕНДА ПОЛИНОМАМИ И АНАЛИЗ ОСТАТКОВ

Выделение региональных закономерностей путем аппроксимации эмпирических данных функцией координат пространства связано с довольно сложными вычислениями, обычно требующими применения ЭВМ. В качестве аппроксимирующих функций используются ортогональные полиномы различных степеней, уравнение Лапласа, тригонометрические полиномы и др.

Ортогональные полиномы обычно применяются в случае равномерной прямоугольной сети наблюдений. При этом тренд определяется как линейная функция географических координат, построенная по совокупности наблюдений таким образом, что сумма квадратов отклонений значений признака от плоскости тренда минимальна. Такая модель представляет собой вариант статистического метода множественной регрессии, в котором функция  $\Phi(x, y) = \hat{u}$ , описывающая поверхность тренда, рассматривается как  $\hat{u} = \beta_0 + \beta_1 x + \beta_2 y$  (где  $x$  и  $y$  — координаты пространства;  $\beta_0$ ,  $\beta_1$  и  $\beta_2$  — полиномиальные

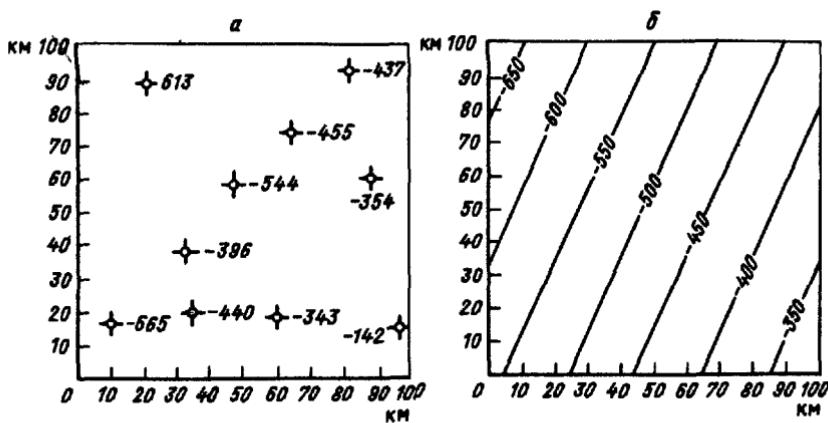


Рис. 48. Карта отметок подошвы меловых отложений в северо-восточной Африке. По Дж. Дэвису:

*a*—расположение скважин с замерами абсолютных отметок подошвы меловых отложений; *б*—поверхность тренда

коэффициенты). Для оценки трех указанных коэффициентов используются уравнения

$$\begin{aligned}\sum u &= \beta_0 n + \beta_1 \sum x + \beta_2 \sum y; \\ \sum xu &= \beta_0 \sum x + \beta_1 x^2 + \beta_2 \sum xy; \\ \sum yu &= \beta_0 \sum y + \beta_1 \sum xy + \beta_2 y^2,\end{aligned}$$

где  $n$ —число точек наблюдения;  $u$ —значения признака в точках наблюдений;  $x$  и  $y$ —координаты точек наблюдений.

Для решения уравнений они записываются в матричной форме:

$$\begin{bmatrix} n & \sum x & \sum y \\ \sum x & \sum x^2 & \sum xy \\ \sum y & \sum xy & \sum y^2 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum u \\ \sum xu \\ \sum yu \end{bmatrix}$$

и решаются относительно  $\beta_0$ ,  $\beta_1$  и  $\beta_2$ . Такой метод нахождения оценок биномодальных коэффициентов называется методом наименьших квадратов.

В качестве примера рассмотрим определения плоскости тренда отметок подошвы меловых отложений, экранирующих нефтяную толщу.

На рис. 48 показана схема расположения пробуренных скважин и условные координаты площадки, которые сведены в табл. 35.

Для построения плоскости тренда вычисляются:

$$\begin{aligned}\sum x &= 539; & \sum y &= 482; & \sum u &= -4579; \\ \sum x^2 &= 36\,934; & \sum y^2 &= 31\,692; & \sum xu &= -211\,098; \\ \sum xy &= 27\,030; & & & \sum yu &= -232\,342.\end{aligned}$$

Эти значения записываются в матричной форме:

$$\begin{bmatrix} 10 & 539 & 482 \\ 539 & 36\,943 & 27\,030 \\ 482 & 27\,030 & 31\,692 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -4579 \\ -211\,098 \\ -232\,342 \end{bmatrix},$$

и после решения матричного уравнения получаем:

$\beta_0 = -621,0$ ;  $\beta_1 = 4,8$ ;  $\beta_2 = -2,0$ . Подставляя эти значения в уравнение  $\hat{u} = \beta_0 + \beta_1 x + \beta_2 y$ , можно вычислить значения отметок плоскости тренда ( $\hat{u}$ ) для каждой скважины и разности ( $u - \hat{u}$ ), характеризующие составляющие случайной изменчивости гипсометрической поверхности.

Оценка степени приближения плотности тренда к наблюденным результатам, т. е. средняя изменчивость их отклонений может быть охарактеризована величиной  $SS_D = SS_T - SS_R = 21\ 463,5$ , где  $SS_T =$

$$= \sum u^2 - \frac{(\sum u)^2}{n} = 215\ 324,9; \quad SS_R = \sum \hat{u}^2 - \frac{(\sum \hat{u})^2}{n} = 193\ 861,4.$$

Таблица 35. Координаты скважин, абсолютные отметки подошвы меловых отложений и оценки плоскости тренда

№ п/п	Координаты		Абс. отметка $u$ , м	Отметка пло- скости тренда, $\hat{u}$ , м	Разность $(u - \hat{u})$ , м
	$x$ , км	$y$ , км			
1	10	17	-665	-606,6	-58,3
2	21	89	-613	-695,7	82,7
3	33	38	-586	-537,8	-48,1
4	35	20	-440	-492,8	52,8
5	47	58	-544	-510,2	-33,7
6	60	18	-343	-369,2	26,2
7	65	74	-455	-455,5	0,5
8	82	93	-437	-411,5	-25,4
9	89	60	-354	-313,0	-40,9
10	97	15	-142	-186,0	44,1

Таким образом, плоскость тренда учитывает 90% общей изменчивости:

$$K^2 = \frac{SS_R}{SS_T} \cdot 100\% = 90\%,$$

а коэффициент множественной корреляции составляет  $R = \sqrt{R^2} = 0,95$ . Полученные изолинии плоскости тренда показаны на рис. 48, б.

В приведенном примере задача аппроксимации поверхности тренда удовлетворительно решается с применением ортогональных полиномов первой степени. В случаях, когда доля случайной изменчивости остается все же достаточно большой после аппроксимации линейными функциями, для выявления закономерной изменчивости более высокого порядка применяются полиномы второй, третьей и реже — более высоких степеней.

Поверхность тренда второго порядка будет описываться уравнением  $\hat{u} = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$ , а число неизвестных полиномиальных коэффициентов увеличится до пяти. Для перехода к уравнению следующего более высокого порядка каждая географическая координата возводится в заданную степень и добавляются соответствующие смешанные произведения.

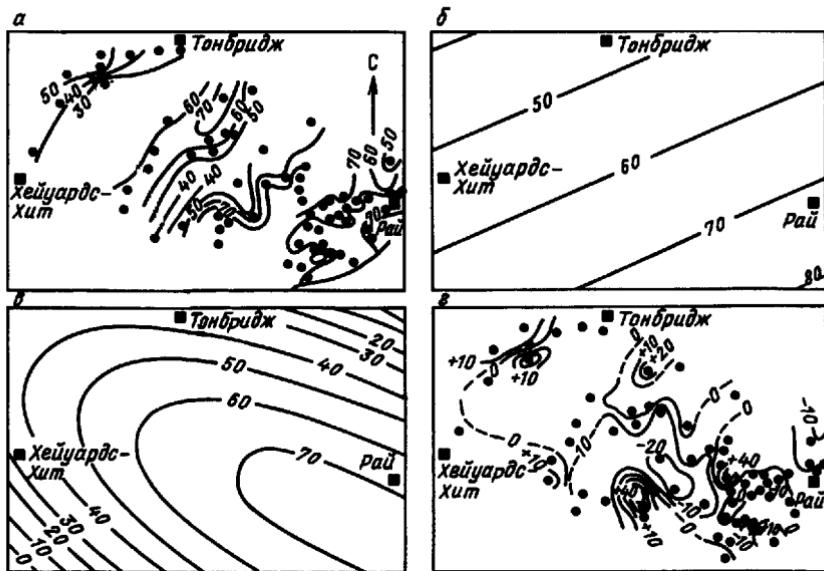


Рис. 49. Изменение размеров зерен циркона в галечниковом слое. По У. Крамбейну:

а—изолинии средних размеров зерен, мк; б—плоскость тренда; в—поверхность тренда второго порядка; г—карта отклонений второго порядка

В качестве примера использования полиномов первой и второй степени приводятся результаты изучения изменений средних размеров зерен циркона во фракциях тяжелых минералов из верхней части галечникового слоя Ашдаун на юго-востоке Англии (рис. 49, а). По характеру поверхности тренда первого порядка (см. рис. 49, б) можно заключить, что циркон на изучаемой территории перемещался главным образом с юго-востока на северо-запад. Однако основной поток циркона был, по-видимому, осложнен поступлениями некоторого количества этого минерала с северо-востока, на что указывает характер изолиний поверхности тренда второго порядка (см. рис. 49, в).

Выбор степени аппроксимирующего полинома и оценка значимости выявленных закономерностей могут осуществляться с помощью дисперсионного анализа (см. гл. 2). Для этого подсчитываются средние квадраты отклонений эмпирических значений исследуемого признака в точках замера от среднего арифметического и от аппроксимирующих поверхностей разного порядка, а также средние квадраты отклонений от среднего арифметического самих аппроксимирующих поверхностей. Значимость закономерностей, описываемых полиномами определенного порядка, проверяется с помощью критерия Фишера.

Расчет значений критерия Фишера для полиномов первой, второй и третьей степени, описывающих изменение размеров зерен циркона в галечниках слоя Ашдаун, приведен в табл. 36.

Высокие значения критерия Фишера для линейной и квадратичной модели указывают на реальность описываемых ими закономерностей.

В то же время для кубической поверхности  $F=0,15$ , что свидетельствует о несущественном влиянии компоненты третьего порядка.

Таблица 36. Дисперсионный анализ тренда размеров зерен циркона

Тип изменчивости	Сумма квадратов	Число степеней свободы	Средний квадрат	Критерий Фишера
Линейная модель	4982,01	2	2491,00	11,89
Отклонения от линейной модели	13 403,93	64	209,44	
Квадратичная модель	1266,19	3	422,06	2,12
Отклонения от квадратичной модели	12 137,74	61	198,98	
Кубическая модель	123,38	4	30,84	0,15
Отклонения от кубической модели	12 014,36	57	210,78	
Общая	18 385,94	66		

В геологической практике региональные закономерности обычно удовлетворительно описываются полиномами не выше третьей степени.

Аппроксимация тригонометрическими полиномами позволяет описывать закономерные периодические колебания свойств геологических объектов.

Из всех возможных аппроксимирующих функций выбирается та, которая точнее описывает имеющиеся данные и содержит наименьшее число параметров. Однако вид такой функции нельзя предсказать заранее, что существенно затрудняет практическое использование данных моделей. Аппроксимирующие функции координат пространства как модели геологических объектов имеют и некоторые другие недостатки:

- допускают существование нереальных значений изучаемых переменных, например, отрицательных значений содержания химических элементов в породах или мощностей рудных тел;

- не учитывают резких, скачкообразных изменений значений изучаемого свойства по геологическим границам, вследствие чего при моделировании рудных тел высокие содержания полезного компонента иногда распространяются на заведомо безрудные породы, например, на пострудные дайки;

- непригодны при использовании их для описания прерывистых объектов (например, рудных тел с прерывистым характером оруденения), так как происходит сглаживание исходных данных и искажается представление о степени прерывистости (увеличивается коэффициент рудоносности).

Выделение аномальных значений изучаемого свойства имеет в геологии большое практическое значение, так как с «аномалиями» часто связаны тела полезных ископаемых и другие наиболее интересные геологические объекты.

Задача отделения аномальных значений от фоновых не имеет строгого математического решения. В каждом конкретном случае используются различные эмпирические подходы, основанные на опыте изучения аналогичных объектов. Существенную помощь при выявле-

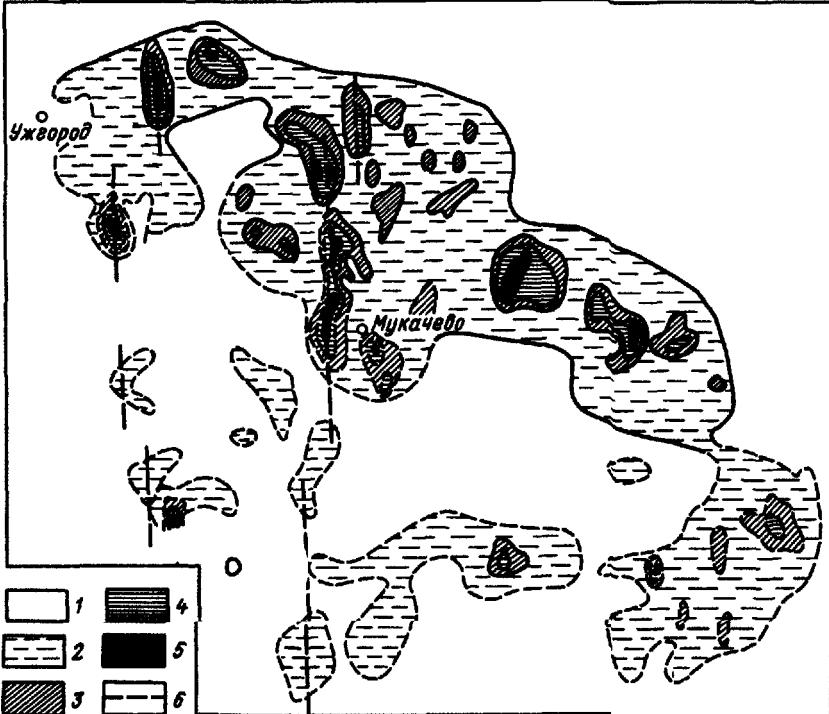


Рис. 50. Схема изменчивости аномалийного магнитного поля Закарпатья. По Г. И. Кругляковой:

1—невозмущенное поле; 2—5—возмущенность поля, % (2—0—20, 3—20—40, 4—40—60, 5—>60); 6—разломы древнего заложения

ния аномалий и установлении их геологической природы могут оказать карты «остатков» от тренда, которые строятся путем вычитания вычисленных значений тренда из наблюдаемых значений поля в каждой точке (см. рис. 49, 2). Подобные карты могут быть построены и для случайной составляющей после сглаживания исходных данных скользящим окном или с помощью преобразования какого-либо другого вида.

### 5.9. ТРАНСФОРМАЦИЯ ГЕОЛОГИЧЕСКИХ ПРОСТРАНСТВЕННЫХ ПЕРЕМЕННЫХ

При изучении изменчивости непрерывных геологических полей часто возникают задачи описания их структур с учетом не только тренда, но также влияния всей совокупности аномалий. В конечном итоге они сводятся к описанию статистических свойств локальных отклонений от фона с использованием частотных и амплитудных статистических характеристик. Конкретные характеристики степени изменчивости полей выбираются в зависимости от целей и задач исследований.

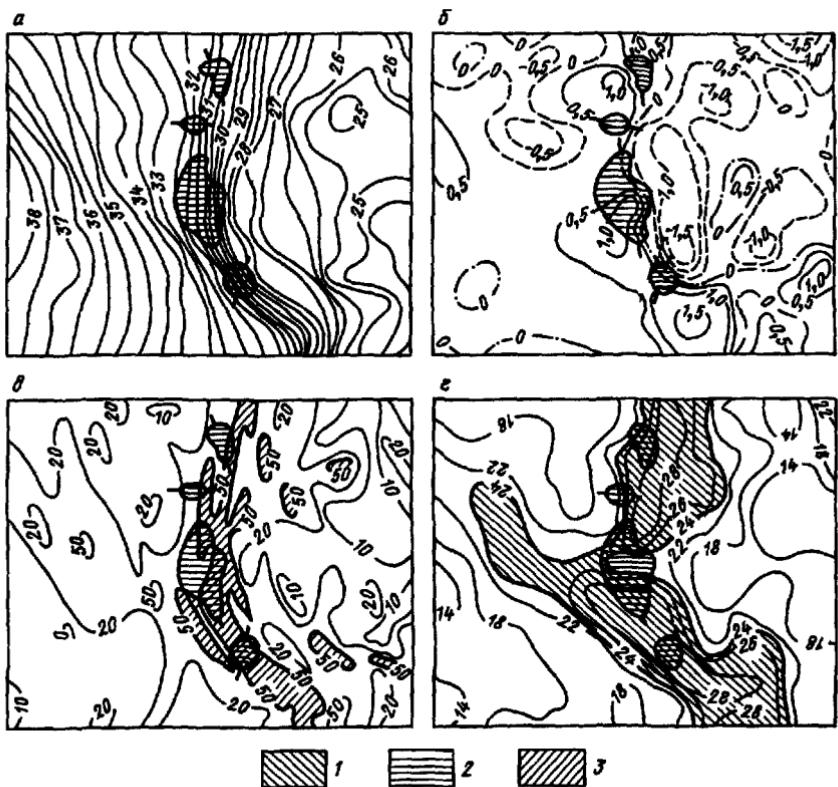


Рис. 51. Различные способы описания гравитационного поля. По Н. Н. Боровко:  
 а—гравитационное поле в изолиниях, б—локальные аномалии, в—горизонтальные градиенты, г—карта энтропии. 1, 3—аномальные значения горизонтальных градиентов и энтропии; 2—вторичные ореолы свинца

Они служат основой для построения новых трансформированных геологических переменных, наиболее ярко и наглядно отражающих особенности структуры, характер изменчивости или анизотропию изучаемых геологических полей.

Например, когда размеры аномалий близки к размерам ячейки сети наблюдений, а их количество велико, исследуются обычно не отдельные аномалии, а их совокупность, определяющая характер изменчивости изучаемого поля пространственной переменной. Для описания изменчивости предложен ряд произвольных характеристик. Для оценки степени «изрезанности» поля используют коэффициент «аномалийности», который определяется по относительной разнице между длиной графика изучаемого свойства на участке профиля постоянной длины и длиной этого участка. На рис. 50 повышенные значения коэффициента аномалийности магнитного поля четко фиксируют крупные разрывные структуры.

При прослеживании геологических границ по геофизическим данным применяются горизонтальные градиенты или энтропия поля.

Горизонтальные градиенты определяются по числу изолиний поля, попадающих на единичный отрезок, перпендикулярный к преобладающему в пределах этого отрезка направлению изолиний. Для расчета энтропии с помощью палетки подсчитываются доли каждого интервала между изолиниями, попадающими в пределы элементарной площадки трансформации.

На рис. 51 видна приуроченность ореолов свинца в делювии к областям повышенных значений горизонтальных градиентов и энтропии гравитационного поля, а наиболее крупный из ореолов приурочен к пересечениям двух зон повышенной энтропии.

## 5.10. ОБЛАСТИ ПРИМЕНЕНИЯ ГОРНО-ГЕОМЕТРИЧЕСКИХ МОДЕЛЕЙ И ТРЕНД-АНАЛИЗА В ГЕОЛОГИИ

Большинство геологических задач относится к числу пространственных исследований и имеет цель выявить особенности размещения изучаемых геологических объектов в структурах земной коры или элементов ее строения. Поэтому методы количественного описания и математического моделирования пространственных геологических закономерностей являются ведущими во всех отраслях геологических наук.

В геологической практике издавна исключительно широко распространены методы горно-геометрического моделирования геологических тел и свойств горных пород и полезных ископаемых.

Графические модели различных свойств природных геологических тел широко используются в структурной геологии, геологии полезных ископаемых, рудничной геологии и методике поисков и разведки полезных ископаемых. Методы горно-геометрического моделирования изучаются в курсе геометризации недр. На принципах П. К. Соболевского были разработаны различные аналитические методы описания изменчивости, использующие для этих целей первые или вторые последовательные разности значений показателей изменчивости по смежным пунктам наблюдений.

С помощью горно-геометрических моделей можно выразить особенности пространственной изменчивости свойств геологических образований, установить значение изучаемого свойства в любой точке исследуемого объекта, получить представление о его морфологии и внутреннем строении.

Гипсометрические планы поверхностей контактов, не выходящих на поверхность интрузивных тел и рудных залежей, графики изолиний содержаний полезных компонентов в рудных телах (рис. 52), карты геохимических и геофизических полей широко используются в геологической практике, так как они обеспечивают наглядность изображения и улучшают пространственное восприятие изучаемых закономерностей. Однако требование непрерывности и плавности изменения изучаемого свойства ограничивает область их практического применения объектами с весьма выдержаными в пространстве свойствами. К таким объектам относятся пласты осадочных пород, границы интрузивных

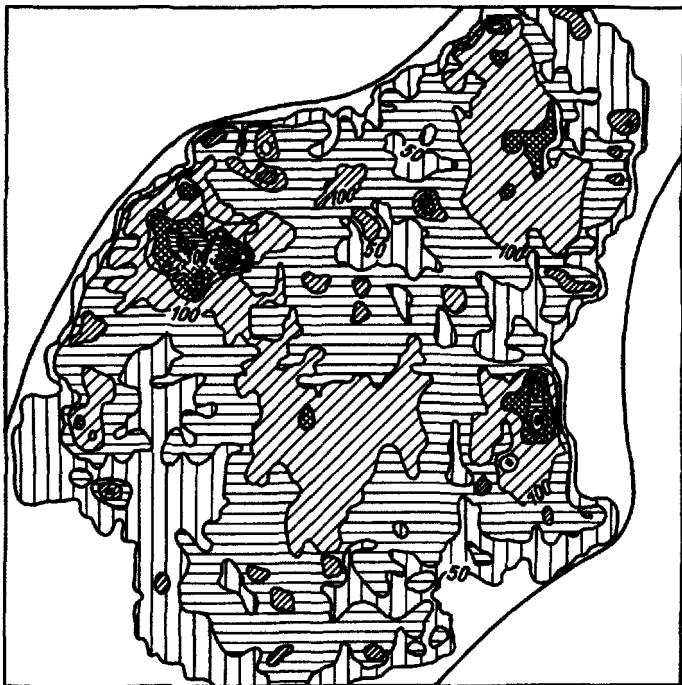


Рис. 52. План изолиний содержаний молибдена в условных единицах по уступу карьера:

1 — < 10; 2 — 10—50; 3 — 50—100; 4 — 100—200; 5 — 200—300; 6 — 300—400; 7 — 400—500; 8 — > 500

образований, рудные тела с простой морфологией и относительно равномерным характером оруденения и т. п.

В связи с бурным внедрением вычислительной техники и ЭВМ, последние десятилетия все шире и многообразнее становится использование в геологической практике методов тренд-анализа исходных данных, в числе которых особенно широкое развитие получили методы сглаживания и аппроксимации поверхностей тренда. Исключительное разнообразие предложенных методов тренд-анализа объясняется стремлением их создателей подобрать в каждом конкретном случае такую модель аппроксимации, которая не только обладала бы минимальными отклонениями наблюдаемых значений признаков от поверхности тренда, но и по возможности, не смещала бы выявленные их экстремальные значения. Однако здесь не учитывалось, что при достаточно сложном характере природной изменчивости признака ограниченности числа наблюдений и редкой их сети, высокая степень приближения аппроксимирующей поверхности ко всей совокупности наблюдений является иллюзорной.

Ранее уже упоминалось, что с помощью любой модели тренд-анализа можно выявить кажущиеся «закономерные» составляющие изменения даже в совокупностях заведомо случайных данных (например,

подобранных по таблицам случайных чисел). Поэтому выявление истинных закономерностей пространственной изменчивости сложно построенных геологических объектов возможно только путем последовательного сгущения сети наблюдений. По редкой их сети выявление многочисленных высокочастотных колебаний значений признака невозможно и никакая, даже самая сложная, модель аппроксимации здесь не поможет.

С позиций системного подхода значительно важнее сначала обеспечить густоту сети наблюдений, достаточную для выявления заданных высокочастотных колебаний признака, а затем выбрать оптимальные размеры слаживающего окна, ориентируясь на уверенную геометризацию элементов неоднородности заданного иерархического уровня строения изучаемого объекта. Нельзя забывать, что достоверные значения «скользящей средней» будут обеспечены лишь при достаточном числе наблюдений в окне (в большинстве случаев более девяти, а в оптимальных вариантах около 25). При малом же числе точек, доверительные вероятности значений «скользящих средних» увеличиваются настолько, что их уже нельзя рассматривать как детерминированные величины, а следовательно, и как закономерные составляющие изменчивости. С этой точки зрения мало эффективны и результаты использования сложных полиномов «взвешенного слаживания». Разные долевые участия значений признака приводят к искажениям оценок истинных его средних значений в окне, а лучшая аппроксимация эмпирических поверхностей оказывается фактически иллюзорной. Применительно к решению многих геологических задач, сравнительные результаты моделирования с использованием различных моделей тренд-анализа свидетельствуют в пользу невзвешенного слаживания, но с тщательным и объективным подбором размеров скользящего окна и числа наблюдений в нем, обеспечивающих выявление структуры изучаемого геологического объекта на заданном иерархическом уровне его строения.

## 5.11. МОДЕЛИРОВАНИЕ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ПОЛЕЙ

Дискретные случайные поля используются для изучения особенностей расположения объектов в пространстве. Наиболее часто в геологической практике возникает задача выявления закономерностей в расположении точечных объектов на плоскости или относительно линий и объектов произвольной формы, имеющих конечные площади. Точечным считается объект, размеры которого очень малы по сравнению с исследуемой территорией.

Выделяют общую задачу, когда проверяется гипотеза о случайному расположении исследуемых объектов, и локальную, когда требуется определить области относительного сгущения и разрежения точечных объектов.

Для решения общей задачи при изучении распределения точечных объектов на плоскости исследуемая территория разбивается на квадратные ячейки постоянного размера. При этом часть ячеек ( $p$ )

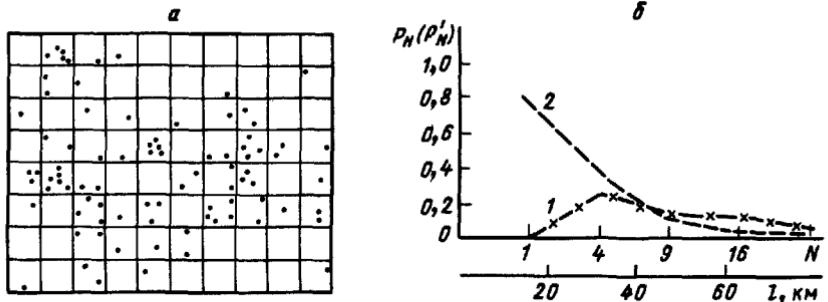


Рис. 53. Проверка гипотезы о случайному расположении медных месторождений в штате Аризона:

а—схема расположения медных месторождений на территории западной части штата;  
б—зависимость доли пустых ячеек от длины стороны квадрата (1—фактическая,  
2—теоретическая)

будет содержать хотя бы один точечный объект, а остальные ( $1-p$ ) окажутся пустыми. После этого исходные ячейки группируются в новые квадраты по  $N$  исходных ячеек в каждом. При случайному расположении точечных объектов вероятность того, что новый квадрат окажется пустым, будет равна  $P_N = (1-p)^N$ . Рассматриваются все возможные варианты объединения исходных ячеек в квадраты по  $N=4, 9, 16$  шт. и т. д. и находятся доли пустых ячеек  $p'_N$  для каждого  $N$ . Повышенная доля  $p'_N$  пустых ячеек по сравнению с теоретической вероятностью указывает на группировку «роение» объектов на отдельных участках, а пониженная — на регулярность их расположения на плоскости.

*Пример.* Проверяется гипотеза о случайному расположении медных месторождений в западной части шт. Аризона (США). Для этого исследуемая площадь была разделена на исходные квадраты площадью  $250 \text{ км}^2$ , составляющие 0,25 части от квадратных ячеек, изображенных на рис. 53, а. График изменения величины для случайного расположения точек в зависимости от размеров ячеек приведен на рис. 53, б. Фактическое значение числа пустых ячеек существенно отличается от теоретического, что указывает на неслучайный характер распределения месторождений на исследуемой территории.

При большом количестве точечных объектов дискретную модель обычно заменяют непрерывной путем подсчета количества точечных объектов в элементарных ячейках.

Для решения локальной задачи пользуются специальными палетками в виде концентрических окружностей или квадратов, центр которых последовательно помещается в разные точки изучаемой площади. При каждом положении палетки подсчитывается число объектов в пределах меньшей фигуры  $m$  и большей  $n$ . Избыточная плотность расположения объектов  $v$  при каждом положении палетки определяется по формуле

$$v = \frac{m}{p} - n, \quad (30)$$

где  $p$  — отношение площади меньшей фигуры к большей.

Вероятность случайного попадания не менее чем  $m$  объектов из их общего числа  $n$  в область с относительными размерами  $p$  описывается биномиальным законом



Рис. 54. Расположение радиометрических аномалий в горно-таежном районе.

1 — радиометрические аномалии; 2 — изолинии избыточной плотности аномалий; 3 — изолинии недостаточной плотности аномалий; 4 — участки сгущения, значимо отличающиеся от площадей со случайным расположением аномалий

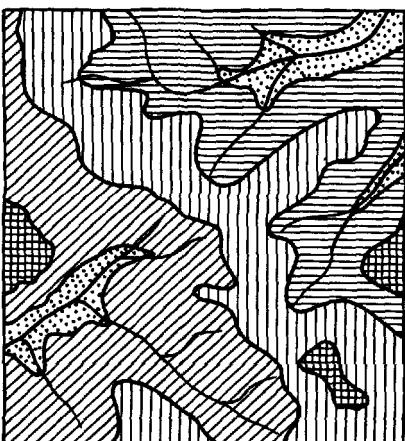


Рис. 55. Схема морфологически-ландшафтного районирования горно-таежного района.

Рельеф: 1 — аллювиальный; 2 — крутых склонов; 3 — водоразделов; 4 — склонов средней крутизны; 5 — аккумулятивный

$$P_1 = \sum_{i=m}^n C_n^i p^i (1-p)^{n-i}. \quad (31)$$

Вероятность случайного попадания не более чем  $m$  объектов при тех же условиях может быть определена по формуле

$$P_2 = \sum_{i=0}^m C_n^i p^i (1-p)^{n-i}.$$

Задавшись определенной доверительной вероятностью, по полученным значениям  $p_1$  и  $p_2$  можно выделить области относительного сгущения или разрежения, значимо отличные от областей со случайным расположением точечных объектов.

*Пример.* На территории одного из горно-таежных районов в результате поисков был выявлен ряд радиометрических аномалий (рис. 54). Для расчета избыточной плотности аномалий использовалась квадратная палетка с соотношением площадей квадратов  $p = \frac{1}{9}$  (см. рис. 54). Расчет осуществлялся по формуле (30). Например, для точки, помеченной на рис. 54,  $a$  крестом, при  $m=2$ ,  $n=3$  и  $p=\frac{1}{9}$   $v=9 \cdot 2 - 3 = 15$ .

С помощью формулы (31) были установлены и оконтурены участки сгущения аномалий, значимо отличающиеся от площадей со случайным расположением точечных объектов при доверительной вероятности 0,95 (см. рис. 54). Сопоставление изолиний избыточной плотности аномалий со схемой ландшафтного районирования данной территории (рис. 55) позволяет сделать следующие выводы. Площади с избыточной плотностью радиометрических

аниомалий отчетливо приурочены к ландшафтам крутых склонов и в виде локальных участков к ландшафтам водоразделов. В то же время ландшафты склонов средней крутизны, и особенно аккумулятивные и аллювиальные ландшафты, характеризуются отрицательными значениями  $v$ . Это дает основание заключить, что частота выявления радиометрических аномалий на данной территории существенно зависит от природных условий ведения поисковых работ.

О зависимости расположения  $n$  точечных объектов относительно объектов с конечной суммарной площадью, составляющей часть  $p$  от всей исследуемой площади, можно судить по нарушению равенства  $\frac{m}{n} = p$ , где  $m$  — количество точечных объектов, попадающих на площадные объекты. Вероятность случайного попадания  $m$  объектов из  $n$  в область с относительными размерами  $p$  оценивается по формуле (31).

*Пример.* В одном из районов Южного Урала медно-колчеданные месторождения и рудопроявления приурочены к площадям развития кислых вулканитов (рис. 56). Эта тенденция настолько очевидна, что не нуждается в статистической проверке. Однако на некоторых участках развития благоприятных пород рудопроявления отсутствуют. Было высказано предположение, что в пределах площадей развития кислых вулканитов размещение проявлений медно-колчеданной минерализации определяется локальными аномалиями тектонической напряженности. Тектоническая напряженность оценивалась по произведению количества трещин на их суммарную длину в пределах скользящего ока площадью  $4 \text{ км}^2$ . Для проверки этой гипотезы примем за исследуемую площадь развития кислых вулканитов, а в качестве площадных объектов — участки аномалий тектонической напряженности в пределах этих пород. Отношение суммарной площади аномалий в пределах кислых вулканитов к общей площади этих пород ( $p$ ) составляет 0,46, а отношение количества рудопроявлений и месторождений в кислых вулканитах в пределах аномалий ( $m$ ) к общему их числу в данных породах ( $n$ ) равно

$$\frac{m}{n} = \frac{24}{44} \approx 0,55.$$

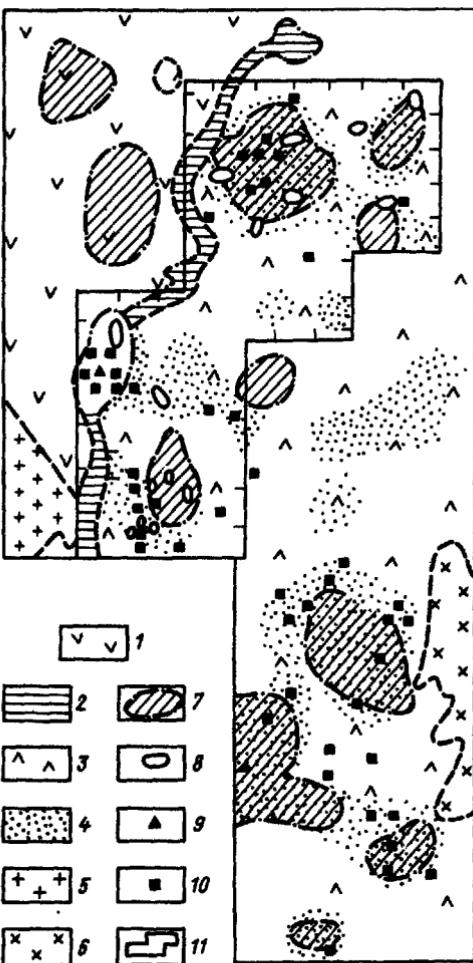
Таким образом, концентрация проявлений минерализации в пределах аномалий несколько выше, чем по кислым вулканитам в целом. Однако разница между величинами  $p$  и  $\frac{m}{n}$  невелика. Вероятность такого расхождения при случайному расположении точек минерализации определенная по формуле (31), составляет 0,36. Это не позволяет считать связь медно-колчеданных рудопроявлений и месторождений с аномалиями тектонической напряженности доказанной.

Аналогичным образом проверяется гипотеза о зависимости расположения точечных объектов относительно линейных. Под линейными в данном случае понимаются объекты, шириной которых можно пренебречь — тектонические нарушения, контакты пород и т. п. Вдоль линейных объектов строятся полосы различной ширины, которые можно рассматривать как объекты с конечной площадью и применять к ним методы, рассмотренные выше.

Более полно проблема математического моделирования пространственных геологических переменных рассмотрена в работах Н. Н. Боровко, У. Крамбейна и Ф. Грейбила. Способы проверки гипотез о наличии тренда подробно описаны в книге Р. Миллера и Д. Кана.

Рис. 56. Расположение медно-колчеданных месторождений в одном из районов Южного Урала. По В. У. Емельяненко и Н. В. Левиту:

1—порфириты основного состава и их туфы; 2—кремнистые сланцы, яшмы; 3—вулканиты от основного до кислого состава; 4—кислые вулканиты; 5—граниты; 6—гранодиориты; 7—локальные аномалии поля тектонической напряженности; 8—установленные геофизическими работами суббульканические тела; 9—месторождения; 10—рудопроявления; 11—площадь детальных геофизических исследований



#### Контрольные вопросы

1. Приведите примеры непрерывных и дискретных геологических пространственных переменических, скалярных и векторных полей.
2. Что такое область воздействия эксперимента? Какое влияние эти воздействия оказывают на характеристики наблюдаемой изменчивости свойств полезных ископаемых?
3. Какие геологические, методические и технические факторы обусловливают появление закономерной и случайной составляющей наблюдаемой изменчивости?
4. Дайте определение понятиям *фон* и *аномалия*. Каковы их взаимоотношения?
5. Какие геологические задачи решаются с помощью тренд-анализа?
6. Что такое *топографункции* и при каких условиях они могут быть выявлены по данным наблюдений?
7. В чем заключаются основные недостатки и преимущества способов «сглаживания» исходных данных?
8. В чем различия между невзвешенным и взвешенным способами стягивания исходных данных?

9. Охарактеризуйте важнейшие статистические гипотезы проверки наличия тренда и условия их применения.

10. В чем заключается метод аппроксимации поверхностей тренда полиномами? Для решения каких геологических задач используется этот метод?

11. Перечислите важнейшие области применения горно-геометрических моделей и тренд-анализа в геологии.

## ГЛАВА 6

# МОДЕЛИРОВАНИЕ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ С ПОМОЩЬЮ СЛУЧАЙНЫХ ФУНКЦИЙ

---

## 6.1. ПРИНЦИПЫ МОДЕЛИРОВАНИЯ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ С ПОМОЩЬЮ СЛУЧАЙНЫХ ФУНКЦИЙ

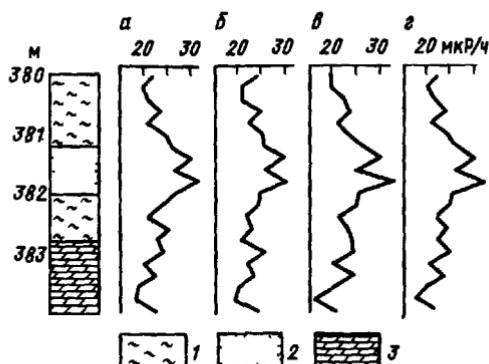
При решении многих геологических задач возникает необходимость количественной оценки пространственной изменчивости свойств геологических объектов. Методы статистики случайных величин для этих целей малопригодны, так как все статистические модели не учитывают пространственного расположения точек наблюдений. Для оценки пространственной изменчивости геологических признаков может быть использован математический аппарат теории случайных функций, широко применяемый в различных областях науки и техники.

Случайной называется такая функция, которая в результате опыта может принять тот или иной конкретный вид, неизвестно заранее, какой именно. Аргументами при этом могут быть время (случайные процессы) или координаты пространства (случайные поля или последовательности).

Случайные процессы используются в качестве моделей прибрежноморских гидродинамических процессов, для описания сезонных колебаний уровней грунтовых вод, при изучении гидродинамических свойств пород, при опытных откачках из скважин и т. п. Гораздо чаще геологу приходится иметь дело с моделями случайных полей и последовательностей, описывающими пространственную изменчивость самых различных геологических признаков в одномерном, двумерном или трехмерном пространствах.

Конкретный вид, который принимает функция в результате опыта, называется ее реализацией. Так, например, каротажная диаграмма по скважине может рассматриваться как реализация случайной функции, описывающей изменение изучаемого физического свойства пород на пробуренном интервале. Совокупность нескольких контрольных каротажных диаграмм образует семейство или ансамбль реализаций случайной функции (рис. 57). Каждая реализация отражает общие тенденции в изменении изучаемого свойства, но они несколько отличаются друг от друга. В данном примере это обусловлено аппаратурными погрешностями, некоторым несовпадением точек замеров и другими случайными факторами.

Рис 57 Результаты гамма-каротажа (в мкР/ч) по скважине  
 а—основной каротаж, б—контрольный, в—усредненная каротажная кривая, 1—алевролиты, 2—песчаники, 3—известняки



Совокупность значений случайной функции  $X(l)$  при любом фиксированном значении аргумента  $l$  называется сечением случайной функции и является обычной, случайной величиной (см рис 57, а, б, в), т. е. такая функция совмещает в себе черты случайной величины и функции. Если зафиксировать значение аргумента, она превращается в обычную случайную величину, а отдельная реализация может рассматриваться как обычная (неслучайная) функция.

Модель типа случайной функции, как и статистическая, основана на положениях теории вероятностей. Процесс моделирования сводится в конечном итоге к получению выборочных значений и к их статистической оценке путем проверки статистических гипотез с помощью критериев согласия. При оперировании со случайными функциями широко используются основные понятия математической статистики о функциях и моментах распределения, уровнях значимости и доверительных вероятностях событий.

Главными характеристиками случайной функции являются ее математическое ожидание  $Mx(l)$ , дисперсия  $D_x(l)$  и корреляционная функция  $K_x(l)$ . В отличие от числовых характеристик случайных величин, представляющих собой определенные числа, характеристики случайных функций представляют собой не числа, а функции.

Математическим ожиданием случайной функции  $X(l)$  называется неслучайная функция  $Mx(l)$ , которая при каждом значении аргумента  $l$  равна математическому ожиданию соответствующего сечения случайной функции. Оценка математического ожидания по ансамблю из  $N$  реализаций рассчитывается по формуле

$$\bar{X}(l) = \sum_{k=1}^N X_k(l)/N$$

и является геометрическим местом точек средних значений изучаемого свойства по сечениям. В приведенном выше примере (см. рис. 57) оценкой математического ожидания служит усредненная каротажная диаграмма, рассчитанная по трем реализациям (см. рис. 57, г). При изучении случайных многократно повторяющихся процессов оценкой математического ожидания является усредненный график параметра процесса, построенный путем расчета средних значений

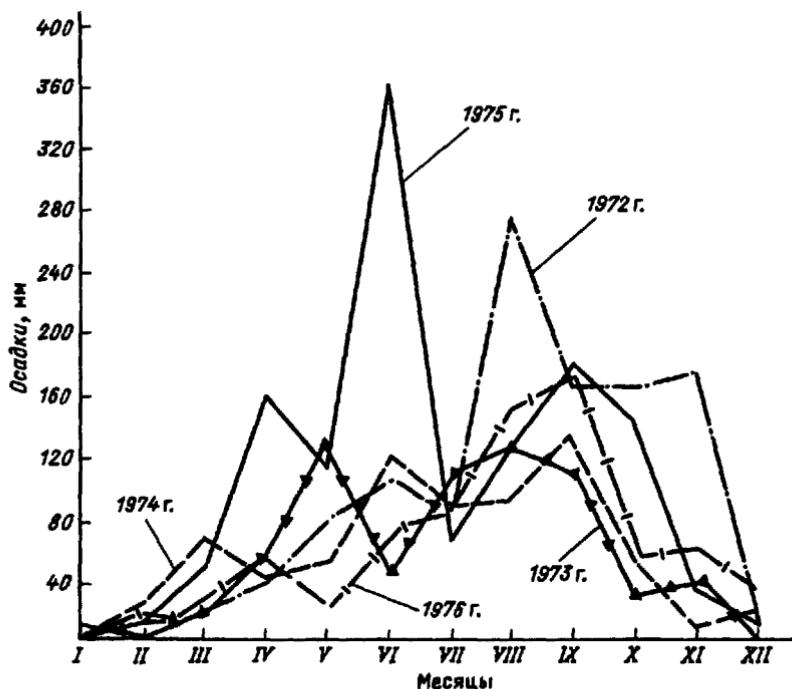


Рис. 58. Среднемесячное количество осадков в районе Дальнегорского месторождения

этого параметра для фиксированных значений или интервалов времени с начала процесса. Примером такого графика может быть приведенная на рис. 58 кривая изменения среднемесячного количества атмосферных осадков в районе Дальнегорского месторождения за пятилетний период.

Дисперсией случайной функции  $X(l)$  называется неслучайная функция  $D_x(l)$ , значения которой для каждого  $l$  равны дисперсии соответствующего сечения случайной функции. Она характеризует ширину полосы разброса возможных реализаций случайной функции относительно ее математического ожидания. Оценка дисперсии случайной функции рассчитывается по формуле

$$D_x(l) = \frac{\sum_{k=1}^N [x_k(l) - \bar{x}(l)]^2}{N-1}.$$

Для описания основных свойств случайной функции математического ожидания и дисперсии недостаточно. Ни одна из данных характеристик не отражает особенностей внутренних связей между отдельными сечениями такой функции. С этой целью используется специальная характеристика — корреляционная функция, отражающая степень зависимости между сечениями случайной функции при различных  $l$ .

Корреляционной функцией случайной функции  $X(l)$  называется неслучайная функция двух аргументов  $K_x(l', l'')$ , которая при каждой паре значений аргумента равна корреляционному моменту соответствующих сечений случайной функции

$$K_x(l', l'') = M[\dot{x}(l') \dot{x}(l'')].$$

В этом выражении  $\dot{x}(l')$  и  $\dot{x}(l'')$  — значения центрированной случайной функции при значениях аргумента  $l'$  и  $l''$ . Центрирование случайной функции осуществляется путем вычитания из значений случайной функции ее математического ожидания

$$\dot{x}(l) = x(l) - Mx(l).$$

В частном случае, когда оба аргумента случайной функции совпадают, т. е.  $l' = l''$ , корреляционная функция  $K_x(l', l'')$  обращается в дисперсию случайной функции

$$K_x(l', l'') = M\{[x(l')]^2\} = D_x(l').$$

При использовании случайных функций для целей геолого-математического моделирования геологические объекты рассматриваются как случайные поля, состоящие из элементарных участков, каждый из которых соответствует зоне экстраполяции единичного наблюдения. В пределах каждого элементарного участка изучаемое свойство рассматривается как случайная величина, изменяющаяся относительно некоторого ее среднего значения. Изменения математических ожиданий и дисперсий этих случайных величин можно описать неслучайными функциями координат центров элементарных участков и рассматривать эти функции как математическое ожидание и дисперсию случайной функции.

Упорядоченный ряд значений признака, полученных путем опробования по заданному направлению в определенных точках внутри данного поля, представляет собой одну из возможных реализаций случайной функции, аргументом которой является расстояние между точками наблюдений. Группа точечных значений параметра, расположенных в одной плоскости, представляет собой реализацию плоскостного сечения случайного поля, а вся совокупность точечных замеров — одну из его объемных реализаций.

Принципиальная особенность данной модели заключается в том, что она отражает не только специфику природной изменчивости геологических объектов, но и методику их изучения. Модели одного и того же участка недр, построенные по совокупности экспериментальных данных, могут существенно различаться в зависимости от объема проб и густоты сети наблюдений. Для описания изменчивости любого свойства практически достаточно знать характеристики случайных функций по трем взаимно ортогональным направлениям, совпадающим с осями анизотропии природных геологических тел.

Для описания характера изменчивости двумерных геологических полей используют *двумерную автокорреляционную функцию* (ДАКФ), которая выражает силу корреляционной связи между значениями признака в точках поля, расположенных на разном расстоянии друг от друга относительно двух координат пространства  $X$  и  $Y$ .

Использование аппарата теории случайных функций для решения геологических задач сопряжено с рядом сложностей и ограничений. Главное из них заключается в том, что такая функция, подобно случайной величине, является совокупностью частных реализаций, в то время как результаты геологических наблюдений чаще всего представляют собой одну-единственную реализацию случайной функции. Оценка случайной функции по одной реализации возможна только в том случае, если она обладает свойствами стационарности и эргодичности. Функции же, характеризующие наблюдалемую изменчивость свойств природных геологических образований, как правило, не обладают этими свойствами. Поэтому при использовании моделей типа случайных функций вводится ряд допущений.

Для расчета важнейших характеристик эмпирических случайных функций необходимо обеспечить массовость исходных экспериментальных данных. Поэтому широкое применение модели рационально только в тех случаях, когда количество единичных наблюдений достаточно велико, а их плотность обеспечивает отчетливое проявление автокорреляции наблюдалемых значений признака.

## 6.2. ИСПОЛЬЗОВАНИЕ АВТОКОРРЕЛЯЦИОННЫХ ФУНКЦИЙ ДЛЯ РЕШЕНИЯ ГЕОЛОГИЧЕСКИХ ЗАДАЧ

Случайная функция  $X(l)$  называется стационарной, если все ее вероятностные характеристики не зависят от  $l$  (точнее, не изменяются при любом сдвиге аргументов по оси  $l$ ).

Стационарная случайная функция отличается постоянством математического ожидания, постоянством дисперсии и зависимостью ее корреляционной функции не от двух аргументов, а только от расстояния  $r$  между первым и вторым аргументами:  $K_x(l, l+r) = K_x(r)$ .

Следовательно, автокорреляционная функция стационарной случайной функции представляет собой функцию не двух, а одного аргумента, что сильно упрощает математические операции над ними.

Корреляционная функция стационарной случайной функции выражается через корреляционный момент соответствующих сечений случайной функции и характеризует степень связи между ними в зависимости от значений аргумента. В общем виде она может быть выражена формулой

$$K_x(r) = \frac{1}{L-r} \int_0^{L-r} [f(x_i) - Mx] [f(x_{i+r}) - Mx] dx,$$

где  $L$  — длина исследуемого профиля (участка);  $r$  — расстояние между точками наблюдений, выраженное числом интервалов между ними;  $f(x_i)$  — переменная величина (например, значения содержаний металла в пробах в ряду, который начинается значением  $x_1$  и заканчивается значением  $x_{n-r}$ );  $f(x_{i+r})$  — переменная величина (например, значение содержаний металла в пробах в ряду, который начинается значением  $x_i$ , и заканчивается значением  $x_n$ );  $Mx$  — среднее значение переменной величины  $f(x)$  в интервале от 0 до  $L$ .

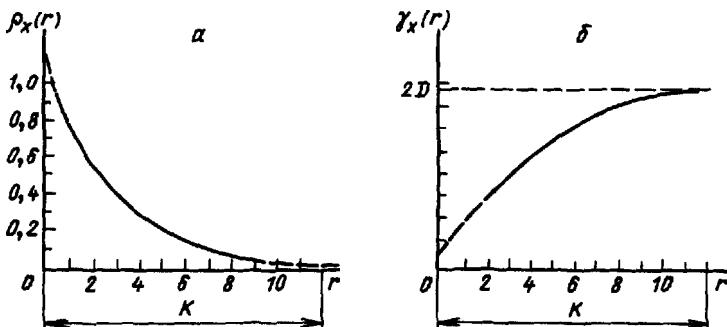


Рис. 59. Корреляционная (а) и структурная (б) функции

На практике часто пользуются корреляционной функцией, нормированной по дисперсии  $\rho_x(r) = K_x(r)/D_x$ .

Значения функции  $\rho_x(r)$  представляют собой коэффициенты корреляции между сечениями случайной функции, разделенными расстояниями  $r$ . Очевидно, что значение  $\rho_x(0)$  равно 1.

Для количественного описания характера изменчивости стационарных случайных функций используется также структурная функция (вариограмма), которая рассчитывается через квадраты разностей значений признака в точках, отстоящих друг от друга на расстоянии  $r$ :

$$\gamma_x(r) = \frac{1}{L-r} \int_0^{L-r} [f(x_{i+r}) - f(x_i)]^2 dx.$$

Между структурной и корреляционной функциями существует почти строгая зависимость при  $r$  заведомо меньшем, чем  $L$ :  $\gamma_x(r) = 2[K_x(0) - K_x(r)]$ .

Корреляционная функция стационарной случайной функции постепенно убывает с увеличением аргумента до определенного его значения, называемого пределом корреляции, и равна нулю при значениях аргумента, больших предела корреляции (рис. 59, а). Ее структурная функция возрастает в том же интервале значений  $r$ , а при расстоянии большем предела корреляции равна удвоенному значению статистической дисперсии исследуемого свойства (рис. 59, б).

При расчетах эмпирических, корреляционных и структурных функций знаки интегралов заменяются знаками суммы, поскольку сеть наблюдений дискретна. При равномерной дискретной сети наблюдений эмпирические корреляционные и структурные функции рассчитываются по формулам

$$K_x(r) = \frac{\sum_{i=1}^{n-r} (x_i - \bar{x})(x_{i+r} - \bar{x})}{(n-r)}. \quad (32)$$

$$\gamma_x(r) = \frac{\sum_{i=1}^{n-r} (x_i - x_{i+r})^2}{(n-r)},$$

где  $i$  — порядковый номер замера в исследуемом ряду значений  $x$ ;  $n$  — общее число замеров;  $r$  — последовательно принимает значения  $1, 2, 3, \dots, (n-1)$ .

Расчет значений эмпирической нормированной корреляционной функции заключается в вычислении коэффициентов корреляции между значениями исследуемого ряда с порядковыми номерами от 1 до  $(n-r)$  и значениями того же ряда, сдвинутыми относительно исходного на  $r$  значений, т. е. ряда с порядковыми номерами от  $r$  до  $n$ .

Пример расчета первой точки нормированной корреляционной функции по результатам гамма-каротажа скважины (см. рис. 57) приведен в табл. 37.

Если одна реализация стационарной случайной функции на достаточно большом интервале эквивалентна большому числу реализаций на ограниченном интервале, считается, что функция обладает эргодическим свойством. В этом случае ее среднее значение на достаточно большом участке приближенно равно среднему по множеству наблюдений. Не все стационарные случайные функции обладают этим свойством. Различают два подкласса стационарных случайных функций — эргодические и неэргодические.

**Таблица 37. Расчет первой точки нормированной корреляционной функции значений интенсивности гамма-излучения по скважине**

$i$	$x_i$	$x_i - \bar{x}$	$x_{i+1}$	$x_{i+1} - \bar{x}$	$(x_i - \bar{x})(x_{i+1} - \bar{x})$	$(x_i - \bar{x})^2$
1	21,5	-2,0	20,0	-3,5	7,00	4,00
2	20,0	-3,5	21,0	-2,5	8,75	12,25
3	21,0	-2,5	23,5	0	0	6,25
4	23,5	0	21,0	-2,5	0	0
5	21,0	-2,5	24,5	1,0	-2,50	1,00
6	24,5	1,0	26,0	2,5	2,50	6,25
7	26,0	2,5	30,0	6,5	16,25	6,25
8	30,0	6,5	26,5	3,0	19,50	42,25
9	26,5	3,0	31,5	8,0	24,00	9,00
10	31,5	8,0	26,5	3,0	24,00	64,00
11	26,5	3,0	24,5	1,0	3,00	9,00
12	24,5	1,0	21,0	-2,5	-2,50	1,00
13	21,0	-2,5	25,5	2,5	-6,25	6,25
14	25,5	2,0	22,5	-1,0	-2,00	4,00
15	22,5	-1,0	24,0	0,5	-0,50	1,00
16	24,0	0,5	20,00	-3,5	-1,75	0,25
17	20,0	-3,5	22,0	-1,5	5,25	12,25
18	22,0	-1,5	18,5	-5,0	7,50	2,25
19	18,5	-5,0	18,0	-5,5	27,50	25,00
20	18,0	-5,5	22,5	-1,0	5,50	30,25
21	22,5	-1,0	—	—	—	1,00
$\Sigma$		490,5			135,25	243,50

$$\bar{x} = \frac{490,5}{21} = 23,5; S^2 = \frac{243,5}{20} = 12,175; \rho_x(1) = \frac{135,25}{20 \cdot 12,175} = 0,555.$$

Каждая реализация эргодической случайной функции  $X_1(l)$  обладает примерно одинаковыми средним значением, средней амплитудой и средним периодом колебаний. Если мысленно продлить любую реализацию этой функции на достаточно большое расстояние по оси  $l$ , то осредненные значения основных характеристик случайной функции по этой одной реализации совпадут со средними характеристиками, рассчитанными по ансамблю реализаций.

Реализации неэргодической случайной функции  $X_2(l)$  заметно отличаются друг от друга по среднему значению и амплитуде колебаний, поэтому оценки математического ожидания и дисперсии для этого типа случайных функций могут быть получены только по множеству реализаций.

Свойствами стационарности и эргодичности обладают только однородные поля геологических переменных, в пределах которых отсутствуют тенденции к увеличению значений исследуемых свойств в каких-либо направлениях от центра к периферии поля или периодические колебания с периодами, сопоставимыми с размерами поля. В однородных полях закономерная составляющая изменчивости может проявляться только в виде локальных колебаний со случайными периодами, по крайней мере, на порядок меньшими размеров изучаемого поля.

Примерами однородных полей могут служить магнитные поля крупных геотектонических структур, единых по своему строению и происхождению (Русская платформа, Западно-Сибирская плита и т. п.). Средние размеры локальных аномалий обычно не превышают 100 км, т. е. несоизмеримо меньше размеров этих регионов.

О стационарности случайной функции можно судить по виду ее корреляционной функции, которая в этом случае стремится к нулю при увеличении  $r$ , т. е. близка по виду к теоретической корреляционной функции, изображенной на рис. 58, а. Наличие в пределах исследуемого поля крупных элементов неоднородности приводят к тому, что эмпирические корреляционные функции переходят в область отрицательных значений или характеризуются незатухающей периодичностью.

С помощью корреляционных или структурных функций могут быть выражены основные особенности пространственной изменчивости изучаемого признака:

- общий уровень колебаний его значений;
- предельные расстояния между соседними точками наблюдений, начиная с которых значения изучаемых признаков можно считать практически взаимонезависимыми;
- соотношения и уровни случайной и закономерной составляющих изменчивости в зависимости от густоты сети наблюдений.

Перечисленные свойства позволяют использовать корреляционную и структурную функции для решения следующих геологических задач.

1. Проверка гипотез о наличии закономерности (тренда) в последовательности значений изучаемого свойства. С этой

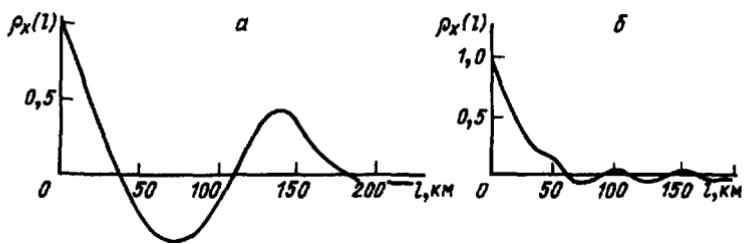


Рис. 60. Корреляционные функции аномального магнитного поля.  
По В. Н. Луговенко:  
а — по профилю 23; б — по профилю «Анабарский»

целью рассматриваются первые две-три точки нормированной корреляционной функции изучаемой последовательности и оценивается статистическая значимость их отличия от нуля по формуле (32).

Этот способ наиболее надежный по сравнению с рассмотренными в гл. 5 статистическими способами проверки гипотезы о наличии тренда, так как позволяет выявлять слабо проявленные закономерности.

*Пример.* Проверим гипотезу о наличии закономерности в изменении интенсивности гамма-излучения по скважине (см. табл. 36)

$$t = \frac{\rho_x(1)}{\sqrt{1 - \rho_x^2(1)}} \sqrt{n-2} = \frac{0,555}{\sqrt{1 - 0,555^2}} \sqrt{20-2} = 2,83.$$

Расчетное значение критерия Стьюдента превышает табличное для доверительной вероятности 0,95 и числа степеней свободы 18, следовательно, гипотеза об отсутствии закономерности в изучаемой последовательности должна быть отвергнута.

2. Классификация геологических объектов по характеру пространственной изменчивости их признаков. Например, по виду корреляционных функций аномальных магнитных полей (АМП) можно классифицировать участки земной коры с различной историей тектонического развития. Для участков молодой океанической коры характерна упорядоченная линейная система гребней или разломов срединно-океанических хребтов, обусловливающая закономерные, периодические колебания АМП, которые отчетливо видны на корреляционных функциях (рис. 60, а). Вдали же от рифтов упорядоченность топографии океанического дна выалируется из-за наложения нескольких этапов тектонической активизации. Корреляционные функции АМП для участков с длительной историей развития океанической коры имеют вид функций стационарного типа, характерных для континентальной коры (рис. 60, б).

Эталонами и основой для классификации промышленных типов месторождений важнейших видов минерального сырья могут служить типизированные графики характеристик случайных функций важнейших свойств залежей полезных ископаемых, составленные по обобщенным экспериментальным данным.

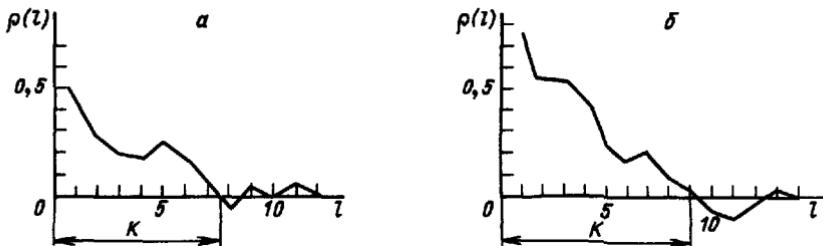


Рис. 61. Корреляционные функции линейных запасов металла по рудной залежи:  
а — по простиранию ( $r=10$  м); б — по падению ( $r=15$  м)

По виду корреляционных функций, рассчитанных для различных интервалов разрезов толщ метаморфических пород, можно обоснованно проводить их расчленение на более мелкие стратиграфические подразделения.

3. Оценка предельных расстояний между точками наблюдений, при которых сохраняются корреляционные связи между значениями исследуемых свойств, используется для обоснования способов количественной обработки результатов и геометризации недр, поскольку лишь в пределах интервала корреляции правомерна интерполяция экспериментальных данных. Значения же признаков за пределами корреляции можно рассматривать как совокупности случайных, независимых величин и они должны обрабатываться только обычными статистическими методами.

Соотношения расстояний предельной корреляции ( $K$ ), установленные по линиям, ориентированным в различных направлениях, могут быть использованы в качестве показателей анизотропии изучаемого свойства по этим направлениям.

*Пример.* Расстояния предельной корреляции значений линейных запасов (произведений мощности на содержание полезного компонента) по простиранию ( $K_1$ ) и падению ( $K_2$ ) рудной залежи равны соответственно 75 и 135 м (рис. 61). Следовательно, показатель анизотропии данного свойства в плоскости залежи будет равен

$$A = K_2/K_1 = \frac{135}{75} = 1,8.$$

Величина коэффициента анизотропии может быть использована в качестве критерия оптимальности выбранной геометрии сети наблюдений. Так, при разведке данного месторождения длинная сторона ячейки разведочной сети в плоскости рудоносной залежи должна располагаться по направлению падения и должна быть в 1,8 раза длиннее расстояния между разведочными пересечениями по простиранию залежи. При этом для надежной интерполяции разведочных данных между пересечениями сеть должна быть не реже  $75 \times 135$  м.

4. Разделение общей дисперсии признака на случайную и закономерную составляющие для заданных расстояний между точками наблюдений.

Эмпирические корреляционные функции стационарных последовательностей могут быть аппроксимированы экспонентой вида

$$\rho_x(r) = e^{-\alpha r}.$$

В этом случае для определения дисперсии случайной составляющей ( $\sigma_{\text{случ}}^2$ ) можно использовать формулу

$$\sigma_{\text{случ}}^2 = \sigma_0^2 [1 - \rho_x^2(r)], \quad (33)$$

где  $\sigma_0^2$  — общая статистическая дисперсия изучаемого признака;  $\rho_x(r)$  — значение нормированной корреляционной функции для расстояния между точками наблюдения, равного  $r$ . Дисперсия закономерной составляющей находится из равенства

$$\sigma_{\text{зак}}^2 = \sigma_0^2 - \sigma_{\text{случ}}^2.$$

При наличии в изменчивости изучаемого свойства закономерной составляющей в статистические формулы для расчета интервальных оценок целесообразно вводить поправки за связь. Для этого в формулу (18) вместо общей дисперсии подставляется дисперсия случайной составляющей, определенная по формуле (33).

Оценивая величины поправок за связь для различных расстояний между точками наблюдений, можно выявить эффект, получаемый от разрежения сети наблюдений.

*Пример.* С помощью корреляционной функции, изображенной на рис. 61, определим долю случайной составляющей в изменчивости линейного запаса по падениюрудной залежи для расстояний 15, 75 и 120 м:

$$l = 15 \text{ м}, \sigma_{\text{случ}}^2 = \sigma_0^2 (1 - 0,78^2) \approx 0,38 \sigma_0^2;$$

$$l = 75 \text{ м}, \sigma_{\text{случ}}^2 = \sigma_0^2 (1 - 0,25^2) \approx 0,94 \sigma_0^2;$$

$$l = 120 \text{ м}, \sigma_{\text{случ}}^2 = \sigma_0^2 (1 - 0,10^2) \approx 0,99 \sigma_0^2.$$

Таким образом, при расстоянии между соседними разведочными пересечениями, равном 15 м, более 60% дисперсии линейного запаса обусловлено закономерными изменениями, поэтому в данном случае при расчете интервальной оценки поправка за связь будет весьма существенной. Однако разрежение сети наблюдений в три раза приводит к резкому снижению доли закономерной составляющей (до 6%), в результате чего точность вычисления среднего значения линейного запаса значительно снизится даже в том случае, если количество наблюдений остается постоянным. Начиная с расстояния 120 м, точность расчета среднего значения линейного запаса практически полностью будет зависеть только от количества наблюдений.

Отношения дисперсий случайной и закономерной составляющих, наблюданной изменчивости могут служить также показателями степени разведенности месторождений.

При использовании характеристик стационарных случайных функций для описания изменчивости геологоразведочных параметров необходимо учитывать законы их статистических распределений. В частности, расчеты эмпирических корреляционных и структурных функций изменчивости логнормально распределенных свойств следует проводить для логарифмов их значений.

В тех случаях, когда геологический объект изучен по нескольким профилям, для решения перечисленных выше задач используются усредненные корреляционные или структурные функции. При этом неоднородности, присущие каждому единичному профилю, нивелируются и вид усредненных характеристик приближается к таковому для стационарных случайных функций.

### 6.3. ОБЛАСТЬ ПРИМЕНЕНИЯ ВЗАИМНЫХ КОРРЕЛЯЦИОННЫХ ФУНКЦИЙ И ДВУМЕРНЫХ АВТОКОРРЕЛЯЦИОННЫХ ФУНКЦИЙ В ГЕОЛОГИИ

Для оценки корреляционных свойств двух случайных процессов или последовательностей используются *взаимные корреляционные функции* (ВКФ). Для эргодических функций ВКФ можно вычислить по отдельным реализациям —  $X_1$  и  $X_2$ . В качестве таких реализаций могут быть взяты данные по двум профилям наблюдений одного свойства или замера двух разных свойств по одному профилю. Вычисление ВКФ производится по формуле

$$B_x(r) = \frac{\sum_{i=1}^{n-r} (x_{1i} - \bar{x}_1)(x_{2(i+r)} - \bar{x}_2)}{n - |r|},$$

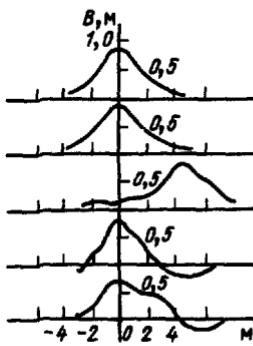
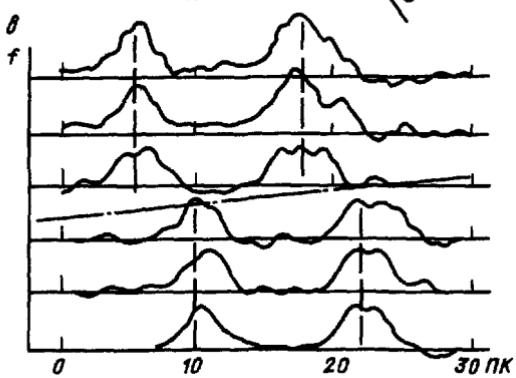
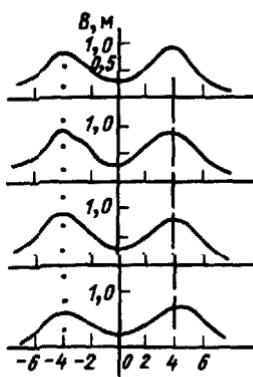
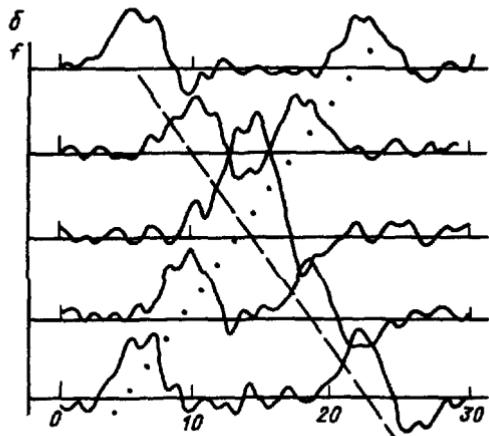
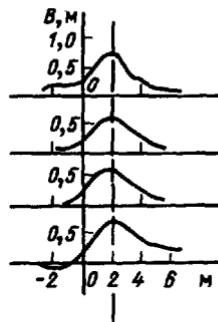
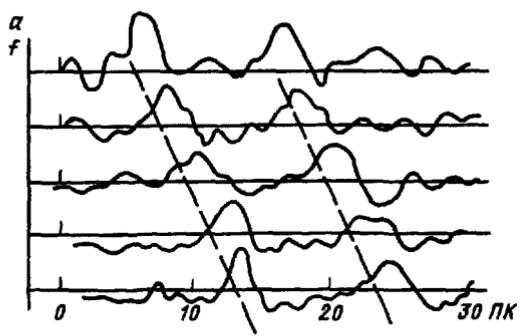
где  $x_{1i}$  — значения по первой реализации, например, содержание определенного химического элемента по профилю геохимической съемки;  $x_{2(i+r)}$  — содержание того же элемента по другому профилю в точке, смещенной относительно замера  $x_{1i}$  на  $r$  пикетов. Смещение при этом производится как в одну, так и в другую сторону.

В практике геологоразведочных работ ВКФ чаще всего используются для решения двух типов задач.

1. При анализе геохимических и геофизических полей с помощью ВКФ, рассчитанных по соседним профилям, можно определить характер ориентировки линейных геохимических и геофизических аномалий (рис. 62). При субпараллельной ориентировке осей аномалий на графике ВКФ наблюдается один максимум, смещенный относительно начала осей координат на величину, равную смещению максимумов аномалий в соседних профилях за счет косой ориентировки осей аномалий относительно направления профилей (см. рис. 62, а). При непараллельной ориентировке аномалий на графике ВКФ появляется два максимума (см. рис. 62, б), а в случае смещения осей аномалий по тектоническому нарушению, наблюдается смещение максимума ВКФ для профилей, между которыми проходит нарушение (см. рис. 62, в).

2 В процессе изучения геохимической зональности метасоматических образований и ореолов рассеяния ВКФ позволяет оценить относительную подвижность химических элементов. Для этого рассчитывают ВКФ для содержаний разных пар химических элементов по одному профилю. Максимумы концентрации более подвижных элементов удалены от рудоконтролирующих структур (контактов интрузивных пород, тектонических разрушений и т. п.) на большее расстояние по сравнению с менее подвижными. По смещению максимума ВКФ относительно начала осей координат можно надежно оценить величину смещения одного элемента относительно другого.

Если геологическое поле изучено по равномерной сети наблюдений, для решения некоторых из перечисленных выше задач целесообразно рассчитать *двумерную автокорреляционную функцию*.



1 — оси аномалий; 2 — тектонические нарушения

Рис. 62. График исходных значений поля и ВКФ в случае аномалий одного простирания (а), различного простирания (б), при смещении аномалий (с). По А. А. Никитину:  
1 — оси аномалий; 2 — тектонические нарушения

чию (ДАКФ). Исследуемое геологическое поле при этом можно представить в виде матрицы

$$U = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1i} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2i} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \dots & x_{ji} & \dots & x_{jn} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{k1} & x_{k2} & \dots & x_{ki} & \dots & x_{kn} \end{bmatrix},$$

где столбцы и строки соответствуют замерам по взаимно перпендикулярным направлениям, например меридиональному и широтному;  $n$  — количество меридиональных профилей;  $k$  — количество широтных профилей.

ДАКФ в этом случае будет представлять собой матрицу ковариаций между всеми возможными парами значений поля, замеренными в точках, расположенных на различных расстояниях друг от друга как в меридиональном, так и в широтном направлении.

Обозначив расстояния между точками наблюдений, выраженное числом интервалов между соседними точками по меридиональному направлению через  $p$ , а по широтному — через  $r$ , получим следующие формулы для расчета значений ДАКФ.

При  $p=0$  ДАКФ равна усредненной одномерной автокорреляционной функции, рассчитанной по сумме автокорреляционных функций всех широтных профилей:

$$R_x(0, r) = \frac{1}{k} \left[ \frac{1}{n-|r|} \sum_{i=1}^{n-r} (x_{1i} - \bar{x}_{1.})(x_{1(i+r)} - \bar{x}_{1.}) + \frac{1}{n-|r|} \sum_{i=1}^{n-r} (x_{2i} - \bar{x}_{2.}) \times \right. \\ \left. \times (x_{2(i+r)} - \bar{x}_{2.}) + \dots + \frac{1}{n-|r|} \sum_{i=1}^{n-r} (x_{ki} - \bar{x}_{k.})(x_{k(i+r)} - \bar{x}_{k.}) \right].$$

При  $p=1$  ДАКФ равна усредненной одномерной функции взаимной корреляции (ВКФ), рассчитанной по сумме ВКФ для всех пар соседних профилей:

$$R_x(1, r) = \frac{1}{k-1} \left[ \frac{1}{n-|r|} \sum_{i=1}^{n-r} (x_{1i} - \bar{x}_{1.})(x_{2(i+r)} - \bar{x}_{2.}) + \right. \\ \left. + \frac{1}{n-|r|} \sum_{i=1}^{n-r} (x_{2i} - \bar{x}_{2.})(x_{3(i+r)} - \bar{x}_{3.}) + \dots + \right. \\ \left. + \frac{1}{n-|r|} \sum_{i=1}^{n-r} (x_{(k-1)i} - \bar{x}_{(k-1).})(x_{k(i+r)} - \bar{x}_{k.}) \right].$$

При  $p=2$  ДАКФ рассчитывается по сумме ВКФ для всех пар широтных профилей, расположенных через один профиль, при  $p=3$  — через два профиля и т. д. При  $p=0$  и  $r=0$  значение ДАКФ равно дисперсии признака.

Для ДАКФ справедливо равенство  $R_x(p, r) = R_x(-p, -r)$ , т. е. она симметрична, поэтому при расчетах обычно вычисляют значения либо только для  $R_x(p, \pm r)$ , либо для  $R_x(\pm p, r)$ .

Таблица 38. Мощность зоны охристых руд по отработанному участку месторождения (в м)

№ профиля	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1								2,0	4,0	4,5				
2								2,2	3,7	4,4	3,3	4,9	1,2	1,0
3					4,2	5,2	1,5	2,0	7,0	1,5	4,0	8,5	6,0	4,0
4			5,1	3,4	3,7	7,2	4,7	3,8	2,0	1,5	1,6	6,0	2,0	
5	4,1	1,0	4,4	4,4	8,5	6,7	5,4	5,1	5,3	4,3	1,1	2,9	4,3	
6	5,0	4,3	7,3	4,3	10,2	8,7	9,1	9,6	6,6	1,0	3,6	1,0		
7	7,0	10,9	9,6	4,5	10,8	10,6	9,5	9,9	7,7	3,6	1,3	1,5		
8	2,5	2,7	12,1	8,0	8,2	7,2	9,7	9,9	9,3	3,7	8,2	4,0		
9	2,0	5,1	11,1	8,0	11,6	14,5	13,2	12,6	8,6	6,5	5,5	5,2	2,0	
10					7,7	11,7	21,8	15,5	9,1		5,0		2,7	
11	1,0	3,0	3,0	5,8	6,1	15,8	6,3							
12	4,3			1,0	8,0									

ДАКФ изображается либо в виде матрицы, либо в виде графика в изолиниях. Графическое изображение более наглядно отражает основные свойства геологических полей. По графику ДАКФ можно определить направления максимальной и минимальной изменчивости изучаемого поля, которые совпадают с направлениями максимального и минимального градиента изменения ДАКФ, а отношение градиентов по этим направлениям является оценкой коэффициента анизотропии.

В связи с этим ДАКФ используется для выбора формы, ориентировки и густоты сети наблюдений при проектировании геологоразведочных работ.

При расчете ДАКФ осредняются данные по всем профилям, поэтому по ней можно получить достаточно надежные количественные характеристики изменчивости изучаемого поля даже в тех случаях, когда количество замеров по каждому профилю невелико.

Обычно ДАКФ нормируют по дисперсии. В этом случае ее значения имеют размерность коэффициентов корреляции.

*Пример.* По отработанной части одного из месторождений никеля, в корне выветривания ультрамафитов были подсчитаны средние мощности зоны охристых руд по участкам площадью 100 × 100 м. Изменение этого параметра в плане изображено в виде матрицы исходных данных в табл. 38 и в виде топографии на рис. 63. Количество замеров по широтным и меридиональным профилям колеблется от 2 до 13, поэтому судить о характере изменчивости по отдельным профилям нельзя. В то же время по матрице (табл. 39) и графику (рис. 64) нормированной ДАКФ можно уверенно сделать следующие выводы:

- в пределах участка мощность зоны изменяется закономерно, так как коэффициенты корреляции между соседними замерами по широтным и меридиональным профилям равны соответственно 0,62 и 0,69, т. е. явно отличаются от нуля;

- изменчивость мощности в плане имеет анизотропный характер, так как изолинии ДАКФ вытянуты в одном направлении;

- минимальная изменчивость характерна для направления северо-запад — юго-восток, а максимальная для направления юго-запад — северо-восток;



Рис. 63. Мощность зоны охристых руд (в м) на месторождении никеля.  
По данным У. Лаваут

— коэффициент анизотропии, определенный по соотношению длинной и короткой оси поля ДАКФ, ограниченного изолинией 0,2, составляет:

$$K = \frac{L}{l} \simeq \frac{10}{5} \simeq 2;$$

— оптимальной сетью наблюдений для изучения мощности зоны охристых руд будет прямоугольная сеть с соотношением сторон ячейки 2:1, ориентированная длинной стороной ячейки в северо-западном направлении.

Таблица 39. Матрица нормированной ДАКФ мощности зоны охристых руд для значений  $R_x(\pm p, r)$

$p$	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	$r \rightarrow$	0	1	2	3	4	5	6	7
-6	-0,05	-0,26	-0,44	-0,52	-0,62	-0,62	-0,29	-0,33														
-5	0,06	-0,12	-0,25	-0,33	-0,59	-0,65	-0,58	-0,38														
-4	0,13	0,09	-0,04	-0,17	-0,49	-0,67	-0,54	-0,26														
-3	0,36	0,28	0,06	-0,10	-0,38	-0,41	-0,40	-0,30														
-2	0,43	0,30	0,16	0,05	-0,17	-0,28	-0,31	-0,32														
-1	0,69	0,52	0,31	0,15	-0,12	-0,25	-0,14	-0,29														
0	1,00	0,62	0,41	0,21	-0,01	-0,20	-0,15	-0,34														
1	0,69	0,52	0,42	0,31	-0,01	-0,20	-0,19	-0,22														
2	0,43	0,45	0,38	0,37	0,06	-0,13	-0,10	-0,11														
3	0,36	0,34	0,33	0,36	0,15	-0,04	-0,11	-0,11														
4	0,13	0,22	0,16	0,12	-0,04	-0,21	-0,29	-0,15														
5	0,06	0,08	-0,01	-0,04	-0,36	-0,35	-0,22	-0,02														
6	0,05	-0,09	-0,05	-0,08	-0,23	-0,04	-0,31	-0,11														

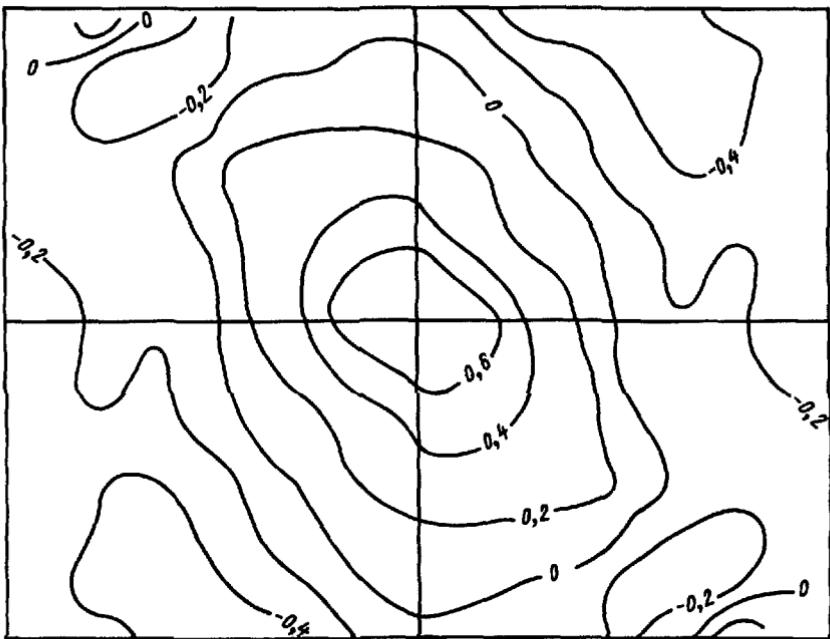


Рис. 64. Двумерная автокорреляционная функция мощности зоны охристых руд

#### 6.4. ОБЛАСТЬ ПРИМЕНЕНИЯ ПОЛИГАРМОНИЧЕСКИХ СЛУЧАЙНЫХ ФУНКЦИЙ В ГЕОЛОГИИ

Многие геологические объекты отличаются периодическим характером изменчивости свойств. Периодические колебания установлены в изменчивости пористости нефтегазоносных осадочных толщ, минеральном составе метаморфических пород. Определенная ритмичность отмечается также в расположении тектонических нарушений как в региональном, так и в локальном масштабе. При этом, однако, закономерные периодические колебания обычно осложняются нерегулярными случайными флюктуациями.

В качестве математической модели изменчивости подобных геологических объектов может быть использована полигармоническая случайная функция, т. е. случайная функция, математическое ожидание которой выражается тригонометрическим полиномом вида

$$Mx(l) = A_0 + \sum_{k=1}^{k=v} A_k \cos(\omega_k l + \phi_k), \quad (34)$$

где  $v$  — количество гармоник;  $A_k$ ,  $\omega_k$ ,  $\phi_k$  — соответственно амплитуды, частоты и фазы каждой гармоники;  $A_0$  — константа.

С помощью этой модели любой ряд значений исследуемого признака при равном расстоянии  $r$  между точками наблюдений может быть описан функцией

$$X(r) = Mx(r) + hx(r), \quad (35)$$

где  $hx(r)$  — случайная составляющая, осложняющая периодические колебания значений изучаемого признака.

Количественное описание изменчивости такого типа представляет собой широко распространенную в технике задачу выявления скрытых периодичностей на фоне случайных помех. Она заключается в разделении изменчивости на две составляющие и отыскании периодов и амплитуд гармоник закономерной составляющей.

К настоящему времени разработано большое количество методов выявления скрытых периодичностей, в которых используются различного вида преобразования исходного ряда наблюдения  $X(r)$ , позволяющие усилить в преобразованном ряде  $X'(r)$  роль периодической составляющей. Наиболее приемлемым из них является метод, основанный на оценке спектральной плотности дисперсии  $S_x(\omega)$ , получаемой в результате разложения в ряд Фурье корреляционной функции

$$S_x(\omega) = \frac{2}{\pi} \int_0^{\infty} K_x(r) \cos \omega r dr.$$

При изучении изменений параметра дискретной сетью наблюдений спектральная плотность заменяется линейчатым энергетическим спектром. Значения энергетического спектра  $d_k$  показывают, каким образом общая дисперсия признака распределяется между отдельными гармониками.

Энергетический спектр случайной полигармонической последовательности вида (34), рассчитанный по ограниченному количеству наблюдений, представляет собой сочетание резких пиков на частотах, соответствующих закономерным гармоническим колебаниям, с относительно постоянными для всех других частот значениями  $d_k$ , обусловленными случайной составляющей (рис. 65, в). Количество пиков соответствует числу гармоник  $v$ , а их высота равна половине квадрата амплитуды гармоники данной частоты, т. е.  $d_k = \frac{1}{2} A_k^2$ .

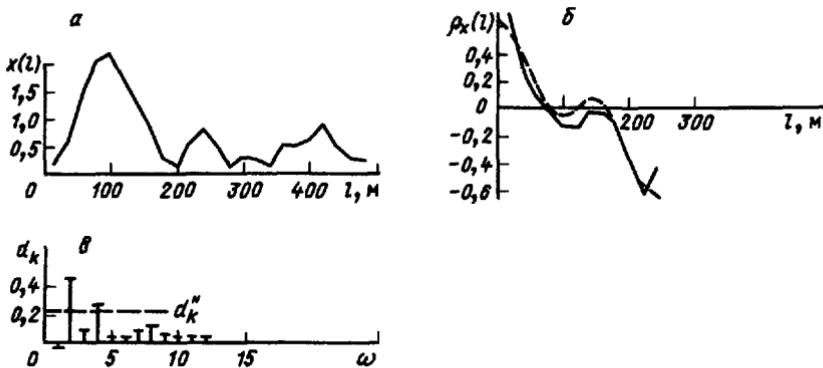


Рис. 65. Характеристики изменчивости линейного запаса железа вкрест профиля прибрежно-морской россыпи ильменит-магнетитовых песков:  
а — изменение линейного запаса по профилю; б — корреляционная функция; в — энергетический спектр

Величину периодов закономерных колебаний можно определить по формуле

$$T_k = (N-1)r/\omega_k,$$

где  $N$  — общее количество замеров по профилю;  $r$  — расстояние между соседними пунктами наблюдений.

Значения  $d_k$ , обусловленные случайной составляющей, имеют бета-распределение Фишера, плотность которого задается формулой  $P(x) = (n-1)(1-x)^{n-2}$ , где  $n$  — общее число значений спектра.

Математическое ожидание и стандартное отклонение этого распределения соответственно равны

$$\bar{d}_k = \frac{1}{n}; \quad \sigma_d = \sqrt{\frac{n-1}{n^2(n+1)}}.$$

Используя эти формулы, можно с заранее заданной доверительной вероятностью проверить гипотезу о принадлежности тех или иных пиков спектра к случайным флюктуациям.

Пики, превышающие предельное значение (например,  $\bar{d}_k + 3\sigma_d = d'_k$  при доверительной вероятности 99%), объясняются наличием гармоник и могут быть использованы для оценки периодов и амплитуд гармонической составляющей. Доля закономерной составляющей в общей дисперсии признака определяется путем суммирования аномальных значений спектра, из которых предварительно вычитается исправленное значение фона, определяемое по формуле

$$\bar{d}_{k(\text{испр})} = \frac{1 - \sum_{k=1}^m d'_k}{n-m}, \quad (36)$$

где  $d'_k$  — сумма аномальных значений спектра;  $m$  — количество аномальных значений спектра.

Периоды и амплитуды аномальных значений спектра служат основой для аппроксимации корреляционной функции и исходного ряда значений тригонометрическими полиномами.

*Пример.* Прибрежно-морская россыпь ильменит-магнетитовых песков разведана скважинами и шурфами по сети  $200 \times 20$  м. Характер изменения линейных запасов железа (произведенний содержания металла на мощность песков) по одному из разведочных профилей показан на рис. 65, а. По этому профилю рассчитана автокорреляционная функция (см. рис. 65, б) и энергетический спектр (см. рис. 65, в) линейного запаса.

Фоновое значение спектра равно

$$\bar{d}_k = \frac{1}{n} = 0,083,$$

а минимальное аномальное при доверительной вероятности 0,95 составляет

$$d'_k = 0,083 + 2 \sqrt{\frac{11}{12^2(12+1)}} \approx 0,235.$$

Таким образом, аномальными следует считать значения спектра на частотах 1 и 3, которые соответствуют колебаниям значений линейного запаса с периодами 480 и 160 м.

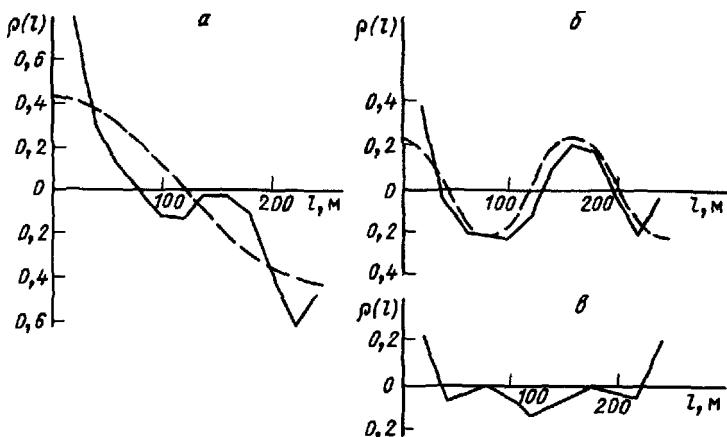


Рис. 66. Последовательное вычитание (а, б) из эмпирической корреляционной функции отдельных гармоник

Исправленное значение фона будет равно

$$\bar{d}_{k(\text{испр})} = \frac{1 - (0,460 + 0,260)}{12 - 2} = 0,028,$$

а сумма аномальных значений спектра после вычитания из них исправленного фона составляет 0,664. Это значит, что более половины дисперсии линейного запаса обусловлено закономерными колебаниями. Эмпирическая корреляционная функция (см. рис. 65, б) удовлетворительно аппроксимируется полиномом (рис. 66):

$$p_x(r) = 0,432 \cos \frac{2\pi r}{480} + 0,232 \cos \frac{2\pi r}{160}.$$

Однако модель, основанная на разделении наблюдаемой изменчивости на полигармоническую составляющую и случайную составляющую, отождествляемую со случайной величиной («белым шумом»), является несколько упрощенной. Наряду с периодическими изменениями для изменчивости геологических объектов обычно характерны закономерные, но непериодические локальные изменения. Поэтому обычно  $hx(r)$  в формуле (35) представляет собой не случайную величину, а реализацию стационарной случайной функции.

Приближенная оценка корреляционной функции составляющей  $hx(r)$  может быть получена путем последовательного вычитания из эмпирической корреляционной функции гармонических составляющих.

Для приведенного выше примера остаточный график корреляционной функции члена  $hx(r)$ , построенный путем последовательного вычитания гармонических колебаний с частотами 1 и 3, характеризуется незначительными отклонениями от нуля почти для всех значений  $r$  (рис. 66, в). Исключение составляют только крайние точки графика. Отклонение в последней точке, по-видимому, объясняется малым количеством замеров в исследуемом ряду, а отклонение в первой точке указывает на наличие небольших по амплитуде высокочастотных колебаний с переменным периодом.

Таким образом, с помощью спектра амплитуд наблюдаемую изменчивость любого свойства геологических объектов можно разделить на три составляющие:

— координированную, обусловленную закономерными на всем протяжении изучаемого участка изменениями свойства и описываемые с помощью тригонометрического полинома;

— коррелированную, описывающую закономерные в локальной области изменения свойства с помощью корреляционной функции, зависящей только от расстояния между точками наблюдений;

— случайную, описывающую незакономерные случайные колебания свойства, которые могут рассматриваться как реализация случайной величины («белый шум»).

Координированная составляющая, описывая общие закономерности изменения свойств в пространстве, отражает главные особенности строения геологических объектов. С помощью координированной составляющей можно:

— выделять элементы неоднородности в строении объектов исследования на различных стадиях их изучения;

— предсказать наиболее вероятные значения свойства в любой точке исследуемого профиля;

— оценивать степень выявления закономерных пространственных изменений в зависимости от густоты сети наблюдений.

Коррелированную закономерную составляющую нельзя использовать для геометризации элементов неоднородности, так как она характеризует лишь средний градиент изменений изучаемого свойства в пределах расстояния корреляции. Она используется для решения задач, связанных с распространением эмпирических данных на сравнительно малый объем изучаемого объекта.

Случайная составляющая отражает незакономерные флюктуации параметров и должна учитываться при расчете точности средних значений параметров.

Способ выражения закономерностей (координированной) составляющей изменчивости свойств геологических объектов с помощью спектрального разложения имеет ряд существенных преимуществ по сравнению с наиболее распространенными способами сглаживания и тренд-анализа:

1) математическая модель изменчивости, на которой основан этот способ, отвечает интуитивному представлению о наличии в изменчивости реальных геологических объектов закономерностей различного порядка;

2) способ не требует априорных допущений о количестве этих порядков, так как их число определяется самим методом;

3) спектральное разложение корреляционной функции является эффективным селективным преобразованием для выделения закономерностей любого масштаба;

4) спектр амплитуд позволяет разделить закономерные изменения параметров на периодические и непериодические, а обуславливающие их геологические факторы на общие и частные;

5) спектр амплитуд показывает, каким образом общая дисперсия признака распределяется между закономерностями разного масштаба и позволяет оценить значимость выявленных закономерностей.

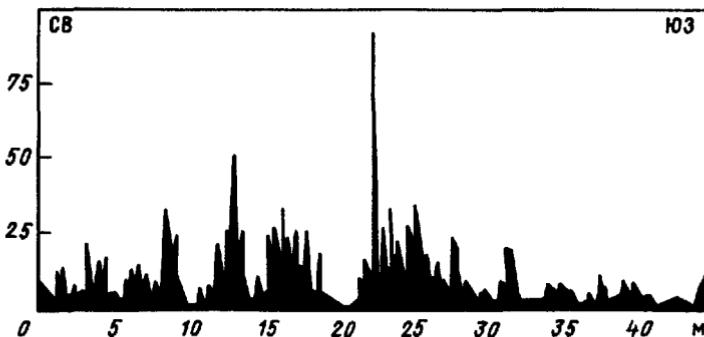


Рис. 67. Распределение трещин в интрузии. По В. А. Апродову.  
По горизонтали отложены расстояния от контакта интрузии, по вертикали — количество трещин, приходящихся на 1 м<sup>2</sup> обнажения

*Пример.* На основании изучения закономерностей распределения трещин в некоторых интрузиях В. А. Апродовым было высказано предположение о волновой природе явлений трещиноватости. Им было подмечено, что изменение густоты трещин от центра интрузии к ее контакту происходит волнообразно с определенным периодом (рис. 67). Это явление В. А. Апродов объяснил отражением упругих волн от контактных поверхностей интрузии и появлением эффекта так называемых «стоячих волн». В кучностях «стоячих волн» развивается более интенсивная трещиноватость, а в узлах порода сохраняет относительную монолитность.

На графике спектра амплитуд плотности трещин (рис. 68) отчетливо выделяются два пика на частотах 1 и 13. Пик на частоте 1 соответствует общей тенденции и уменьшению плотности трещин от центральной части исследуемого профиля к его концам, а пик на частоте 13 соответствует колебаниям с периодом

$$T = \frac{L}{\omega} = \frac{50 \text{ м}}{13} \simeq 4 \text{ м.}$$

Повышенное значение  $d_k$  отмечается также на частоте 4, соответствующей колебаниям с периодом 11 м. Однако для уверенного отнесения этого значения к аномальным необходимо выполнить расчеты по формуле (36). В данном примере фоновое значение спектра амплитуд

$$\bar{d}_k = \frac{1}{74} \simeq 0,0135,$$

а стандартное отклонение значений спектра от фонового при условии отсутствия закономерной составляющей будет

$$\sigma_d = \sqrt{\frac{73}{74^2 \cdot 75}} = 0,0133.$$

Следовательно, минимальное аномальное значение при доверительной вероятности 0,99 составит

$$d'_k = \bar{d}_k + 3\sigma_d = 0,0135 + 0,0399 = 0,0534.$$

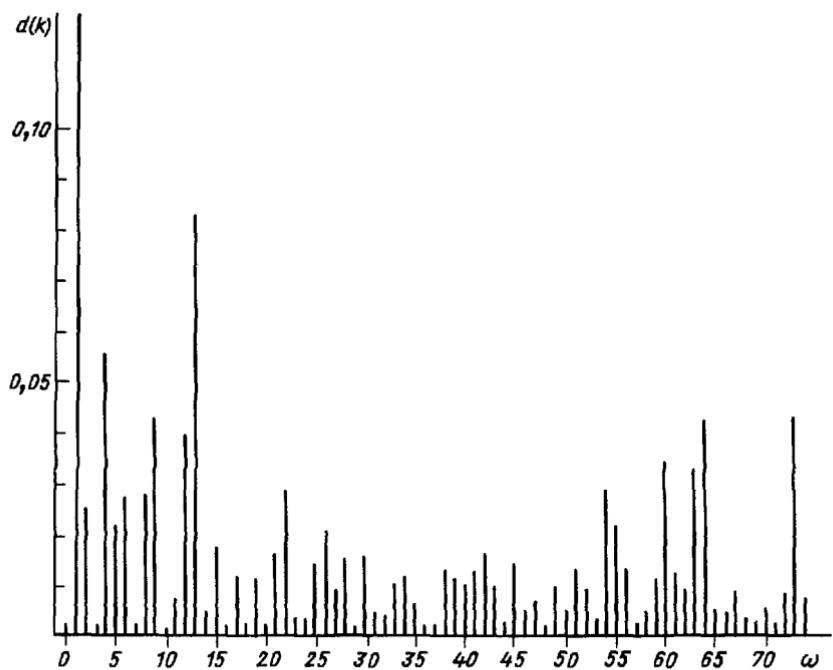


Рис. 68. Энергетический спектр распределения трещин в интрузии

Значение спектра на частоте 4 немножко превышает минимальное аномальное, что позволяет признать реальным существование периодических колебаний плотности трещин с периодом 11 м. Спектр позволил также установить, что ведущая роль в трещиноватости принадлежит тенденции к снижению количества трещин по мере приближения к контакту интрузии, а из периодических изменений более отчетливо проявились высокочастотные колебания с периодом 4 м, в то время как колебания с частотой 11 выражены слабее.

Эмпирические спектры амплитуд позволяют находить вероятность и величину возможных ошибок в определении средних значений за счет периодических колебаний, кратных или равных шагу основной сети наблюдений, а также выявлять запрещенные разведочные сети, шаг которых близок к периодам колебаний, доминирующими на том или ином уровне строения месторождения.

Для более глубокого ознакомления с математическим аппаратом теории случайных функций может быть рекомендована книга М. Кендалла и А. Стьюарта; математический аппарат гармонического анализа подробно рассмотрен в книге А. А. Никитина. Примеры практического использования случайных функций при геологическом картировании приведены в работе Б. М. Роненсона, В. М. Ройзенмана и Н. А. Ортенберга. Опыт использования данной модели при решении задач разведки месторождений обобщен в работе А. Б. Каждана.

## *Контрольные вопросы*

1. Что такое случайный процесс, случайная последовательность и случайное поле?
2. Почему случайные функции можно использовать для описания свойств геологических объектов и процессов?
3. Какие свойства геологических образований и процессов описывают автокорреляционные функции?
4. Какие геологические задачи можно решать с помощью автокорреляционных функций?
5. Что такое расстояние предельной корреляции? Для чего используется этот показатель?
6. Какие геологические задачи можно решать с помощью взаимных корреляционных функций (ВКФ)?
7. Какие геологические задачи можно решать с помощью двумерных автокорреляционных функций (ДАКФ)?
8. Свойства каких геологических объектов можно описывать полигармоническими случайными функциями?
9. Что такое энергетический спектр и спектральная плотность дисперсии?
10. Как можно выявить аномальные значения энергетического спектра?
11. Как можно с помощью энергетического спектра разделить наблюдаемую изменчивость на случайную и закономерную составляющие?
12. Какие геологические задачи можно решать с помощью гармонического анализа?

## **ГЛАВА 7** **ФАКТОРЫ, ОПРЕДЕЛЯЮЩИЕ ВЫБОР** **И ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ** **МАТЕМАТИЧЕСКИХ МЕТОДОВ**

---

Выбор и эффективность использования математических методов при решении геологических задач зависят от многих факторов, в числе которых ведущее значение имеют:

- характер решаемых геологических задач;
- свойства изучаемых геологических объектов;
- методика изучения геологических объектов;
- полнота и объективность геологического анализа при выборе геологической модели и интерпретации полученных результатов.

При этом характер решаемой задачи обычно определяет тип используемой математической модели, а свойства геологических объектов и методика их изучения влияют главным образом на эффективность тех или иных конкретных методов решения.

### **7.1. ВЛИЯНИЕ ТИПА ГЕОЛОГИЧЕСКОЙ ЗАДАЧИ НА ВЫБОР МАТЕМАТИЧЕСКОЙ МОДЕЛИ**

С позиций системного подхода все геологические задачи, в известной степени условно, можно разделить на две группы.

*В первую группу* целесообразно объединить задачи, решение которых основано на расчете средних характеристик свойств (параметров) геологической системы. При этом принимается, что геологическая система однородна, а особенность взаимосвязи между ее

элементами не учитывается. Ведущую роль при решении таких задач играют одномерные статистические модели.

Во вторую группу можно выделить задачи, при решении которых используются сведения о *характере взаимосвязи элементов* в геологической системе, т. е. о ее *внутреннем строении (структуре)*. Если в качестве элементов системы рассматриваются ее отдельные свойства (параметры), то используются главным образом двумерные и многомерные статистические модели. Когда элементами системы являются отдельные участки геологических объектов или определенные временные отрезки геологических процессов, т. е. свойства изучаемых объектов рассматриваются как пространственные или временные переменные, используются горно-геометрические модели, методы тренд-анализа или случайные функции.

Однако для решения большинства задач в той или иной степени требуются сведения как о средних параметрах геологической системы, так и о ее структуре. Поэтому часто их нельзя решать в рамках одной математической модели.

Необходимо помнить, что любая математическая модель является упрощенным аналогом изучаемого объекта и исходит из определенных априорных допущений относительно его свойств. Поэтому нередко выбор наиболее эффективного метода решения требует предварительного изучения этих свойств с помощью различных моделей.

На рис. 69 приведена схема возможных сочетаний различных математических методов при решении наиболее типичных геологических задач. Сплошными линиями отмечены основные методы решения уже формализованной (математической) задачи, а пунктирными — методы, которые применяются на этапах предварительного анализа исходных данных при переходе от геологической к опробуемой и выборочной совокупности, выборе наиболее адекватной математической модели, а также на этапе интерпретации полученных результатов (см. разд. 1.6).

Наиболее простыми являются задачи, связанные с *нахождением точечных оценок средних параметров* геологических объектов. В эту группу попадают задачи, связанные с оценкой качества и подсчетом запасов полезных ископаемых, определением прочностных свойств горных пород и руд для инженерных расчетов при выборе оптимальной технологии добычи и переработки полезных ископаемых и т. п. Средние характеристики свойств геологических объектов не зависят от их структуры, поэтому такие задачи можно решать в рамках одномерной статистической модели. Свойства изучаемого объекта и методы его изучения в данном случае влияют лишь на выбор конкретного вида модели (статистического закона распределения). Необходимость в привлечении других моделей может возникнуть лишь на этапе проверки адекватности геологической и опробуемой совокупности. Для суждения о случайному расположении выборочных наблюдений относительно элементов внутреннего строения геологической системы, представляющей собой определенный объем недр, целесообразно использовать простейшие модели, описывающие пространственные геологические переменные — графические (карты, проекции, разрезы) или планы изолиний (топографии).

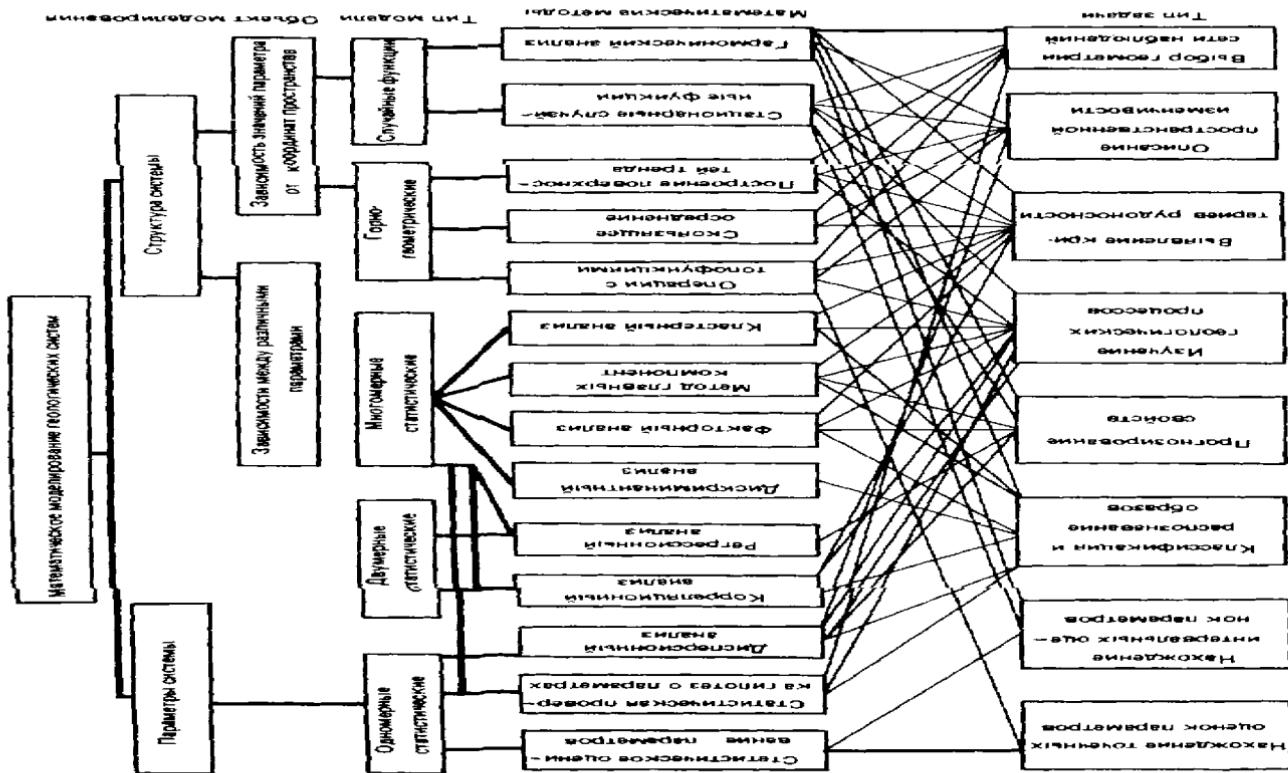


Рис. 69. Схема возможных сочетаний математических методов геологических задач

Гораздо более сложными являются задачи *нахождения интервальных оценок параметров* геологических объектов.

Если выборочная совокупность сформирована заведомо случайным образом и не содержит информации о пространственной структуре изучаемой геологической системы, то величина возможной погрешности средних параметров будет зависеть только от интенсивности изменчивости (дисперсии) изучаемого свойства и объема выборки. В этих условиях интервальные оценки можно найти с помощью методов, рассмотренных в гл. 2. Такая ситуация возникает, например, при расчете случайных среднеквадратичных погрешностей анализов проб, характеризующих качество работы лабораторий.

Однако в большинстве случаев изучаемые геологические объекты представляют собой тела, имеющие определенный объем, и их свойства должны рассматриваться как пространственные переменные. При этом выборочная совокупность обычно представляет собой серию замеров изучаемого свойства в определенных точках координат пространства и содержит сведения не только об интенсивности наблюдаемой изменчивости, но и о ее характере — соотношении случайной и закономерной составляющей. В этом случае для нахождения интервальной оценки необходимо с помощью методов тренд-анализа, случайных функций или гармонического анализа проверить гипотезу об отсутствии в наблюдаемой изменчивости закономерной составляющей. Если эта гипотеза отвергается, то с помощью тех же методов оценивается величина закономерной составляющей и в статистические формулы, приведенные в разд. 2.6, вносится соответствующая поправка.

Решение многих геологических задач основано на принципе аналогии, когда недостаток информации об объекте изучения восполняется сведениями, полученными при изучении сходных объектов. Для определения объектов-аналогов используются различные классификации и группировки. Поэтому задачи классификации и распознавания образов очень широко распространены в геологической практике. В качестве классификационных признаков могут выступать различные свойства геологических систем, характеризующие как параметры, так и структуру. Поэтому методы решения данного типа задач весьма разнообразны.

Под *классификацией* подразумевается разделение совокупности изучаемых объектов на естественные сообщества (классы), т. е. совокупности объектов, обладающих хотя бы одним общим признаком. Этот признак обычно задается априорно, исходя из существа геологической задачи, решаемой на основе классификации. Классификационным признаком могут служить: средние значения, медианы или моды одного или нескольких свойств геологических объектов, характеристики рассеяния — дисперсии или коэффициенты вариации, тип корреляционной связи между различными свойствами объектов, характер изменчивости свойств объектов в пространстве или процессов во времени и т. д. Классифицироваться могут совокупности объектов или различные свойства одного объекта. Количество классов сгруппирования при этом может быть заранее неизвестно.

Под *распознаванием образов* понимается процесс классификации, когда изучаемый объект относится к одному из эталонных классов с заранее известными свойствами.

Простейшие задачи классификации заключаются в том, что сравниваются два объекта по одному признаку. Их решение основано на проверке гипотез о равенстве параметров распределения этого признака с помощью статистических критериев согласия. Внутренняя структура сравниваемых объектов при этом не учитывается. Для решения этих задач используются одномерные статистические модели (см. разд. 2.8). Необходимость в привлечении других моделей может возникнуть лишь на этапе определения границ геологической совокупности и при оценке ее адекватности опробуемой совокупности.

Для сравнения двух геологических объектов по комплексу свойств применяются многомерные статистические модели. При этом используются многомерные аналоги параметрических и непараметрических критериев согласия, рассмотренных в разд. 2.8,—критерии Пури—Сена—Тамуры, Джеймса—Сю и др. В качестве меры сходства или различия в этом случае можно использовать также коэффициенты сходства (Сокала—Миченера, Рао—Рассела и др.) или коэффициенты расстояния (Махalanобиса, Минковского и др.), которые рассчитываются с помощью алгоритмов кластерного анализа.

Иногда необходимо сравнить два или более объекта по характеру корреляционных связей между их различными свойствами. Такая ситуация возникает, например, при оценке комплексных геохимических аномалий, определении глубины эрозионного среза месторождений, выделении рудных формаций и т. д. Задачи такого типа решаются путем проверки гипотез о равенстве ковариационных матриц с помощью многомерных критериев согласия (Пури—Сена, Кульбака и др.) или методами кластерного анализа путем сравнения дендрографов и дендрограмм (см. разд. 4.6).

Значительно реже в качестве классификационных признаков используются количественные характеристики пространственной изменчивости геологических объектов или изменения параметров геологических процессов во времени. Пример использования автокорреляционной функции аномального магнитного поля для классификации крупных блоков земной коры приведен в разд. 6.2. Автокорреляционные функции и спектры амплитуд используются также при морфологической классификации тел полезных ископаемых и сопоставлении разрезов метаморфических толщ.

Задачи распознавания образов, связанные главным образом с прогнозной оценкой территорий при поисках месторождений полезных ископаемых (см. разд 4.8), имеют достаточно надежные решения лишь при использовании многомерных наблюдений. Для их решения широко используются методы дискриминантного и факторного анализов.

Таким образом, для решения задач классификации и распознавания образов применяются практически все типы математических моделей. Поэтому при выборе математического метода необходимо четко представлять целевое назначение классификации, а наиболее

эффективный метод нередко можно найти только путем оценки результатов, полученных разными методами.

Некоторые свойства геологических объектов иногда не поддаются непосредственному изучению или их изучение затруднено по техническим или экономическим причинам. Поэтому в геологической практике широко распространены приемы *прогнозирования таких свойств* на основе использования обычно существующих зависимостей между элементами (свойствами) геологических систем.

Предсказание неизвестного признака по значениям одного или нескольких известных признаков осуществляется с помощью регрессионного анализа, рассмотренного в главах, посвященных двумерным и многомерным статистическим моделям (см. разд. 3.3, 4.7). Основным условием эффективного применения уравнений регрессии для целей предсказания является однородность изучаемой геологической совокупности. Однако это условие соблюдается далеко не всегда. Характер корреляционных связей между содержаниями химических элементов в пределах рудных тел или интрузивных образований, например, может изменяться в связи с их геохимической, минералогической и текстурной зональностями. Поэтому для успешного применения регрессионных методов необходимо предварительно изучить характер пространственной изменчивости свойств геологических объектов с помощью горно-геометрических моделей или используя случайные функции.

К прогнозированию можно отнести приемы экстраполяции измеренных значений свойств за пределы изученной области, а также приемы нелинейной интерполяции параметров между точками наблюдений, основанные на количественном описании закономерной пространственной изменчивости. Для их решения используются методы гармонического анализа, цепи Маркова, аппроксимация полиномами (тренд-анализ).

При *изучении геологических процессов* используются методы как статического, так и динамического моделирования.

Применение статического моделирования для этих целей основано на том, что особенности геологических процессов в той или иной мере отражаются в структуре геологических образований. Поэтому математические модели, описывающие зависимости между различными свойствами геологических образований или пространственную изменчивость этих свойств, можно использовать и для ретроспективного изучения прошедших геологических процессов. Для этого применяются методы двумерного и многомерного корреляционного анализа, факторный анализ и метод главных компонентов, кластерный анализ, тренд-анализ, гармонический анализ и т. д.

Так, например, с помощью кластерного или факторного анализа удается выявить геохимические и минеральные ассоциации в рудах, что позволяет судить о физико-химических условиях их образования, а методы, описывающие пространственную изменчивость (тренд-анализ, случайные функции и т. п.), дают возможность проследить, как эти условия менялись в пространстве — с глубиной или при удалении от основных рудоконтролирующих структур. При изучении

осадочных образований этот же комплекс математических методов позволяет по минеральному и гранулометрическому составу обломочной фракции определить направление сноса обломочного материала, область питания бассейна осадконакопления, палеогеографическую и палеотектоническую обстановки во время осадконакопления и т. д.

С помощью гармонического анализа удается выявить периодичности в изменении свойств геологических образований, что часто указывает на циклический характер геологических процессов.

При динамическом моделировании, основанном на сравнении теоретических реализаций геологического процесса с фактическими свойствами реальных геологических объектов, применяются методы статистической проверки гипотез с помощью одномерных или многомерных критериев согласия, а также методы корреляционного, кластерного и факторного анализов.

Решение многих задач методики поисков и разведки месторождений полезных ископаемых основано на использовании пространственной, а иногда и генетической связи полезных ископаемых с определенными, обычно легко картируемыми, геологическими образованиями — горными породами определенного состава или возраста, тектоническими нарушениями и т. д. Для выявления таких образований, называемых *поисковыми критериями и рудоконтролирующими факторами*, применяются математические методы, позволяющие установить зависимость между частотой встречаемости рудных образований или их параметрами и различными признаками, характеризующими геологическое строение тех или иных участков недр. С этой целью используются методы однофакторного, двухфакторного и многофакторного дисперсионного анализа, двумерный и многомерный корреляционные анализы, факторный анализ, а также методы описания пространственных переменных — топофункции, тренд-анализ, методы моделирования дискретных случайных полей (см. разд. 5.11).

Нередко в геологической практике возникает необходимость иметь общую картину изменения свойств геологических объектов в пространстве. Например, планы, отражающие изменения содержания полезных компонентов в рудах на различных гипсометрических уровнях месторождения, используются для подсчета запасов, проектирования горнодобывающего предприятия, планирования добычи, а графические изображения геофизических полей в виде изолиний служат основой для построения геологических карт и выделения аномалий, соответствующих рудным объектам. Построение таких графических документов производится с помощью топофункций или методами тренд-анализа. Однако для проверки правомерности использования этих методов привлекаются также статистические модели (проверка гипотез о наличии тренда) и характеристики случайных функций, а при выборе оптимального размера сглаживающего окна или вида аппроксимирующего полинома используется гармонический анализ.

С задачами выбора оптимальной сети наблюдений геолог сталкивается постоянно. От правильного выбора геометрии и густоты сети наблюдений зависит эффективность геологоразведочных работ.

Применение математических методов для решения таких задач имеет длительную историю, а предлагавшиеся методы весьма разнообразны. Вначале предпринимались попытки использовать одномерные статистические модели и топофункции. Однако многолетний опыт их применения убедительно доказал, что надежные решения этой задачи можно получить только на основе математических моделей, позволяющих количественно описать как случайную, так и закономерную составляющие наблюдаемой изменчивости свойств изучаемых геологических объектов. Поэтому в последние годы для их решения используются характеристики случайных функций — одномерные и двумерные автокорреляционные функции, а также энергетические спектры и спектры амплитуд (см. разд. 6.3, 6.4).

Рассмотренные в данном разделе типовые задачи и методы их решения отражают далеко не все направления применения математики в геологии. Однако даже из приведенных примеров ясно, что выбор наиболее эффективного математического метода решения геологической задачи является довольно сложной проблемой, для решения которой кроме знания различных разделов математики необходима основательная профессиональная подготовка.

## 7.2. ВЛИЯНИЕ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ НА ВЫБОР И ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ МАТЕМАТИЧЕСКИХ МЕТОДОВ

Основными свойствами геологических объектов, влияющими на выбор математических методов и эффективность их использования, являются:

- структура (внутреннее строение) объекта;
- размерность пространства, в котором изучаются свойства объекта, и количество одновременно изучаемых признаков;
- тип переменных величин, отражающих свойства объекта, и вид их частотных распределений.

Многие математические методы разработаны исходя из определенных априорных предположений о *структуре изучаемого объекта*.

В основу большинства математических моделей и методов заложено предположение об *однородности изучаемого объекта*. Принято считать, что структура объекта принципиально не меняется в различных его частях. Нарушение этого условия снижает эффективность математических методов, а иногда приводит к неправильным решениям.

В случае применения одномерных статистических моделей неоднородность изучаемого объекта часто является причиной возникновения сложных полимодальных частотных распределений признаков, что исключает возможность применения при проверке геологических гипотез параметрических критериев согласия, обладающих большей мощностью по сравнению с их непараметрическими аналогами.

При решении геологических задач с помощью двумерных статистических моделей неоднородность исследуемого объекта приводит к появлению сложных нелинейных зависимостей (см. разд. 3.3) и ложной корреляции. Например, при изучении геохимических полей

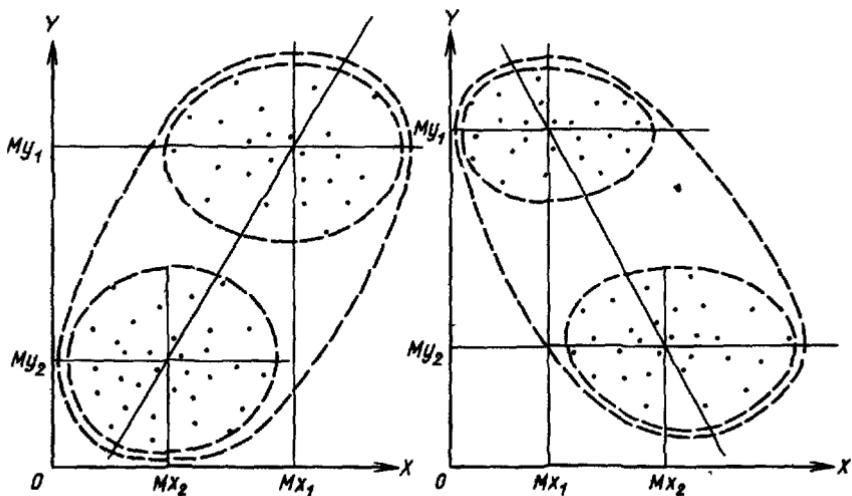


Рис. 70. Возникновение ложной корреляции в неоднородных совокупностях: слева — прямая ложная корреляция; справа — обратная ложная корреляция

ложная корреляция возникает за счет наложения эпигенетической (аномальной) составляющей на сингенетическую (фоновую) составляющую (рис. 70).

Для выбора оптимальной сети наблюдений часто используют автокорреляционные функции и энергетические спектры, рассчитанные по участкам детализации, где сеть наблюдений значительно гуще, чем по всему объекту. При этом предполагается, что характер изменчивости изучаемого свойства по участку детализации адекватно отражает структуру всего объекта. Неоднородность объекта в этих условиях может привести к значительным ошибкам аналогии.

Наиболее простые и поэтому широко распространенные алгоритмы корреляционного регрессионного, дискриминантного, факторного и кластерного анализа основаны на предположении о линейном характере зависимостей между изучаемыми свойствами, а при описании пространственной изменчивости объектов с помощью топофункций или поверхностей тренда предполагается, что соблюдается условие однозначности и конечности изучаемой величины, а ее изменения имеют непрерывный и плавный характер.

Такие предположения о структуре реальных геологических объектов соблюдаются далеко не всегда. Поэтому выбор математического метода решения геологической задачи всегда должен базироваться на проверке соответствия реальных свойств геологического объекта тем допущениям, на которых основан тот или иной метод. Для проверки гипотез о статистической однородности выборочных данных, линейности корреляционных зависимостей, плавности и непрерывности изменения свойств в пространстве можно использовать простые и наглядные графические способы, рассмотренные в разд. 2.2, 2.3, 5.5, 5.7, подкрепляя их аналитическими расчетами в задачах, связанных с принятием наиболее ответственных решений.

При изучении свойств геологических объектов, как пространственных переменных, выбор методов их описания зависит от размерности изучаемого пространства.

Сложность математического моделирования и вычислительных операций заметно увеличивается при переходе от одномерного пространства (случайные последовательности) к двумерному (случайные поля), а тем более к трехмерному. Поэтому в процессе геолого-математического моделирования следует стремиться к снижению размерности пространства, прежде всего путем перехода к осредненным числовым характеристикам или разделения общей задачи на несколько частных.

Так, например, при изучении пространственного размещения содержаний полезного компонента в объеме рудного тела обычно выделяют две самостоятельные задачи: 1) описание изменчивости содержаний по каждому разведочному пересечению по интервальным пробам (реализации случайных последовательностей) и 2) описание их изменчивости в плоскости рудного тела по осредненным значениям содержаний в каждом сквозном пересечении (реализация двумерного случайного поля). Аналогичная проблема возникает и при изучении объектов, характеризующихся несколькими свойствами, каждое из которых условно принимается в качестве координаты признакового пространства. При решении многих геологических задач описание изучаемых объектов с помощью многих признаков и их численных значений становится весьма сложным, но даже большое число разнообразных признаков не гарантирует учета всех, наиболее важных свойств изучаемого объекта. Опыт применения подобных многомерных моделей свидетельствует о том, что основная часть признаков почти не содержит полезной информации, а все они в совокупности создают «информационный шум», на фоне которого часто теряются полезные сигналы от единичных, наиболее информативных признаков. В связи с этим возникает задача снижения размерности геологических переменных путем отбраковки большинства неинформативных признаков.

В практике геолого-математического моделирования в настоящее время используются весьма разнообразные методы снижения размерности и определения информативности признаков. Ведущее место среди них занимают методы факторного анализа. Сведение многомерных случайных величин к сочетаниям трех-, а еще лучше двумерных моделей позволяет применять не только более простые аналитические методы, но и широко использовать более наглядные и простые графические модели исследования. Так, например, при изучении рудных тел, в качестве пространственной переменной исследуется линейный запас, т. е. произведение мощности на содержание по разведочному пересечению, а в практической геохимии — коэффициент зональности, вычисленный как отношение сумм или произведений, содержаний «надрудных» и «подрудных» элементов — индикаторов оруденения.

В геологической практике для описания свойств геологических объектов используются случайные величины различного типа — скалярные и векторные, непрерывные и дискретные. Тип переменной

величины определяет вид применяемой статистической модели, т. е. закон распределения, методы решения геологических задач и употребляемые при этом статистические критерии. Однако в ряде случаев определенное свойство геологического объекта можно выразить величинами различного типа. При увеличении объема выборочных совокупностей свойства распределений дискретных величин приближаются к распределениям непрерывных величин. Поэтому, например, при нахождении интервальных оценок коэффициента рудоносности можно пользоваться нормальной аппроксимацией биномиального закона. При изучении закономерностей распределения точечных объектов на плоскости (дискретное случайное поле) можно перейти к полю непрерывной величины — избыточной плотности объектов, путем подсчета количества объектов в элементарных ячейках (см. разд. 5.11).

Область применения статистических методов, основанных на использовании параметрических критериев согласия, ограничиваются объектами, для которых частотные распределения изучаемых признаков согласуются с определенными законами. Однако несоответствие эмпирического распределения теоретическому не всегда является непреодолимым препятствием для применения таких методов. В ряде случаев эмпирические распределения можно приблизить и теоретическим путем соответствующих преобразований — логарифмирования, прибавления или вычитания константы и т. д. (см. разд. 2.4).

При решении задач предсказания неизвестных свойств с помощью уравнений регрессии логарифмирование или возвведение в степень изучаемых случайных величин часто позволяет устраниТЬ нелинейность связи между ними, что значительно упрощает расчеты (см. разд. 3.3).

### 7.3. ВЛИЯНИЕ МЕТОДИКИ ИЗУЧЕНИЯ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ НА ВЫБОР И ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ МАТЕМАТИЧЕСКИХ МЕТОДОВ

Геологические объекты обычно изучаются выборочными методами, поэтому при математической обработке полученных данных необходимо учитывать не только их природные свойства, но и особенности методики их изучения:

- принятую схему опробования;
- геометрию сети наблюдений;
- геометрию проб;
- объем выборочной совокупности;
- шкалу измерения изучаемой величины.

Из всех применяемых в геологии схем опробования (см. разд. 1.2) наиболее распространенным является равномерное опробование. Оно обеспечивает одинаковую детальность изучения всех частей опробуемого объекта и наиболее эффективно для выявления направленных изменений свойств при наличии в их изменчивости закономерных составляющих. Равномерное расположение точек наблюдений или проб существенно упрощает вычислительные операции при тренд-анализе, расчете основных числовых характеристик моделей,

использующих случайные функции, и при построении горно-геометрических моделей. Однако следует помнить, что при наличии периодической составляющей в изменчивости изучаемых признаков, равномерная сеть может привести к значительным ошибкам в оценке средних параметров, если расстояние между наблюдениями будет близким или кратным по отношению к длине периода закономерных колебаний. Для выявления таких «запрещенных» сетей целесообразно использовать наряду с равномерным многостадийное опробование, когда на отдельных участках, представительно характеризующих весь объект, сеть наблюдений сгущается в несколько раз.

При случайному опробовании пункты наблюдений или места отбора проб располагаются без видимой пространственной закономерности. Для целей случайногого опробования координаты точек наблюдений или отбора проб выбираются с помощью таблиц случайных чисел или другими способами, исключающими появление систематических пространственных составляющих. Случайное опробование применяется в тех случаях, когда исследователя не интересуют пространственные закономерности изменения изучаемых свойств, а задача сводится только к оценке их средних статистических характеристик. Так, например, случайный способ отбора может использоваться при составлении выборок проб для целей их контрольных анализов.

Детальность представлений о строении изучаемого геологического объекта зависит от густоты сети дискретных наблюдений. Чем гуще эта сеть, т. е. меньше расстояния между смежными пунктами наблюдений, тем более глубокий уровень в строении природных геологических образований может быть выявлен путем анализа и соответствующего группирования результатов единичных наблюдений.

Если группа единичных наблюдений располагается в пределах одного структурного элемента, то между значениями признака в смежных пунктах наблюдений возникают более или менее отчетливые автокорреляционные связи, а в характеристике наблюданной изменчивости признака отчетливо проявляется неслучайная (закономерная) составляющая. Если же проводить наблюдения по более редкой сети, так чтобы смежные пункты приходились на различные элементы неоднородности, корреляционные связи между значениями изучаемого признака не проявятся, а в характеристике изменчивости признака будет преобладать случайная ее составляющая. Поэтому изменчивость свойств одного и того же природного геологического объекта, например участка рудной залежи, может быть оценена как случайная при редкой сети и как неслучайная при густой сети наблюдений (рис. 71).

В связи с этим в практике геологоразведочных работ нередко возникают ситуации, когда на ранних стадиях изучения свойства геологических объектов описываются с помощью статистических моделей, а на более поздних, когда появляется информация о пространственной структуре этих объектов, возникает необходимость в использовании более сложных моделей и методов — случайных функций, тренд-анализа, гармонического анализа.

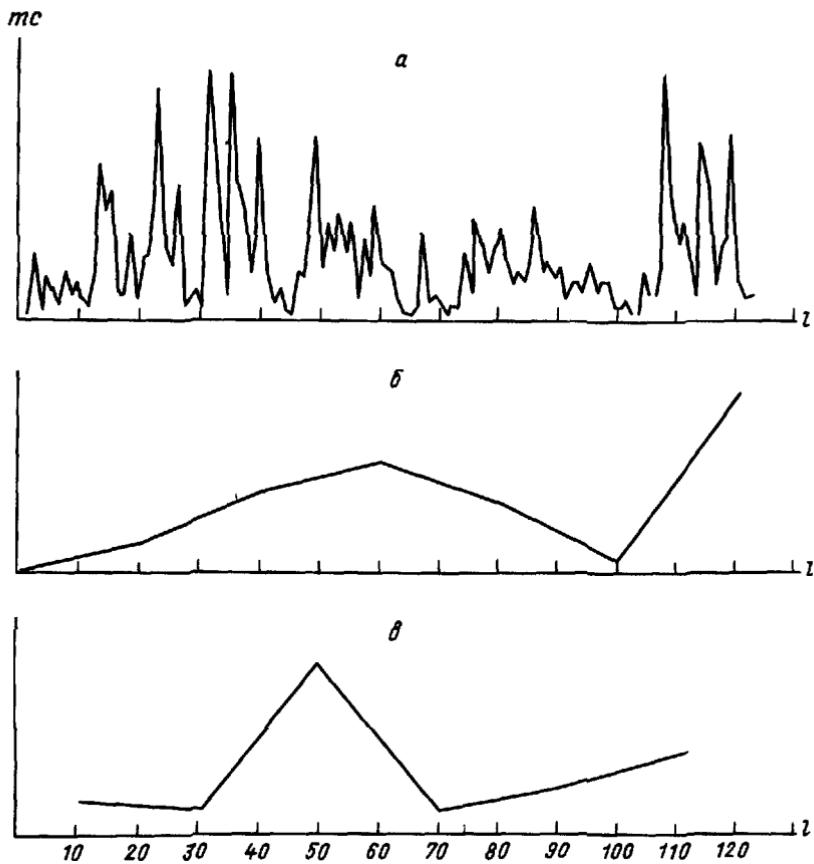


Рис. 71. Графики изменения линейного запаса  $ms$  по простиранию рудной залежи:  
а—расстояние между замерами 5 м; б, в—расстояние между замерами 100 м

При проведении статистической обработки экспериментальных геологических данных всегда следует помнить о том, что характеристики изменчивости, а также функции распределения изучаемых свойств отражают не только природную изменчивость, но и условия проводимого эксперимента и зависят от геометрии проб. Так, например, гиперболовидное распределение золота по секционным бороздовым пробам сменяется асимметричным логнормальным для «сквозных» проб и почти симметричным для «площадных» проб (рис. 72).

При использовании методов корреляционного и регрессионного анализов в геологии и разведке также необходимо учитывать влияние геометрии проб—их объемов, масс, размеров и ориентировки, на количественные параметры выборочных оценок связи. Так, например, при выявлении типа и силы связи между содержаниями двух элементов в рудах какого-либо месторождения следует пользоваться данными анализов таких проб, геометрия которых отвечает условиям

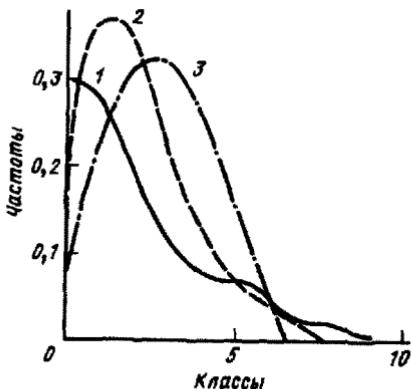


Рис. 72. Изменение асимметрии кривых распределения золота в зависимости от размера проб. По Д. А. Зенкову:

1—по секционным пробам; 2—по сквозным пробам; 3—по плоскостным пробам

решаемой практической задачи: если характеристики связи определяются для целей подсчета запасов одного элемента по содержанию другого в конкретных блоках месторождения, то должны использоваться средние содержания исследуемых элементов в подсчетных блоках. Если же исследователя интересует характер связи между элементами в отдельных кусках отбитой руды (например, для целей покусовой сортировки), то для целей корреляционного анализа должны использоваться данные о содержаниях элементов в штуфных пробах.

На рис. 73 приводится график зависимости выборочных значений коэффициента корреляции между содержаниями основного и сопутствующего элемента в рудах одного из редкometалльных месторождений в зависимости от геометрических характеристик анализируемых проб. Из рассмотрения графика видно, что значение выборочного коэффициента корреляции изменяется в широком диапазоне — от статистической значимой прямой до статистически значимой обратной корреляции. Наблюдаемая тенденция изменения выборочного коэффициента корреляции отражает неоднородность руд месторождения на различных уровнях его строения и, следовательно, имеет убедительное геологическое объяснение.

На выбор наиболее эффективного математического метода решения геологической задачи большое влияние оказывает также объем выборочной совокупности. Это связано с тем, что мощность многих

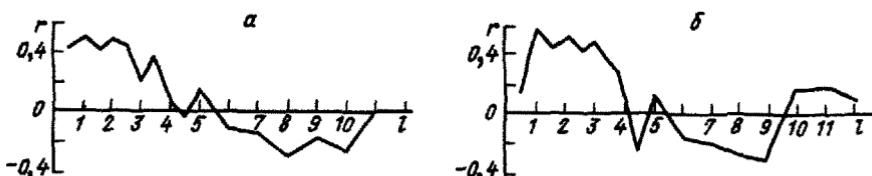


Рис. 73. Изменения выборочных значений коэффициента корреляции между содержаниями основного и сопутствующего компонента в зависимости от длин бороздовых проб:

а—для логарифмов содержаний полезных компонентов; б—для логарифмов метропроцентов полезных компонентов

статистических критериев в условиях малых выборок резко снижается. Однако чувствительность тех или иных критериев к объему выборки различна.

При выборе математических методов необходимо учитывать и шкалу измерений, с помощью которой получены количественные характеристики свойств геологических объектов (см. разд. 1.3). Для обработки результатов наблюдений, измеренных с помощьюnominalной или порядковой шкалы, например, обычно нельзя применять методы, в которых используются параметрические статистические критерии согласия.

#### 7.4. РОЛЬ ГЕОЛОГИЧЕСКОГО АНАЛИЗА ПРИ ВЫБОРЕ ГЕОЛОГО-МАТЕМАТИЧЕСКОЙ МОДЕЛИ

При решении вопроса о том, насколько правомерно применение той или иной математической модели, прежде всего выясняется ее соответствие типовой геологической модели объекта. Достоверность же геологической модели полностью зависит от правильности и полноты геологических представлений об изучаемом объекте на данном этапе исследований. Поскольку полнота геологических представлений зависит от уровня внутреннего строения объекта, достигнутого на данном этапе изучения, очевидно, что для каждого уровня строения необходимо создание самостоятельной геолого-математической модели. С учетом перечисленных обстоятельств очевидна ведущая роль геологического анализа на всех стадиях использования математических методов.

Геологический анализ лежит в основе сбора, анализа и интерпретации исходной цифровой информации, поскольку в процессе геологических исследований размещение пунктов наблюдений и проб, а также интерпретация полученных результатов производится с максимальным учетом геологических данных.

Перед началом математической обработки полученных данных они должны быть систематизированы и оценены не только с позиций статистического анализа, но и с геологической точки зрения. Все изучаемые объекты—массивы горных пород, скопления полезных ископаемых, их участки или блоки—должны быть проверены с позиции их геологической однородности, а для совокупностей, не однородных в геологическом отношении, должны быть установлены их границы. Для каждой выделяемой совокупности цифровых данных должны быть проверены однотипность условий эксперимента и степень представительности результатов отдельных испытаний (например, идентичность способов пробоотбора, размеров, ориентировки и объемов проб, а также методическая правильность пространственного расположения отдельных проб). Необходимо также убедиться в геологической однородности внутреннего строения изучаемых объектов в пределах участков, объединяющих цифровую информацию одного статистического коллектива. Так, например, для рудных скоплений следует убедиться в однотипности структур и закономерностей распределения полезного компонента по минеральным составляющим.

Было бы неправильным оценивать средние значения и определять показатели изменчивости важнейших свойств рудной залежи в целом, если ее верхняя часть располагается в пористых песчаниках, а нижняя — в плотных трещиноватых известняках. Учитывая геологическую неоднородность верхней и нижней частей залежи, методически правильнее выделить по литологическому признаку два самостоятельных блока и провести математическую обработку цифровой информации раздельно по каждому из них. Нельзя также вычислять общие показатели, характеризующие изменчивость содержаний полезного компонента, например золота, в пределах опробованного участка, если известно, что в одном объеме золото встречается только как примесь к сульфидам, а в другом — в самородном виде. В таких случаях для целей математической обработки необходимо провести раздельную обработку исходных данных, используя, например, результаты рациональных анализов, позволяющие судить о содержании сульфидного и самородного золота в каждой пробе.

Изменение размеров (геометрии) проб может отражаться не только на дисперсии изучаемых признаков, но и на других моментах их статистических распределений — средних значениях и коэффициентах асимметрии. Поэтому в процессе геолого-математического моделирования необходимо выяснить влияние геологической природы объектов на проявления тех или иных «масштабных эффектов». М. В. Рац предлагает различать масштабные эффекты первого, второго и третьего рода.

Масштабный эффект первого рода — уменьшение среднего значения признака с увеличением размера пробы (образца) проявляется только при изучении некоторых специфических свойств пород, как, например, их хрупкая прочность. Масштабные эффекты второго и третьего рода типичны для многих геологоразведочных параметров, в том числе для содержаний компонентов, значений пористости, трещиноватости, объемных масс и других свойств пород и руд. Если наблюдаемые свойства горных пород или руд описываются нормальным законом статистического распределения, то их дисперсии изменяются по «независимой» схеме, обратно пропорционально корню квадратному из числа элементов эффективной неоднородности (размеры элементов эффективной неоднородности должны быть на 1—2 порядка меньше размеров проб.). По «независимой» схеме изменяются дисперсии значений объемных масс, пористости и некоторых других свойств. Дисперсии содержания многих полезных компонентов, асимметричные распределения которых описываются логнормальным или другими статистическими законами, изменяются по более сложной схеме и зависят не только от числа элементов эффективной неоднородности, но и от силы корреляционных связей между ними.

Таким образом, использование математического анализа для решения конкретных геологических задач возможно только на достоверной геологической основе. В качестве основы математического моделирования путем геологического анализа создается геологическая модель изучаемого объекта, адекватная ему на соответствующем уровне изучения. По образцу геологической модели строится математическая модель.

тическая модель, с помощью которой обрабатывается исходная цифровая информация.

Результаты математической обработки данных, полученных на ранних стадиях исследований, используются в совокупности с уточненными геологическими данными для совершенствования геологической модели объекта, а по усовершенствованной геологической модели соответственно изменяется и математическая модель.

В свете изложенных положений вряд ли необходимо приводить доказательства опасности отрыва математического анализа от геологической основы. *Многочисленные примеры увлечения математикой в отрыве от конкретной геологической обстановки неизбежно приводили к получению абсурдных результатов и к необоснованным выводам о невозможности применения математических методов для решения геологических задач.*  
*Контрольные вопросы*

1. Какие факторы определяют выбор и эффективность использования математических методов?
2. Почему большинство геологических задач нельзя решать в рамках одной математической модели?
3. Какие математические модели и методы используются при нахождении точечных и интервальных оценок средних параметров геологических объектов?
4. Какие математические модели и методы применяются для решения задач классификации и распознавания образцов?
5. Какие приемы математического моделирования используются при изучении геологических процессов?
6. С помощью каких математических методов решаются задачи прогнозирования свойств геологических объектов?
7. Какие математические модели применяются для изучения изменчивости свойств геологических объектов в пространстве и выбора оптимальной сети наблюдений?
8. Какие свойства геологических объектов влияют на выбор и эффективность использования математических методов?
9. Почему при математическом моделировании геологических объектов необходимо учитывать методику их изучения?
10. Какие особенности методики изучения геологических объектов влияют на выбор математических методов решения геологических задач?
11. Как влияют на эффективность применения математических методов густота сети наблюдений и их общее количество?
12. Почему при математическом моделировании свойств геологических объектов необходимо учитывать геометрию проб?
13. В чем заключается роль геологического анализа при выборе математической модели?

## ГЛАВА 8 ПРИМЕНЕНИЕ ЭВМ В ГЕОЛОГИИ

---

В последние 15—20 лет наблюдается широкое внедрение современных средств обработки информации и ЭВМ в практику геологических исследований и геологоразведочных работ. Это, с одной стороны, вызвано изменением характера и резким увеличением объема геологической информации, а с другой,—совершенствованием вычислительной техники, существенно расширявшим возможности ее применения.

В ряде производственных геологических объединений РСФСР, Казахстана, Украины и Грузии уже накоплен опыт обработки геологоразведочных данных на ЭВМ. Аналогичный опыт имеют и многие производственные организации стран СЭВ—Геоиндустрися (ЧСФР), Венгерская алюминиевая корпорация (ВР), и другие, а также ряд государственных организаций и фирм США, Канады, Франции, Австралии, ФРГ и Японии.

Все исследования по применению ЭВМ в геологии можно в известной степени условно разделить на три направления:

— автоматизация трудоемких расчетных операций по предварительной обработке исходных данных;

— разработка методов решения конкретных геологических задач на основе математического моделирования геологических процессов и объектов;

— создание автоматизированных систем сбора, хранения и обработки геологической информации на базе ЭВМ.

## 8.1. АВТОМАТИЗАЦИЯ ПЕРВИЧНОЙ ОБРАБОТКИ ДАННЫХ

Широкое внедрение геофизических и геохимических методов на всех стадиях геологоразведочных работ привело к резкому увеличению объема получаемой информации и коренным образом изменило ее характер. Геологоразведочная информация стала в основном количественной. Это вызвало необходимость разработки принципиально новых способов ее обработки и анализа.

Внедрение высокопроизводительных комплексных, в том числе и дистанционных, геофизических методов поисков резко увеличило объем первичных данных, что вызвало необходимость автоматизации процедуры ее записи и предварительной интерпретации. Так, например, широко используемые в настоящее время комплексные аэрогеофизические станции включают в себя пятиканальный гамма-спектрометр, протонный магнитометр, электrorазведочную аппаратуру по методу индукций, курсо- и высотограф. Эта аппаратура обладает высокой чувствительностью, однако для правильной интерпретации полученных данных необходимо введение целого ряда поправок: на высоту полета, космическую составляющую остаточного фона, экранирующее действие лесного покрова, влажность почв и т. д. Вследствие этого предварительная интерпретация выполняется с помощью ЭВМ, а для оперативного обнаружения и проверки аномалий используются специализированные счетные устройства, установленные непосредственно на борту самолета.

Резкое увеличение объема количественной информации при разведке месторождений связано с комплексным использованиемrud и применением ядерно-физических методов их опробования, одним из преимуществ которых является возможность получения данных о содержании полезного компонента в пределах элементарно малых участков линейных проб длиной 10—15 см. Однако количественная интерпретация результатов ядерно-физического опробования, как пра-

вило, требует введения целого ряда поправок. Поэтому для повышения достоверности результатов опробования широко практикуется комплексирование этих методов, обеспечивающее их взаимный контроль и уточнение поправок, а также получение таких сведений о руде, которые не выявляются каждым методом в отдельности. Все это приводит к тому, что количественная интерпретация ядерно-физических методов опробования представляет собой весьма трудоемкую вычислительную процедуру, часто основанную на решении системы уравнений, реализовать которую без применения ЭВМ практически невозможно.

Результаты поисковых и разведочных работ обычно представляют собой совокупности замеров тех или иных свойств изучаемых объемов недр в точках, расположенных по определенной сети. Поэтому они могут рассматриваться как геологические, геохимические, геофизические и другие поля. Для более наглядного представления о свойствах полей исходные данные подвергаются различным преобразованиям путем скользящего осреднения или аппроксимации эмпирических данных на всей исследуемой площади единой функцией координат пространства (тренд-анализ) (см. гл. 5).

Автоматизация этих расчетных операций позволяет также оперативно производить различные преобразования исходных полей.

Для описания характера изменчивости полей в скользящем окне рассчитываются дисперсии, средние амплитуды локальных аномалий, коэффициенты аномальности, характеристики отражающие изменение частот и амплитуд колебаний, градиенты изменения поля, энтропия и т. д. При этом значения исходных полей в определенных точках могут суммироваться (аддитивные поля), перемножаться (мультипликативные поля) или делиться друг на друга. В качестве производных пространственных переменных используются также коэффициенты корреляции между значениями различных полей (см. разд. 5.6). Поля производных характеристик в ряде случаев более тесно коррелируются срудовмещающими структурами, что позволяет отбраковывать некоторые породные, эвапорационные и другие неперспективные аномалии (см. разд. 5.9, 7.2).

В процессе разведки и оценки месторождений полезных ископаемых в виде полей пространственных переменных изображаются параметры тел полезных ископаемых: мощность и содержание полезных компонентов, физические и химические свойства вмещающих пород, степень их метасоматической проработки, рельефрудовмещающих структур и т. д. При этом широко используются приемы трансформации исходных данных методом скользящего осреднения и рассчитываются производные характеристики: коэффициенты вскрыши — отношения мощностей перекрывающих пород к мощностям тел полезных ископаемых; коэффициентырудоносности — отношения суммарных мощностей рудных интервалов к общей мощности рудоносных зон; линейные запасы (метропроценты) — произведения мощностей рудных тел на содержания полезного компонента и т. д. Большой объем вычислений требуется также для введения поправок за счет искривления скважин и расчета средних значений параметров при подсчете запасов месторождений полезных ископаемых.

Массовое применение ЭВМ для автоматизации операций по первичной обработке геологоразведочных данных стало возможным после появления мини- и микроЭВМ типа СМ-2, СМ-4, «Искра-226», «Электроника-85», ДВК и др., на базе которых стали создаваться вычислительные комплексы непосредственно в геологоразведочных экспедициях и партиях.

## 8.2. РЕШЕНИЕ ГЕОЛОГИЧЕСКИХ ЗАДАЧ С ПОМОЩЬЮ ЭВМ

В настоящее время ЭВМ широко применяются при моделировании геологических процессов, прогнозировании и оценке геологических ресурсов и подсчете запасов месторождений полезных ископаемых, поиске оптимальных решений в процессе проектирования геологоразведочных работ, а также для решения других прикладных и научных задач в различных областях геологии.

Попытки использовать ЭВМ для моделирования геологических процессов предпринимаются начиная с 60-х годов, причем интерес к этой проблеме неуклонно возрастает.

Предлагались различные модели процессов седиментации, образования ритмичных слоистых толщ, складчатых и разрывных тектонических нарушений, процессов кристаллизации пород из расплавов, рудообразования и т. д. Н. Г. Клушин и Н. Н. Абрамович предприняли попытку математического моделирования тектономагматического развития зон субдукции.

Для моделирования геологических процессов используется главным образом математический аппарат теории случайных функций. В последние годы высказываются предложения для этих целей использовать теорию двумерных и трехмерных марковских цепей. Продолжают совершенствоваться и методы динамического моделирования, когда на ЭВМ просчитываются варианты хода геологических процессов при изменении исходных состояний и параметров моделируемых природных систем. Примером исследований этого направления является работа канадских геологов, моделировавших на ЭВМ развитие надвиговых структур.

Широкое применение ЭВМ для прогнозирования и оценки ресурсов полезных ископаемых обусловлено необходимостью комплексного учета разнообразной информации, получаемой в процессе геологических, геохимических и геофизических съемок изучаемых территорий. Разработанные к настоящему времени математические методы решения прогнозных задач основаны на использовании многомерных статистических моделей (см. гл. 3) и связаны обычно с весьма трудоемкими вычислительными операциями. В связи с большим объемом привлекаемой информации для решения прогнозных задач разработаны специальные автоматизированные системы прогнозирования: АСОД-Прогноз, РЕГИОН, ПРОСПЕКТОР и др.

Совершенствование методики подсчета запасов на основе использования ЭВМ в настоящее время идет по двум направлениям:

— разрабатываются автоматизированные технологии подсчета запасов традиционными способами;

— создаются новые способы подсчета, основанные на более полном использовании разведочной информации.

Государственной Комиссией по запасам СССР неоднократно рассматривались выполненные с помощью ЭВМ подсчеты запасов месторождений различных видов полезных ископаемых. Наиболее рациональной признана технология, разработанная опытно-методической экспедицией ГПО Казгеология, в которой автоматизированы практически все операции по предварительной подготовке исходных данных и определению цифр запасов в подсчетных блоках.

После введения в ЭВМ результатов опробования, инклинометрии и маркшейдерских замеров, а также кондиционных показателей, в автоматическом режиме производятся следующие операции:

— пересчет стволовых мощностей в истинные, горизонтальные или вертикальные;

— пересчет в условное содержание одного компонента всех основных компонентов комплексных руд;

— расчет выхода керна в процентах и теоретических масс проб для каждого опробованного интервала;

— расчет производных характеристик (метропроцентов, отношений содержаний различных компонентов и т. д.);

— выделение кондиционных интервалов по совокупности заданных кондиционных показателей;

— расчет средних содержаний, мощностей, метропроцентов и других параметров по каждому кондиционному интервалу;

— выделение сортов и типов руд;

— расчет параметров по единичным разведочным пересечениям (коэффициентов вскрыши, коэффициентов рудоносности и т. п.).

При необходимости в результаты анализов вводятся поправки.

Результаты предварительной обработки могут быть выданы на печать и размножены в виде таблиц произвольной формы, а также в виде планов изолиний любого масштаба.

Увязка кондиционных рудных интервалов в рудные тела и выделение подсчетных блоков осуществляется геологом с учетом геологической ситуации.

Дальнейшие операции по подсчету запасов в пределах подсчетных блоков производятся способом блоков или разрезов в автоматическом режиме. При этом расчет средних содержаний и других подсчетных параметров может осуществляться путем взвешивания на мощности, объемные массы, площади влияния разведочных пересечений и т. п.

Опыт использования этой технологии показал, что переход на автоматизированный подсчет запасов экономически выгоден, когда количество проб по месторождению составляет несколько тысяч, а для месторождений комплексных руд — около тысячи.

Автоматизация процедуры подсчета запасов позволяет:

— уменьшить срок выполнения расчетов по крупным объектам;

— практически исключить технические ошибки;

— оперативно выполнять подсчет запасов по различным вариантам кондиций.

Максимальный экономический эффект от использования ЭВМ при подсчете запасов может быть получен, если автоматизированная процедура будет применена еще на этапе расчета постоянных кондиций.

Накопление разведочной информации в памяти ЭВМ позволяет также решать в автоматическом режиме целый ряд самостоятельных задач не связанных непосредственно с подсчетом запасов:

- рассчитывать статистические характеристики изменчивости геологоразведочных параметров;
- производить анализ густоты разведочной сети;
- проверять различные геологические гипотезы с помощью статистических критериев;
- выявлять характер и силу корреляционных связей между различными параметрами;
- анализировать результаты контроля пробоотбора, обработки и анализов проб.

Дальнейшее совершенствование методики подсчета запасов в этом направлении возможно лишь по линии создания более экономичных технологий, использующих возможности ЭВМ новых поколений, что является сугубо технической задачей.

Поэтому более перспективным представляется второе направление — разработка принципиально новых способов подсчета запасов, позволяющих уменьшить ошибки, обусловленные дискретностью разведочной сети и относительно небольшим количеством разведочных пересечений в каждом подсчетном блоке.

Работами Д. Криге, Ж. Матерона и других было доказано, что величины и знаки ошибок аналогии, возникающие за счет распространения результатов опробования разведочных скважин и горных выработок на зоны их влияния, зависят от характера изменчивости оцениваемого параметра и его значения, установленного по данному разведочному пересечению. Опыт эксплуатации месторождений свидетельствует о том, что высокие значения параметров по разведочным пересечениям, как правило, оказываются завышенными по отношению к его истинному среднему значению в зоне влияния, а низкие значения — заниженными.

Различные способы минимизации ошибок аналогии, получившие название способов «крайгинга», основаны на использовании таких зависимостей для определения поправочных коэффициентов, вводимых на каждый замер. В условиях действующего предприятия для этого можно применять регрессионные кривые, построенные по результатам сравнения данных разведки и отработки, а в процессе разведки использовать данные по участкам детализации. Французскими геологами с этой же целью применяются специальные номограммы, построенные на основе характеристик изменчивости изучаемого параметра, либо расчет коэффициентов крайгинга производится путем решения систем уравнений.

Применение способов крайгинга с одной стороны устраниет так называемые «ураганные» значения, а с другой уменьшает ошибки оконтуривания.

Перспективы массового применения ЭВМ при решении различных геологических задач связаны с оснащением геологических организаций персональными компьютерами. Они обычно снабжены устройствами ввода и вывода как цифровой и буквенной, так и графической информации, что существенно расширяет их возможности при решении геологических задач. В состав программного обеспечения современных ЭВМ обычно входит большое количество пакетов прикладных программ, ориентированных на управление базами данных (СУБД): их статистическую обработку и преобразование. Специальные программы дают также возможность пользователям непосредственно общаться с ЭВМ на обычном для них языке, минуя этапы подготовки программ и данных на машинных носителях — перфолентах или магнитных дисках. Диалоговый режим использования ЭВМ позволяет геологу оперативно пополнять и корректировать массивы исходных данных, осуществлять различные варианты их преобразований, анализировать промежуточные результаты и корректировать ход дальнейших вычислительных операций.

В дальнейшем совершенствовании алгоритмов решения геологических задач на ЭВМ наметилась тенденция к созданию достаточно простых программных комплексов для мии- и микроЭВМ, широкому использованию цифровых и графических периферийных технических средств и стремлению максимально упростить процесс общения специалиста-геолога с ЭВМ.

### **8.3. АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ ОБРАБОТКИ ГЕОЛОГИЧЕСКИХ ДАННЫХ**

Опыт применения ЭВМ при решении отдельных геологических задач, накопленный в 60-е годы, позволил выявить возникающие при этом трудности и наметить основные пути их преодоления.

Одним из главных препятствий для широкого внедрения ЭВМ в геологическую практику явилась трудоемкость предварительной подготовки данных к вводу в ЭВМ. Затраты времени на кодирование и перевод информации на машинные носители составляли обычно 80—90% от всего процесса решения задачи. В то же время информация, подготовленная для решения конкретной задачи по одному алгоритму, оказывалась непригодной при использовании другого алгоритма, а тем более для решения другой задачи. В связи с этим начали интенсивно разрабатываться автоматизированные системы хранения, накопления и поиска информации (АИС), позволяющие при однократном введении геологического-геофизических данных в память ЭВМ использовать их для решения широкого круга задач.

В 1970 г. в США в информационных системах на базе ЭВМ уже хранились данные по 600 тыс. нефтяных и газовых скважин.

В Канаде были созданы крупные базы данных по рудным месторождениям — ОПСЕП, МИНДЕП и др., содержащие разностороннюю информацию, которая используется для решения вопросов региональной металлогенеза, поисков, разведки и промышленной оценки месторождений.

В СССР первые разработки по созданию автоматизированных систем хранения и обработки геологических данных были начаты в 70-х годах на базе ЭВМ второго поколения. Однако широкий характер они приобрели лишь в последние 10 лет, чему способствовала унификация ЭВМ, выпускаемых странами-членами СЭВ и создание системного программного обеспечения ЕС ЭВМ. Совокупности системных программ (*операционные системы*) выполняют функции, связанные с преобразованием исходных программ в машинные коды, исправлением в них ошибок, загрузкой преобразованных программ и исходных данных в память ЭВМ, вызовом стандартных программ из библиотек, управлением внешними устройствами—дисплеями, принтерами, графопостроителем, накопителями на магнитных дисках, а также обеспечивают диалог пользователя с ЭВМ.

К концу 80-х годов в СССР различными научными и производственными организациями были созданы автоматизированные информационные системы (АИС), на основе которых формируются базы геологических данных различного назначения. Союзгеофондом создана система КАДАСТР, предназначенная для хранения и анализа информации о состоянии минерально-сырьевой базы СССР. При обработке результатов геохимических съемок широко используется автоматизированная информационно-поисковая система ГЕОНавТ, осуществляющая операции ввода данных в ЭВМ, накопление и хранение первичных данных в долговременной памяти ЭВМ, поиск и формирование целевых массивов данных для решения конкретных задач. В ряде производственных геологических организаций действуют АИС, предназначенные для создания территориальных баз геологических данных: АИПС—Геология (ПГО Севказгеология), БГД—ЗСНП (ЗапСибНИГНИ) и др.

На основе объединения региональных баз данных создаются мощные общегосударственные и международные информационные системы. Примером такой системы является информационно-аналитическая система ИНТЕРГЕОРЕСУРС, предназначенная для прогнозирования и управления перспективным развитием минерально-сырьевой базы стран-членов СЭВ. Для решения теоретических, методологических и организационных проблем, возникающих при создании таких систем, созданы международная комиссия по геологической документации (КОГЕОДОК) и международная организация ГЕОИНФОРМ.

К настоящему времени в СССР и за рубежом создан ряд проблемно ориентированных автоматизированных систем обработки геологических данных (АСОД), позволяющих решать широкий круг прикладных задач. Накоплен опыт эксплуатации человека-машинной системы геологического прогнозирования «Регион», а также АСОД—ГППИ, предназначенной для анализа геохимической информации при поисках полезных ископаемых. При обработке аэрокосмических снимков с целью геологического картирования успешно применяется АСОД АКС. Различные специализированные АСОД широко применяются при обработке геофизических данных.

Казахстанской опытно-методической экспедицией ГПО Казгеология разработана АСОД МСПВ, осуществляющая операции по подсчету запасов твердых полезных ископаемых, и АСП ГЕО, пред-

назначенная для моделирования процесса геолого-экономической оценки месторождений.

Наиболее совершенные АСОД в последние годы приобретают черты «экспертных», т. е. «советующих» систем. Их составной частью являются банки знаний и банки моделей, в которых накапливаются и обобщаются знания специалистов по различным направлениям наук о Земле. К системам такого типа можно отнести человеко-машинную систему геологического прогнозирования ПРОСПЕКТОР (США), рассчитанную на решение задач по выделению перспективных площадей, оценке прогнозных ресурсов полезных ископаемых и выбору мест заложения поисковых скважин, а также одну из последних версий системы РЕГИОН—«РЕГИОН-СКАНДИНГ».

Несмотря на различное целевое назначение, структуры и общие принципы функционирования всех АСОД во многом сходны. Рассмотрим их на примере человеко-машинной системы геологического прогнозирования РЕГИОН.

Система РЕГИОН состоит из следующих основных частей (рис. 74):

- блока кодирования и ввода геологической информации всех видов, в том числе и картографической;
- регионального банка геологических данных и моделей, включающего как исходную, так и трансформированную информацию;
- блока разработки вариантов решений, содержащего пакеты программ статистической обработки исходной и трансформированной информации, алгоритмов распознавания образцов, безэталонной классификации (таксономии) и имитационного моделирования;
- блока обеспечения диалога пользователей с системой.

Блок кодирования обеспечивает формирование территориальных баз как числовых, так и картографических данных, полученных в процессе геологической, геохимической, геофизической и аэрокосмической съемок исследуемой территории.

Исходные данные геохимических и геофизических съемок обычно представлены в виде цифр, и затруднений при их вводе в ЭВМ не возникает. В то же время при геологическом картировании и проведении космо- и аэрофотометрических исследований информация представляется в графическом виде, и для ввода в ЭВМ ее необходимо преобразовать в дискретный вид. Для этого выделяют *картографические факторы*—конкретные геологические образования, изображенные в легенде карты в виде самостоятельных единиц. Преобразование картографической информации в дискретный вид заключается в разбиении карты на элементарные ячейки по определенной геометрической сети и присвоении знака 1 ячейкам, в которых данный фактор присутствует, и знака 0 ячейкам, где его нет. При этом геологическая карта заменяется набором бинарных матриц, количество которых соответствует числу картографических факторов. Перевод графической информации на машинные носители можно осуществлять как вручную, так и в полуавтоматическом режиме с помощью специальных кодирующих устройств—дигитайзеров. В первом случае фактор кодируется построчно, а во втором контуры геологических факторов обводятся электронным курсором. Дигитайзеры в комплекте

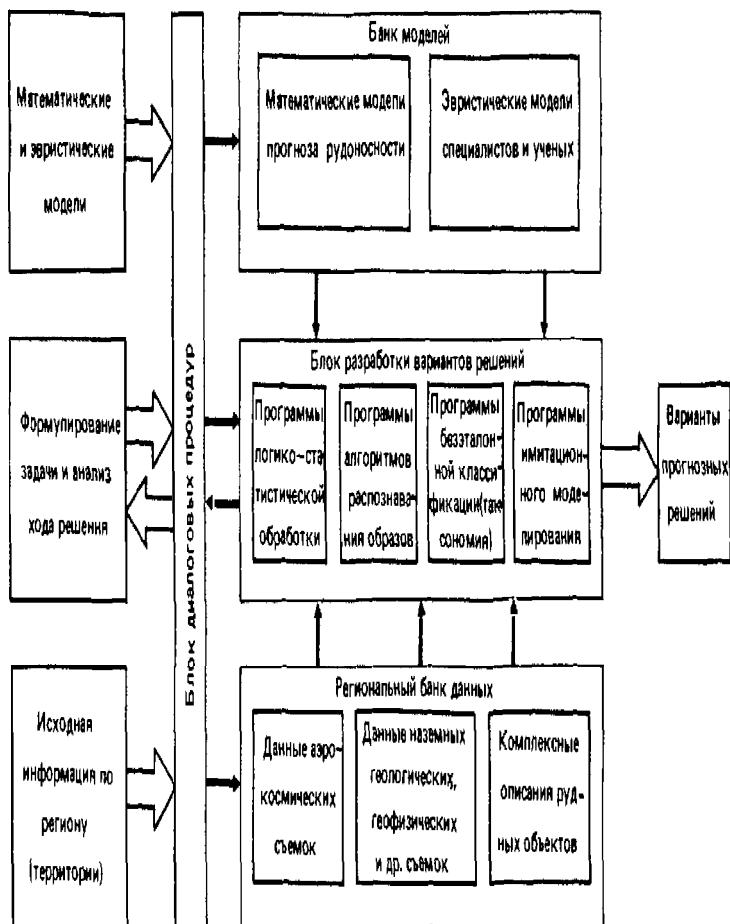


Рис. 74. Структура системы «Регион»

с процессором, дисплеем, графопостроителем и соответствующим программным обеспечением образуют *интерактивную графическую систему*, позволяющую контролировать и редактировать вводимую информацию, а также выполнять ее преобразования — масштабирование, сдвиги, повороты и т. п.

*Банк данных и моделей* представляет собой систему наполнения, хранения и первичной обработки геолого-геофизической информации в ЭВМ. Он обеспечивает использование этой информации при решении различных задач, дает возможность обновлять и уточнять информацию и осуществлять поиск любых хранимых в нем сведений.

Изображение картографических факторов в виде бинарных матриц не учитывает расположение отдельных точек фактора относительно его контура, т. е. контакта с другими факторами. Так, например, ячейке в зоне эндоконтакта интрузивного массива и ячейке в его центральной части в бинарной матрице этого фактора будет присвоен одинаковый знак — 1. Поэтому при анализе картографической информации массив исходных данных (бинарные матрицы) преобразуется в массив *картографических признаков*, который представляет собой числовой массив расстояний от любой точки карты до ближайших границ всех факторов, имеющихся на данной карте. Трансформация массива исходных данных в массив признаков производится с помощью ЭВМ по специальным программам. В памяти ЭВМ хранятся массивы как картографических факторов, так и признаков.

С помощью программ блока *разработки вариантов решений* геолог в процессе взаимодействия с системой может анализировать характер исходных и трансформированных данных выводя на дисплей, принтер или графопостроитель отдельные картографические факторы или их сочетания, а также рассчитывая частоты встречаемости тех или иных факторов и их сочетаний на различных участках исследуемой территории. При наличии эталонных участков, где выявлены рудные объекты он может оценивать информативность тех или иных факторов.

Для решения прогнозных задач в системе РЕГИОН используются методы распознавания образов (алгоритм «Гиперпласт»), таксономической классификации и эвристического моделирования.

Блок разработки вариантов решений позволяет также оценивать прогнозные ресурсы полезных ископаемых и проверять различные геологические гипотезы и концепции на основе имитационного моделирования.

Программы блока *диалоговых процедур* позволяют геологу:

- получать справки о наличии в базе данных интересующей его информации;
- формировать личные целевые массивы данных по конкретной территории;
- выбирать методы решения содержательных задач;
- анализировать промежуточные результаты в процессе решения и вносить необходимые корректизы.

Человеко-машинная система РЕГИОН позволяет соединить творческий потенциал геологов с огромными возможностями современных ЭВМ по обработке многофакторной информации.

Разнообразие применяемых в настоящее время в геологической практике АИС и АСОД обусловлено не только разнообразием решаемых проблем и видов исходной информации, но и отсутствием должной координации на первом этапе работ по их созданию. Поэтому в 1987 г. были начаты работы по созданию единой Государственной автоматизированной системы сбора и обработки данных геологических, геофизических и геохимических исследований на четырех уровнях наблюдений: космос — воздух — земля — скважина. Для реализации программы работ создан межотраслевой научно-технический комплекс «ГЕОС», в состав которого включены коллективы из различных организаций Министерства геологии СССР и других ведомств, связанных с изучением и использованием минерально-сырьевых ресурсов.

В задачи Геосистемы входит:

- интегральная обработка геоинформации;
- математическое моделирование геологических процессов;
- оптимизация ведения геологоразведочных работ на различных стадиях;
- контроль за рациональной разработкой месторождений полезных ископаемых.

Геосистема позволит резко повысить эффективность геологоразведочных работ и научных исследований во всех областях геологии.

#### *Контрольные вопросы*

1. Для чего используются ЭВМ в геологии?
2. Какие расчетные операции целесообразно автоматизировать с помощью ЭВМ?
3. Какие типы геологических задач решаются с помощью ЭВМ?
4. Почему ЭВМ широко применяются при прогнозировании полезных ископаемых?
5. Как ЭВМ используются при подсчете запасов полезных ископаемых?
6. В чем преимущество подсчета запасов с помощью ЭВМ?
7. Какие функции выполняют автоматизированные информационные системы (АИС) в геологии?
8. Каковы принципы функционирования проблемно-ориентированных автоматизированных систем обработки данных (АСОД) в геологии?
9. Для чего используется человеко-машинная система РЕГИОН?
10. Какова структура системы РЕГИОН?
11. Как вводится в ЭВМ цифровая и картографическая информация?
12. Для чего в СССР создается Государственная Геосистема?

## СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

---

1. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики.— М.: Наука, 1982.
2. Бондаренко В. Н. Статистические решения некоторых задач геологии.— М.: Недра, 1970.
3. Боровко Н. Н. Статистический анализ пространственных геологических закономерностей.—Л.: Недра, 1971.
4. Вистелус А. Б. Основы математической геологии.—Л.: Наука, 1980.
5. Давид М. Геостатистические методы при оценке запасов руд/Пер. с англ. О. А. Лутковской.—Л.: Недра, 1980.
6. Дэвис Дж. Статистика и анализ геологических данных/Пер. с англ. В. А. Голубевой.—М.: Мир, 1977.
7. Каждан А. Б. Поиски и разведка месторождений полезных ископаемых. Научные основы поисков и разведки.—М.: Недра, 1984.
8. Каждан А. Б., Гуськов О. И., Шиманский А. А. Математическое моделирование в геологии и разведке полезных ископаемых.—М.: Недра, 1979.
9. Кендалл М. Дж., Стюарт А. Многомерный статистический анализ и временные ряды.—М.: Наука, 1976.
10. Коган Р. И. Интервальные оценки в геологических исследованиях.— М.: Недра, 1986.
11. Коган Р. И., Белов Ю. П., Родионов Д. А. Статистические ранговые критерии в геологии.—М.: Недра, 1983.
12. Крамбейн У., Грейбилл Ф. Статистические модели в геологии.—М.: Мир, 1969.
13. Мардза К. Статистический анализ угловых наблюдений/Пер. с англ. Э. В. Хамаладзе.—М.: Наука, 1978.
14. Марченко В. В. Человеко-машинные методы геологического прогнозирования.—М.: Недра, 1988.
15. Матерон Ж. Основы прикладной геостатистики/Пер. с франц. Ю. В. Рощина.—М.: Мир, 1968.
16. Миллер Р. Л., Кан Дж. С. Статистический анализ в геологических науках/Пер. с англ. Д. А. Родионова.—М.: Мир, 1965.
17. Никитин А. А. Теоретические основы обработки геофизической информации.—М.: Недра, 1986.
18. Поллард Дж. Справочник по вычислительным методам статистики/Пер. с англ. В. С. Зандворова.—М.: Финансы и статистика, 1982.
19. Родионов Д. А. Статистические решения в геологии.—М.: Недра, 1981.
20. Роненсон Б. М., Ройзенман В. М., Ортенберг Н. А. Методы расчленения метаморфических комплексов.—Л.: Недра, 1976.
21. Смирнов Б. Н. Корреляционные методы при парагенетическом анализе.—М.: Недра, 1981.
22. Справочник по математическим методам в геологии/Д. А. Родионов, Р. И. Коган, В. А. Голубева и др.—М.: Недра, 1987.
23. Теоретические основы инженерной геологии. Механико-математические основы/Под ред. акад. Е. М. Сергеева.—М.: Недра, 1986.
24. Тьюокс Дж. Анализ результатов наблюдений. Разведочный анализ/Пер. с англ. А. Ф. Кущириа.—М.: Мир, 1981.
25. Харбух Дж., Бонэм-Карттер Г. Моделирование на ЭВМ в геологии/Пер. с англ. Д. А. Родионова.—М.: Мир, 1974.
26. Харченко А. Г. Принципы и методы прогнозирования минеральных ресурсов.—М.: Недра, 1987.
27. Чумаченко Б. А., Власов Е. П., Марченко В. В. Системный анализ при геологической оценке перспектив рудоносности территорий.—М.: Недра, 1980.
28. Электронные вычислительные машины. Кн. 1. Введение в ЭВМ. Кн. 2. Основы информатики/Под ред. А. Я. Савельева.—М.: Высшая школа, 1987.

## ПРИЛОЖЕНИЯ

### Приложение 1

#### Критические значения статистики $\bar{R}$ критерия равномерности Релея

$n$	$\alpha$				
	0,10	0,05	0,025	0,01	0,001
5	0,677	0,754	0,816	0,879	0,991
6	0,618	0,690	0,753	0,825	0,940
7	0,572	0,642	0,702	0,771	0,891
8	0,535	0,602	0,660	0,725	0,847
9	0,504	0,569	0,624	0,687	0,808
10	0,478	0,540	0,594	0,655	0,775
11	0,456	0,516	0,567	0,627	0,743
12	0,437	0,494	0,544	0,602	0,716
13	0,420	0,475	0,524	0,580	0,692
14	0,405	0,458	0,505	0,560	0,669
15	0,391	0,443	0,489	0,542	0,649
16	0,379	0,429	0,474	0,525	0,630
17	0,367	0,417	0,460	0,510	0,613
18	0,357	0,405	0,447	0,496	0,597
19	0,348	0,394	0,436	0,484	0,583
20	0,339	0,385	0,425	0,472	0,569
21	0,331	0,375	0,415	0,461	0,556
22	0,323	0,367	0,405	0,451	0,544
23	0,316	0,359	0,397	0,441	0,533
24	0,309	0,351	0,389	0,432	0,522
25	0,303	0,344	0,381	0,423	0,512
30	0,277	0,315	0,348	0,387	0,470
35	0,256	0,292	0,323	0,359	0,436
40	0,240	0,273	0,302	0,336	0,409
45	0,226	0,257	0,285	0,318	0,386
50	0,214	0,244	0,270	0,301	0,367
100	0,15	0,17	0,19	0,21	0,26
$2n\bar{R} \approx \chi^2_2$	4,605	5,991	7,378	9,210	13,816

### Приложение 2

#### Параметр концентрации $k$ распределения Минзеса

$\bar{R}$	$k$	$\bar{R}$	$k$	$R$	$k$	$\bar{R}$	$k$
0,00	0,00000	0,25	0,51649	0,50	1,15932	0,76	2,45490
0,01	0,02000	0,26	0,53863	0,51	1,19105	0,77	2,54686
0,02	0,04001	0,27	0,56097	0,52	1,22350	0,78	2,64613
0,03	0,06003	0,28	0,58350	0,53	1,25672	0,79	2,75382
0,04	0,08006	0,29	0,60625	0,54	1,29077	0,80	2,87129
0,05	0,10013	0,30	0,62922	0,55	1,32570	0,81	3,00020
0,06	0,12022	0,31	0,65242	0,56	1,36156	0,82	3,14262
0,07	0,14034	0,32	0,67587	0,57	1,39842	0,83	3,30114
0,08	0,16051	0,33	0,69958	0,58	1,43635	0,84	3,47901
0,09	0,18073	0,34	0,72356	0,59	1,47543	0,85	3,68041
0,10	0,20101	0,35	0,74783	0,60	1,51574	0,86	3,91072
0,11	0,22134	0,36	0,77241	0,61	1,55738	0,87	4,17703
0,12	0,24175	0,37	0,79730	0,62	1,60044	0,88	4,48876
0,13	0,26223	0,38	0,82253	0,63	1,64506	0,89	4,85871

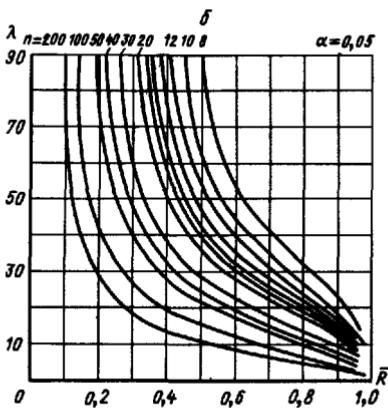
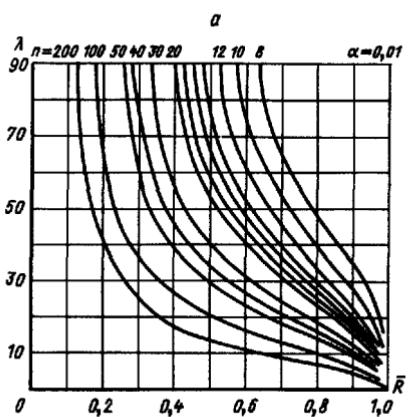
R	R	R	R	R	R	R	R
0,14	0,28279	0,39	0,84812	0,64	1,69134	0,90	5,3047
0,15	0,30344	0,40	0,87408	0,65	1,73945	0,91	5,8522
0,16	0,32419	0,41	0,90043	0,66	1,78953	0,92	6,5394
0,17	0,34503	0,42	0,92720	0,67	1,84177	0,93	7,4257
0,18	0,36599	0,43	0,95440	0,68	1,89637	0,94	8,6104
0,19	0,38707	0,44	0,98207	0,69	1,95357	0,95	10,2716
0,20	0,40828	0,45	1,01022	0,70	2,01363	0,96	12,7661
0,21	0,42962	0,46	1,03889	0,71	2,07685	0,97	16,9266
0,22	0,45110	0,47	1,06810	0,72	2,14359	0,98	25,2522
0,23	0,47273	0,48	1,09788	0,73	2,21425	0,99	50,2421
0,24	0,49453	0,49	1,12828	0,74	2,28930	1,00	$\infty$

## Приложение 3

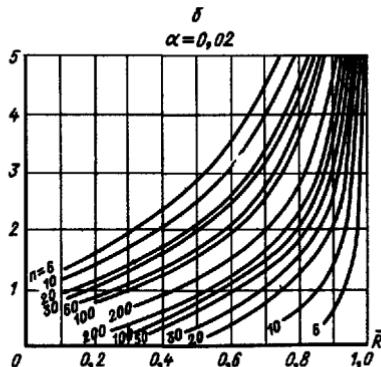
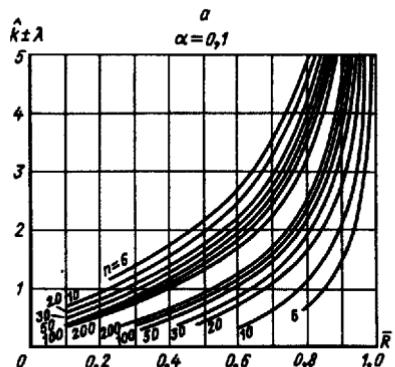
Значения функции  $\Psi_N(t)$  по Ачисону и Брауну

t	N								
	10	20	30	40	50	60	70	80	90
0,05	1,0458	1,0458	1,0494	1,0499	1,0502	1,0504	1,0505	1,0506	1,0507
0,10	0933	0992	1012	1022	1028	1032	1034	1037	1038
0,15	1427	1521	1553	1569	1579	1585	1590	1593	1596
0,20	1938	2072	2118	2142	2156	2166	2173	2178	2182
0,25	2468	2648	2710	2742	2761	2774	2784	2791	2796
0,30	3018	3248	3329	3370	3395	3413	3424	3434	3441
0,35	3587	3874	3975	4028	4060	4081	4097	4108	4117
0,40	4177	4527	4652	4716	4756	4782	4801	4816	4827
0,45	4788	5207	5359	5437	5485	5517	5540	5558	5571
0,50	5421	5917	6097	6191	6248	6287	6315	6336	6352
0,55	6076	6657	6869	6980	7048	7094	7127	7152	7172
0,60	6754	7428	7676	7806	7886	7940	7979	8009	8031
0,65	7457	8231	8519	8670	8763	8826	8871	8906	8933
0,70	8184	9068	9399	9574	9681	9754	9807	9847	9879
0,75	1,8936	1,9940	2,0319	2,0519	2,0643	2,0727	2,0788	2,0834	2,0870
0,80	9714	2,0848	1279	1508	1650	1746	1816	1869	1911
0,85	2,0519	1794	2283	2542	2703	2813	2893	2954	3001
0,90	1352	2779	3330	3624	3807	3932	4022	4091	4145
0,95	2214	3804	4424	4755	4962	5103	5206	5284	5345
1,00	3104	4872	5565	5938	6170	6330	6445	6534	6603
1,05	4025	5984	6757	7174	7435	7614	7745	7844	7922
1,10	2,4977	2,7141	2,8002	2,8467	2,8759	2,8959	2,9106	2,9217	2,9305
1,15	2,5961	2,8345	2,9300	2,9818	3,0144	3,0368	3,0532	3,0656	3,0755
1,20	2,6978	2,9597	3,0655	3,1231	3,1594	3,1849	2026	2165	2275
1,25	2,8028	3,0901	2069	2707	3110	3388	3591	3746	3868
1,30	9114	2257	3544	4250	4696	5005	5230	5403	5539
1,35	3,0235	3668	5084	5862	6356	6697	6947	7139	8290
1,40	1393	5135	6689	7547	8092	8469	8746	8958	9125
1,45	2589	6661	8364	9307	9908	4,0324	4,0630	4,0864	4,1049
1,50	3824	8247	4,0111	4,1146	4,1807	2266	2603	2861	3065
1,55	5099	9897	1933	3068	3793	4297	4669	4953	5178
1,60	6415	4,1612	3832	5074	5870	6424	6832	7145	7393
1,65	7774	3394	5819	7171	8042	8649	9097	9441	9714
1,70	9176	5247	7878	9360	5,0113	5,0978	5,1469	5,1847	5,2146

$t$	$N$								
	10	20	30	40	50	60	70	80	90
1,75	4,0623	7173	5,0031	5,1646	2687	3415	3953	4366	4694
1,80	2116	9174	2275	4034	5170	5965	6559	7005	7365
1,85	3657	5,1253	4614	6527	7764	8832	9275	9700	6,0163
1,90	5246	3413	7052	9129	6,0477	6,1423	6,2124	6,2665	3094
1,95	6885	5657	9592	6,1847	3312	4342	5107	5896	6165
2,00	8575	7988	6,2239	4684	6276	7396	8229	8871	9383



Приложение 4. Номограммы для нахождения интервальных оценок кругового среднего направления  $\mu$ :  
 $a$ —для уровня значимости 0,01;  $b$ —для уровня значимости 0,05



Приложение 5. Номограмма для нахождения интервальных оценок параметра концентрации  $k$ :  
 $a$ —для уровня значимости 0,1;  $b$ —для уровня значимости 0,02

Приложение 6

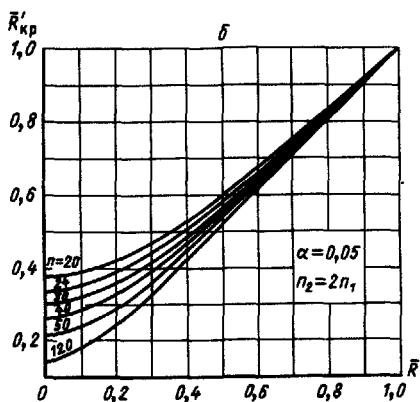
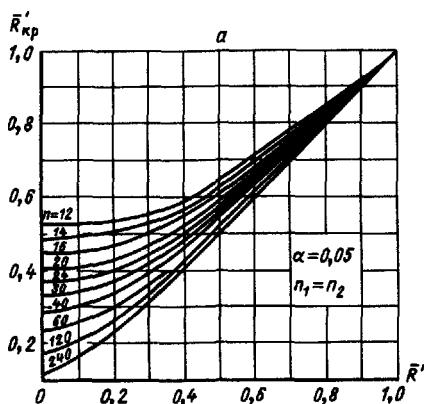
Критические значения статистики Х-критерия Ван-дер-Вардена для односторонних границ

N	$\alpha=0,025$			$\alpha=0,001$		
	$m=0$ $m=1$	$m=2$ $m=3$	$m=4$ $m=5$	$m=0$ $m=1$	$m=2$ $m=3$	$m=4$ $m=5$
8	2,40	2,30	—	1,42	1,37	1,23
9	2,38	2,20	—	1,56	1,48	1,30
10	2,60	2,49	2,30	1,71	1,67	1,57
11	2,72	2,58	2,40	1,83	1,77	1,64
12	2,86	2,79	2,68	1,98	1,94	1,87
13	2,96	2,91	2,78	2,09	2,03	1,93
14	3,11	3,06	3,00	2,22	2,19	2,12
15	3,24	3,19	3,06	2,33	2,28	2,20
16	3,39	3,36	3,28	2,44	2,42	2,36
17	3,49	3,44	3,36	2,54	2,51	2,44
18	3,63	3,60	3,53	2,65	2,64	2,59
19	3,73	3,69	3,61	2,76	2,72	2,66
20	3,86	3,84	3,78	2,85	2,84	2,80
21	3,96	3,92	3,85	2,95	2,92	2,87
22	4,08	4,06	4,01	3,05	3,04	3,00
23	4,18	4,15	4,08	3,14	3,12	3,06
24	4,29	4,27	4,23	3,23	3,22	3,19
25	4,39	4,36	4,30	3,33	3,29	3,26
26	4,50	4,48	4,44	3,41	3,39	3,37
27	4,59	4,56	4,51	3,49	3,47	3,43
28	4,69	4,68	4,64	3,57	3,57	3,54
29	4,78	4,76	4,72	3,66	3,64	3,60
30	4,88	4,87	4,84	3,74	3,73	3,70
31	4,97	4,95	4,91	3,82	3,80	3,76
32	5,07	5,06	5,03	3,89	3,88	3,86
33	5,15	5,13	5,10	3,96	3,95	3,92
34	5,25	5,24	5,21	4,05	4,05	4,02
35	5,33	5,31	5,28	4,12	4,11	4,08
36	5,42	5,41	5,38	4,19	4,19	4,16
37	5,50	5,48	5,45	4,26	4,25	4,24
38	5,59	5,58	5,55	4,33	4,33	4,32
39	5,67	5,65	5,62	4,40	4,39	4,38
40	5,75	5,74	5,72	4,48	4,48	4,46
41	5,83	5,81	5,79	4,54	4,53	4,50
42	5,91	5,90	5,88	4,62	4,62	4,59
43	5,99	5,97	5,95	4,68	4,67	4,66
44	6,06	6,06	6,04	4,76	4,74	4,73
45	6,14	6,12	6,10	4,81	4,80	4,78
46	6,21	6,21	6,19	4,88	4,86	4,86
47	6,29	6,27	6,25	4,93	4,93	4,90
48	6,36	6,35	6,34	5,00	5,00	4,99
49	6,43	6,42	6,39	5,07	5,05	5,04
50	6,50	6,50	6,48	5,14	5,13	5,11

**Критические значения статистик  $W_1$  и  $2MW$  критерия Вилькоксона для односторонних границ**

$N_1$	$N_2$	$W_1$		$2MW$	$N_1$	$N_2$	$W_1$		$2MW$
		$\alpha=0,025$	$\alpha=0,010$				$\alpha=0,025$	$\alpha=0,010$	
10	10	78	87	210	11	11	96	106	253
	11	81	91	220		12	99	110	264
	12	84	94	230		13	103	114	275
	13	88	98	240		14	106	118	286
	14	91	102	250		15	110	123	297
	15	94	106	260		16	113	127	308
	16	97	109	270		17	117	131	319
	17	100	113	280		18	121	135	330
	18	103	117	290		19	124	139	341
	19	107	121	300		20	128	144	352
	20	110	125	310		21	131	148	363
	21	113	128	320		22	135	152	374
	22	116	132	330		23	139	156	385
	23	119	136	340		24	142	161	396
	24	122	140	350		25	146	165	407
	25	126	144	360	13	13	136	149	351
12	12	115	127	300		14	141	154	364
	13	119	131	312		15	145	159	377
	14	123	136	314		16	150	165	390
	15	127	141	336		17	154	170	403
	16	131	145	348		18	158	175	416
	17	135	150	360		19	163	180	429
	18	139	155	372		20	167	185	442
	19	143	159	384		21	171	190	455
	20	147	164	396		22	176	195	468
	21	151	169	400		23	180	200	481
14	22	155	173	420	15	24	185	205	494
	23	159	178	432		25	189	211	507
	24	163	183	444		15	184	200	465
	25	167	187	456		16	190	206	480
	14	160	174	406		17	195	212	495
	15	164	179	420		18	200	218	510
	16	169	185	434		19	205	224	525
	17	174	190	448		20	210	230	540
	18	179	196	462		21	216	236	555
	19	183	202	476		22	221	242	570
16	20	188	207	490		23	226	248	585
	21	193	213	504		24	231	254	600
	22	198	218	518		25	237	260	615
	23	203	224	532		17	240	259	595
	24	207	229	543		18	246	266	612
	25	212	235	560		19	252	273	629
	16	211	229	528		20	258	280	646
17	217	235	544		21	264	287	663	
18	222	242	560		22	270	294	680	
19	228	248	576		23	276	300	697	
20	234	255	592						
21	239	261	608						
22	245	267	624						

$N_1$	$N_2$	$W_1$		$2MW$	$N_1$	$N_2$	$W_1$		$2MW$
		$\alpha=0,025$	$\alpha=0,010$				$\alpha=0,025$	$\alpha=0,010$	
18	23	251	274	640	19	24	282	307	714
	24	256	280	656		25	288	314	731
	25	262	287	672					
	18	270	291	666		19	303	325	741
	19	277	299	684		20	309	333	760
	20	283	306	702		21	316	341	779
	21	290	313	720		22	323	349	798
	22	296	321	738		23	330	357	817
	23	303	328	756		24	337	364	836
	24	309	335	774		25	344	372	855
20	25	316	343	792	21	21	373	399	903
						22	381	408	924
	20	337	361	820					
	21	344	370	840		23	388	417	945
	22	351	378	860		24	396	425	966
	23	359	386	880		25	404	434	987
22	24	366	394	900	23	23	451	481	1081
	25	373	403	920		24	459	491	1104
						25	468	500	1127
	22	411	439	990					
	23	419	448	1012		24	492	525	1176
24	24	427	457	1034	25	25	501	535	1200
	25	435	467	1056		25	536	570	1275



Приложение 8. Критические значения критерия Ватсона — Вильямса при  $n_1 = n_2(a)$  и  $n_2 = 2n_1(b)$

Приложение 9

Критические значения критерия Краскала—Уоллиса

$N_1$	$N_2$	$N_3$	Уровень значимости			$N_1$	$N_2$	$N_3$	Уровень значимости		
			$\alpha=0,10$	$\alpha=0,05$	$\alpha=0,01$				$\alpha=0,10$	$\alpha=0,05$	$\alpha=0,01$
2	2	2	4,57	—	—	5	3	1	4,01	4,96	—
3	2	2	4,50	4,71	—	5	3	2	4,65	5,25	6,82
3	3	2	4,55	5,36	—	5	3	3	4,53	5,34	6,98
3	3	3	4,62	5,60	7,20	5	4	1	3,98	4,98	6,95
4	2	2	4,37	5,33	—	5	4	2	4,54	5,27	7,11
4	3	2	4,51	5,44	6,44	5	4	3	4,54	5,63	7,44
4	3	3	4,70	5,72	6,74	5	4	4	4,61	5,61	7,76
4	4	1	4,16	4,96	6,66	5	5	1	4,10	5,12	7,30
4	4	2	4,55	5,45	7,03	5	5	2	4,50	5,33	7,33
4	4	3	4,54	5,59	7,14	5	5	3	4,54	5,70	7,57
4	4	4	4,65	5,69	7,65	5	5	4	4,52	5,66	7,82
5	2	2	4,37	5,16	6,53	5	5	5	4,56	5,78	7,98

Приложение 10

Критические значения критерия Фридмана

$Q$	$P$	Уровень значимости			
		$\alpha=0,05$	$\alpha=0,02$	$\alpha=0,01$	$\alpha=0,005$
3	3	6,000	—	—	—
3	4	6,500	8,000	8,000	8,000
3	5	6,400	6,400	8,400	10,000
3	6	7,000	8,333	9,000	10,333
3	7	7,143	8,000	8,857	10,286
3	8	6,250	7,750	9,000	9,750
3	9	6,222	8,000	8,667	10,667
4	2	6,000	—	—	—
4	3	7,400	8,200	9,000	9,000
4	4	7,800	8,400	9,600	10,200

Приложение 11

Значения функции нормального распределения с параметрами 0 и 1 (для отрицательных значений  $Z$ )

$Z$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-0,0	0,5000	0,4920	0,4929	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
-0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
-0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
-0,3	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
-0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
-0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
-0,7	2420	2398	2358	2327	2297	2266	2236	2206	2177	2148
-0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
-1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
-1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
-1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	9985
-1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
-1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
-1,5	0668	0655	0643	0630	0618	0606	0,594	0582	0570	0559
-1,6	0548	0537	0526	0515	0505	0495	0485	0475	0465	0455
-1,7	0,4046	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0359	0351	0344	0336	0329	0322	0314	0307	0300	0294
-1,9	0287	0281	0274	0268	0262	0256	0250	0244	0238	0233
-2,0	0227	0222	0217	0212	0207	0202	0197	0192	0188	0183
-2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
-2,2	0,139	0135	0132	0219	0125	0122	0119	0116	0113	0110
-2,3	0,107	0,104	0102	0099	0096	0094	0091	0089	0087	0084
-2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
-2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0047
-2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
-2,7	0035	0034	0033	0032	0030	0030	0029	0028	0027	0026
-2,8	0026	0025	0024	0023	0023	0022	0021	0020	0020	0019
-2,9	0019	0018	0017	0017	0016	0016	0015	0015	0014	0014
-3,0	0013	0013	0012	0012	0012	0011	0011	0011	0010	0010
-3,1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
-3,3	0005	0005	0004	0004	0004	0004	0004	0004	0004	0003
-3,4	0003	0003	0003	0003	0003	0003	0003	0003	0002	0002
-3,5	0002	0002	0002	0002	0002	0002	0002	0002	0002	0002
-3,6	0002	0002	0001	0001	0001	0001	0001	0001	0001	0001
-3,7	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
-3,8	0001	0001	0001	0001	0001	0001	0001	0001	0000	0000

Примечание. Значения функции для положительных  $Z$  находятся вычитанием из 1 значений функции для  $(-Z)$ . Пример: для  $Z=072 P=1-0,2358=0,7642$ .

## Приложение 12

Функция распределения Мизеса  $F(\theta/180^\circ, k)$  с круговым средним направлением  $\mu = 180^\circ$

$\theta$	$F(\theta)$							
	$k=0$	$k=0,2$	$k=0,4$	$k=0,6$	$k=0,8$	$k=1,0$	$k=1,2$	$k=1,4$
0	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
5	0,01389	0,01126	0,00895	0,00699	0,00536	0,00404	0,00301	0,00221
10	0,02778	0,02254	0,01793	0,01400	0,01074	0,00811	0,00604	0,00444
15	0,04167	0,03385	0,02697	0,02108	0,01620	0,01225	0,00913	0,00672
20	0,05556	0,04522	0,03608	0,02826	0,02175	0,01647	0,01230	0,00907
25	0,06944	0,05665	0,04531	0,03557	0,02744	0,02083	0,01559	0,01153
30	0,08333	0,06816	0,054467	0,04304	0,03329	0,02535	0,01903	0,01411
35	0,09722	0,07978	0,06420	0,05071	0,03936	0,03007	0,02266	0,01686
40	0,11111	0,09152	0,07392	0,05861	0,04567	0,03504	0,02650	0,01981
45	0,12500	0,10338	0,08386	0,06679	0,05228	0,04029	0,03062	0,02299
50	0,13889	0,11540	0,09405	0,07527	0,05921	0,04587	0,03505	0,02647
55	0,15278	0,12757	0,10452	0,08409	0,06653	0,05184	0,03985	0,03028
60	0,16667	0,13992	0,11529	0,09331	0,07428	0,05825	0,04509	0,03450

$\theta$	$F(\theta)$							
	$k=0$	$k=0,2$	$k=0,4$	$k=0,6$	$k=0,8$	$k=1,0$	$k=1,2$	$k=1,4$
65	0,18056	0,15246	0,12639	0,10295	0,08251	0,06517	0,05082	0,03919
70	0,19444	0,16520	0,13784	0,11306	0,09128	0,07265	0,05711	0,04442
75	0,20833	0,17815	0,14968	0,12368	0,10064	0,08078	0,06407	0,05030
80	0,22222	0,19132	0,16192	0,13485	0,11066	0,08962	0,07176	0,05690
85	0,23611	0,20471	0,17460	0,14662	0,12139	0,09925	0,08028	0,06436
90	0,25000	0,21834	0,18772	0,15901	0,13289	0,10975	0,08974	0,07277
95	0,26389	0,23222	0,20130	0,17206	0,14522	0,12122	0,10025	0,08228
100	0,27778	0,24633	0,21537	0,18582	0,15844	0,13372	0,11191	0,09302
105	0,29167	0,26069	0,22992	0,20030	0,17260	0,14734	0,12483	0,10514
110	0,30556	0,27529	0,24498	0,21554	0,18774	0,16217	0,13913	0,11876
115	0,31944	0,29014	0,26054	0,23154	0,20392	0,17825	0,15491	0,13405
120	0,33333	0,30522	0,27659	0,24832	0,22114	0,19566	0,17226	0,15112
125	0,34722	0,32053	0,29314	0,26587	0,23944	0,21444	0,19125	0,17009
130	0,36111	0,33606	0,31017	0,28420	0,25882	0,23460	0,21194	0,19106
135	0,37500	0,35180	0,32766	0,30327	0,27926	0,256165	0,23435	0,21408
140	0,38889	0,36774	0,34559	0,32306	0,30073	0,27909	0,25849	0,23918
145	0,40278	0,38385	0,36392	0,34353	0,32319	0,30334	0,28431	0,26633
150	0,41667	0,40013	0,38263	0,36463	0,34656	0,32883	0,31172	0,29544
155	0,43056	0,41655	0,40166	0,38628	0,37077	0,35546	0,34060	0,32638
160	0,44444	0,43309	0,42098	0,40841	0,30570	0,38309	0,37079	0,35897
170	0,47222	0,46644	0,46025	0,45379	0,44722	0,44066	0,43423	0,42800
175	0,48611	0,48321	0,49009	0,47684	0,47353	0,47022	0,46696	0,46381
180	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000

## Приложение 13

Допустимые значения критерия Стьюдента при данном объеме выборки  $N$  и уровне значимости  $\alpha$

$N$	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,02$	$\alpha=0,01$	$\alpha=0,001$	$N$	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,02$	$\alpha=0,01$	$\alpha=0,001$
1	6,31	12,71	31,82	63,66	636,62	18	1,73	2,10	2,55	2,88	3,92
2	2,92	4,30	6,97	9,93	31,60	19	1,73	2,09	2,54	2,86	3,88
3	2,35	3,18	4,54	5,84	12,94	20	1,73	2,09	2,53	2,85	3,85
4	2,13	2,78	3,75	4,60	8,61	21	1,72	2,08	2,52	2,83	3,82
5	2,02	2,57	3,37	4,03	6,86	22	1,72	2,07	2,51	2,82	3,79
6	1,94	2,45	3,14	3,71	5,96	23	1,71	2,07	2,50	2,81	3,77
7	1,90	2,37	3,00	3,50	5,41	24	1,71	2,06	2,49	2,80	3,75
8	1,86	2,31	2,90	3,36	5,04	25	1,71	2,06	2,49	2,79	3,73
9	1,83	2,26	2,82	3,25	4,78	26	1,71	2,06	2,48	2,78	3,71
10	1,81	2,23	2,76	3,17	4,59	27	1,70	2,05	2,47	2,77	3,69
11	1,80	2,20	2,72	3,11	4,44	28	1,70	2,05	2,47	2,76	3,67
12	1,78	2,18	2,68	3,06	4,32	29	1,70	2,05	2,46	2,76	3,66
13	1,77	2,16	2,65	3,01	4,22	30	1,70	2,04	2,46	2,75	3,65
14	1,76	2,15	2,62	2,98	4,14	40	1,68	2,02	2,42	2,70	3,55
15	1,75	2,13	2,60	2,95	4,07	60	1,67	2,00	2,39	2,66	3,46
16	1,75	2,12	2,58	2,92	4,02	120	1,66	1,98	2,36	2,62	3,37
17	1,74	2,11	2,57	2,90	3,97	$\infty$	1,65	1,96	2,33	2,58	3,29

## ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

---

### А

- Автокорреляционная функция 187
- Автоматизированные информационные системы (АИС) 229
- Автоматизированные системы обработки данных (АСОД)
  - 226, 229
- Анализ гармонический 200
  - дисперсионный 97
  - кластерный 131
  - корреляционный 106, 112, 127
  - поверхностей тренда 170
  - регрессионный 105, 134
  - факторный 150
- Анизотропия изменчивости 193, 198
  - поля 158, 198
- Аномалия 92, 161
- Антимода 42

### Б

- База данных 229
- Банк данных 229, 231

### В

- Вариограмма 189
- Вероятность доверительная 75
  - случайного события 32
- Взаимокорреляционная функция (ВКФ) 195
- Винзоризация 68

### Г

- Генеральная совокупность 25
- Гистограмма 34
- Главные компоненты 146
- Граф 130

### Д

- Двумерная автокорреляционная функция (ДАКФ) 195
- Диаграмма розы наблюдений 35
  - схематическая («ящик с усами») 30
  - числовая («стебель с листьями») 29
- Дискриминантная метка 143
  - функция 143
- Дисперсия
  - общая 38
  - остаточная 98, 101

- случайной функции 186
- условная 105
- факторная 98
- Дендрограмма 132
- Дендрограф 133
- Доверительный интервал 69

### 3

Закон распределения 33

### И

- Изменчивость геологических объектов 9, 15
  - закономерная 155, 160
  - наблюдаемая 15
  - природная 15
  - пространственная 155
  - случайная 155, 160
- Информативность признака 142

### К

- Ковариация 107
- Корреляционная зависимость 105
- Корреляционное отношение 107
  - поле точек 107
- Коэффициент анизотропии 193
  - аномалийности 176
  - вариации 38, 89
  - дистанционный 131
  - корреляции, множественный 128
  - корреляции, парный 114
  - корреляции, ранговый Спирмена 115
  - корреляции, условный 128
  - корреляции, частный 128
  - рудоносности 47
- Коэффициенты уравнения регрессии 122
- Критерий согласия 75
  - двусторонние 77
  - непараметрические 77
  - односторонние 77
  - параметрические 77
- Критерий Ван дер Вардена 81
  - Вилера — Ватсона — Ходжеса 85
  - Вилкоксона 83
  - Краскала — Уоллиса 102
  - Махаланобиса 145
  - Пирсона ( $\chi^2$ ) 57, 70
  - равномерности Релея 5
  - Родионова 79
  - Сиджела — Тьюки 89

- Смирнова 93
- Стыодента 79
- Фергюсона 93
- Фишера 88
- Фридмана 102
- Круговая дисперсия направлений 42
  - медиана 42
  - мода 42
- Круговое среднее направление 41
- Кумулята 34

## Л

- Лапласа, преобразование 44
- Линия регрессии 105

## М

- Математическое ожидание 36
  - случайной величины 36
  - случайной функции 185
- Медиана 38
- Метод главных компонент 146
  - максимального правдоподобия 66
- Моделирование геолого-математическое 23

## О

- Опробование многостадийное 12
  - равномерное 12
  - случайное 12
  - стратифицированное 12
- Оценка выборочная 62
  - интервальная 63
  - максимально-правдоподобная 65
  - максимально-эффективная 63
  - несмещенная 63
  - состоятельная 63
  - точечная 63
- Ошибка второго рода 75
  - первого рода 75

## П

- Параметр концентрации 50
- Переменная пространственная 157
- Период гармонических колебаний 201, 202
- Плотность распределения 33
- Погрешность аналогии 13
  - измерения 13
- Показатель асимметрии 39
  - эксцесса 39
- Полиномы ортогональные 170

— тригонометрические 170

Предел корреляции 193

Преобразование Фурье 201

## P

Размерность пространства 214

— случайной величины 214

Распознавание образов 137

Распределение асимметричное 44

— биномиальное 47

— логнормальное 44

— логнормальное отраженное 46

— логнормальное трехпараметрическое 47

— логнормальное усеченное 46

— Мизеса 50

— нормальное 43

— Пирсона 53, 57

— Пуассона 50

— равномерное 50, 58

— симметричное 43

— Стьюдента 79

— условное 105

— Фишера 88

Регрессия линейная 105, 120

— множественная 134

— нелинейная 123

## C

Сглаживание исходных данных 163

Системный подход 10

Случайная величина 32

— дискретная 32

— непрерывная 32

— угловая 32

Случайная последовательность 167, 184

Случайная функция 185

— полигармоническая 200

— стационарная 188

— эргодическая 190

Случайной функции дисперсия 186

— корреляционная функция 187

— математическое ожидание 185

— реализация 187

— сечение 186

Случайный процесс 184

Совокупность выборочная 17

— генеральная 25

— геологическая 11

— опробуемая 12

Спектр амплитудный 201  
— энергетический 201  
Спектральная плотность 201  
Среднее арифметическое 63  
— взвешенное 64  
— гармоническое 64  
— геометрическое 64  
— квадратичное 63  
— круговое направление 69  
Стандартное отклонение 38

## Т

Таксономия 231  
Типофункция 162  
Тренд-анализ 161, 170, 177

## У

Уровень значимости 69  
— строения 160, 161, 221

## Ф

Факторные нагрузки 150  
Факторный анализ 150  
Фон 92, 161

## Ч

Частость 33  
Частота 33  
Человеко-машинные системы 230

## Ш

Шкала измерения интервальная 18  
— номинальная 18  
— относительная 18  
— порядковая 18

## Э

Эксцесс 39  
Энтропия 176

# ОГЛАВЛЕНИЕ

---

Предисловие .....	3
Введение .....	4
<b>Глава 1. Особенности использования математических методов в геологии....</b>	9
1.1. Специфика геологических образований и процессов как объектов изучения .....	9
1.2. Методы изучения геологических объектов .....	11
1.3. Характер геологической информации .....	16
1.4. Моделирование в геологии .....	18
1.5. Типы геолого-математических моделей .....	20
1.6. Принципы и методы геолого-математического моделирования .....	22
1.7. Место и роль математических методов в процессе изучения геологических объектов .....	24
<b>Глава 2. Одномерные статистические модели .....</b>	25
2.1. Сущность и условия применения .....	25
2.2. Простейшие преобразования количественной геологической информации .....	28
2.3. Статистические характеристики, используемые в геологии .....	32
2.4. Основные статистические законы распределения, используемые в геологии .....	43
2.5. Проверка гипотез о законе распределения параметров геологических объектов .....	51
2.6. Точечные и интервальные оценки свойств геологических объектов.....	62
2.7. Статистическая проверка геологических гипотез .....	75
2.8. Проверка гипотез о равенстве средних (математических ожиданий) ....	78
2.9. Проверка гипотез о равенстве дисперсий .....	88
2.10. Анализ однородности выборочных геологических совокупностей .....	92
2.11. Однофакторный и двухфакторный дисперсионный анализ в геологии .....	97
<b>Глава 3. Двумерные статистические модели .....</b>	103
3.1. Сущность и условия применения двумерных статистических моделей .....	103
3.2. Простейшие преобразования двумерных случайных величин .....	107
3.3. Проверка гипотез о наличии корреляционной связи .....	112
3.4. Использование корреляционных связей для предсказания свойств геологических объектов .....	120
<b>Глава 4. Многомерные статистические модели .....</b>	126
4.1. Сущность и условия применения многомерных статистических моделей .....	126
4.2. Многомерный корреляционный анализ .....	127
4.3. Статистические методы выделения ассоциаций химических элементов .....	129
4.4. Кластер-анализ (дендограммы и дендрографы) .....	131
4.5. Множественная регрессия и ее использование для предсказания свойств геологических объектов .....	134
4.6. Задачи распознавания образов в геологии .....	137
4.7. Статистическая оценка перспективрудоносности территорий .....	140
4.8. Оценка информативности геологических признаков .....	142
4.9. Линейные дискриминантные функции .....	143
4.10. Метод главных компонент .....	146
4.11. Область применения многомерных статистических моделей в геологии .....	154

<b>Глава 5. Моделирование пространственных переменных .....</b>	<b>155</b>
5.1. Геологические объекты, как поля пространственных переменных .....	157
5.2. Элементы неоднородности, изменчивость и анизотропия геологических полей .....	158
5.3. Геологические, методические и технические факторы, обусловливающие появление закономерной и случайной составляющих наблюдаемой изменчивости .....	160
5.4. Фон, аномалии и поверхность тренда .....	161
5.5. Геометрические методы выявления закономерных составляющих признаков .....	162
5.6. Способы сглаживания случайных полей .....	163
5.7. Статистические методы проверки гипотез о наличии тренда .....	167
5.8. Аппроксимация поверхностей тренда полиномами и анализ остатков .....	170
5.9. Трансформация геологических пространственных переменных .....	175
5.10. Области применения горно-геометрических моделей и тренд-анализа в геологии .....	177
5.11. Моделирование дискретных случайных полей .....	179
<b>Глава 6. Моделирование геологических объектов с помощью случайных функций .....</b>	<b>184</b>
6.1. Принципы моделирования свойств геологических объектов с помощью случайных функций .....	184
6.2. Использование автокорреляционных функций для решения геологических задач .....	188
6.3. Область применения взаимных корреляционных функций и двумерных автокорреляционных функций в геологии .....	195
6.4. Область применения полигармонических случайных функций в геологии .....	200
<b>Глава 7. Факторы, определяющие выбор и эффективность использования математических методов .....</b>	<b>207</b>
7.1. Влияние типа геологической задачи на выбор математической модели .....	207
7.2. Влияние свойств геологических объектов на выбор и эффективность использования математических методов .....	214
7.3. Влияние методики изучения геологических объектов на выбор и эффективность использования математических методов .....	217
7.4. Роль геологического анализа при выборе геолого-математической модели .....	221
<b>Глава 8. Применение ЭВМ в геологии .....</b>	<b>223</b>
8.1. Автоматизация первичной обработки данных .....	224
8.2. Решение геологических задач с помощью ЭВМ .....	226
8.3. Автоматизированные системы обработки геологических данных .....	229
<b>Список рекомендуемой литературы .....</b>	<b>235</b>
<b>Приложения .....</b>	<b>236</b>
<b>Предметный указатель .....</b>	<b>245</b>

УЧЕБНОЕ ИЗДАНИЕ

Каждан Алексей Борисович  
Гуськов Олег Иванович

## МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОЛОГИИ

Заведующий редакцией *В. А. Крыжановский*  
Редактор издательства *А. П. Хуповка*  
Технические редакторы *В. А. Лыкова, Л. А. Миронова*  
Корректор *М. В. Дроздова*

ИБ 7684

---

Сдано в набор 04.05.90. Подписано в печать 04.10.90. Формат 60×90<sup>1</sup>/<sub>16</sub>. Бумага офсетная № 1. Гарнитура Таймс. Печать офсетная. Усл. печ. л. 16,0. Усл. кр.-отт. 16,0. Уч.-изд. л. 19,93. Тираж 6000 экз. Заказ 1094/1757-2 Цена 1 руб.

---

Ордена «Знак Почета» издательство «Недра»,  
125047 Москва, пл. Белорусского вокзала, 3.

Набрано в Ордена Октябрьской Революции и ордена Трудового Красного Знамени  
МПО «Первая Образцовая типография» Государственного комитета СССР по печати.  
113054, Москва, Валовая, 28.

Отпечатано в московской типографии № 6 Государственного комитета СССР по печати  
109088, Москва, Ж-88, Южнопортовая ул., 24