

А. Б. КАЖДИН, О. И. ГУСЬКОВ, А. А. ШИМАНСКИЙ

МАТЕМАТИЧЕСКОЕ
МОДЕЛИРОВАНИЕ
В ГЕОЛОГИИ
И РАЗВЕДКЕ
ПОЛЕЗНЫХ
ИСКОПАЕМЫХ

*Допущено
Министерством высшего и среднего специального
образования СССР в качестве учебного пособия
для студентов геологических специальностей вузов*



МОСКВА «НЕДРА» 1979

Каждан А. Б., Гуськов О. И., Шиманский А. А. Математическое моделирование в геологии и разведке полезных ископаемых. Учебное пособие. М., «Недра», 1979, 168 с.

В книге рассматриваются основы геолого-математического моделирования строения и свойств природных минеральных образований применительно к решению различных геологических задач. Излагаются принципы и методы геолого-математического моделирования как одного из методов познания недр. Описываются сущность и условия применения одномерных, двумерных и многомерных статистических моделей, горно-геометрические методы моделирования и модели типа случайных функций. Приводятся примеры использования этих моделей при различных геологических, минералого-петрографических исследованиях в процессе изучения геологии и разведки месторождений полезных ископаемых.

Книга является учебным пособием для студентов геологических специальностей вузов. Она может быть полезна и для широкого круга специалистов по геологии и разведке месторождений полезных ископаемых.

Табл. 33, ил. 57, список лит. — 58 назв.

Рецензенты:

Кафедра геологии месторождений и методики разведки Красноярского института цветных металлов, А. В. Канцель.

Стремительно возрастающие возможности использования ЭВМ при обработке полевых и экспериментальных геологических наблюдений вызвали необходимость внесения существенных изменений в учебные планы подготовки инженеров-геологов. Наряду с корректировкой программы преподавания математических дисциплин, возникла необходимость в создании специального курса, знакомящего студентов с особенностями геологических процессов и образований как объектов математического моделирования, со спецификой геологических задач, решаемых с помощью математических методов, и факторами, влияющими на эффективность использования этих методов. Такой курс, получивший название «Основы геолого-математического моделирования», с 1973 г. введен в МГРИ в программу обучения студентов по специальности геология и разведка месторождений полезных ископаемых. Специфика данного курса предполагает для специалистов-геологов, ведущих преподавание, полное овладение методами геолого-математического моделирования. Данный курс изучается на пятом семестре после завершения курса «Высшая математика» и одновременно с изучением основ программирования и получением навыков работы на ЭВМ. К началу прохождения курса студенты владеют знаниями по общей геологии, палеонтологии, минералогии, частично по петрографии и структурной геологии. Это обстоятельство ограничивает круг геологических задач и конкретных примеров, рассматриваемых в отдельных разделах курса. Так как полученные знания будут использоваться студентами в последующие годы при прохождении специальных дисциплин, в числе примеров рассматриваются и некоторые наиболее простые задачи поисков и разведок месторождений полезных ископаемых. На шестом семестре по данному курсу предусмотрена курсовая работа, при выполнении которой расчетные операции ведутся на малых ЭВМ.

В основу данного учебного пособия положены лекции по курсу «Основы геолого-математического моделирования», читавшиеся в 1973—1975 гг. профессором А. Б. Кажданом и в 1976—1978 гг. доцентом О. И. Гуськовым, методические разработки по лабораторным занятиям преподавателей кафедры методики поисков и разведки месторождений полезных ископаемых — О. И. Гуськова, П. И. Кушнарера, В. Н. Мошкина, Е. А. Сидоркова, М. В. Шумилина и П. П. Ясковского, а также опыт использования многомерных статистических моделей при

решении геологических задач доцента Иркутского политехнического института А. А. Шиманского — автора соответствующей главы книги. Остальные разделы пособия написаны А. Б. Кажданом и О. И. Гуськовым, под общей редакцией А. Б. Каждана.

Кроме студентов геологических вузов учебным пособием могут воспользоваться геологи производственных организаций для ознакомления с наиболее общими вопросами применения математических методов при решении геологических задач. Однако при этом необходимо учитывать, что пособие рассчитано на читателя, знакомого с такими разделами математики, как теория вероятностей, математическая статистика, теория случайных функций, матричная алгебра, гармонический анализ.

Для более глубокого изучения специфики решения отдельных геологических задач с помощью математических методов читатель может воспользоваться книгами В. Н. Бондаренко [4], Н. Н. Боровко [6], А. Б. Каждана [23], У. Крамбейна и Ф. Грейбила [27], Ж. Матерона [35], Р. Л. Миллера и Дж. С. Кана [36], Д. А. Родионова [43], Б. М. Роненсона и др. [44].

Авторы отчетливо сознают несовершенство первого издания учебного пособия, обусловленное сравнительно небольшим опытом преподавания данной дисциплины, разнообразием возможных подходов и сложностью решения поставленных задач, а также ограниченным объемом книги. Поэтому они искренне благодарят рецензентов книги А. В. Канцеля и Ю. Г. Шестакова (кафедра геологии месторождений и методики разведки Красноярского института цветных металлов) за ценные замечания, которые способствовали заметному улучшению ее содержания. Авторы будут также весьма благодарны всем, кто пришлет свои замечания, направленные на улучшение содержания и качества учебного пособия.

§ 1. «МАТЕМАТИКА — НОВОЕ ОРУЖИЕ ГЕОЛОГА»

Переживаемый нами период войдет в историю как период научно-технической революции во всех отраслях науки и техники. В связи с расширением возможностей современных физико-химических методов, а также бурным развитием электронно-вычислительной техники последние четверть века наблюдается глубокое проникновение математических методов во все отрасли естественных наук, в том числе и в геологию. Исключительные успехи некоторых естественных наук, например, биологии, теснейшим образом связаны с применением современных математических методов.

Необходимость применять математические методы при обработке и обобщении геологических данных все острее ощущается и в таких дисциплинах, как палеонтология, стратиграфия, структурная геология, литология, петрография и др., которые раньше считались чисто описательными. Это связано с тем, что ежегодно в геологических организациях накапливается огромный эмпирический материал — миллионы количественных определений химического состава различных минералов, химического и минерального состава горных пород и полезных ископаемых, их физических, горно-технических, технологических и других свойств. Из года в год поток количественной информации возрастает, а визуальные методы анализа и обобщения эмпирических данных не обеспечивают извлечения из него всей возможной полезной информации. В результате познавательные возможности проводимых геологических исследований заметно ограничиваются, что приводит к неоправданному снижению эффективности геологоразведочных работ. В то же время анализ количественных характеристик геологических объектов позволяет решать задачи, не разрешимые традиционными методами. Так, например, внедрение количественных методов определения абсолютного возраста и статистических методов обработки геохимических анализов дало возможность расчленить фаунистически неохарактеризованные (немые) толщи осадочных пород и сложные древние метаморфические комплексы, а появление ЭВМ позволило моделировать процессы осадко-накопления и восстанавливать историю геологического развития отдельных бассейнов по характеру изменения состава осадочных пород.

Острую необходимость внедрения математических методов

в практику своей работы испытывают и производственные геологические организации, осуществляющие поиски и разведку месторождений полезных ископаемых. Это обусловлено возросшими требованиями промышленности к объему и детальности геологоразведочных данных и обеспечивается возросшей технической оснащенностью геологоразведочных организаций современной электронно-вычислительной техникой.

Резкое увеличение количественной информации, получаемой в процессе разведки месторождений, вызвало необходимость разработки принципиально новых способов ее хранения, обработки и анализа с помощью ЭВМ. В ряде организаций территориальных геологических управлений Министерства геологии РСФСР — Якутского, Уральского, Северо-Восточного и т. д. — уже накоплен опыт обработки разведочных данных на ЭВМ. Большую исследовательскую и организационную работу по внедрению расчетов на ЭВМ в практику геологоразведочных работ выполняет Комплексная опытно-методическая экспедиция Министерства геологии Казахской ССР, имеющая в своем распоряжении крупный вычислительный центр. Всесоюзным институтом экономики минерального сырья разрабатывается отраслевая автоматизированная система управления (АСУ — Геология). Министерством Геологии СССР выделены значительные ассигнования для оснащения геологоразведочных организаций современной электронно-вычислительной техникой.

Учитывая возросшую роль математических методов и широкое использование ЭВМ в практической деятельности геологических организаций, академик А. В. Сидоренко образно назвал математику «новым оружием геолога» [46]. Но это оружие будет эффективным лишь в том случае, если теоретическими основами математического моделирования геологических процессов и объектов овладеют широкие массы геологов.

§ 2. ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ГЕОЛОГИИ И РАЗВЕДКЕ

Широкое внедрение математических методов в геологическую науку и практику сопряжено с рядом объективных трудностей. Геология принадлежит к числу естественных наук, которые в отличие от точных физико-химических наук принято называть описательными. Как теоретическая дисциплина геология развивалась в условиях практически полного отсутствия экспериментальных исследований, играющих ведущую роль в развитии «точных» физико-химических наук. Экспериментальной основой в геологии являются полевые наблюдения и результаты их камеральной обработки, по данным которых строятся гипотезы и делаются теоретические обобщения. В отличие от физика или химика, геолог, как правило, лишен возможности проверять свои теоретические построения путем

проведения соответствующего эксперимента. Поэтому основные теоретические выводы геологов основываются на интуиции, опираются на представления о генезисе изучаемых явлений, отражая в большей или меньшей степени убеждения авторов. В отличие от точных физико-химических наук, число и мера играют в геологии пока что скромную роль, выполняя в большей степени иллюстративные, нежели познавательные функции.

Системы научных понятий в геологии и в математике не соответствуют друг другу. Объекты математических исследований — пространственные формы и количественные отношения реального мира — абстрактны, однако все понятия в математике однозначны, логически совершенны и предельно лаконичны. С помощью строгой системы определений и доказательств они логически выводятся из принятых за основу исходных положений — аксиом.

Теоретические понятия геологии многоплановы и неоднозначны. Они многословны, сформулированы на языке, полном образных выражений, сравнений и аналогий. Многие из них противоречивы и, как правило, логически несовершенны.

На современном геологическом языке нельзя практически однозначно описать геологический объект или геологическое явление. Установлено [13], что в современной геологической науке существует 39 определений понятия минерал, 49 определений понятия горная порода, 63 определения понятия формация и 112 определений понятия фация. Так, например, В. И. Вернадский считает минерал продуктом химических реакций; Д. П. Григорьев — атомами, соединившимися в соответствующие структуры; А. С. Уклонский — сочетанием атомов или ионов; А. К. Болдырев — частью земной коры; А. Г. Бетехтин — частью горных пород и руд и т. д.

По А. Н. Заварицкому, горная порода — это агрегат минералов, состоящий из петрогенных элементов, а по Ф. Ю. Левинсон-Лессингу, — агрегат минералов, обладающих различным составом и структурой. Г. Розенбуш считает, что порода — это часть земной коры постоянного химического и минерального состава, а В. И. Вернадский понимает под горной породой участок литосферы, отличающийся происхождением, химическим и минеральным составом, а также физической характеристикой. Таким образом, каждый геолог определяет и описывает геологическое явление с позиций собственного подхода к пониманию предмета, вследствие чего описания лишены однозначности, отличаются сложностью и многоплановостью. Для использования подобных понятий в современной математике необходимо было бы из огромного набора геологических понятий выделить основные понятия — аксиомы, из которых путем операций классифицирования и математической логики выводились бы все остальные. Математически строгая формулировка понятия x должна указывать как отделить x от всей совокупности рас-

смаатриваемых понятий, что значит описать *x* и как обеспечить возможность экспериментального осуществления процесса отделения и описания *x*. Рассмотрим для примера принятую в геологии классификацию горных пород.

Все горные породы различаются по совокупности признаков, позволяющих судить о процессах их формирования, на *магматические (изверженные), осадочные, метаморфические*.

Изверженные породы разделяются по структурам и условиям их залегания на: *интрузивные, субвулканические, эффузивные, жильные (асхистовые и диасхистовые)*.

Каждая из этих четырех групп в свою очередь разделяется на ряд групп по минеральному и химическому составу. Количество классификаций велико (С. Мишель-Леви, Г. Розенбуш, Ф. Ю. Левинсон-Лессинг, П. Ниггли, А. Н. Заварицкий и др.), а количество групп различно. Например: *граниты — риолиты, сиениты — трахиты, диориты — андезиты, габбро — базальты, фельдшпатиты — фонолиты и др.*

В каждой из групп (например, в группе гранитов) выделяются семейства по признаку минерального и химического состава: *щелочные граниты* (преобладают калиевые полевые шпаты) и *нормальные граниты* (преобладают плагиоклазы).

Каждое семейство разделяется на виды и разновидности по темноцветным минералам: *биотитовые граниты, эгириновые граниты, риббекитовые граниты, щелочные аляскиты и аляскиты, двухслюдистые граниты, роговообманковые граниты, пироксеновые граниты, чарнокиты (гиперстеновые граниты)* или по структурным признакам: *рапакиви (овоиды, ортоклаза)*.

Нетрудно видеть, что петрографические классификации горных пород не удовлетворяют строгим требованиям математической логики. В основу классификации положен комплекс самых различных признаков, которые сочетаются в сложной последовательности и взаимоотношениях. Поэтому не все члены рассматриваемой системы могут быть выведены из основных.

От химического состава породы можно (в первом приближении) перейти к ее минеральному составу и наоборот. Однако минеральный состав породы зависит не только от химического состава, но и от условий ее образования (например, лейцит образуется только при малых давлениях, а слюды — только при больших). Поэтому в геологических классификациях минеральный состав пород используется дополнительно к химическому как один из ведущих признаков. Выделение групп и семейств магматических пород проводится по химическому составу, однако выделенные группы и семейства относятся одновременно как к интрузивным, так и к эффузивным, субвулканическим и некоторым дайковым породам. Вследствие неясности генетических особенностей, оказывающих существенное влияние на положение конкретной породы в систематической таблице, некоторые горные породы не находят в ней определен-

ного места. Например, эклогит относится по минеральному составу (присутствуют моноклинный пироксен и гранат) к группе пироксенитов, а по происхождению вероятнее всего является метаморфической породой. Таким образом, однозначно описать конкретную породу, т. е. x , с помощью набора символов, характеризующих тот или иной признак, а также отделить x от остального массива символов на основе геологических классификаций невозможно.

Несоответствие математических и геологических систем научных понятий приводит к взаимонепониманию геологов и математиков, в результате чего возникают различные мнения по поводу возможностей и целесообразности широкого внедрения математических методов в геологию.

Многие крупные геологи (например, С. Н. Бубнов и В. В. Белоусов) практически отрицают целесообразность привлечения математических методов в геологию, считая, что математически обоснованный опыт в геологии не играет и не может играть решающей роли, так как математике недоступна сложность природного явления, т. е. его внутренняя нерасчлененность. Другая крайняя точка зрения высказана Э. Э. Фотиади и Ю. А. Ворониным. Она сводится к тому, 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 базой для эффективного внедрения математических методов и ЭВМ в геологию. Наконец, многие геологи считают возможным использовать математические методы для обработки и обобщения экспериментальных данных, причем формализации подвергается не вся геологическая наука, а только объект непосредственного наблюдения в соответствии с поставленной задачей исследования. Такой процесс рассматривается как геолого-математическое моделирование, при выполнении которого должен быть обеспечен комплекс условий, гарантирующих соответствие геологических и математических моделей. Далее мы будем придерживаться именно этой точки зрения.

§ 3. КРАТКАЯ ИСТОРИЧЕСКАЯ СПРАВКА ОБ ИСПОЛЬЗОВАНИИ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ГЕОЛОГИИ

Первые работы по применению математических методов в геологии относятся к 30-м годам XIX в. Они связаны с именем Ч. Ляйеля, который с помощью выборочного метода на основании соотношений различных видов ископаемых раковин в раз-

резах произвел стратиграфическое расчленение третичных отложений. К эоцену он отнес отложения, содержащие менее 3,5% ныне живущих видов моллюсков, к миоцену соответственно от 3,5 до 17%, к нижнему плиоцену — от 17 до 50% и к верхнему плиоцену — более 50%. Работы по применению вероятностно-статистических методов в минералогии и петрографии на рубеже XIX и XX веков связаны с именами Г. Ниггли и Ф. Ю. Левинсона-Лессинга. По результатам статистической обработки химических анализов и микроскопических исследований шлифов ими установлены границы главных семейств горных пород, парагенетические типы породообразующих минералов и связи между составом пород и свойствами породообразующих минералов. При поисках и разведках месторождений полезных ископаемых статистические методы использовались с начала XX в. Н. Псаревым, С. Ю. Деборжинским, Д. А. Зенковым, В. В. Барышевым, В. И. Красниковым, К. Л. Пожарицким, В. Г. Соловьевым и многими другими.

В период с 1930 по 1960 г. широкие статистические исследования по изучению закономерностей распределения содержаний элементов в породах проводились Н. К. Разумовским, Д. А. Родионовым и Л. Аренсом. К этому же периоду относятся первые работы А. Б. Вистелиуса по корреляции немых красноцветных толщ острова Челекен и по анализу пористости нефтеносных пород в кернах скважин с помощью математического аппарата теории случайных функций. В период после 1950 г. появляется масса работ по применению различных математических методов практически во всех отраслях геологии. В их числе исследования М. Ф. Мирчинка и В. П. Бухарцева по выявлению корреляционными методами структурных соотношений в Волго-Уральской области, Д. А. Родионова — по статистическим методам разграничения геологических объектов, В. Н. Бондаренко — по статистическому изучению петрохимических особенностей вулканических пород, А. В. Канцеля — по математическому изучению особенностей процессов рудоотложения, Б. М. Роненсона — по разработке математических методов корреляции немых метаморфических толщ при крупномасштабном картировании, М. В. Раца — по изучению неоднородности инженерных свойств пород и многих других исследователей.

В области поисков и разведок в нашей стране за эти годы опубликованы исследования А. М. Марголина, З. Д. Низгурецкого, П. Л. Каллистова, В. И. Кузьмина. А за рубежом — Д. Криге, Э. Карлье, Ж. Матерона и др. Разработка математических методов применительно к решению геологоразведочных задач проводится в течение ряда лет на кафедре методики поисков и разведки МГРИ (А. Б. Каждан, М. В. Шумилин, В. А. Викентьев, О. И. Гуськов, Е. А. Сидорков и др.).

К настоящему времени в геологии и разведке месторождений

полезных ископаемых наиболее широкое применение находят статистические методы исследования, включающие теорию вероятностей, статистику случайных величин и статистику случайных функций. В последние годы стали применяться более сложные методы гармонического анализа и кибернетики.

§ 4. ПРИНЦИПЫ И МЕТОДЫ ГЕОЛОГО-МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Моделирование как средство познания закономерностей широко используется в самых различных областях науки и техники. Оно представляет собой метод исследования строения сложных природных объектов, динамики процессов или основных особенностей систем. Понятие модели в настоящее время весьма обширно. Это могут быть подобию предметов в измененном или натуральном масштабе (физическое или объемно-макетное моделирование), чертежи (графическое моделирование), символы или формулы (математическое моделирование), воспроизводящие свойства интересующих нас объектов.

Необходимость в моделировании возникает при изучении сложных природных явлений, которые В. В. Налимов [37] предложил рассматривать как «плохо организованные системы». В отличие от «хорошо организованных систем», зависящих от небольшого числа переменных, плохо организованные системы управляются одновременно многими факторами различной физической природы. С хорошо организованными системами имеют дело точные физико-химические науки. По результатам экспериментальных работ они поддаются количественному описанию, на основании которого выводятся законы, вскрывающие взаимосвязи между изучаемыми факторами.

Геологические объекты представляют собой классические плохо организованные природные системы, что объясняется сложностью минералообразующих породо- и рудообразующих процессов. Поэтому при изучении геологических объектов не удается разграничить действия многочисленных переменных различной физической природы. Природные геологические системы не поддаются безупречному количественному описанию, вследствие чего строгое понятие закона заменяется при их изучении более широким, хотя и весьма расплывчатым понятием модели.

В отличие от закона, имеющего на данном уровне знаний характер абсолютной истины, модель обеспечивает лишь приближенное представление о строении объекта (или о динамике изучаемого процесса). При этом любая, даже самая совершенная модель позволяет судить не о всех, а только о тех свойствах системы, для изучения которых осуществлялось моделирование.

Объектами моделирования могут быть отдельные участки земной коры, а также различные свойства природных геологи-

ческих образований — пород, минералов, полезных ископаемых, подземных вод или газов. В процессе моделирования познаются те свойства их строения, знание которых необходимо для решения научных и практических задач. Моделированию могут быть подвергнуты и процессы, происходящие (или происходившие) в земной коре. Результат такого рода моделирования способствует выявлению условий формирования минералов, горных пород и полезных ископаемых в недрах, обеспечивая научную основу геотектоники, петрографии и теории рудообразования.

Для целей количественного описания особенностей строения геологических объектов и процессов, происходящих в земных недрах, могут быть использованы методы физического, графического (в первую очередь горно-геометрического) и математического моделирования.

Физические (химические или физико-химические) модели процессов, происходящих в земной коре, используются в экспериментальной геотектонике, петрографии, геохимии и в других науках о земле. Они специфичны и представляют собой предмет самостоятельного исследования.

Графические, особенно горно-геометрические модели, широко используются в геологии и в горно-маркшейдерском деле. Горно-геометрические графики, планы и карты составляются с применением способа изолиний. Модели в изолиниях признака отражают морфологические свойства и внутреннее строение изучаемых объектов, а также особенности пространственного размещения изучаемых свойств. Методы горно-геометрического моделирования с помощью изолиний изучаются в курсах топографии, маркшейдерии и геометризации недр.

Математические модели также используются при изучении свойств, морфологии и строения геологических образований. В качестве математических моделей используются символы и формулы, описывающие количественные взаимосвязи и закономерности распределения изучаемых геологических признаков.

Природные геологические объекты обладают рядом специфических особенностей, которые в значительной мере определяют методику их изучения:

— горные породы и содержащиеся в них скопления полезных ископаемых скрыты в недрах и практически недоступны для непосредственного наблюдения;

— размеры большинства изучаемых объектов несоизмеримо больше, чем размеры естественных или искусственных обнажений, по которым производится их изучение;

— изучаемые объекты — природные минеральные скопления и многие массивы горных пород — часто обладают сложным внутренним строением.

Так, например, золоторудные минерализованные зоны дробления обычно состоят из отдельных сближенных золоторудных

залежей, разделенных участками слабоминерализованных пород. В свою очередь золоторудные залежи также обладают прерывистым строением и представлены чередованием рудных гнезд или скоплений неправильной формы с участками практически пустых пород, а каждое рудное гнездо состоит из чередующихся золотосодержащих и безрудных минеральных агрегатов. При более точном определении золоторудных агрегатов под микроскопом устанавливается, что и они в свою очередь сложены самородками золота, золотосодержащими сульфидами или кварцем и зернами безрудного кварца (рис. 1).

Для изучения горных пород и природных скоплений полезных ископаемых, скрытых в недрах, следует применять сеть естественных и искусственных обнажений. В качестве искусственных обнажений используются разведочные горные выработки и скважины, по которым проводятся геологические наблюдения, отбираются образцы и пробы для изучения свойств изучаемых объектов.

Перечисленные особенности определяют основные принципы математического моделирования природных геологических образований и их свойств.

1. Математическое описание свойств природных геологических объектов должно проводиться на основе системного подхода к оценке особенностей их внутреннего строения. С этой целью внутреннее строение объектов рассматривается как система. На каждом из уровней структура объекта определяется совокупностью множества условно однородных структурных единиц, которые выступают на данном уровне строения как элементы неоднородности.

Резкие несоответствия размеров природных скоплений полезных ископаемых или массивов изучаемых горных пород и размеров обнажений приводят к тому, что фактические наблюдения обеспечивают информацию о строении изучаемых объектов только на структурных уровнях высших порядков, элементы неоднородности которых заведомо меньше площади обнажения. Для суждения же о структуре изучаемого объекта в промежутках между пунктами наблюдений фактические данные не могут быть получены. Обобщенное суждение о структуре возможно только на основании усредненных характеристик, вычисленных по группам пространственно сближенных единичных наблюдений и моделирования характера изменчивости параметров объекта.

2. Природная сложность и недоступность геологических объектов, несопоставимость их размеров с размерами отбираемых проб, ограниченность экспериментальных данных и прерывистость сети наблюдений не способствуют широкому применению детерминированных математических моделей. В этих условиях экспериментальные данные представляют собой совокупность случайных, а не детерминированных величин,

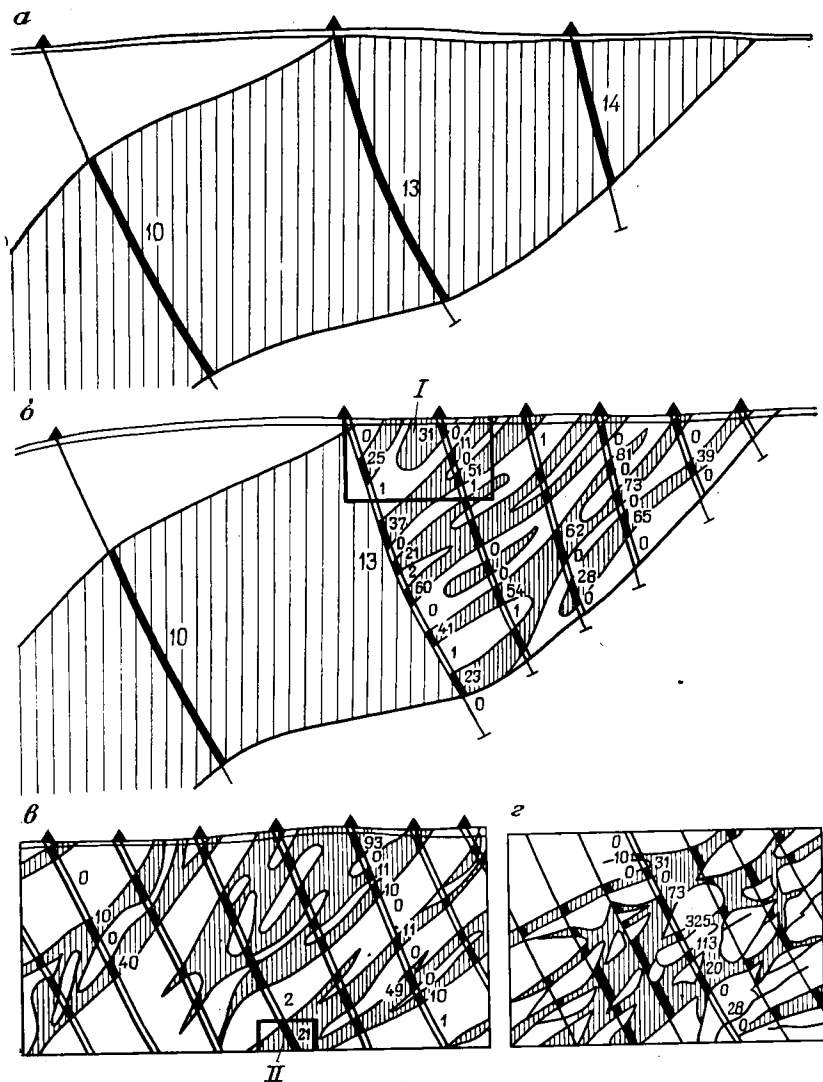


Рис. 1. Уровни строения золоторудного месторождения:
a — по данным сквозных проб (цифрами указаны содержания золота в условных единицах); *б* — детализация разреза по данным секционных проб; *в* — участок детализации I; *г* — участок детализации II

из-за чего большинство математических моделей в геологии строится на вероятностной основе. Однако проявление случайной изменчивости свойств геологических образований указывает не на отсутствие геологических закономерностей, а лишь на неполноту наших знаний об истинной природе явлений на данном этапе их изучения.

3. Выбор наиболее приемлемой математической модели определяется условием соответствия ее свойств свойствам объекта моделирования. Но поскольку полнота представлений о свойствах геологических объектов зависит от их природной сложности и детальности проведенных геологических наблюдений, правильнее говорить о соответствии математической модели тому представлению геологов, которое выработалось у них к моменту моделирования по совокупности эмпирических наблюдений на данном объекте, а также на основе геологической информации, полученной по данным изучения сходных геологических объектов.

Практически это означает, что математические модели свойств геологических образований должны разрабатываться на базе типовых геологических моделей природных объектов и могут поэтому с полным правом называться *геолого-математическими*.

4. В рамках математического моделирования невозможно полное описание свойств изучаемого объекта. Описанию подвергаются только отдельные свойства природных геологических систем и связанные с ними явления, которые и подлежат предварительному обобщению и формализации.

Таким образом, специфическая особенность геолого-математического моделирования заключается в том, что моделируются не истинные геологические структуры и свойства природных объектов, а изменчивость этих свойств, наблюдаемая на изучаемом уровне строения объекта. Характер этой *наблюдаемой изменчивости* зависит от природы явления, а также от методики и детальности проведенных геологических исследований.

ГЛАВА I

ОДНОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

**§ 1. СУЩНОСТЬ И УСЛОВИЯ ПРИМЕНЕНИЯ ОДНОМЕРНОЙ
СТАТИСТИЧЕСКОЙ МОДЕЛИ**

Статистика в широком смысле слова— это обобщение и наглядное представление эмпирических данных большого объема с последующими выводами из этих данных. Статистика позволяет распространить выводы, полученные по ограниченному числу наблюдений (выборке), на весь объект (совокупность). Именно эта задача решается при проведении любых геологических исследований. Она сводится к выборочному изучению состава и свойств горных пород, минералов или полезных ископаемых в зернах, шлифах, образцах и пробах, отобранных в отдельных участках земной коры. Каждое выборочное наблюдение относится к элементарно малому, пространственно разобщенному объему недр (в искусственном или естественном обнажении), а выводы, полученные по наблюдениям, распространяются на весь изучаемый объем — между пунктами наблюдений и за их пределами.

Генеральная совокупность случайных величин в геологии может быть конечной (например, совокупность анализов химических проб) или бесконечной (например, совокупность возможных замеров мощности залежи полезного ископаемого).

Выводы и заключения, сделанные на основании выборки, всегда имеют случайный характер, однако они могут быть в большей или в меньшей степени распространены на всю совокупность, т. е. на изучаемое явление в целом. Методы математической статистики обеспечивают возможность наилучшего использования выборочной информации для получения наиболее надежных результатов и для определения степени надежности полученных выводов.

Для использования случайной величины в качестве статистической модели свойств геологического объекта необходимо убедиться в том, что геологические наблюдения:

— удовлетворяют условию массовости, обеспечивая возможность многократного повторения одного и того же комплекса условий;

— могут быть представлены в виде схемы случайных событий;

— могут быть выражены случайной величиной.

При проведении геологических исследований комплекс условий заключается в замерах значений изучаемых свойств в некоторых произвольно выбранных точках пространства земных недр. Реализация данного комплекса условий может рассматриваться как серия случайных событий, а числовые значения наблюдаемых свойств представляют собой величины случайные, так как их нельзя предсказать заранее. Комплекс реализуемых условий может быть повторен многократно и с этой точки зрения он удовлетворяет условию массовости.

Следовательно, явления, изучаемые в процессе геологических наблюдений, могут рассматриваться как массовые и случайные. При использовании статистической модели для изучения закономерностей распределения важнейших свойств геологического объекта отдельные участки недр уподобляются генеральной статистической совокупности, состоящей из практически бесконечно большого количества элементарных участков. Каждый из таких элементарных участков соответствует по размеру отдельной пробе или единичному замеру изучаемого свойства, который рассматривается как «случайная величина». Среднее значение свойства в объеме всего участка представляет собой математическое ожидание этой случайной величины.

Процесс опробования или замера свойства уподобляется серии массовых испытаний, каждое из которых выявляет его частное значение. Результаты испытаний образуют выборочную статистическую совокупность, которая состоит из обособленных, количественно однородных элементов. Каждый единичный замер или наблюдение принимается за вариант реализации случайной величины. По самому смыслу статистической модели предполагается, что все варианты представляют собой отдельные независимые отклонения величины параметра относительно его среднего значения.

Ввиду того, что в статистических моделях все колебания значений изучаемых свойств рассматриваются как случайные, область их применения ограничена объектами, для которых характерно отсутствие каких-либо закономерных изменений в пространстве, или задачами, при решении которых эти закономерности можно не учитывать.

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

Выборочной оценкой неизвестного параметра или его числовой характеристики (среднего значения признака, его дисперсии, значения коэффициента вариации и др.) называется значение этого параметра, вычисленное на основании выборочных данных. Выборочная оценка всегда является случайной величиной. В задачу математической статистики входит выбор такого метода вычисления оценки, который обеспечил бы ее

наилучшее приближение к оцениваемому параметру, а также определение степени надежности полученной оценки.

Статистические гипотезы проверяют правдоподобность выводов о тех или иных закономерностях, полученных на основании анализа выборочных данных (гипотезы о соответствии эмпирических распределений теоретическим, о равенстве или различии полученных выборочных оценок и др.) Гипотеза считается правдоподобной и принимается, если существование исследуемой закономерности для явления в целом может рассматриваться как событие практически достоверное.

§ 2. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

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 равна нулю. Таким образом, вероятность случайного события выражается числом, лежащим в пределах от 0 до 1.

Случайные величины бывают прерывистыми (дискретными) и непрерывными. Случайная величина, принимающая в испытаниях только определенные числовые значения, называется дискретной. Примером дискретных случайных величин может служить количество зерен определенного минерала при изучении шлифов под микроскопом, количество скважин, вскрывших рудное тело при разведке месторождения и т. д., так как эти величины могут выражаться только целыми числами. Непрерывная случайная величина может принимать в заданном интервале любые значения. Например, содержание свинца в рудах полиметаллического месторождения, определенное по пробам малого объема, может быть любым от нуля до величины содержания свинца в чистом галените.

Число появлений события в серии испытаний называется его частотой, а отношение числа появлений события к общему числу опытов в серии — его частотью. При увеличении числа опытов (объема выборки) частота событий 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 представляет собой наиболее полную характеристику случайной величины, так как устанавливает связь между возможными значениями случайной величины и соответствующими им вероятностями (рис. 2).

Интегральная функция распределения $F(x)$ выражает вероятность того, что выборочное значение случайной величины

ξ окажется меньше некоторого предела, ограниченного x , где x — заданная переменная (рис. 2, а), т. е. вероятность события $\xi \leq x$. (I.1)

Дифференциальная функция распределения (функция плотности распределения) $f(x)$ характеризует вероятность попадания выборочного значения случайной величины ξ в заданный интервал от x до $x + \Delta x$ (рис. 2, б), т. е. вероятность события $x < \xi < x + \Delta x$. (I.2)

Интегральная и дифференциальная функции распределения связаны отношением

$$F(x) = \int_{-\infty}^x f(x) dx, \quad \text{причем} \quad \int_{-\infty}^{+\infty} f(x) dx = 1. \quad (I.3)$$

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

К важнейшим характеристикам положения относятся математическое ожидание (среднее значение), мода и медиана.

Математическое ожидание (Mx) дискретной (прерывистой) случайной величины есть сумма произведения всех возможных значений на соответствующие им вероятности

$$Mx = \sum_{i=1}^n x_i p_i. \quad (I.4)$$

Математическое ожидание непрерывной величины выражается через плотность распределения

$$Mx = \int_{-\infty}^{+\infty} x f(x) dx. \quad (I.5)$$

Мода (M_0) случайной величины указывает ее наиболее вероятное значение (для дискретной величины) или значение с максимальной плотностью вероятности (для непрерывной величины).

Медиана (Me) случайной величины указывает ее значение для вероятности 0,5, т. е. такое значение, для которого вероятности (p) встречи больших и меньших значений в выборке равны

$$p(\xi \leq Me) = p(\xi > Me). \quad (I.6)$$

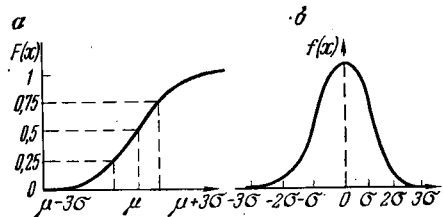


Рис. 2. Функции распределения: а — интегральная; б — дифференциальная

В общем случае математическое ожидание, медиана и мода случайной величины имеют различные значения. Они совпадают только для симметричных распределений.

Характеристиками рассеяния, определяющими степень разброса значений случайной величины вокруг ее математического ожидания, служат центральные моменты различных порядков.

Главной характеристикой рассеяния случайной величины служит центральный момент второго порядка — дисперсия случайной величины

$$\mu_2 = \sigma^2. \quad (I.7)$$

Дисперсия дискретной случайной величины равна

$$\sigma^2 = \sum_{i=1}^n (x_i - Mx)^2 p_i, \quad (I.8)$$

а дисперсия непрерывной случайной величины

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - Mx)^2 f(x) dx. \quad (I.9)$$

Производными характеристиками от дисперсии является стандарт (среднее квадратичное отклонение)

$$\sigma = \sqrt{\sigma^2} \quad (I.10)$$

и коэффициент вариации

$$V = \frac{\sigma}{Mx} 100\%. \quad (I.11)$$

Для характеристики степени асимметрии распределения случайной величины относительно ее математического ожидания используется центральный момент третьего порядка (третий центральный момент), вычисляемый как

$$\mu_3 = \sum_{i=1}^n (x_i - Mx)^3 p_i, \quad (I.12)$$

или

$$\mu_3 = \int_{-\infty}^{+\infty} (x_i - Mx)^3 f(x) dx \quad (I.13)$$

соответственно для дискретных или для непрерывных случайных величин. Для симметричного распределения значение третьего центрального момента равно нулю. Если распределение асимметрично, то его значение отличается от нуля в положительную или отрицательную сторону тем сильнее, чем больше выражена асимметрия.

Коэффициентом асимметрии называют безразмерную величину

$$A = \frac{\mu_3}{\sigma^3}. \quad (I.14)$$

Если $A > 0$, распределение называется положительно асимметричным, а если $A < 0$ — отрицательно асимметричным. Кривые положительно и отрицательно асимметричных распределений показаны на рисунках 2, б и 3.

Центральный момент четвертого порядка (четвертый центральный момент) служит для характеристики эксцессивности (степени островершинности) кривой распределения. Для широ-

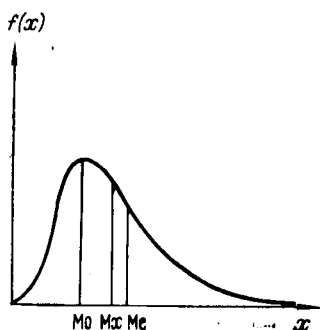


Рис. 3. Мода, медиана и средняя логнормального распределения

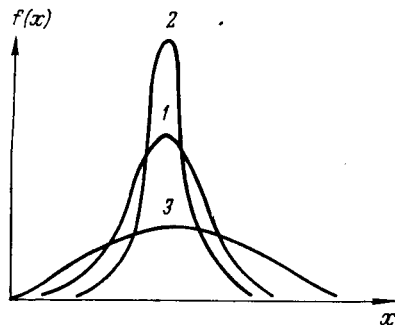


Рис. 4. Кривые плотности вероятности с различными значениями эксцесса: 1 — $E=0$; 2 — $E>0$; 3 — $E<0$

ко распространенного нормального распределения величина отношения

$$\frac{\mu_4}{\sigma^4} = 3. \quad (I.15)$$

Поэтому для характеристики эксцессивности распределения используется показатель E , называемый эксцессом, который рассчитывается по формуле

$$E = \frac{\mu_4}{\sigma^4} - 3. \quad (I.16)$$

Кривые, более островершинные по сравнению с нормальными имеют положительный, а более пологие — отрицательный эксцесс (рис. 4).

§ 3. ЗАКОНЫ СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Для аппроксимации (приближенного описания) эмпирических распределений различных свойств геологических объектов предлагалось использовать самые различные теоретические законы распределений — нормальный и логарифмически нор-

мальный, биномиальный, гамма- и бета-распределения, законы Вейбулла, Пуассона и др. При решении большинства статистических задач в геологии и разведке наиболее широко используются нормальный, логарифмически нормальный и биномиальный законы распределения. Двум первым законам подчиняются распределения большинства изучаемых непрерывных случайных величин, а биномиальный закон распределения используется для описания дискретных случайных величин.

Нормальным законом распределения называется закон, для которого интегральная функция распределения имеет следующий вид:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-Mx)}{2\sigma^2}} dx. \quad (I.17)$$

Величины Mx и σ носят название параметров распределения и представляют собой математическое ожидание (Mx) и стандарт (σ) случайной величины.

Функция плотности нормального распределения имеет следующий вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-Mx)^2}{2\sigma^2}}. \quad (I.18)$$

Кривая плотности нормального распределения симметрична относительно математического ожидания. Мода, медиана и математическое ожидание для нормального распределения совпадают.

Нормальное распределение возникает, когда случайная величина может рассматриваться как сумма очень большого числа независимых случайных величин, влияние каждой из которых на эту сумму является равномерным и малым. Нормальному закону следуют распределения случайных ошибок измерений химических и спектральных анализов, измерений объемного веса, плотности, пористости и влажности, распределение наблюдаемых мощностей геологических тел и реже содержаний минералов и элементов. Нормальный закон распределения является предельным законом, к которому в определенных условиях сходятся многие другие распределения.

Распределение величины $\tau = \frac{x - Mx}{\sigma_x}$ с математическим ожиданием $M_\tau = 0$ и дисперсией $\sigma_\tau = 1$ называется *нормированным*. Переход к нормированному распределению заключается в переносе центра распределения в начало координат с выражением случайной величины в долях ее стандарта.

Значения функции нормированного распределения $\Phi(\tau)$ табулированы (нормированная функция Лапласа).

Логарифмически нормальным (логнормальным) называется

закон распределения, при котором нормально распределены логарифмы значений случайной величины. Логарифмически нормальное распределение является положительно асимметричным и имеет положительный эксцесс (см. рис. 3).

Математическое ожидание случайной величины, мода и медиана логнормального распределения не совпадают, причем $M_0 < M_e < M_x$.

Кривая плотности логарифмически нормального распределения имеет вид нормальной кривой, если по оси абсцисс откладываются значения логарифмов величины. Для логарифмов величин $M_{0 \ln x} = M_{e \ln x} = M_{1 \ln x}$.

Функция распределения для логарифмов случайной величины имеет вид

$$F(\ln \xi \leq \ln x) = \frac{1}{\sigma_{\ln x} \sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(\ln x - M_{1 \ln x})^2}{2\sigma^2}} d \ln x, \quad (I.19)$$

т. е. полностью соответствует функции нормального распределения, где параметры $M_{1 \ln x}$ — математическое ожидание $\ln x$, а $\sigma_{\ln x}$ — стандарт логарифмов x .

Функция плотности логарифмически нормального распределения

$$f^{(\ln x)} = \frac{1}{\sigma_{\ln x} \sqrt{2\pi}} e^{-\frac{(\ln x - M_{1 \ln x})^2}{2\sigma_{\ln x}^2}} \quad (I.20)$$

Величина $M_e = e^{M_{1 \ln x}}$ для логнормально распределенной величины будет соответствовать медиане, а величина $M_0 = e^{M_{1 \ln x} - \sigma_{\ln x}^2}$ будет соответствовать моде.

Условия возникновения логарифмически нормального закона распределения, как и в случае нормального, определяются условием равномерной масти независимых слагаемых в сумме. Однако в качестве таких слагаемых здесь должны рассматриваться логарифмы случайной величины. Проявление же логнормальности связано с отчетливым эффектом пропорциональности. А. Н. Колмогоровым было показано, что логарифмически нормально распределяется диаметр частиц при дроблении их под воздействием некоторой силы. Оно свойственно также содержаниям редких элементов и минералов в горных породах, содержаниям цветных, редких и благородных металлов в рудах, абсолютным отметкам горного рельефа и др.

Биномиальное распределение используется в статистике для описания таких явлений или объектов, при изучении которых в результате каждого испытания может произойти либо событие А, либо событие В. При проведении n независимых испыта-

ний число появлений события А будет случайной величиной X , а при проведении неограниченно большого количества серий по n испытаний в каждой среднее значение величины $x:n$ (т. е. частоты появления события А) будет соответствовать p — вероятности события А. Вероятность альтернативного события В будет равна

$$q = 1 - p. \quad (1.21)$$

Биномиальный закон распределения описывает совокупность вероятностей случайной величины X при $x=0, 1, 2, \dots$ и фиксированных количествах испытаний n . Величины n и p называются параметрами биномиального распределения, а вероятность события x выражается формулой

$$p_n(x) = C_n^x p^x q^{n-x} = C_n^x p^x (1-p)^{n-x}. \quad (1.22)$$

где C_n^x — число сочетаний из n по x . т. е.

$$C_n^x = \frac{n!}{x!(n-x)!}. \quad (1.23)$$

Основные числовые характеристики биномиального распределения:

$$\text{математическое ожидание } Mx = np; \quad (1.24)$$

$$\text{дисперсия } \sigma^2 = np(1-p). \quad (1.25)$$

С увеличением n биномиальное распределение стремится к нормальному.

Биномиальный закон распределения используется в качестве математической модели для определения процентного содержания зерен определенных минералов в шлихах, для оценки коэффициента рудоносности залежей, показывающего, какая часть объема залежи сложена кондиционными рудами.

§ 4. ВЫБОР ВИДА СТАТИСТИЧЕСКОЙ МОДЕЛИ

Выбор вида одномерной статистической модели заключается в подборе теоретического распределения, наилучшим образом описывающего закономерности, наблюдаемые в эмпирическом ряду. Выбрав некоторую теоретическую модель, решают вопрос о согласованности теоретического распределения этой модели с эмпирическим распределением выборочных значений случайной величины. Распределения считаются согласующимися, если расхождение между ними можно отнести за счет случайного различия, обусловленного случайностью самой выборки.

Если выборка достаточно велика, приближенно судить о законе распределения исследуемой случайной величины можно по виду эмпирических графиков частотных распределений (гистограмм). Однако вид гистограммы в значительной степени зависит от случайности выборки, в особенности при не очень большом ее размере.

Более надежным является графический метод проверки согласованности эмпирического и теоретического распределений, основанный на построении спрямленных графиков, накопленных частостей на вероятностной бумаге в нормальном или логнормальном масштабах (рис. 5).

Еще более надежными являются аналитические методы проверки согласованности эмпирического и теоретического распределений. Как отмечалось выше, для нормального распределения значения числовых характеристик

$$A = \frac{\mu_3}{\sigma^3} \quad \text{и} \quad E = \frac{\mu_4}{\sigma^4} - 3$$

равны нулю. Следовательно, выборочные значения этих характеристик могут быть использованы для оценки согласованности эмпирического распределения с нормальным. Этот метод удобен при проверке гипотезы о законе распределения по выборкам среднего (30—50 наблюдений) объема [4].

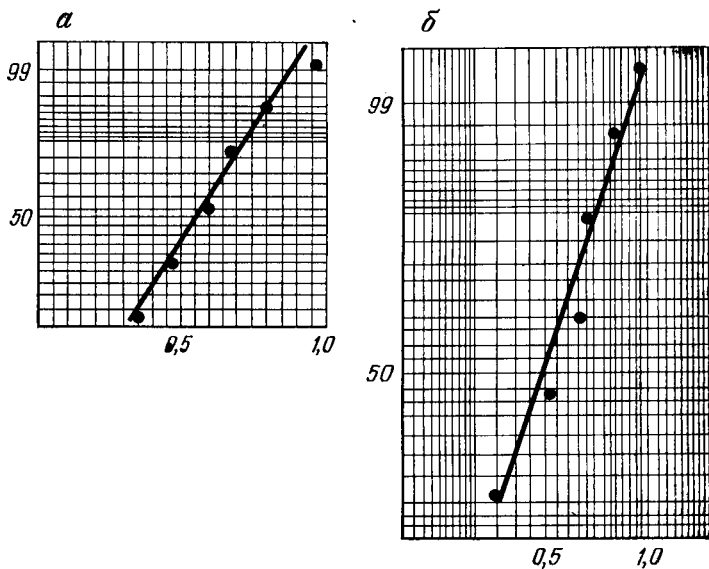


Рис. 5. Графический способ проверки функции распределения на бумагах: а — нормальной; б — логнормальной

Аналогично можно произвести проверку гипотезы и о логнормальном распределении, если оценки A и E вычислять для логарифмов значений случайной величины.

Наиболее точно и полно задача о согласованности данного эмпирического распределения с некоторым теоретическим решается с помощью критерия Пирсона χ^2 . Для этого вычисляются частоты теоретического распределения, параметры которого Mx

и σ^2 равны значениям этих параметров данного эмпирического распределения, а затем эти теоретические частоты сравниваются с частотами, наблюдаемыми в эмпирическом распределении.

Гипотеза о согласии распределений принимается, если вычисленное значение χ^2 оказывается меньше табличного для заданной доверительной вероятности и при данном числе степеней свободы, и отвергается, если вычисленное значение χ^2 больше соответствующего табличного.

Критерий χ^2 может применяться только для выборок значительного объема (60 и более значений). По выборкам меньше 15—20 значений ни один из методов проверки гипотезы о законе распределения не дает надежных результатов.

Принятие гипотезы о соответствии какой-либо теоретической модели эмпирическому распределению не исключает возможного соответствия этих же выборочных данных другому теоретическому распределению.

В геологической практике не редки случаи, когда по выборочным данным не отвергается гипотеза о их соответствии как нормальному, так и логарифмически нормальному закону. Особенно часто это наблюдается для выборок малого объема или выборок из неоднородных генеральных совокупностей (неоднородных геологических объектов). В этих случаях во избежание ошибок необходимо либо увеличить объем выборки, либо разделить изучаемый объект по геологическим признакам на несколько самостоятельных объектов.

§ 5. ОЦЕНКА ЧИСЛОВЫХ ХАРАКТЕРИСТИК СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

В практике геологических исследований особенно часто возникает необходимость статистической оценки средних значений случайных величин и их дисперсий (коэффициентов вариации).

Эти числовые характеристики используются:

— в качестве классификационных признаков для деления геологических объектов на группы. Например, по среднему содержанию SiO_2 решается вопрос о принадлежности интрузивной породы к группе кислых, средних или основных. На средних размерах обломочных частиц основана классификация терригенных пород, а по дисперсии этого параметра можно судить о степени сортированности обломочного материала;

— при оценке месторождений и подсчетах запасов руд. Так, например, основные параметры рудных тел — средняя мощность, среднее содержание полезного компонента, коэффициент рудоносности и т. д. входят в формулу подсчета запасов полезного ископаемого, а сама цифра запасов определяет промышленную ценность месторождения;

— для оценки инженерно-геологических свойств грунтов и пород при строительстве промышленных объектов и сооружений;

— для определения величин возможных ошибок при расчетах средних значений свойств геологических объектов по выборочным данным, определения числа наблюдений для вычисления средних значений с наперед заданной точностью.

Статистические оценки могут быть *точечными* или *интервальными*.

При точечной оценке неизвестная характеристика случайной величины оценивается некоторым числом, а при интервальной оценке указывается некоторый интервал значений, в пределах которого 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 размера выборки.

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

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

Для оценки средних значений наблюдаемых геологических признаков пользуются различными выборочными средними — средней гармонической $\bar{x}_{\text{гарм}} = \frac{n}{\sum_1^n \frac{1}{x}}$, средней геометрической $\bar{lg} x = \frac{\sum_1^n lg x_i}{n}$, средней арифметической $\bar{x} = \frac{\sum_1^n x_i}{n}$, сред-

ней квадратичной $\bar{x}_{\text{кв}} = \sqrt{\frac{\sum_1^n x^2}{n}}$ и др. Наиболее широко

распространена *средняя арифметическая*. В отличие от других средних (например, от средней геометрической, т. е. медианы) при увеличении числа наблюдений средняя арифметическая всегда приближается к неизвестному значению математического ожидания независимо от функции распределения случайной величины (т. е. всегда является состоятельной оценкой). Поэтому для статистической оценки большинства средних значений независимо от функции распределения случайной величины чаще всего используется средняя арифметическая (взвешенная или

невзвешенная). Если распределение резко асимметрично (например, логнормально), средняя арифметическая теряет свойство максимально-эффективной, хотя оба других свойства сохраняются. Это связано с тем, что на величину средней арифметической оказывают сильное влияние единичные большие значения, которые обычно встречаются в малых по объему выборках из совокупностей, имеющих положительно асимметричные распределения.

Иногда в таких случаях пользуются так называемыми максимально правдоподобными оценками средних, которые обеспечивают более высокую эффективность оценок, но в случаях несоответствия эмпирического распределения логнормальному оказываются смещенными и несостоятельными.

Максимально правдоподобную оценку (a) можно вычислить следующим образом:

$$a = e^{\overline{\ln x}} \psi_n(t), \quad (I.26)$$

где

$$t = \frac{1}{2} S_{\ln x}^2,$$

а

$$\psi_n(t) = e^t \left\{ 1 - \frac{t(t+1)}{n} + \frac{t^2(3t^2 + 22t + 21)}{6n^2} + O\left(\frac{1}{n^3}\right) \right\}. \quad (I.27)$$

В этих выражениях $\overline{\ln x}$ — среднее значение натуральных логарифмов изучаемой случайной величины, $S_{\ln x}^2$ — оценка дисперсии натуральных логарифмов, а величина $O\left(\frac{1}{n^3}\right)$ имеет порядок $\frac{1}{n^3}$, т. е. весьма мала, поэтому при расчетах ее обычно не учитывают. Для расчета a используются таблицы функции $\psi_n(t)$ [53] или соответствующие номограммы.

По сравнению со средней арифметической максимально правдоподобная оценка обладает существенно большей эффективностью только при оценке среднего весьма изменчивых логнормально распределенных случайных величин по выборкам малого объема.

Несмещенная, состоятельная и максимально эффективная оценка дисперсии нормально распределенной случайной величины рассчитывается с помощью выражения

$$S^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n-1}. \quad (I.28)$$

Однако для логнормально распределенной случайной величины эта оценка оказывается малоэффективной даже по выборкам

большого объема. В этом случае необходимо пользоваться максимально правдоподобной оценкой дисперсии b^2 , которая рассчитывается следующим образом:

$$b^2 = e^{2 \overline{\ln x}} \left\{ \Psi_n (2S_{\ln x}^2) - \Psi_n \left(\frac{n-2}{n-1} S_{\ln x}^2 \right) \right\}. \quad (I.29)$$

Точечная оценка не содержит информации о точности полученного результата. Чем меньше выборка и чем выше изменчивость признака, тем большей может оказаться ошибка определения точечной оценки средней. Поэтому в условиях малых выборок желательно знать тот интервал значений признака, в который с заданной вероятностью попадет истинное среднее значение изучаемого признака. Для получения интервальной оценки по ограниченному количеству наблюдений необходимо располагать данными о функции статистического распределения выборочных точечных оценок.

Установлено, что оценки средних, полученные по выборкам объемом >60 , как средние арифметические из нормально распределенной генеральной совокупности с параметрами Mx и σ^2 , будут также распределены нормально с математическим ожиданием Mx и дисперсией σ_x^2 , прямо пропорциональной σ^2 и обратно пропорциональной объему выборки n .

Таким образом, оценка дисперсии средней арифметической находится из выражения

$$S_x^2 = \frac{S^2}{n}, \quad (I.30)$$

где S^2 — выборочная оценка дисперсии σ^2 .

Так как выборочные среднеарифметические оценки распределены нормально, то вероятность того, что единичная выборочная оценки \bar{x} будет отличаться от математического ожидания Mx не больше чем на величину

$$u = \frac{\bar{x} - Mx}{\sigma_x}, \quad (I.31)$$

может быть найдена по таблицам интегральной функции Лапласа. Таким образом, интервал значений, в пределах которого с заданной вероятностью находится истинное значение среднего (доверительный интервал), будет

$$\bar{x} \pm u \sigma_x. \quad (I.32)$$

По выборочным данным σ_x заменяется ее оценкой S_x .

Для выборок объемом меньше 60 значений характер распределения величины

$$\frac{\bar{x} - Mx}{\sigma_x} = t \quad (I.33)$$

зависит не только от параметров Mx и $\sigma\bar{x}$, но и от объема выборки n . Распределение величины t для разного объема выборки n называется распределением Стьюдента. По выборкам малого объема доверительный интервал среднего арифметического определяется как

$$\lambda = \bar{x} \pm tS_{\bar{x}} = \bar{x} \pm t \frac{S}{\sqrt{n}}, \quad (I.34)$$

где t — значение критерия Стьюдента, которое находится по соответствующим таблицам для заданной доверительной вероятности p и числа степеней свободы $k = n - 1$. Доверительная вероятность p характеризует вероятность того, что истинное среднее находится в пределах доверительного интервала, т. е. вероятность события

$$\bar{x} - \lambda < Mx < \bar{x} + \lambda. \quad (I.35)$$

Пример. Как известно, вулканические породы разделяются по содержанию SiO_2 на следующие типы:

Типы пород	Содержание SiO_2 , %
Базальты	47,0—52,0
Андезито-базальты	52,0—57,2
Андезиты	57,2—62,1
Андезит-дациты	62,1—63,0
Дациты	63,0—68,5
Липарито-дациты	68,5—70,5
Липариты	70,5 и более

Для определения петрографического типа неогеновых лав одного из районов СССР отобрано и проанализировано на содержание SiO_2 30 образцов (табл. 1). По результатам анализов содержание SiO_2 в образцах «х» изменяется от 56,6 до 73,2%, а среднее арифметическое $\bar{x} = 65,6$ (дациты). Для суждения о надежности этого вывода необходимо вычислить интервальную оценку среднего содержания SiO_2 . Приведенные в табл. 1 расчеты оценок асимметрии и эксцесса показывают, что гипотеза о соответствии эмпирических данных нормальному закону распределения не отвергается. Это позволяет для расчета интервальной оценки содержания SiO_2 в лавах воспользоваться формулой (I.34). При доверительной вероятности 0,95 и числе наблюдений 30 значение критерия Стьюдента t будет равно 2,045 (прил. 1). Интервальная оценка среднего содержания SiO_2 в лавах будет

$$\bar{x} \pm t \frac{S}{\sqrt{n}} = \left(65,5 \pm \frac{2,045 \cdot 4,4}{\sqrt{30}} \right) \% = (65,6 \pm 1,6) \%$$

Поскольку интервальная оценка не выходит за пределы колебаний содержания SiO_2 для дацитов, исследуемые лавы можно уверенно отнести к данному типу пород.

Для нахождения доверительного интервала параметра p биномиального распределения обычно вводится вспомогательная величина

$$\varphi = 2 \arcsin \sqrt{p}, \quad (I.36)$$

Таблица 1

Проверка гипотезы о статистическом законе
распределения содержания SiO_2 в неогеновых лавах

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	59,5	-6,1	37,2	-227,0	1385,0
2	66,8	1,2	1,4	1,7	2,1
3	60,5	-5,1	26,0	-132,6	676,6
4	63,7	-1,9	3,6	-6,9	13,0
5	72,5	6,9	47,6	328,5	2267,0
6	69,2	3,6	13,0	46,7	168,0
7	61,2	-4,4	19,4	-85,2	374,8
8	66,3	0,7	0,5	0,3	0,2
9	79,2	7,6	57,8	439,0	3336,0
10	64,6	-1,0	1,0	-1,0	1,0
11	62,9	-2,7	7,3	-19,7	53,1
12	62,4	-3,2	10,2	-32,8	104,9
13	71,6	6,0	36,0	216,0	1296,0
14	65,8	0,2	0	0	0
15	63,1	-2,5	6,2	-15,6	39,1
16	61,2	-4,4	19,4	-85,2	374,8
17	69,3	3,7	13,7	50,6	187,4
18	64,6	-1,0	1,0	-1,0	1,0
19	67,8	2,2	4,8	10,6	23,4
20	56,6	-9,0	81,0	-729,0	6561,0
21	71,4	5,8	33,6	195,1	1131,6
22	67,7	2,1	4,4	9,3	19,4
23	63,6	-2,0	4,0	-8,0	16,0
24	61,1	-4,5	20,2	-91,1	410,1
25	63,8	-1,8	3,2	-5,8	10,5
26	67,5	1,9	3,6	6,9	13,0
27	65,3	-0,3	0,1	0	0
28	69,9	4,3	18,5	79,5	341,9
29	73,2	7,6	57,8	439,0	3336,0
30	60,7	-4,9	24,0	-117,6	576,5

$$n = 30; \quad \Sigma x_i = 1967,0; \quad \bar{x} = \frac{\Sigma x_i}{n} = 65,6; \quad \Sigma (x_i - \bar{x})^2 = 556,5;$$

$$S^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} = \frac{556,5}{29} = 19,2; \quad S = \sqrt{S^2} = 4,4; \quad S^3 = 85,2;$$

$$S^4 = 374,8; \quad \Sigma (x_i - \bar{x})^3 = 264,7; \quad A = \frac{\Sigma (x_i - \bar{x})^3}{nS^3} = \frac{264,7}{30 \cdot 85,2} = 0,104;$$

$$\Sigma (x_i - \bar{x})^4 = 22719,3; \quad E = \frac{\Sigma (x_i - \bar{x})^4}{nS^4} - 3 =$$

$$= -\frac{22719,3}{30 \cdot 374,8} - 3 = -0,98;$$

$$\sigma_A = \sqrt{\frac{6}{n}} = \sqrt{\frac{6}{30}} = 0,45; \quad \sigma_E = \sqrt{\frac{24}{n}} = \sqrt{\frac{24}{30}} = 0,90;$$

$$\left| \frac{A}{\sigma_A} \right| = 0,23; \quad \left| \frac{E}{\sigma_E} \right| = 1,08; \quad \left| \frac{A}{\sigma_A} \right| < 3; \quad \left| \frac{E}{\sigma_E} \right| < 3.$$

распределение которой близко к нормальному. Стандарт этой величины, выраженной в радианах, приближенно оценивается по формуле

$$S_{\varphi} \approx \frac{1}{\sqrt{n}}. \quad (1.37)$$

Доверительный интервал для φ будет

$$\lambda_{\varphi} = \varphi \pm \frac{u}{\sqrt{n}}, \quad (1.38)$$

где u — вероятностный критерий, который определяется для заданной доверительной вероятности по таблицам функции Лапласа. Переход от φ к p осуществляется с помощью специальных таблиц [3, 52].

§ 6. 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 используются 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 , ока-

жется меньше или больше (в зависимости от формулировки гипотезы) теоретического значения K для аналогичных условий и заданной вероятности p , которое берется по известному распределению. Вероятность p при этом соответствует уровню вероятности практически невозможного события и называется *уровнем значимости*.

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

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

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, согласно мнению многих петрологов [4], средний химический состав лав вулканов и интрузивных пород отражает в общих чертах особенности состава породивших их глубинных магматических очагов. Путем сравнения различных эффузивных и интрузивных пород по среднему содержанию в них химических элементов это позволяет судить о комагматичности (т. е. генетическом родстве) эффузивных 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 пороодообразующих минералов. Различия в наборе и процентных соотношениях этих минералов отражают различия в химическом составе исходных пород, претерпевших метаморфизм [44]. Статистические методы проверки гипотезы о равенстве средних содержаний пороодообразующих 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 [36]. Для выделения нового вида необходимо доказать, что данная группа ископае-

мых организмов существенно отличается по среднему значению какого-либо морфологического признака, например, по степени сферичности (брахиоподы *Pentremites calyx*) или углу между линиями замкового шва и краем вентрального синуса (*Spirifer pennatus*).

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 они характеризуются большой изменчивостью, а объем выборки 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, учитывающие свойства выборочных оценок.

Параметрические критерии согласия выводятся из свойств известных статистических законов распределения. Для их использования необходимо предварительно проверить гипотезу о соответствии выборочных данных теоретическому закону распределения. Непараметрические критерии могут использоваться даже в том случае, если закон распределения сравниваемых случайных величин неизвестен.

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

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (I.39)$$

подчиняется закону распределения Стьюдента с $n_1 + n_2 - 2$ степенями свободы. В формуле (I.39) \bar{x}, \bar{y} — выборочные оценки среднего, а S_1^2, S_2^2 — выборочные оценки дисперсии. Проверка гипотезы о равенстве двух выборочных средних заключается в подстановке в формулу (I.39) оценок \bar{x} и S_1^2 по первой и y и S_2^2 по второй выборке и сравнении полученного значения кри-

терия 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 Д. А. Родионова [42]. Д. А. Родионовым было установлено, что величина

$$t = \frac{|\overline{\lg x} - \overline{\lg y}| + 1,153 (S_{\lg x}^2 - S_{\lg y}^2)}{\sqrt{\frac{S_{\lg x}^2}{n_1} + \frac{S_{\lg y}^2}{n_2} + 2,65 \left(\frac{S_{\lg x}^4}{n_1 - 1} + \frac{S_{\lg y}^4}{n_2 - 1} \right)}} \quad (1.40)$$

распределена асимптотически нормально с математическим ожиданием 0 и дисперсией 1. Поэтому при проверке гипотезы о равенстве средних с помощью этого критерия теоретическое значение величины t находится не по таблице распределения Стьюдента, а по таблице значений интегральной функции Лапласа.

Пример. В районе широко развиты интрузии гранитов средне- и верхнепалеозойского возраста. Со среднепалеозойскими гранитами связаны месторождения редкометалльных пегматитов. Среднепалеозойские граниты отличаются от нерудоносных верхнепалеозойских гранитов повышенным содержанием Na_2O и TiO_2 и пониженным содержанием K_2O (табл. 2). Распре-

Таблица 2

Исходные данные для оценки рудоносности гранитов неизвестного возраста

Возраст гранитов	Количество проб	Na_2O		K_2O		TiO_2	
		\bar{x}	S^2	\bar{x}	S^2	$\overline{\lg x}$	$S_{\lg x}^2$
Средний палеозой	100	3,90	1,21	4,51	1,42	-0,886	0,268
Верхний палеозой	100	3,46	1,52	5,02	1,65	-1,426	0,321
Возраст неизвестен	30	3,38	1,38	4,83	1,88	-1,352	0,225

деление содержаний Na_2O и K_2O в гранитах обоих комплексов соответствует нормальному закону, а содержаний 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.

В нижней строке табл. 2 приведены числовые характеристики содержания окислов по одной из интрузий неизвестного возраста. Ее сопоставление с интрузиями определенного возраста по содержанию Na_2O и K_2O следует провести с помощью критерия Стьюдента, а по содержанию TiO_2 с помощью критерия Родионова.

Сравнение интрузии неизвестного возраста со среднепалеозойскими гранитами

$$\text{по Na}_2\text{O} \quad t = \frac{(3,90 - 3,38)}{\frac{1,21}{100} + \frac{1,38}{30}} = 2,15;$$

$$\text{по K}_2\text{O} \quad t = \frac{(4,51 - 4,83)}{\frac{1,42}{100} + \frac{1,88}{30}} = 1,15;$$

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

Сравнение интрузива неизвестного возраста с верхнепалеозойскими гранитами

$$\text{по Na}_2\text{O} \quad t = \frac{(3,46 - 3,38)}{\frac{1,52}{100} + \frac{1,38}{30}} = 0,32;$$

$$\text{по K}_2\text{O} \quad t = \frac{(5,02 - 4,83)}{\frac{1,65}{100} + \frac{1,88}{30}} = 0,68;$$

$$\text{по TiO}_2 \quad t = \frac{(-1,426 + 1,352) + 1,153(0,321 - 0,225)}{\frac{0,321}{100} + \frac{0,225}{30} + 2,65\left(\frac{0,1030}{99} + \frac{0,0506}{29}\right)} = 1,37.$$

Для доверительной вероятности 0,95 критическое значение критерия Стьюдента при числе степеней свободы $n_1 + n_2 - 2 = 128$ равно 1,98 (прил. 1), а критическое значение критерия Родионова для той же доверительной вероятности равно 1,96.

Таким образом, с вероятностью 95% можно считать, что оцениваемые граниты относятся к верхнепалеозойскому возрасту и, следовательно, не перспективны на обнаружение редкометальных пегматитов, так как по содержанию Na_2O и TiO_2 они значимо отличаются от рудоносных среднепалеозойских гранитов, в то время как отличия по среднему содержанию всех трех окислов от верхнепалеозойских гранитов можно признать несущественными.

Непараметрические критерии: X критерий Ван дер Вардена, Вилкоксона (Манна-Уитни) и других — используются обычно при малом объеме выборок или в тех случаях, когда средние значения рассчитаны по полуколичественным данным, например, по результатам полуколичественного спектрального анализа.

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

$$X = \sum_1^h \Psi\left(\frac{i}{n+1}\right), \quad (I.41)$$

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

При $n > 20$ величина X распределена асимптотически нормально с математическим ожиданием 0 и дисперсией σ_x^2 . Процедура проверки гипотезы сводится к расчету всех значений аргумента $\frac{i}{n+1}$, нахождению по таблицам обратной функции нормального распределения значений функции ψ для этих аргументов, суммированию значений функции ψ и сравнению полученного значения критерия X с табличным для заданного уровня значимости, общего числа наблюдений n и разницы между объемами выборок А и Б [3]. Если расчетное значение X больше табличного, гипотеза о равенстве выборочных средних отвергается.

Другие непараметрические критерии строятся аналогичным образом.

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 количество скважин и шурфов на опытном участке невелико — 13 и 10 штук, для сопоставления вычисленных по ним содержания золота можно использовать непараметрический критерий Ван дер Вардена. Результаты опробования шурфов и скважин приведены в табл. 3. По этим данным составлен общий вариационный ряд (табл. 4).

Таблица 3

**Результаты опробования разведочных выработок
на россыпном месторождении**

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

Ранжированные результаты опробования россыпного месторождения]

i	Содержание, мг/м ³	Выборка	$\frac{i}{n+1}$	$\psi\left(\frac{i}{n+1}\right)$
1	192	А		
2	198	А		
3	221	Б	0,125	-1,15
4	225	А		
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
$n = 23$			$X = \sum \psi\left(\frac{i}{n+1}\right) = 4,19$	

Допустимое значение X критерия Ван дер Вардена для уровня значимости 0,025, числа наблюдений 23 и разности между объемами сравниваемых выборок — 3 равно 4,15 [3]. Таким образом, гипотеза о равенстве средних значений содержания золота по скважинам и шурфам отвергается. До выяснения причин систематического занижения содержания золота по скважинам использовать их для разведки данного месторождения не рекомендуется.

ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ ДИСПЕРСИИ

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.

Различия в дисперсиях свойств аналогичных по составу геологических объектов может указывать и на различия 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 более резко проявились процессы дифференциации [4].

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 обычно используется F -критерий Фишера. Р. Фишером было установлено, что в случае равенства дисперсий двух нормально распределенных случайных величин, величина

$$F = \frac{S_1^2}{S_2^2} \quad \text{при} \quad S_1^2 > S_2^2 \quad (I.42)$$

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 и степенях свободы $k_1=n-1$ и $k_2=n_2-1$ (см. прил. 2 и 3). Если вычисленное значение критерия Фишера превышает табличное, то гипотеза о равенстве двух дисперсий отвергается.

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.

ПРОВЕРКА ГИПОТЕЗ ОБ ОДНОРОДНОСТИ ИЗУЧАЕМОГО ОБЪЕКТА

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.

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

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

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

Для выделения аномальных значений совокупность результатов наблюдений рассматривается как выборка из двух различных генеральных совокупностей — «фоновой» и «аномальной». При этом аномальные 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 Н. В. Смирнова, Т. С. Фергюссона и других.

Н. В. Смирновым было установлено, что если максимальный по значению член выборочной совокупности не является аномальным, то величина

$$t = \frac{x_{\max} - \bar{x}}{S_{\text{см}}^2} \quad (\text{I.43})$$

имеет распределение, называемое его именем. В данной формуле x_{\max} — максимальный член выборки, \bar{x} — среднее арифметическое, $S_{\text{см}}^2$ — смещенная оценка дисперсии, которая рассчитывается через несмещенную оценку дисперсии S^2 по формуле

$$S_{\text{см}}^2 = S^2 \left(\frac{n-1}{n} \right). \quad (\text{I.44})$$

Если рассчитанное значение критерия больше допустимого, определенного по таблицам распределения Н. В. Смирнова [3] для заданной доверительной вероятности и n степеней свободы, то максимальное значение выборки следует считать аномальным.

Критерий Т. С. Фергюссона 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 σ_A^2 .

Таблицы распределения величины A приведены в ряде работ [3, 52]. 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 $\bar{X} \pm 3\sigma$ 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.

Разделение неоднородных выборочных совокупностей позволяет выбрать наиболее рациональный комплекс геохимических и геофизических методов при геологическом картировании, а также выделить наиболее контрастные по своим свойствам литологические разности (маркирующие горизонты). При решении этой задачи каждая разновидность пород рассматривается как самостоятельная генеральная совокупность, а результаты геохимических и геофизических исследований — как смешанная выборка из нескольких генеральных совокупностей. Решение этой задачи статистическими методами возможно лишь в том случае, если каждая разновидность пород представлена в выборке достаточно большим количеством замеров.

Простейшие способы разделения неоднородных выборочных совокупностей основаны на анализе графиков эмпирических кривых распределения. На неоднородность выборки может указывать наличие на графике нескольких максимумов (рис. 6). Неоднородные выборочные совокупности можно разделить с помощью специальных палеток [44].

Палетки (рис. 7) представляют собой набор кривых функции плотности распределения нормированной и центрированной случайной величины, умноженной на объем выборки. Для поль-

зования палетками необходимо, чтобы вертикальный масштаб исследуемой эмпирической кривой распределения соответствовал масштабу палетки. В первую 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

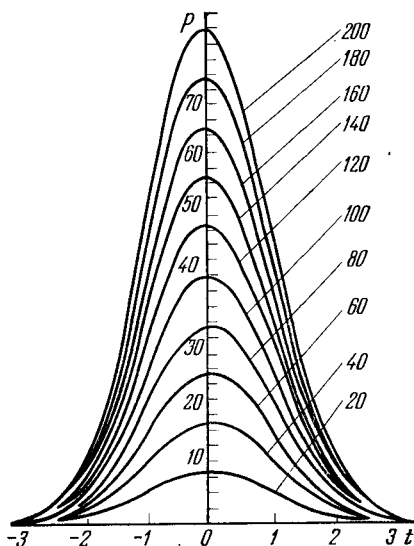
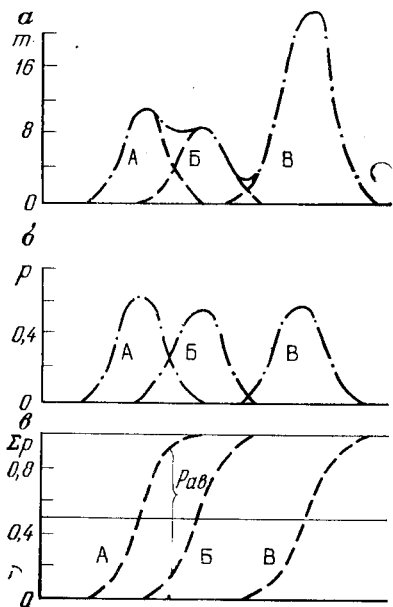


Рис. 6. Классификация амфиболитов по содержанию роговой обманки: а — экспериментальная кривая распределения содержания роговой обманки; б — плотность вероятностей; в — интегральные функции распределения

Рис. 7. Палетки для подбора эталонных кривых плотности вероятностей

стандарту. По полученным числовым характеристикам однородных совокупностей можно построить графики интегральных функций распределений (см. рис. 6, в), с помощью которых удобно оценивать надежность выделения одних однородных объектов на фоне других. Надежность (вероятность) разделения будет определяться по расстоянию $p_{ав}$ между интегральными функциями распределения вдоль линии, проходящей через точку пересечения X_0 кривых плотности вероятностей (см. рис. 6). При расчленении комплексов горных пород с помощью геофизических методов этот способ позволяет решить вопрос о том, какие породы в районе выделяются наиболее надежно,

т. е. дает возможность выбрать маркирующие горизонты, а с другой стороны, позволяет выбрать наиболее эффективный комплекс геофизических методов для картирования каждого типа пород.

Оценка степени влияния различных факторов на характер изменчивости (неоднородность) свойств геологических объектов основана на так называемом дисперсионном анализе и заключается в разделении совокупности выборочных значений какого-либо свойства на группы по определенному принципу, например по положению точки замера или составу вмещающих пород, и сравнении дисперсий свойства внутри групп с общей дисперсией. С помощью дисперсионного анализа оценивается роль различных рудоконтролирующих факторов, поисковых критериев и признаков, выявляются элементы зональности геологических объектов, выясняются причины случайных и систематических ошибок опробования анализов и т. п.

По количеству оцениваемых факторов дисперсионный анализ подразделяется на однофакторный, двухфакторный и многофакторный.

Однофакторный дисперсионный анализ заключается в разделении совокупности из n замеров изучаемого свойства объекта на p групп по какому-либо фактору. После разделения производится расчет оценок дисперсии между группами

$$S_1^2 = \frac{1}{p-1} \sum_{i=1}^p n_i (\bar{x}_i - \bar{x})^2 \quad (I.45)$$

и внутри групп

$$S_2^2 = \frac{1}{n-p} \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2. \quad (I.46)$$

В приведенных формулах n_i и \bar{x}_i — количество замеров и среднее значение свойства по каждой группе, а \bar{x} — общее среднее по всей совокупности. Если фактор, по которому было проведено группирование, не влияет на изменчивость изучаемого свойства, то отношение дисперсий S_1^2 и S_2^2 будет распределено по закону Фишера с $p-1$ и $n-p$ степенями свободы. Гипотеза о влиянии данного фактора на изменчивость свойства отвергается, если

$$\frac{S_1^2}{S_2^2} \leq F, \quad (I.47)$$

где F — табличное значение критерия Фишера для принятой доверительной вероятности и числа степеней свободы $p-1$ и $n-p$.

При двухфакторном дисперсионном анализе одновременно рассматривается влияние двух факторов, например, выясняется, влияют ли на содержание полезного компонента в рудном теле состав вмещающих пород и гипсометрическое положение места отбора пробы. По результатам опробования составляются таблицы (табл. 5), где по строкам сгруппированы данные опробования

Таблица 5

Результаты опробования рудных жил полиметаллического месторождения на разных горизонтах

Содержание Рb во вмещающих породах разного состава, %						
Отметка горизонта горных выработок	Граниты	Кварциты	Глинистые сланцы	Известняки	Доломиты	Среднее содержание по горизонтам
+320	2,95	2,66	2,50	3,15	2,25	2,70
+280	2,25	2,71	2,38	2,56	1,81	2,34
+240	2,38	1,78	1,98	1,21	2,19	1,91
+200	2,18	1,65	2,46	2,10	1,32	1,94
По всем горизонтам	2,44	2,20	2,33	2,26	1,89	2,22

вания по одному гипсометрическому уровню, а в столбцах приведены средние содержания полезного компонента по каждой разновидности вмещающих пород. Компоненты дисперсии рассчитываются по следующим формулам:

$$S_1^2 = \frac{q}{p-1} \sum_{i=1}^p (\bar{x}_i - \bar{x})^2;$$

$$S_2^2 = \frac{p}{q-1} \sum_{j=1}^q (\bar{x}_j - \bar{x})^2; \tag{I.48}$$

$$S_3^2 = \frac{1}{(p-1)(q-1)} \sum_{i=1}^p \sum_{j=1}^q (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2,$$

где \bar{x}_i — средние значения по строкам;

\bar{x}_j — средние значения по столбцам;

q и p — количество столбцов и строк.

Проверка гипотезы о влиянии данных факторов на изменчивость содержания производится по критериям

$$F_A = \frac{S_1^2}{S_3^2} \quad \text{и} \quad F_B = \frac{S_2^2}{S_3^2}. \tag{I.49}$$

Если значения критериев F_A и F_B не превышают табличные для принятой доверительной вероятности и числа степеней сво-

боды $p-1$, $q-1$ и $(p-1)$, $(q-1)$, то гипотеза о влиянии анализируемых факторов на изменчивость данного свойства отвергается.

Пример. Рудные жилы на полиметаллическом месторождении залегают в породах разного состава и вскрыты горными выработками на четырех горизонтах. По результатам опробования этих выработок (см. табл. 5) необходимо выяснить, влияют ли на содержание свинца в рудах состав вмещающих пород и глубина залегания жил. Эта задача может быть решена по схеме двухфакторного дисперсионного анализа [51].

$$S_1^2 = \frac{q}{p-1} \sum_{i=1}^p (\bar{x}_i - \bar{x})^2 = \frac{5 \cdot 0,418}{3} = 0,695;$$

$$S_2^2 = \frac{p}{q-1} \sum_{j=1}^q (\bar{x}_j - \bar{x})^2 = \frac{3 \cdot 0,1714}{4} = 0,129;$$

$$S_3^2 = \frac{1 \cdot p}{(p-1)(q-1)} \sum_{i=1}^p \sum_{j=1}^q (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = \frac{1,9116}{3 \cdot 4} = 0,159;$$

$$F_A = \frac{S_1^2}{S_3^2} = \frac{0,695}{0,159} = 4,371; \quad F_B = \frac{S_2^2}{S_3^2} = \frac{0,129}{0,159} = 0,811.$$

Критические значения F -критерия Фишера для доверительной вероятности 0,95 (см. прил. 2) равны:

$$F_{кр} = 8,74 \text{ при числе степеней свободы } p-1=3 \text{ и } (p-1)(q-1)=12;$$

$$F_{кр} = 5,91 \text{ при числе степеней свободы } q-1=4 \text{ и } (p-1)(q-1)=12.$$

Таким образом, гипотеза о влиянии состава вмещающих пород и глубины залегания жил на содержание свинца в рудах отвергается.

Использованный в данной главе математический аппарат подробно описан в специальных учебных пособиях по теории вероятностей и математической статистике, например, в книге Б. Л. Ван дер Вардена [8]. Применению одномерных статистических моделей при геохимических исследованиях посвящены работы В. Н. Бондаренко [4] и Д. А. Родионова [43]. Примеры использования этой модели в различных областях геологии приведены в работе Р. Л. Миллера и Дж. С. Кана [36].

ГЛАВА II

ДВУМЕРНЫЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

§ 1. СУЩНОСТЬ И ОСНОВНЫЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДВУМЕРНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ

При решении разнообразных геологических задач часто необходимо совместное рассмотрение не одной, а нескольких случайных величин. 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. Так, например, в редкометалльных пегматитах тантал и ниобий входят в состав только одного минерала — танталит-колумбита $(\text{Fe}, \text{Mn}) (\text{Nb}, \text{Ta})_2\text{O}_6$. Между содержаниями этих элементов в рудах всегда наблюдается прямая корреляционная зависимость: чем больше тантала, тем больше ниобия, а между их содержаниями в мономинеральных фракциях — обратная корреляционная зависимость. 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. Однако гораздо чаще связь между различными свойствами геологических объектов 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 опоискованных 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 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 определенное значение величины Y , а некоторый набор ее значений $y_1, y_2, y_3, \dots, y_n$, каждое из которых характеризуется определенной вероятностью p_1, p_2, \dots, p_n . Функция распределения величины Y , соответствующая значению $X=x$, характеризуется математическим ожиданием \bar{y}_x и дисперсией σ_y^2 .

Распределения величины Y , соответствующие выбранным значениям величины X , называются условными распределениями, а дисперсии σ_y^2 — условными дисперсиями. Геометрическое место точек, соответствующих центрам условных распределений \bar{y}_x , 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 σ_x^2 .

Системе из двух случайных величин всегда будут соответствовать две линии регрессии: $y_x=f(x)$ — регрессия y по x и $x_y=f(y)$ — регрессия x по y (рис. 8). 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:

$$y = a_1 + b_1x \text{ (регрессия } y \text{ на } x);$$

$$x = a_2 + b_2y \text{ (регрессия } x \text{ на } y).$$

(II.1)

Уравнения нелинейной регрессии зависят от вида кривой. Например, для параболической регрессии:

$$y = a_1 + b_1x + c_1x^2; \tag{II.2}$$

$$x = a_2 + b_2y + c_2y^2.$$

Регрессия может быть однозначно описана, если известен вид уравнения и значения коэффициентов a , b , c и т. д. В системе двух уравнений линейной регрессии коэффициенты a_1 и a_2 определяют положения начальных точек линий регрессии.

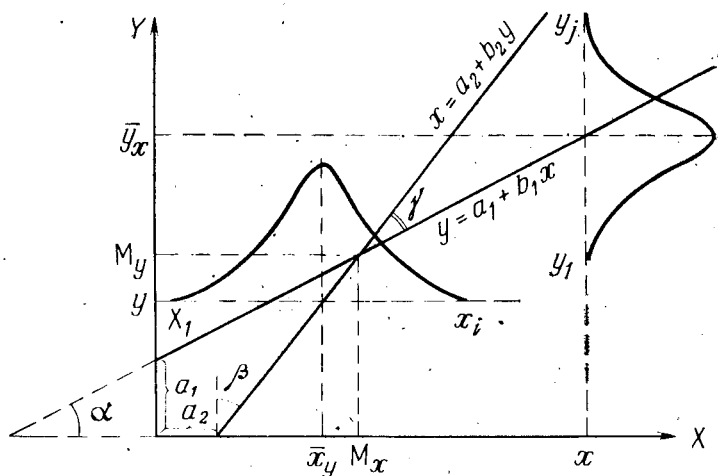


Рис. 8. Графическое изображение регрессии y по x и x по y :
 \bar{y}_x — центр условного распределения Y для $X=x$; \bar{x}_y — центр условного распределения X для $Y=y$; M_x и M_y — математические ожидания величин X и Y ; $b_1 = \text{tg } \alpha$; $b_2 = \text{tg } \beta$

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 углов наклона прямой регрессии $y = a_1 + b_1x$ к оси абсцисс (угол α) и прямой регрессии $x = a_2 + b_2y$ к оси ординат (угол β).

В общем случае прямые регрессии пересекаются в точке, координаты которой равны математическим ожиданиям величин X и Y , а угол γ между ними изменяется от 0 до 90°. Чем меньше величина угла γ , 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 коэф-

коэффициент будет отличаться от нуля. Следовательно, степень связи между величинами будет зависеть 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 не существует.

Стохастическая связь между случайными величинами вызывается тем, что среди действующих на них факторов имеются как общие факторы, так и факторы, влияющие только на величину X и только на величину Y . Методами корреляционного анализа 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 величин 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

$$\text{cov}(x; y) = M[(x - Mx)(y - My)] = M[(\bar{x}_y - M_{\bar{x}})(\bar{y}_x - M_{\bar{y}})]. \quad (\text{II.3})$$

Коэффициент корреляции представляет собой ковариацию, нормированную по стандартам

$$\rho = \frac{\text{cov}(x; y)}{\sigma_x \sigma_y}. \quad (\text{II.4})$$

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

$$y = a_1 + b_1x$$

и

$$(II.5)$$

$$x = a_2 + b_2y,$$

то коэффициент корреляции

$$\rho = \sqrt{b_1 b_2}. \quad (II.6)$$

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:

$$\eta_{y/x} = \frac{\sigma_{y_x}}{\sigma_y}; \quad \eta_{x/y} = \frac{\sigma_{x_y}}{\sigma_x}. \quad (II.7)$$

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. Характерные случаи значений коэффициентов корреляции ρ и корреляционных отношений η показаны на рис. 9.

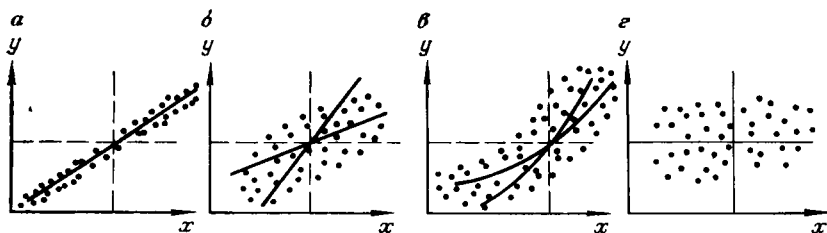


Рис. 9. Типичные случаи значений коэффициентов корреляции и корреляционных отношений:

$$a - \eta = \rho = 1; \quad б - \eta_{y/x} = \eta_{x/y} = |\rho|; \quad в - \eta_{y/x} \neq \eta_{x/y}; \quad г - \eta = \rho = 0$$

§ 2. ПРОВЕРКА ГИПОТЕЗ О НАЛИЧИИ КОРРЕЛЯЦИОННОЙ СВЯЗИ

Выявление корреляционных связей между различными свойствами геологических объектов способствует решению многих геологических задач. Так, например, наличие корреляционных связей между петрогенными и редкими элементами способствует оценке роли процессов дифференциации магмы и ассимиляции ею вмещающих пород; между концентрацией рудных элементов в породах и рудах — выяснению источников рудного вещества; между физическими свойствами и минеральным

составом пород — дешифрированию геофизических аномалий при геологическом картировании и т. д.

При отсутствии корреляционной связи коэффициент корреляции, условные коэффициенты линейной регрессии и корреляционные отношения равны нулю. Поэтому проверка гипотезы о наличии корреляционной связи заключается в расчете выборочных оценок этих характеристик и оценки значимости их отличия от нуля.

Выборочная оценка коэффициента корреляции может быть рассчитана по формуле

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1) S_x S_y}, \quad (\text{II.8.})$$

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

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] \left[\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right]}}. \quad (\text{II.9.})$$

Когда математическое ожидание выборочного коэффициента корреляции равно нулю, величина

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \quad (\text{II.10.})$$

имеет распределение Стьюдента с $n-2$ степенями свободы. Если рассчитанное по этой формуле значение величины t превышает табличное значение критерия Стьюдента (см. прил. 1) для принятой доверительной вероятности и числа степеней свободы $n-2$, гипотеза об отсутствии корреляционной связи отвергается, т. е. считается, что связь существует.

Приближенная оценка коэффициента корреляции может быть получена графическим методом. Для этого каждая пара значений исследуемой двумерной величины X , Y изображается в виде точки с координатами x и y (рис. 10). При наличии корреляционной связи поле точек вытянуто вдоль линии, ориентированной под углом к осям координат. Поле точек разделяется на четыре квадранта линиями, соответствующими медианам величин X и Y . При наличии корреляционной связи количество

точек в разных квадрантах будет различным. Для оценки коэффициента корреляции используется формула

$$r = \frac{n_1 - n_2}{n_1 + n_2}, \quad (\text{II.11})$$

где n_1 — число наблюдений в квадрантах I и III, а n_2 — в квадрантах II и IV.

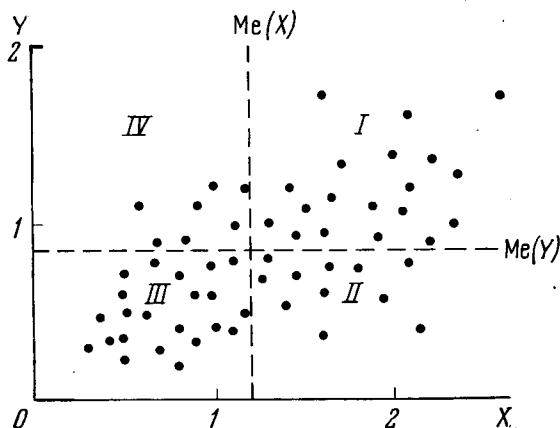


Рис. 10. Корреляционное поле точек содержания серы в нефти (Y) и во вмещающих породах (X) в %

Пример. На рис. 10 изображено корреляционное поле точек содержания серы в пробах нефти (Y) и вмещающих породах-коллекторах (X) по одному из нефтяных месторождений.

Оценка коэффициента корреляции между содержаниями серы в нефти и вмещающих породах по формуле (II.11) будет

$$r = \frac{41 - 19}{60} = 0,37.$$

Значение критерия t , рассчитанное по формуле (II.10), равно

$$t = \frac{0,37}{\sqrt{1 - 0,37^2}} \cdot \sqrt{60 - 2} = 3,03.$$

Табличное значение критерия Стьюдента для доверительной вероятности 0,95 и числа степеней свободы 58 равно 2,0 (см. прил. 1), т.е. меньше расчетного. Это позволяет принять гипотезу о наличии корреляционной связи между содержанием серы в нефти и вмещающих породах и предположить, что повышенное содержание серы в нефти связано с частичным переходом серы в нефть из вмещающих пород-коллекторов.

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 логнормальных или каких-либо других преобразований.

Пример. На редкометальном месторождении проверялась гипотеза о наличии стохастической связи между значениями мощности залежи и содержанием одного из металлов. При обработке результатов были получены данные, приведенные в табл. 6.

Таблица 6

Статистические характеристики мощности рудной залежи (в м)
среднего содержания металла (в %) по редкометальному месторождению

Признак	Среднее значение	Дисперсия		Коэффициент асимметрии		Коэффициент эксцесса	
		признака	логарифма признака	признака	логарифма признака	признака	логарифма признака
Мощность	30,6	851,0	2,05	3,18	-2,33	1,59	0,25
Содержание	0,27	0,049	0,66	4,54	-0,27	5,67	-0,86

Как видно из таблицы, ряды значений обоих признаков значительно лучше согласуются с логнормальной, чем с нормальной моделью.

Расчет коэффициента корреляции непосредственно по значениям признака приводит к величине $r = +0,22$, которая при 37 парах значений в выборке оценивается как незначимо отличная от 0.

$$t = \frac{0,22 \sqrt{37-2}}{\sqrt{1-0,22^2}} = 1,28.$$

При $p=0,95$ величина критерия t для $n=37$ не должна превышать 2,01, а $1,28 < 2,01$. Для логарифмов значений признака расчет коэффициентов корреляции приводит к величине $r_{ln} = +0,85$, которая при том же числе наблюдений оценивается, безусловно, значимой, так как $10,0 > 2,01$.

$$t = \frac{0,85 \sqrt{37-2}}{\sqrt{1-0,85^2}} = 10,0.$$

Таким образом, при проверке гипотез о стохастической связи случайных величин по коэффициенту корреляции необходимо учитывать функцию их эмпирических распределений.

Если не удастся 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

$$r = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (\text{П.12})$$

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 $\frac{1}{n-1}$,

Для проверки значимости рангового коэффициента корреляции можно использовать величину

$$r_k = \frac{\Psi(p)}{\sqrt{\frac{1}{n-1}}}, \quad (\text{П.13})$$

где $\Psi(p)$ — значение обратной функции нормального распределения при доверительной вероятности p .

Если расчетное значение коэффициента Спирмена (r) больше критического (r_k), 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 (ρ_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 ρ_k и m_r можно воспользоваться ранговым коэффициентом корреляции (табл. 7).

Значение рангового коэффициента корреляции будет равно

$$r = 1 - \frac{6 \cdot 104}{12(144 - 12)} = 1 - 0,39 = 0,61.$$

Значение обратной функции нормального распределения для доверительной вероятности 0,95 равно 1,64, а критическое значение рангового коэффициента корреляции при этой доверительной вероятности и объеме выборки в 12 значений будет

$$r_k = \frac{1,64}{\sqrt{12-1}} = 0,49.$$

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 7

Результаты вычисления рангового коэффициента
корреляции для значений ρ_k и m_r

№ скважины	m_r		ρ_k		d_i	d_i^2
	значение в %	ранг	значение в омм	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
$\Sigma d_i^2 = 104$						

Коэффициент корреляции может служить надежной характеристикой стохастической связи лишь в условиях линейности обоих уравнений регрессии, так как он характеризует стохастическую связь величин только «в среднем». 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. 11). Это объясняется тем, что в бедных вкрапленных рудах галенит первой генерации тесно ассоциирует с золотосным 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. 12) [9].

О характере связи судят по виду эмпирических линий регрессии. Если они заметно отличаются от прямой линии, гипотезу о наличии корреляционной связи следует проверить с помощью корреляционного отношения.

Для вычисления оценок корреляционных отношений выборочные данные группируются в классы по значениям одного из исследуемых свойств. По каждому классу рассчитываются

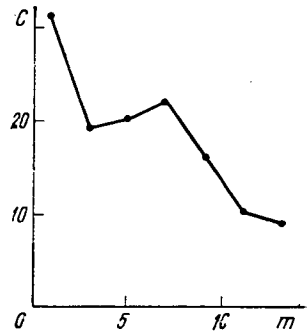
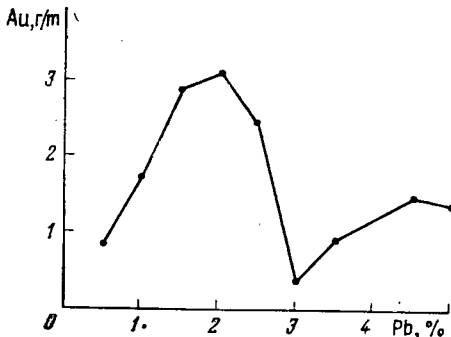


Рис. 11. Зависимость содержания золота от содержания свинца в рудах свинцового месторождения

Рис. 12. Распределение среднего содержания полезного компонента C (в условных единицах) по классам мощностей рудного тела m (в м)

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

$$S_{\bar{y}_i} = \sqrt{\frac{1}{N} \sum_{i=1}^m n_i (\bar{y}_i - \bar{y})^2};$$

$$S_{\bar{x}_i} = \sqrt{\frac{1}{N} \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2},$$
(II.14)

где n_i — число наблюдений в i -ой группе;

m — число групп;

N — общее число наблюдений.

Выборочные значения корреляционных отношений определяются из выражений

$$\eta_{y/x} = \frac{S_{\bar{y}_i}}{S_y};$$

$$\eta_{x/y} = \frac{S_{\bar{x}_i}}{S_x},$$
(II.15)

где S_x и S_y — оценки общего стандартного отклонения исследуемых случайных величин.

Статистическая значимость отличия корреляционного отношения $\eta_{y/x}$ от нуля проверяется с помощью критерия

$$\theta_y = \frac{\eta_{y/x}^2 (N - m - 2)}{(1 - \eta_{y/x}^2) (m - 2)} \sqrt{\frac{(m - 2) (N - m - 4)}{2 (N - 4)}} \quad (\text{II.16})$$

При равенстве истинного корреляционного отношения нулю величина θ_y распределена нормально с математическим ожиданием 0 и дисперсией 1, что позволяет определять критические значения θ_y для заданных доверительных вероятностей по таблицам нормального распределения. Если расчетное значение θ_y превышает критическое, гипотеза об отсутствии корреляционной связи отвергается. Аналогично проверяется гипотеза о наличии корреляционной связи по $\eta_{x/y}$, но вместо $\eta_{y/x}$ вводится $\eta_{x/y}$.

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. 8). Однако, учитывая нелинейный характер графика линии регрессии (рис. 13), гипотезу о наличии корреляционной связи следует проверить повторно по корреляционному отношению (табл. 9). С этой целью по формуле

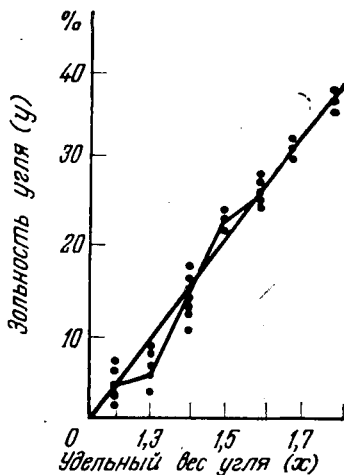


Рис. 13. Пример графического изображения результатов регрессионного анализа

(II.14) рассчитываем стандартное

$$S_{y_i} = \sqrt{\frac{1}{60} \cdot 33,479} = 0,747.$$

Затем, используя данные табл. 8, рассчитываем стандартное отклонение величины y по формуле

$$S_y = \sqrt{\frac{1}{N-1} \left[\sum_i^N y_i^2 - \frac{\left(\sum_i^N y_i \right)^2}{N} \right]}, \quad (\text{II.17})$$

$$S_y = \sqrt{\frac{1}{59} \left(250,98 - \frac{107,11^2}{60} \right)} = 1,006.$$

Расчет коэффициента корреляции между содержаниями
золота (Y в г/т) и свинца (X в %)

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^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

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
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
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
	114,55	107,11	198,05	324,26	250,98

$$r = \frac{198,05 - \frac{1}{60} \cdot 114,55 \cdot 107,11}{\sqrt{\left(324,26 - \frac{1}{60} \cdot 114,55^2\right) \left(250,98 - \frac{1}{60} \cdot 107,11^2\right)}} =$$

$$= \frac{-6,44}{79,44} = -0,081;$$

$$t = \frac{0,081 \sqrt{58}}{\sqrt{1-0,0066}} = 0,62.$$

Таблица 9

Расчет корреляционного отношения между содержаниями
золота (Y) и свинца (X)

i	Классы группировки по X	\bar{y}_i	n_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
						33,479

Оценка корреляционного отношения $\eta_{y/x}$ для содержания золота по содержанию свинца составит

$$\eta_{y/x} = \frac{S_{\bar{y}_i}}{S_y} = \frac{0,747}{1,006} = 0,743.$$

Для проверки статистической значимости отличия корреляционного отношения от нуля по формуле (II.16) рассчитываем значение величины θ_y

$$\theta_y = \frac{0,743^2 (60 - 10 - 2)}{(1 - 0,743^2) (10 - 2)} \sqrt{\frac{(10 - 2) (60 - 10 - 4)}{2 (60 - 4)}} = 13,69.$$

Поскольку табличное значение величины θ_y 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.

В 1964 г. А. Б. Вистелиусом [12] предложен существенно иной способ оценки зависимости между содержаниями элементов. Смысл его сводится к тому, что статистически оценивается количество информации, которую дает результат анализа одного элемента о другом. Получаемый коэффициент корреляции в отличие от обычного назван *информационным*. По мнению А. Б. Вистелиуса, информационный коэффициент корреляции позволяет устанавливать связи любого вида, а не только линейные связи между элементами. Поэтому оценки информационного коэффициента корреляции, как правило, выше оценок обычных коэффициентов корреляции. Информационный коэффициент корреляции равен:

$$r_I(x, y) = \sqrt{1 - e^{-2\hat{T}(x, y)}}, \quad (\text{II.18})$$

где $\hat{T}(x, y)$ — несмещенная оценка количества информации об исследуемых параметрах, которая в свою очередь определяется по формуле

$$\hat{T}(x, y) = \Sigma \Sigma \overline{\hat{f}(x, y)} \ln \frac{\overline{\hat{f}(x, y)}}{g(x) h(y)} - \frac{(k-1)(l-1)}{2N}. \quad (\text{II.19})$$

В этом выражении $\overline{\hat{f}(x, y)}$ — оценка совместной плотности вероятности для величин X и Y ; $g(x)$ и $h(y)$ — безусловные плотности вероятности величин X и Y , а k и l — соответственно число строк и столбцов в таблице совместных частот для значений величин X и Y .

*Пример**. Для выбора наиболее информативной морфологической характеристики раковин *Bifarina adetae* L. проверяется гипотеза о наличии корреляционной зависимости между длиной раковины (x) и длиной ее одной из частей (y). Результаты замеров этих величин по 50 раковинам приведены в табл. 10).

Обе исследуемые величины группируются по классам значений. Определяются частоты попадания значений двумерной величины в каждый класс (табл. 11).

* Пример заимствован из работы В. Н. Бондаренко «Статистические решения некоторых задач геологии». М., 1970.

Результаты замеров длины раковины
и длины ее однорядной части

i	Длина раковины, x_i	Длина однорядной части y_i	i	Длина раковины x_i	Длина однорядной части y_i	i	Длина раковины x_i	Длина однорядной части y_i	i	Длина раковины x_i	Длина однорядной части y_i
1	0,625	0,150	14	0,725	0,200	26	0,950	0,550	39	0,900	0,400
2	0,750	0,300	15	0,850	0,350	27	0,900	0,280	40	1,075	0,625
3	0,700	0,375	16	1,000	0,450	28	1,025	0,500	41	0,870	0,400
4	0,650	0,150	17	0,920	0,320	29	0,875	0,500	42	0,900	0,400
5	0,720	0,300	18	1,125	0,625	30	0,925	0,500	43	0,950	0,400
6	0,525	0,150	19	0,800	0,275	31	0,780	0,350	44	0,675	0,150
7	0,500	0,200	20	0,700	0,175	32	1,050	0,500	45	0,625	0,175
8	0,575	0,200	21	0,800	0,250	33	0,870	0,550	46	0,720	0,175
9	0,620	0,220	22	0,525	0,125	34	0,750	0,300	47	0,550	0,150
10	0,920	0,370	23	0,525	0,175	35	0,870	0,375	48	0,625	0,175
11	1,050	0,350	24	0,625	0,145	36	1,025	0,400	49	0,750	0,275
12	1,000	0,425	25	1,000	0,500	37	0,900	0,400	50	0,720	0,165
13	0,800	0,175				38	0,875	0,425			

Таблица 11

Частоты (числитель) и относительные частоты (знаменатель)
распределения величин X и Y

j	i	1	2	3	4	5	6	7	8	n_j	
		X	0,401	0,501	0,601	0,701	0,801	0,901	1,001		1,101
		Y	0,500	0,600	0,700	0,800	0,900	1,000	1,100		1,200
1	0,101—0,200	$\frac{1}{0,020}$	$\frac{5}{0,100}$	$\frac{7}{0,140}$	$\frac{4}{0,080}$					$\frac{17}{0,340}$	
2	0,201—0,300			$\frac{1}{0,020}$	$\frac{5}{0,100}$	$\frac{2}{0,040}$				$\frac{8}{0,160}$	
3	0,301—0,400			$\frac{1}{0,020}$	$\frac{1}{0,020}$	$\frac{6}{0,120}$	$\frac{3}{0,060}$	$\frac{2}{0,040}$		$\frac{13}{0,260}$	
4	0,401—0,500					$\frac{2}{0,040}$	$\frac{4}{0,080}$	$\frac{2}{0,040}$		$\frac{8}{0,160}$	
5	0,501—0,600					$\frac{1}{0,020}$	$\frac{1}{0,020}$			$\frac{2}{0,040}$	
6	0,601—0,700							$\frac{1}{0,020}$	$\frac{1}{0,020}$	$\frac{2}{0,040}$	
	n_{i_0}	$\frac{1}{0,020}$	$\frac{5}{0,100}$	$\frac{9}{0,180}$	$\frac{10}{0,200}$	$\frac{11}{0,220}$	$\frac{8}{0,160}$	$\frac{5}{0,100}$	$\frac{1}{0,020}$	$\frac{50}{1,000}$	

Значения величины $\frac{\overline{f(x, y)}}{g(x) h(y)}$ для каждого класса рассчитываются по формуле

$$\frac{\overline{f(x, y)}}{g(x) h(y)} = \frac{n_{ij}N}{n_i n_j} \quad (II.20)$$

Результаты расчетов этой величины и ее логарифмов приведены в табл. 12, а значения величины

$$\overline{f(x, y)} \ln \frac{\overline{f(x, y)}}{g(x) h(y)}$$

в табл. 13.

Количество информации определяется суммированием величин

$$\overline{f(x, y)} \ln \frac{\overline{f(x, y)}}{g(x) h(y)}$$

по строкам и столбцам:

$$\begin{aligned} \overline{T(x, y)} &= \sum_{i=1}^k \sum_{j=1}^l \frac{n_{ij}}{N} \ln \left(\frac{n_{ij}N}{n_i n_j} \right) - \frac{(k-1)(l-1)}{2N} = \\ &= 0,7309 - \frac{(8-1)(6-1)}{2 \cdot 50} = 0,7309 - 0,35 = 0,3809. \end{aligned}$$

Таблица 12

Значения величины $\frac{\overline{f(x, y)}}{g(x) h(y)}$ (числитель) и ее логарифмов (знаменатель)

j	i							
	1	2	3	4	5	6	7	8
1	2,941	2,941	2,287	1,177				
	1,078	1,078	0,828	0,161				
2			0,700	3,125	1,138			
			-0,357	1,139	0,131			
3			0,427	0,384	2,100	1,142	1,538	
			-0,851	-0,958	0,742	0,365	0,432	
4					1,138	3,166	2,500	
					0,131	1,151	0,916	
5					2,272	3,125		
					0,820	1,139		
6							5,000	25,000
							1,609	3,128

Результаты вычисления величины $f(x, y) \ln \frac{\bar{f}(x, y)}{g(x) h(y)}$

j	i								
	1	2	3	4	5	6	7	8	9
1	0,0216	0,1078	0,1159	0,0129					0,2582
2			-0,0071	0,1139	0,0042				0,1110
3			-0,0170	-0,0192	0,0890	0,0219	0,0173		0,0920
4					0,0052	0,0921	0,0366		0,1339
5					0,0164	0,0228			0,0392
6							0,0322	0,0644	0,0966
	0,0216	0,1078	0,0918	0,1076	0,1148	0,1368	0,0861	0,0644	0,7309

Информационный коэффициент корреляции между длиной раковин и длиной однорядной части раковин *Bifarina adalae* L.

$$r_{I(x,y)} = \sqrt{1 - e^{-2 \cdot 0,3809}} = \sqrt{1 - 0,468} = 0,795,$$

что подтверждает предположение о существенной зависимости между этими величинами.

Недостатком информационного коэффициента корреляции является отсутствие надежного критерия для оценки значимости его выборочных оценок.

§ 3. ИСПОЛЬЗОВАНИЕ КОРРЕЛЯЦИОННЫХ СВЯЗЕЙ ДЛЯ ПРЕДСКАЗАНИЯ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

Если для двух величин на основании представительной выборки доказано наличие корреляционной связи, определен ее вид и подобран описывающее его уравнение, то создается возможность прогноза значений одной из случайных величин по значениям другой. Подобные задачи часто возникают в геологической практике. Например, определение средних содержаний сопутствующих компонентов по содержаниям главных или прогнозная оценка содержаний металла в рудном теле в связи с изменением его мощности. При решении задач такого рода могут исследоваться связи между однородными признаками (например, между содержаниями минералов различного состава, между содержаниями различных элементов в минералах или в породах) и связи между неоднородными признаками (например, между содержаниями металлов в рудах и мощностями рудных тел или эффективной пористостью вмещающих пород).

В первом случае рассматриваются парагенетические ассоциации величин, а во втором — более сложные статистические связи. Однако в обоих случаях нельзя считать, что корреля-

ция между двумя признаками указывает хотя бы на частичную зависимость одного признака от другого. Связь между двумя признаками обычно указывает только на наличие некоторой причины, общей для обеих измеряемых величин.

В геологической практике весьма 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.

Основная предпосылка регрессионного анализа заключается в том, что одна переменная (X) рассматривается как независимая, а вторая (Y) — как зависимая от первой и имеющая нормальное распределение с математическим ожиданием и дисперсией, не зависящими от X (см. рис. 13).

Регрессионный анализ применяется для:

— оценки содержания сопутствующих компонентов по содержаниям основных компонентов в рудах. Например, кадмий в полиметаллических рудах обычно входит в состав сфалерита, а рений в молибденсодержащих медно-порфиновых рудах в состав молибденита. Непосредственное определение редких и рассеянных элементов в рудах затруднительно ввиду их низких концентраций, сопоставимых с точностью анализов. Надежные количественные оценки содержания 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 руды. На месторождениях железа, свинца, хрома, барита и других полезных ископаемых, где руды характеризуются высоким содержанием полезного минерала, значительно отличающегося по удельной массе от вмещающих 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 за-

мере 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 оказывает влияние ряд других факторов, не всегда поддающихся 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 бедных — заниженными [29, 55]. 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 в случае линейной связи коэффициент корреляции и корреляционное отношение совпадают по абсолютной величине.

Подходящим критерием для проверки поставленной гипотезы является критерий Фишера

$$F = \frac{\eta_{y/x}^2 - r^2}{1 - \eta_{y/x}^2} \cdot \frac{Nm}{m - 2}, \quad (\text{II.21})$$

где $\eta_{y/x}^2$ — корреляционное отношение признака Y по классам группирования X ;

m — число классов группирования;

N — число членов выборки.

Аналогично вычисляется критерий F и для корреляционного отношения $\eta_{x/y}$. Полученные значения F сравниваются с табличными значениями для заданного уровня значимости при $(m-2)$ и $(N-m)$ степенях свободы (см. прил. 2, 3). Корреля-

ция должна считаться нелинейной, если $F > F_{\text{табл}}$ для большего из корреляционных отношений.

Пример. Проверим гипотезу о линейности корреляционной связи между содержаниями золота и свинца в рудах свинцового месторождения (исходные данные в табл. 8) с помощью критерия Фишера:

$$F = \frac{0,552 - 0,0066}{1 - 0,552} \cdot \frac{60 - 10}{10 - 2} = 10,65.$$

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:

$$y = a_1 + b_1x;$$

$$x = a_2 + b_2y.$$

Оценки коэффициентов линейной регрессии определяются по формулам:

$$b_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}; \quad (\text{II.22})$$

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad (\text{II.23})$$

$$a_2 = \bar{y} - b_1\bar{x}; \quad (\text{II.24})$$

$$a_1 = \bar{x} - b_2\bar{y}. \quad (\text{II.25})$$

Нелинейная связь может быть аналитически выражена с помощью уравнения параболы

$$\bar{y}_x = a + bx + cx^2. \quad (\text{II.26})$$

Коэффициенты этого уравнения находятся путем группирования одного параметра по классам значений другого и решения системы уравнений

$$\left. \begin{aligned} an + b\sum n_x x + c\sum n_x x^2 &= \sum n_x \bar{y}_x \\ a\sum n_x x + b\sum n_x x^2 + c\sum n_x x^3 &= \sum n_x x \bar{y}_x \\ a\sum n_x x^2 + b\sum n_x x^3 + c\sum n_x x^4 &= \sum n_x x^2 \bar{y}_x \end{aligned} \right\} \quad (\text{II.27})$$

где n_x — количество проб в классах группирования;

\bar{y}_x — частное среднее значение величины Y в данном классе.

Расчет уравнения регрессии для содержаний золота и свинца во вкрапленных рудах

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\frac{(x_i - \bar{x})}{(y_i - \bar{y})}$	$(x_i - \bar{x})^2$
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
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
N $\sum_{i=1}$	28,39	48,35			32,036	6,660	3,900

$$\bar{x}=0,95; \bar{y}=1,61;$$

$$b_1 = \frac{6,600}{3,900} = 1,69;$$

$$a_1 = 1,61 - 1,69 \cdot 0,95 \approx 0; \quad y = 1,69x.$$

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:

$$y = a + b \lg x; \quad (\text{II.28})$$

$$\lg y = a + bx; \quad (\text{II.29})$$

$$\lg y = a + b \lg x. \quad (\text{II.30})$$

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 не-

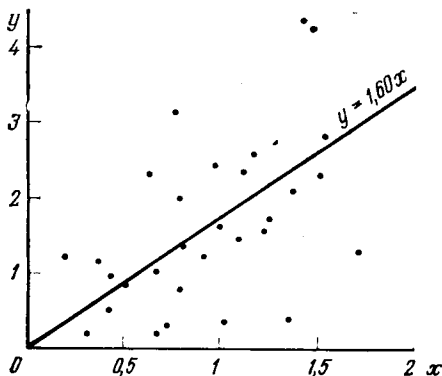


Рис. 14. Линия регрессии содержаний свинца x (в %) и золота y (в г/т) для вкрапленных руд свинцового месторождения

сколько совокупностей, однородных с генетических позиций. Так, выделив из данных, приведенных в табл. 8, только пробы, характеризующие вкрапленные руды (табл. 14), можно убедиться, что связь между содержаниями золота и свинца для этого типа руд близка к линейной (рис. 14) и выражается уравнением

$$y = 1,69x.$$

Вопросы применения корреляционного и регрессионного анализов в геологии и геохимии рассмотрены в работах В. Н. Бондаренко [4], Р. Л. Миллера и Дж. С. Кана [36], а также в сборнике «Математическая обработка данных в поисковой геохимии» [34]. Интересные примеры использования двумерной статистической модели для решения некоторых задач разведки месторождений приведены в работах Д. Г. Криге [29, 55].

Каждое геологическое явление или объект характеризуется множеством признаков, которые можно наблюдать и измерять. Наблюдаемые значения признаков обязаны в большинстве случаев действию не одного, а целого ряда причин, находящихся друг с другом в различных временных и пространственных взаимоотношениях. Изучаемые природные системы в геологии, как правило, зависят от большого числа факторов и требуют для своего описания соответствующего признакового пространства. Так, например, угол наклона прибрежной части морского дна зависит от основных гидродинамических характеристик волн — их высоты, периодов колебаний, углов подхода волн в прибрежной зоне и глубины морского дна, а в качестве промежуточных контролирующих факторов могут быть использованы характеристики береговых течений и средний размер зерен прибрежно-морских отложений. Взаимоотношения этих факторов схематически показаны на рис. 15. Эта схема упро-

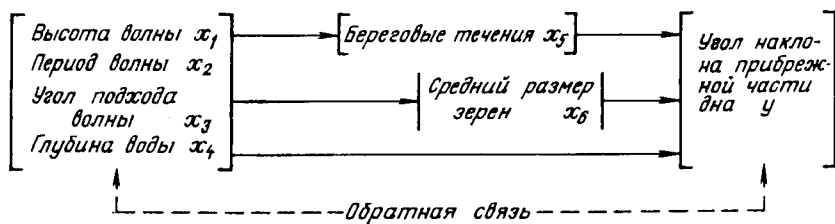


Рис. 15. Схема зависимости между переменными, изучаемыми в процессе исследования угла наклона прибрежной части дна. По У. Крамбейну

щена по сравнению с реальным процессом, но тем не менее она наглядно показывает его сложность [28].

В качестве другого примера можно привести задачу распознавания пород сходного минерального состава, принадлежащих к различным формационным типам [30]. Внешне сходные породы различных магматических формаций обладают некоторыми специфическими петрохимическими особенностями, которые не так значительны, чтобы быть заметными с первого взгляда. Они могут быть уверенно выявлены только при статистической обработке большого количества химических анализов. Приведенные примеры убеждают в необходимости совместного рассмотрения комплекса признаков при решении подобных задач.

В качестве математической модели значений комплекса признаков обычно рассматривается многомерная случайная величина или, как ее называют, случайный вектор. Обозначим число признаков через m , а одномерные случайные величины, соответствующие каждому признаку в отдельности, через $x_1, x_2, x_3, \dots, x_m$. Тогда m -мерную случайную величину X можно рассматривать как вектор-столбец

$$X = \begin{vmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{vmatrix} \quad (\text{III.1})$$

или как вектор-строку

$$X' = |x_1, \dots, x_m|. \quad (\text{III.2})$$

Совокупность векторов наблюдений, записанных в последовательности по строкам, составляет матрицу наблюдений

$$X = \begin{vmatrix} x_{11} & \cdot & \cdot & \cdot & \cdot & \cdot & x_{1n} \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ x_{1m} & \cdot & \cdot & \cdot & \cdot & \cdot & x_{nm} \end{vmatrix}, \quad (\text{III.3})$$

которая представляет собой прямоугольную таблицу с определенным расположением образующих ее элементов. Любой элемент матрицы обозначается x_{ij} , где первый нижний индекс i обозначает номер строки, а второй j — номер столбца. Число строк и столбцов определяют порядок матрицы, который для приведенного примера будет $n \times m$.

Частный случай матрицы — скаляр. Он воспринимается как отдельное число. Матрица порядка 1×1 есть скаляр. Так, например, единичное значение содержания олова в геохимической пробе представляет собой скалярную величину, но если геохимическая проба проанализирована на несколько химических элементов и таких проб много, то их набор значений можно записать в виде матрицы.

Транспонированный (переставленный) вектор-столбец X обозначается штрихом — X' , как в случае формулы (III.2). Вообще транспозицией матрицы X порядка $n \times m$ будет матрица X' порядка $m \times n$, образуемая путем замены строк матрицы X на столбцы

$$X' = \begin{vmatrix} x_{11} & \dots & \dots & \dots & x_{1n} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ x_{1m} & \dots & \dots & \dots & x_{nm} \end{vmatrix}. \quad (\text{III.4})$$

Статистические свойства исходной матрицы задаются ковариационной матрицей (предполагается, что x -переменные центрированы), т. е.

$$L = \frac{X'X}{N-1} = \frac{1}{N-1} \begin{vmatrix} \Sigma x_1^2 \Sigma x_1 x_2 & \dots & \dots & \dots & \Sigma x_1 x_m \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \Sigma x_m x_1 \Sigma x_m x_2 & \dots & \dots & \dots & \Sigma x_m^2 \end{vmatrix}. \quad (\text{III.5})$$

Ее диагональные элементы являются дисперсиями, внедиагональные — ковариациями. Если вычислить ковариационную матрицу для стандартизированных переменных, то получится корреляционная матрица

$$R = \begin{vmatrix} 1, & r_{x_1 x_2} & \dots & \dots & r_{x_1 x_m} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ r_{x_m x_1}, & r_{x_m x_2} & \dots & \dots & 1 \end{vmatrix}. \quad (\text{III.6})$$

Здесь по диагонали стоят единицы, а недиагональные элементы представляют собой обычные парные корреляционные коэффициенты.

Для изучения внутренней структуры и статистических свойств ковариационной (или корреляционной) матрицы применяются различные методы многомерной статистики — регрессионный, дискриминантный, дисперсионный, факторный и другие анализы [27, 28 и др.].

§ 1. МЕТОД МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Метод множественной регрессии позволяет решать классификационные и прогнозные задачи. Задача классификации является типичной задачей распознавания образов. Основные положения в теории распознавания исходят из представления, что входные величины (факторы) образуют p -мерное пространство, в котором каждому образу соответствует множество точек. Задача распознавания сводится к построению гипер-

плоскости, разделяющей эти множества. Функция, описывающая разделяющую гиперплоскость, строится по разным законам. В пределе задача сводится к нахождению формы математического описания многофакторного объекта.

Уравнение множественной регрессии можно записать как линеаризованную функцию классификации вида

$$y = a_0 + \sum_{i=1}^m a_i x_i, \quad (\text{III.7})$$

где x_i — значения i -ого признака.

Найденное уравнение множественной регрессии (III.7) наилучшим образом (в смысле метода наименьших квадратов) соответствует имеющимся точкам и позволяет определять совместное влияние многих параметров на зависимые переменные. Кроме того, имеется возможность определить относительное влияние каждого параметра и измерить полный эффект с помощью коэффициентов a_i .

Таким образом, задача множественной регрессии сводится к вычислению значений коэффициентов регрессии a_0, a_1, \dots, a_k по совокупности N наблюдений переменных x_1, x_2, \dots, x_m и зависимой переменной Y при условии, что вычисленные значения переменной Y максимально приближены к соответствующим наблюдаемым значениям этой же переменной.

Решение задачи достигается путем выполнения следующих вычислительных процедур.

1. Вычисление сумм взаимных произведений отклонений всех переменных

$$S_{jk} = \sum_{i=1}^n (x_{ij} - T_j)(x_{ik} - T_k) - \frac{\sum_{i=1}^n (x_{ij} - T_j) \sum_{i=1}^n (x_{ik} - T_k)}{n}, \quad (\text{III.8})$$

где $j=1, 2, 3, \dots, m;$
 $k=1, 2, 3, \dots, m;$

$$T_j = \frac{\sum_{i=1}^m x_{ij}}{m};$$

$$T_k = \frac{\sum_{i=1}^m x_{ik}}{m}.$$

2. Вычисление средних для всех переменных

$$x_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad (\text{III.9})$$

где $j=1, 2, 3, \dots, m.$

3. Вычисление коэффициентов корреляции для всех возможных пар переменных

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}} \sqrt{S_{kk}}}, \quad (\text{III.10})$$

где $j=1, 2, \dots, m$;
 $k=1, 2, 3, \dots, m$.

4. Вычисление стандартных отклонений для всех переменных

$$S_j = \frac{\sqrt{S_{jj}}}{\sqrt{n-1}}, \quad (\text{III.11})$$

где $j=1, 2, 3, \dots, m$.

5. Обращение матрицы парных корреляционных коэффициентов. Обратной по отношению к данной называется матрица, которая при умножении как справа, так и слева на данную матрицу дает единичную матрицу, диагональ которой представлена единицами, а остальные члены матрицы — нулями

$$RR^{-1} = R^{-1}R = E.$$

6. Вычисление коэффициентов регрессии

$$a_j = \frac{S_y}{S_j} \sum r_{ij} r_{ij}^{-1}, \quad (\text{III.12})$$

где S_y — стандартное отклонение зависимой переменной;
 S_j — стандартное отклонение j -ой независимой переменной;
 r_{iy} — парная корреляция i -ой независимой переменной с зависимой переменной;

r_{ij}^{-1} — обратная величина парной корреляции независимых переменных $j=1, 2, 3, \dots, k$.

7. Вычисление свободного члена

$$a_0 = Y - \sum_{j=1}^k a_j \bar{x}_j, \quad (\text{III.13})$$

где Y — среднее значение зависимой переменной y ;

\bar{x}_j — среднее значение j -ой независимой переменной.

8. Вычисление множественного коэффициента корреляции

$$R = \sqrt{1 - \frac{|L|}{a_{11}|L'|}}, \quad (\text{III.14})$$

где $|L|$ — определитель ковариационной матрицы;

a_{11} — первый член ковариационной матрицы;

$|L'|$ — определитель ковариационной матрицы без первого столбца и первой строки.

Пример использования метода множественной регрессии. Рассмотрим результаты статистического изучения особенностей распределения содержания рубидия, цезия, лития и калия в микроклинах крупного пегматитового тела. Многими исследователями было отмечено, что калиевые полевые

Таблица 15

Пространственное положение проб, интенсивность окраски и содержания элементов в микролинах

Относительная высота (h), м	Относительное расстояние от контакта (l), %	Цветовой индекс (цв)	Содержание, %			
			K	Rb	Cs	Li
2	50	3	9,75	0,758	0,030	0,0700
55	50	4	9,55	0,437	0,043	0,0173
128	33	3	8,90	0,615	0,032	0,0280
15	50	3	8,50	0,562	0,035	0,0467
47	22	4	8,50	0,562	0,031	0,0117
112	2	2	10,20	0,615	0,028	0,0262
112	6	2	8,12	0,690	0,032	0,0692
112	50	3	11,50	0,775	0,032	0,0724
288	50	2	8,70	1,070	0,078	0,0149
160	4	2	11,00	1,380	0,042	0,0678
160	10	2	9,55	0,775	0,020	0,0196
160	35	3	9,55	0,740	0,042	0,0318
160	50	2	11,20	0,675	0,024	0,0818
162	30	3	8,30	0,562	0,018	0,0047
175	50	3	9,55	0,890	0,060	0,0584
20	50	3	12,00	0,890	0,035	0,0336
17	50	4	10,00	0,740	0,047	0,0678
13	23	2	9,32	0,588	0,015	0,0027
3	3	2	11,00	1,020	0,016	0,0332
95	50	3	10,70	1,100	0,023	0,0262
142	50	3	10,70	1,230	0,054	0,0588
200	10	2	10,20	1,410	0,029	0,0135
200	10	3	9,75	1,070	0,029	0,0093
160	50	2	9,32	0,510	0,049	0,0294
278	13	2	9,32	0,710	0,037	0,0164
295	50	2	10,00	1,210	0,049	0,0514
280	12	3	8,12	0,417	0,045	0,0136
257	30	2	9,10	0,991	0,052	0,0304
276	14	2	8,50	1,150	0,338	0,0136
312	28	2	9,32	2,700	0,234	0,0215
320	13	2	12,60	1,730	0,076	0,1308
260	50	2	11,20	1,320	0,052	0,0107
210	50	1	10,20	1,050	0,026	0,0023
190	13	1	8,30	0,588	0,021	0,0240
190	10	2	8,12	0,562	0,044	0,0050
210	50	2	10,70	1,170	0,066	0,0369
235	50	2	9,32	0,630	0,170	0,0864
317	50	1	9,55	1,040	0,044	0,0304
328	50	2	8,90	2,180	0,091	0,0229
327	50	2	10,20	0,890	0,117	0,0271
327	50	2	8,50	0,850	0,105	0,0552
325	20	1	10,00	0,660	0,035	0,0042
315	13	2	10,50	0,975	0,105	0,0467
318	50	1	11,20	0,890	0,056	0,0794
15	36	3	10,20	0,490	0,040	0,0467
15	3	2	10,70	0,338	0,032	0,0140
15	13	3	8,12	0,562	0,041	0,0444
20	12	2	9,32	0,331	0,017	0,0047
19	28	2	9,76	0,338	0,015	0,0005

Относительная высота (h), м	Относительное расстояние от контакта (l), %	Цветовой индекс (цв)	Содержание, %			
			K	Rb	Cs	Li
18	50	3	11,50	0,500	0,027	0,0397
276	13	2	11,41	0,725	0,043	0,0232
20	50	3	11,75	0,910	0,138	0,0186
318	50	2	11,00	0,870	0,071	0,0148
190	50	2	11,16	0,725	0,072	0,0381
20	50	3	10,91	0,645	0,035	0,0506
18	50	3	11,08	0,910	0,026	0,0645

шпаты, взятые из различных частей пегматитовых жил, характеризуются разным содержанием щелочей, причем наблюдается закономерное уменьшение содержания 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 простираанию, так и по падению: наличие раздувов, пережимов, изгибов и апофиз), отличается крупными размерами и характеризуется довольно ясно выраженным симметрично-зональным строением. Призальбантовую часть тела слагает эндоконтактная мусковит-кварцевая оторочка шириной от 10 до 40 см. Эта оторочка сменяется зоной мелкозернистого существенно альбититового пегматита мощностью от 1 до 3 м. Центральная часть жилы сложена неравномернозернистым средне- и крупнозернистым, часто блоковым, неравномерно альбитизированным сподумен-микроклин-кварцевым пегматитом директивной структуры, среди которого 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. Результаты анализов приведены в табл. 15.

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 (см. табл. 15). Содержания элементов (S_{Rb} , S_{Cs} , S_{Li} , S_K) были выражены 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 централь-

ной 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 самих случайных величин и для логарифмов этих величин. Но так как проверка гипотезы о равенстве коэффициентов корреляции с помощью преобразования Фишера показала, что различия между оценками этих коэффициентов для самих случайных величин и для их логарифмов являются несущественными (при 5%-ном уровне значимости), результаты вычислений для логарифмов не приводятся.

В результате проведенных вычислений были получены основные статистические характеристики коррелируемых параметров (табл. 16), их парные

Таблица 16

Статистические характеристики коррелируемых параметров микроклин из пегматитовой жилы

Признак	Среднее арифметическое, \bar{x}	Среднее квадратическое отклонение, σ_x	Среднее арифметическое логарифмов, $\lg \bar{x}$	Среднее квадратическое отклонение логарифмов, $S_{\lg x}$	Среднее геометрическое, \bar{x}
<i>56 проб</i>					
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)	11429,46	1849,26	4,0140	0,138	10330,0
рубидия (C_{Rb})	891,62	426,28	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* (цв.)	2,34	0,71	0,3470	0,144	2,22
<i>34 пробы</i>					
Относительная высота (h)	163,03	117,89	1,9730	0,574	91,83
Содержание (в $n \cdot 10^{-3}\%$):					
калия (C_K)	10145,00	959,84	4,0062	0,041	10140,0
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
Цв.	2,44	0,74	0,3649	0,148	2,32

* Далее здесь и в аналогичных таблицах этот признак будет дан сокращенно.

и множественные коэффициенты корреляции (табл. 17 и 18) и составлены в ряде случаев уравнения множественной регрессии.

Анализируя полученные данные, можно сделать следующие выводы.

1. Сравнение статистических характеристик проб только из центральной части жилы ($n=34$) с характеристиками всего набора проб ($n=56$) показывает (см. табл. 16), что исключение из расчета приконтактных проб (у которых $l > 25\%$) вызывает некоторое увеличение средних содержаний рубидия, цезия, лития, уменьшение содержаний калия и величины отношения 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. 17) показывают, что при всех вариантах расчета наблюдается значимая линейная связь между параметром h и такими параметрами

Т а б л и ц а 17

Парные коэффициенты корреляции параметров
для 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	(0,017)	-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)	
K/Rb	-0,658				
Rb/Cs	(-0,223)				
l					
Цв.	-0,749				

* Здесь и в табл. 20 в скобках даны корреляционные коэффициенты, величины оценок которых при 5%-ном уровне значимости несущественно отличаются от нуля. Оценка проведена по преобразованию Фишера.

Параметры	K/Rb	Rb/Cs	l	Цв.
h	-0,574	(-0,262)	(0,013)	0,579
C_{Rb}	-0,795	(0,176)	(0,096)	(-0,249)
C_{Cs}	-0,366	-0,521	(0,008)	(-0,168)
C_{Li}	(-0,078)	(0,071)	(0,133)	(0,035)
C_K	0,409	(0,133)	-0,312	-0,404
K/Rb		(0,224)	(-0,145)	(-0,225)
Rb/Cs	(-0,186)		(-0,172)	(-0,112)
l				(0,204)
Цв.	(0,411)	(-0,094)		

Таблица 18

Множественные корреляционные коэффициенты параметров
для микроклинов по двум выборкам проб

Признаки	$n=56$			
	6*	5	4	3
Относительная высота (h)	0,5440	0,5436	0,5297	
Содержание:				
рубидия (C_{Rb})	0,5937	0,5825	0,5494	0,5488
цезия (C_{Cs})	0,5253	0,5215	0,5034	0,5010
лития (C_{Li})	(0,1256)	(0,1209)	(0,0636)	
калия (C_K)	(0,3751)	(0,3059)		
Относительное расстояние от контакта (l)	(0,2682)			

* Число признаков, участвовавших в расчете.

Признаки	n=34			
	5	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)			

рами, как содержания рубидия, цезия (прямая связь), величина отношения калия к рубидию и интенсивность окраски (обратная связь), т.е. чем выше взята проба микроклина, тем больше в ней рубидия и цезия, меньше величина отношения калия к рубидию и светлее окраска. Существенная и надежная положительная связь наблюдается между содержаниями рубидия и цезия, тогда как между содержаниями рубидия и лития, так же как и между содержаниями цезия и лития, связь отсутствует. Влияние расстояния от контакта (связь с параметром l) проявилась только для содержания калия (значимая отрицательная связь). Изменение величины отношения калия к рубидию определяется в основном изменением содержания рубидия, а величины отношения рубидия к цезию — изменением содержания цезия.

Оценки парных коэффициентов корреляции, рассчитанные для проб только из центральной части жилы ($n=34$), обладают, как правило, большей абсолютной величиной, чем оценки коэффициентов проб из всей жилы ($n=56$), но различия эти, как показывает проверка с помощью F -преобразования Фишера, являются несущественными.

3. Результаты определения оценок коэффициентов множественной корреляции (см. табл. 18) 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 уравнения множественной регрессии типа

$$\left. \begin{aligned} h &= 48,06 + 0,10C_{Rb} + 0,66C_{Cs} - 0,18C_{Li} \\ h &= 79,74 + 0,08C_{Rb} + 0,20C_{Cs} \end{aligned} \right\} n = 56;$$

$$\left. \begin{aligned} h &= 43,08 + 0,08C_{Rb} + 1,01C_{Cs} - 0,27C_{Li} \\ h &= 25,48 + 0,08C_{Rb} + 1,05C_{Cs} \end{aligned} \right\} n = 34.$$

Полученные уравнения могут быть использованы для примерной оценки величины эрозионного среза и для прогноза жильного тела на глубину.

§ 2. МЕТОД ЛИНЕЙНОЙ ДИСКРИМИНАНТНОЙ ФУНКЦИИ

Ковариационные матрицы используются и в задачах дискриминации, когда выборку, задаваемую многомерным вектором наблюдений $X' = (x_1, x_2, \dots, x_p)$, нужно отнести к одной из

p -мерных нормально распределенных генеральных совокупностей.

Подобные задачи часто встречаются в повседневной деятельности геологов. При множестве наблюдений, заданных многомерными векторами, нужно разбить их на группы так, чтобы была достигнута максимальная однородность внутри групп и минимальная — между группами.

Метод линейной дискриминантной функции, решающий классификационную задачу, заключается в комплексном использовании нескольких признаков, объединенных в линейную комбинацию таким образом, чтобы получить наилучшее разделение сравниваемых групп. Физический смысл линейной дискриминантной функции состоит в том, что она представляет собой уравнение гиперплоскости в p -мерном пространстве признаков, проведенной так, чтобы по одну сторону от нее оказалось максимальное количество объектов, относящихся к первой группе, а по другую — максимальное количество объектов, принадлежащих второй группе. В том случае, когда имеется больше двух групп объектов, распознавание достигается путем дихотомического деления [1, 30].

Дискриминантный анализ разработан для случая, когда совокупности признаков описываются многомерным нормальным распределением. При этом используются два типа дискриминантных функций — линейные и квадратные. Линейная дискриминантная функция строится при условии равенства ковариационных матриц обеих совокупностей. Она выражает уравнение гиперплоскости, наилучшим образом делящей эти совокупности. Степень пересечения совокупностей характеризуется ошибками первого и второго рода. При неравных ковариационных матрицах возникают некоторые трудности в построении дискриминантных функций. В этих условиях выражение функции усложняется, и поэтому классификация объектов может проводиться только с использованием ЭВМ.

Пусть имеются два геологических объекта, которые представлены двумя p -мерными статистическими совокупностями с соответствующими плотностями вероятностей. Каждому объекту соответствует свое значение вектора среднего \bar{X} и значение ковариационной матрицы. Положим, что ковариационные матрицы для первого и второго объектов равны, т. е.

$$L_1 = L_2 = L.$$

Такое, не слишком сильное ограничение ведет к большой простоте классификационного правила.

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

$$U = \begin{pmatrix} u_{11}, & u_{12}, & u_{13}, & \dots, & u_{1n_1} \\ u_{21}, & u_{22}, & u_{23}, & \dots, & u_{2n_1} \\ \vdots & & & & \\ \vdots & & & & \\ u_{k1}, & u_{k2}, & u_{k3}, & \dots, & u_{kn_1} \end{pmatrix}. \quad (\text{III.15})$$

Обозначим через v_{ij} результат измерения признака с номером i в пробе с номером j , взятой из второй совокупности, и получим матрицу V порядка $k \times n_2$

$$V = \begin{pmatrix} v_{11}, & v_{12}, & \dots, & v_{1n_2} \\ v_{21}, & v_{22}, & \dots, & v_{2n_2} \\ \vdots & & & \\ \vdots & & & \\ v_{k1}, & v_{k2}, & \dots, & v_{kn_2} \end{pmatrix}. \quad (\text{III.16})$$

Используя эти данные, вычисляем элементы выборочной ковариационной матрицы B , которую можно представить в матричной записи

$$B = \frac{1}{n_1 + n_2 - 2} (S_U + S_V), \quad (\text{III.17})$$

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

$$D(x) = \sum_{j=1}^p a_j \bar{x}_j. \quad (\text{III.18})$$

Чтобы отнести предъявленный объект с признаками x_1, x_2, \dots, x_k к одной из двух групп, следует руководствоваться следующим правилом: если $D(x) \geq 0$, то объект относят к первой группе, в противном случае объект относят к второй группе.

С дискриминантной функцией связана величина, которая служит мерой надежности принимаемых решений

$$D^2 = \frac{1}{|R|} \sum_{i=1}^n \sum_{j=1}^p R_{ij} \frac{\bar{x}_{1i} - \bar{x}_{2i}}{\sigma_i} \frac{\bar{x}_{1j} - \bar{x}_{2j}}{\sigma_j}, \quad (\text{III.19})$$

где R — матрица, образованная оценками коэффициентов корреляции;

R_{ij} — алгебраическое дополнение элемента, стоящего на пересечении i -й строки и j -го столбца.

Статистика D^2 носит название «обобщенного расстояния» или критерия Махаланобиса. 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 ошибки классификации определяется из соотношения

$$P = 1 - \Phi(D^2), \quad (III.20)$$

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 и в классификации не участвуют.

Пример. Классификация кимберлитовых пород одного из месторождений Сибири по петрохимическим данным.

На месторождении выделены три типа кимберлитов — афировые туфы северо-восточной части месторождения, кластопорфиновые 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:

$$45,157MnO - 1,236FeO + 1,642CaO - 33,489Na_2O + 21,809K_2O + \\ + 43,213P_2O_5 - 1486S_{\text{сульф}} - 1,040CO_2 - 61,386 = D(x).$$

Если $D(x) > 0$, то кимберлитовая порода относится к афировым кимберлитам северо-востока, при $D(x) < 0$ — к афировым туфам юго-запада.

Т а б л и ц а 19

Теоретическая и эмпирическая ошибки неправильной классификации для кимберлитовых пород одного месторождения

Афировые туфы юго-запада	Кластопорфиновые брекчии	Тип кимберлитов
$\frac{0,1\%}{0,00 (18)}$	$\frac{0,3\%}{0,00 (42)}$	Афировые туфы (северо-восточная часть)
	$\frac{2,0\%}{10,0\% (44)}$	Афировые туфы (юго-западная часть)
	$\frac{5\%}{8,0\% (52)}$	Все туфы

Примечание. В числителе — теоретическая ошибка классификации, в знаменателе — эмпирическая. В скобках — количество проэкзаменованных проб.

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 исходных данных (табл. 19).

Анализ данных, приведенных в табл. 19, показывает, что полученные дискриминантные функции дают возможность достаточно надежно определить тип кимберлитовых пород по их химическому составу.

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.

В работах [17, 18] показано, что при решении классификационных задач в геологии может быть с успехом применен метод максимального правдоподобия, имеющий неоспоримые преимущества перед методом дискриминантного анализа.

Сущность метода максимального правдоподобия заключается в следующем.

Пусть η — случайная величина и $P(x, \theta)$ — соответствующая ей плотность вероятности, вид которой известен, а значение параметра θ неизвестно. Произведена выборка объема n и получено n значений случайной величины η , которые обозначим x_1, x_2, \dots, x_n . Результаты выборки можно рассматривать как значения n независимых одинаково распределенных случайных величин $\eta_1, \eta_2, \dots, \eta_n$ с плотностью $P(x, \theta)$.

Функцией правдоподобия $L(x_1, x_2, \dots, x_n)$ называется совместная плотность распределения случайных величин $\eta_1, \eta_2, \dots, \eta_n$ при фиксированных значениях x_1, x_2, \dots, x_n . Пусть R — множество значений параметра θ . Каждому значению $\theta \in R$ при заданных x_1, x_2, \dots, x_n можно поставить в соответствие значение функции $L(x_1, x_2, \dots, x_n; \theta)$. Представляется вполне естественным, как считает Д. А. Родионов [43], из всех $\theta \in R$ при данных x_1, x_2, \dots, x_n выбрать то значение θ , которому соответствует максимум функции правдоподобия.

В связи с тем, что случайные величины $\eta_1, \eta_2, \dots, \eta_n$ предполагаются независимыми

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n P(x_i; \theta), \quad (\text{III.21})$$

для отыскания соответствующей максимально правдоподобной оценки параметра θ достаточно решить уравнение

$$\frac{dL(x_1, x_2, \dots, x_n; \theta)}{d(\theta)} = 0 \quad (\text{III.22})$$

относительно θ .

Часто для удобства его заменяют уравнением

$$\frac{d \ln L(x_1, x_2, \dots, x_n; \theta)}{d(\theta)} = 0. \quad (\text{III.23})$$

При решении классификационных задач метод максимального правдоподобия заключается в отнесении объекта r к тому классу R_i , для которого максимальна величина $P\left(\frac{x}{R_i}\right)$. При классификации объектов непосредственно по значениям плотностей вероятности нет необходимости строить C_N^2 дискриминантных функций и достаточно лишь N сравнений этих плотностей для различных i . При этом точность не теряется с увеличением числа используемых признаков и существенно упрощаются вычисления по сравнению с вычислениями для построения дискриминантных функций (особенно при больших N).

К необходимости вычисления непосредственного отношения величин $P\left(\frac{x}{R_i}\right)$ для классификации многомерных нормальных совокупностей при неравных ковариационных матрицах пришли многие авторы. Такое отношение рассматривается ими как построение нелинейных дискриминантных функций [1], но по существу это усложненный принцип правдоподобия. Величины $P\left(\frac{x}{R_i}\right)$ более удобны для классификации, чем их отношения.

При использовании для классификации непосредственно величин $P\left(\frac{x}{R_i}\right)$ или $\omega = \ln P\left(\frac{x}{R_i}\right)$ возможно наглядное графическое решение задачи. Для этого в прямоугольной системе координат ω_1, ω_2 отображаются классы обучения R_1 и R_2 , а также классифицируемые объекты. Это позволяет судить о взаимном положении классов, их области пересечения и определить вероятную ошибку классификации. Графический способ дает возможность замены части трудоемких вычислительных операций простыми построениями.

Пример. На рис. 16 показано использование дискриминантного анализа для выявления нефтегазоносных структур в Казахстане по средним значениям концентрации метана, гелия, урана, меди, свинца, хрома, никеля, бора, стронция и окислительно-восстановительного потенциала eH .

В качестве облачающего материала были взяты эти признаки по 18 скважинам, пробуренным по газопродуктивной структуре Жетбай (R_1) и по такому же числу скважин на структуре Шалабай-Сенк (R_2), лишенной нефти и газа.

На оси абсцисс (см. рис. 16) отложены значения линейной дискриминантной функции. По оси ординат в виде гистограммы показаны частоты P полученных оценок $D(x)$ для структур обучающего материала. Под осью абсцисс показаны значения $D(x)$, полученные для различных скважин клас-

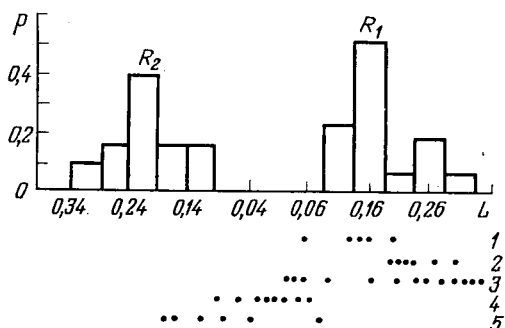


Рис. 16. Классификация структур с использованием дискриминантного анализа. По Н. М. Франковской.

Структуры: 1 — Ушкююн; 2 — Южный Тасбулат; 3 — Тасбулат; 4 — Кокумбай; 5 — Кокше

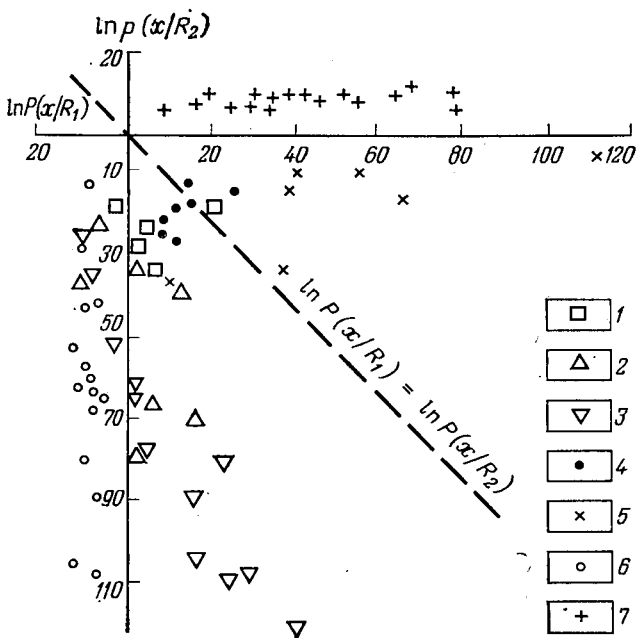


Рис. 17. Классификация структур Южного Мангышлака методом максимального правдоподобия. По Н. М. Франковской.

Скважины в структурах: 1 — Ушкююн; 2 — Южный Тасбулат; 3 — Тасбулат; 4 — Кокумбай; 5 — Кокше; 6 — Жетбай; 7 — Шалабай-Сенк

сифицируемых структур. Скважина распознается как принадлежащая структуре типа Шалабай-Сенк или Жетыбай, если соответствующая ей точка расположена левее или правее значений $D(x)=0$.

На рис. 17 показаны результаты классификации скважин непосредственно по величинам $P\left(\frac{x}{R_i}\right)$. Распределения $P\left(\frac{x}{R_i}\right)$ аппроксимированы нормальным законом, но без ограничения равенства ковариационных матриц.

В пространстве, показанном на рис. 17, любая точка на биссектрисе угла, образованного координатными осями, имеет $P(x/R_1)=P(x/R_2)$, т. е. равновероятно относится к R_1 и R_2 . Точки, лежащие ближе к оси $\ln P(x/R_2)$, имеют значения $P(x/R_1)$ больше значения $P(x/R_2)$, т. е. относятся к первому классу и наоборот. Как видно, структура Кокумбай, относившаяся по дискриминантному анализу примерно равновероятно к R_1 и R_2 , теперь уверенно относится к R_1 . Структура Кокше отмечается точ-

ками, сильно разбросанными в пространстве низких значений $\ln P\left(\frac{x}{R_i}\right)$. Поэтому можно предполагать, что она принадлежит к классу, не представленному в материале обучения.

Приведенный пример свидетельствует о предпочтительности метода максимального правдоподобия для классификации многомерных объектов.

§ 3. МЕТОД ГЛАВНЫХ КОМПОНЕНТ

С увеличением размерности пространства признаков быстро растет трудность изучения характеризующих ими объектов. Возникает задача замены многочисленных наблюдаемых признаков меньшим числом переменных без существенной потери информации. В этом отношении заслуживает внимания метод главных компонент [5, 27, 36, 48]. Основные его достоинства следующие.

1. Метод характеризуется единой универсальной вычислительной процедурой, гарантирующей объективность решения.

2. Он обеспечивает получение наиболее рациональной комбинации наблюдаемых признаков, каждый из которых вносит вклад, соответствующий его информативности и степени согласия с другими признаками.

3. Полученные комбинации признаков выступают в качестве относительно немногочисленных новых (синтезированных) переменных, взаимно независимых. Их суммарная дисперсия равна дисперсии исходных данных, т. е. они несут практически всю полезную информацию, заключенную в многочисленных исходных данных. При этом может быть оценена информативность каждой из новых переменных в отдельности и устанавливается их соответствие с независимыми природными (геологическими) факторами.

Для перехода к новым переменным используются особенности внутренней структуры ковариационной (корреляционной) матрицы.

В методе главных компонент для j -й главной компоненты находят некоррелированные нормированные линейные комбинации, дисперсии которых максимизированы

$$\gamma_j = \sum_{i=1}^k \omega_{ji} x_i, \quad j, i = 1, 2, \dots, k. \quad (\text{III.24})$$

Здесь ω_{ji} — факторный вес j -й компоненты в i -й переменной наблюдения x . Ковариационная (корреляционная) матрица оказывается расщепленной на k ортогональных компонент [1, 31, 37].

Нахождение главных компонент геометрически выражается переходом к новой ортогональной системе координат: первая координатная ось проводится так, чтобы дисперсия соответствующей ей линейной формы составляла возможно большую часть общей дисперсии; далее вытискивается следующая (ортогональная первой) ось с аналогичным требованием по отношению к оставшейся части общей дисперсии и т. д.

От новых координат можно вернуться к старым, записав (с учетом ортогональности преобразования)

$$x_i = \sum_{j=1}^k \omega_{ji} \gamma_j. \quad (\text{III.25})$$

Формула (III.25) выражает основное соотношение метода главных компонент, в котором k главных компонент исчерпывают всю дисперсию исходных переменных.

Реализация метода главных компонент при наличии современной вычислительной техники сравнительно несложна и сводится к использованию двух вычислительных процедур.

1. Формирование ковариационной матрицы S . Для этого матрица наблюдений X размерности $m \times p$, где m — количество наблюдений по p признакам, умножается слева на ее транспонированный аналог

$$S = X'X. \quad (\text{III.26})$$

2. Вычисление собственных значений и векторов ковариационной матрицы способом вращений (метод Якоби). В результате получается:

$$S = W' \Lambda W. \quad (\text{III.27})$$

Здесь W — матрица размерности $p \times p$, столбцы которой являются собственными векторами матрицы S

$$W = \begin{bmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1p} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p1} & \omega_{p2} & \dots & \omega_{pp} \end{bmatrix}, \quad (\text{III.28})$$

где w_{ji} — факторный вес j -й компоненты в i -й переменной наблюдения x ;

Λ — диагональная $p \times p$ матрица, главная диагональ которой представлена собственными значениями матрицы S ,

$$\Lambda = \begin{vmatrix} \lambda_1^2 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^2 & 0 & \dots & 0 \\ \vdots & & & & \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \lambda_p^2 \end{vmatrix} \quad \lambda_1^2 > \lambda_2^2 > \dots > \lambda_p^2. \quad (\text{III.29})$$

Величина $\lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2 = S_p^2$ равна суммарной дисперсии всех наблюдаемых переменных.

Первая главная компонента представляет собой комбинацию

$$\gamma_1 = w_{11}x_1 + w_{21}x_2 + \dots + w_{p1}x_p, \quad (\text{III.30})$$

а ее дисперсия равна λ_1^2 . Вторая главная компонента — соответственно

$$\gamma_2 = w_{12}x_1 + w_{22}x_2 + w_{32}x_3 + \dots + w_{p2}x_p, \quad (\text{III.31})$$

ее дисперсия равна λ_2^2 и т. д.

Замена наблюдаемых переменных главными компонентами называется разложением на ортогональные составляющие, так как главные компоненты между собой не коррелируются из-за ортогональности.

Если сумма дисперсий первых k главных компонент составляет существенную долю от общей суммы S_p^2 , то данные главных компонент могут представлять весь комплекс изучаемых элементов.

После нормирования столбцов матрицы w по соответствующим собственным значениям (каждый элемент j -го столбца после возведения в квадрат умножается на λ_j^2) получим матрицу разложения дисперсии

$$F = \begin{vmatrix} f_{11} & f_{12} & \dots & f_{1p} \\ f_{21} & f_{22} & \dots & f_{2p} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ f_{p1} & f_{p2} & \dots & f_{pp} \end{vmatrix}. \quad (\text{III.32})$$

Если λ_j^2 считать дисперсией, обусловленной воздействием j -го фактора на значения всех признаков, то величина f_{ji} должна рассматриваться как дисперсия, которую вносит j -й фактор

в значение i -го элемента. При интерпретации факторов рассматриваются абсолютные значения и знаки элементов матрицы в пределах первых k столбцов.

Пример использования метода главных компонент при интерпретации геохимических данных. На одном из месторождений редкометалльных пегматитов было установлено, что первичные ореолы рассеяния многих элементов отличаются сложным строением и нечеткими границами [21]. Характер распределения концентраций олова, бериллия и редких щелочей в коренных породах иллюстрируется на рис. 18.

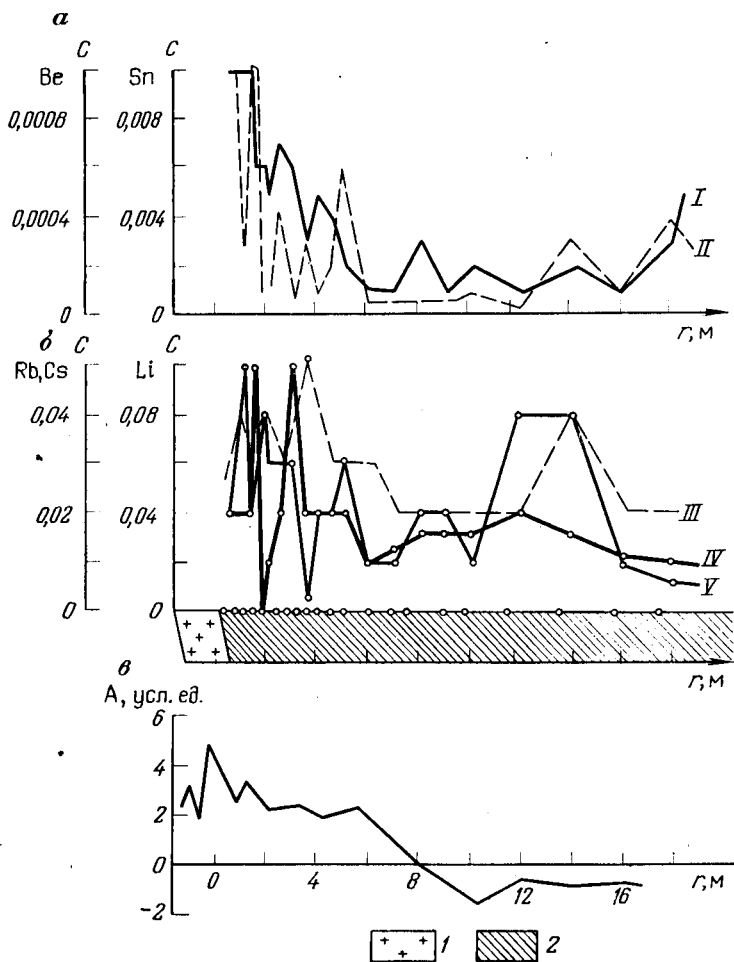


Рис. 18. Характер распределения элементов-индикаторов редкометалльного оруденения в первичных ореолах рассеяния:

a — распределение олова (I), бериллия (II); $б$ — распределение рубидия (III), лития (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 (табл. 20).

Т а б л и ц а 20

Результаты корреляционного и компонентного анализа геохимических данных по месторождению редкометалльных 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 $r > 0,32$.

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 от пегматитовой жилы и положительно с концентрациями исследуемых элементов. Эта компонента отражает общий характер размещения элементов-индикаторов редкометалльного оруденения в первичных ореолах рассеяния и может быть с успехом использована для их выделения (см. рис. 18, в).

Рассмотрим пример совместного использования метода главных компонент и тренд-анализа, которое представляется весьма плодотворным при интерпретации результатов геохимических съемок (см. главу IV).

На одной из площадей была проведена геохимическая съемка по вторичным ореолам рассеяния. По системе профилей отобраны и проанализированы на 12 элементов 208 проб (рис. 19).

Участок исследований сложен полем гнейсов и крупной гранитной интрузией (см. рис. 19, а). Попытка построить тренд-поверхности по отдельным элементам успеха не имела, так как доля отражаемой ею систематической составляющей оказалась слишком малой (см. рис. 19, б). Проведенный компонентный анализ дал результаты, показанные в табл. 21.

Первая главная компонента учитывает 40% от суммарной дисперсии. Все коэффициенты (факторные веса) положительны и существенно отличаются от нуля. Геохимический смысл пер-

вой компоненты может быть определен как уровень общего заражения рыхлых образований всеми двенадцатью элементами. Поверхность тренда, построенная по этой компоненте, отражает распределение суммы всех элементов по площади. Так, для гранитной интрузии отмечается более низкий уровень содержания элементов, чем в поле развития гнейсов (см. рис. 19, в).

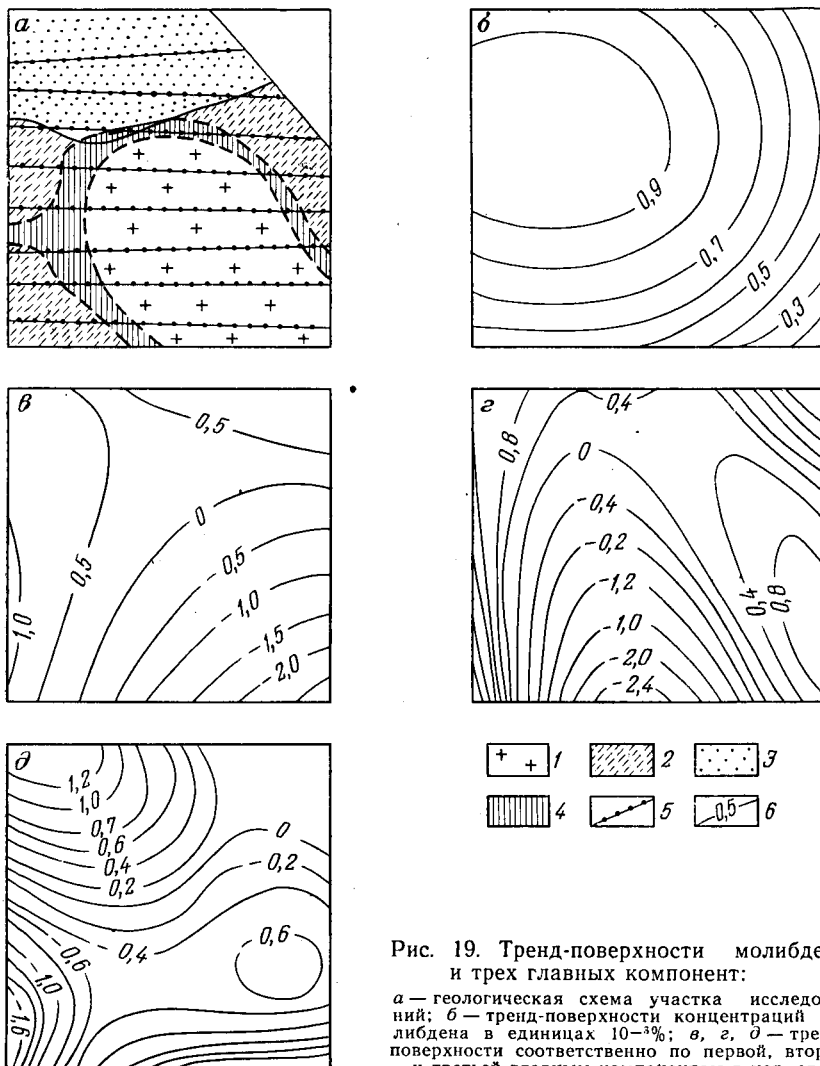


Рис. 19. Тренд-поверхности молибдена и трех главных компонент:

а — геологическая схема участка исследований; б — тренд-поверхности концентраций молибдена в единицах $10^{-3}\%$; в, г, д — тренд-поверхности соответственно по первой, второй и третьей главным компонентам в усл. ед.

1 — граниты; 2 — гнейсы биотитовые, слабо мигматизированные; 3 — биотитовые гнейсы, мигматизированные, с горизонтами гранат-биотитовых гнейсов; 4 — мигматизированные гнейсы и амфибол-кварцевые породы; 5 — места отбора литогеохимических проб; 6 — изолинии со значениями концентраций молибдена

Результаты компонентного анализа

Элемент	Факторный вес		
	первый	второй	третий
Титан (Ti)	0,351	—0,207	—0,031
Ванадий (V)	0,319	—0,188	—0,431
Хром (Cr)	0,232	0,461	—0,182
Марганец (Mn)	0,314	0,016	0,413
Кобальт (Co)	0,356	0,107	—0,226
Никель (Ni)	0,363	0,148	—0,255
Медь (Cu)	0,256	1,428	—0,124
Цинк (Zn)	0,258	0,200	0,342
Молибден (Mo)	0,228	—0,215	0,246
Олово (Sn)	0,263	—0,365	0,158
Барий (Ba)	0,314	—0,090	0,387
Свинец (Pb)	0,116	—0,514	—0,355
Дисперсия факторов, %	40,0	13,0	9,5

Вторая главная компонента учитывает 13% общей дисперсии. Она отражает поведение двух ассоциаций элементов. Первая из них включает хром, медь, цинк, никель, кобальт, а вторая — свинец, олово, молибден, титан, ванадий и барий. Тренд-поверхность, построенная по этой главной компоненте, отражает повышенные концентрации первой ассоциации для поля гнейсов. В гранитах же повышенными концентрациями характеризуется вторая ассоциация элементов (см. рис. 19, з).

На долю третьей главной компоненты приходится всего 9,5% дисперсии. Она представляет собой две ассоциации элементов. Их расположение на площади геохимической съемки четко определяется тектоническим нарушением широтного простиранья. При этом нулевая изолиния компоненты практически совпадает с тектоническим нарушением (см. рис. 19, д). Итак, три первые главные компоненты учитывают 62,5% от суммарной дисперсии и могут представлять всю совокупность из двенадцати элементов.

Первая главная компонента отражает общую закономерность для данного участка: содержание рудных элементов в гнейсах больше, чем в гранитной интрузии. Вторую и третью главные компоненты можно рассматривать как деталлизации, уточняющие эти общие закономерности с учетом геолого-структурных особенностей участка.

Использование метода главных компонент рассмотрим на примере решения прогнозных задач.

Нерюндинское железорудное месторождение является крупнейшим месторождением Ангаро-Катского района Иркутской области. В его геологическом строении принимают участие туфогенные отложения корвучанской

Матрица парных коэффициентов корреляции и факторных нагрузок*

Признаки	Fe _{60m}	Fe ₂ O ₃	FeO	SiO ₂	Al ₂ O ₃	CaO	MgO	MnO	TiO ₂	P	S	K ₂ O	Факторные нагрузки		
													1-я ком-понента	2-я ком-понента	3-я ком-понента
Fe _{60m}	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 ₂ O 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	-0,52	0,77	0,17	-0,02	0,14	-0,78	0,11	0,00
Al ₂ O 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 ₂ O												1,00	0,00	0,04	-0,31
Дисперсия факторов, %													42,00	18,70	9,30

* Коэффициенты корреляции и факторные нагрузки отличаются от нуля с надежностью 0,95 при их абсолютных значениях $r > 0,25$.

Первая компонента, учитывающая 42% от суммарного воздействия, изученные признаки всей совокупности, определяет общий характер орнения. 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 окислов.

Вторая и третья компоненты раскрывают роль кальциево-магниевого калиевого метасоматоза, предшествующего рудообразованию.

Построенная на разрезе тренд-поверхность первой главной компоненты отражает особенности размещения железного оруденения (рис. 20). Положительные значения главной компоненты характеризуют 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. 20 четко прослеживается тенденция увеличения значений первой главной компоненты ниже скв. 21. Пробуренные буровые скважины 88 и 92, данные которых в расчетах не участвовали, подтвердили наличие руд с высоким содержанием железа.

В заключение целесообразно показать, что метод главных компонент может быть использован для решения классификационных задач. Для классификации полезно строить графики распределения признаков (или объектов) изучения в зависимости от значений главных компонент. Осями координат при этих построениях служат главные компоненты.

Используя данные предыдущего примера, был построен график значений первой и второй главных компонент (рис. 21). На построенном графике четко видно разделение руд Нерюдинского месторождения по содержанию железа.

Возможности метода главных компонент и сфера практического приложения в настоящее время далеко не изучены и не

свиты, песчано-аргиллитовые породы буркуглинской свиты и аргиллиты катской свиты. Широко развиты на месторождении интрузии траппов. В структуре месторождения главную роль играет крупный разлом широтного простирания, пересекающий вулканические трубки, картируемые по наличию витро- и литокластических туфов, прослеживающихся до глубин 800 м и более. Рудные тела месторождения, представленные метасоматическими залежами брекчиевидно-вкрапленных руд и магнетитовыми жилами, залегают в породах, существенно измененных пневмогидротермальными растворами и преобразованными 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 элементах и окислах.

По общепринятой схеме факторного анализа были получены парные корреляционные коэффициенты, главные компоненты, факторные нагрузки, дисперсии (табл. 22) 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 окислов.

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. 20). Положительные значения главной компоненты характеризуют 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. 20 четко прослеживается тенденция увеличения значений первой главной компоненты ниже скв. 21. Пробуренные буровые скважины 88 и 92, данные которых в расчетах не участвовали, подтвердили наличие руд с высоким содержанием железа.

В заключение целесообразно показать, что метод главных компонент может быть использован для решения классификационных задач. Для классификации полезно строить графики распределения признаков (или объектов) изучения в зависимости от значений главных компонент. Осями координат при этих построениях служат главные компоненты.

Используя данные предыдущего примера, был построен график значений первой и второй главных компонент (рис. 21). На построенном графике четко видно разделение руд Нерюдинского месторождения по содержанию железа.

Возможности метода главных компонент и сфера практического приложения в настоящее время далеко не изучены и не

Матрица парных коэффициентов корреляции и факторных нагрузок*

Признаки	Fe ₂ O ₃	FeO	SiO ₂	Al ₂ O ₃	CaO	MgO	MnO	TiO ₂	P	S	Факторные нагрузки			
											1-я ком- понента	2-я ком- понента	3-я ком- понента	
Fe _{общ}	1,00	0,63	0,80	-0,75	-0,80	-0,06	0,33	-0,58	-0,15	0,30	-0,08	0,90	0,02	0,05
Fe ₂ O ₃		1,00	0,64	-0,72	-0,58	-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
SiO ₂				1,00	0,92	0,28	0,46	-0,52	0,77	0,17	-0,02	0,14	-0,78	0,11
Al ₂ O ₃					1,00	0,17	0,37	-0,45	0,68	0,06	-0,14	0,18	-0,59	0,09
CaO						1,00	-0,46	0,01	0,14	-0,02	0,02	-0,08	-0,42	-0,37
MgO							1,00	-0,48	0,52	0,25	-0,50	0,09	-0,04	0,67
MnO								1,00	-0,46	-0,28	0,14	0,16	0,22	-0,23
TiO ₂									1,00	0,37	-0,14	0,07	-0,51	0,22
P										1,00	-0,08	-0,29	-0,06	0,04
S											1,00	0,02	0,03	-0,05
K ₂ O												1,00	0,00	0,04
Дисперсия факторов, %													42,00	18,70
														9,30

* Коэффициенты корреляции и факторные нагрузки отличаются от нуля с надежностью 0,95 при их абсолютных значениях $r > 0,25$.

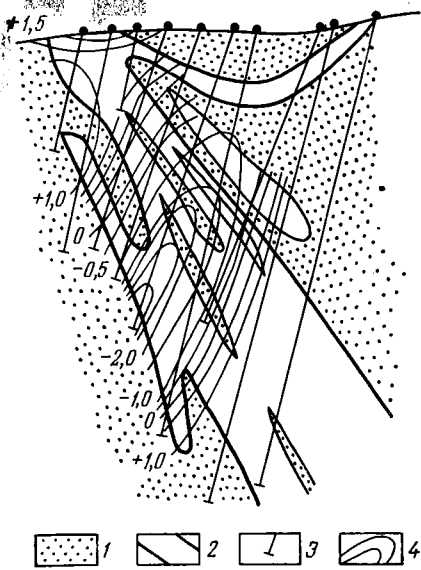


Рис. 20. Тренд-поверхности первой главной компоненты третьего порядка, отражающие размещение железного оруденения по разрезу Нерюндинского месторождения:

1 — вмещающие породы; 2 — рудное тело; 3 — буровые скважины; 4 — изолинии значений первой главной компоненты в усл. ед.

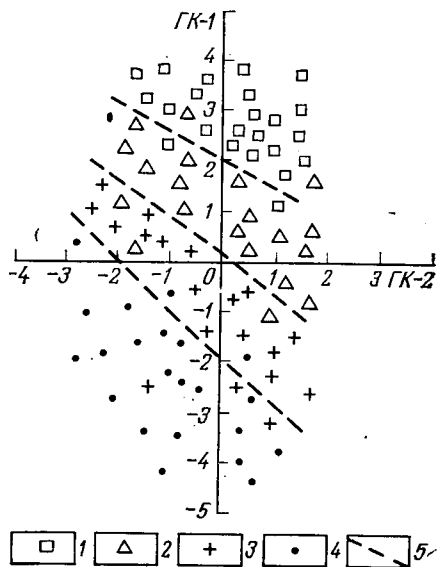


Рис. 21. Распределение типов железных руд по содержанию железа в зависимости от значений первой и второй главных компонент. Содержание железа:

1 — более 30%; 2 — 25—30%; 3 — 20—25%; 4 — менее 20%; 5 — границы полей развития различных типов руд

исчерпываются приведенными примерами. Применение этих методов для решения задач поисков и разведки месторождений только начинается.

Математический аппарат многомерного статистического анализа изложен в книгах Т. Андерсена [1], У. Крамбейна и Ф. Грейбилла [27], Д. Лоули и А. Максвелла [31]. Примеры практического использования многомерной статистической модели приведены в работах Р. И. Дубова [17, 18], Дж. Харбуха и Г. Бонэм-Картера [49].

§ 1. ГЕОЛОГИЧЕСКИЕ ОБЪЕКТЫ КАК ПОЛЯ ПРОСТРАНСТВЕННЫХ ПЕРЕМЕННЫХ

При изучении изменчивости свойств геологических образований горных пород, минеральных ассоциаций или полезных ископаемых геолога интересуют две стороны изменчивости — ее интенсивность и характер.

Интенсивность изменчивости выражается размахом (уровнем) колебаний наблюдаемых значений признака в данной выборке, а ее характер — особенностями пространственного изменения признака в изучаемом объеме недр.

Для целей количественного описания интенсивности изменчивости используется аппарат математической статистики случайных величин. Однако статистическая модель не обеспечивает получения данных о характере изменчивости, поскольку любой статистический показатель отражает только средний уровень изменчивости изучаемого свойства, но не учитывает пространственного размещения точек наблюдений. Для иллюстрации этого положения рассмотрим графики, приведенные на рис. 22 и 23. Наблюдаемая изменчивость признака на обоих графиках обладает одинаковой интенсивностью (коэффициент вариации в обоих случаях равен 0,35), однако по характеру изменчивости оба графика несопоставимы. Кроме того, статистические показатели объективно характеризуют уровень наблюдаемой изменчивости признака только в тех случаях, когда выборочные данные представляют собой совокупность независимых случайных величин. Если же между значениями признака в смежных пунктах наблюдений проявятся корреляционные связи, то статистические характеристики изменчивости позволяют определить только верхние пределы возможных погрешностей, а для оценки реальных погрешностей средних значений потребуются введение поправок «за связь».

Для изучения характера изменчивости признаков используются горно-геометрические и аналитические методы моделирования.

В обоих методах свойства геологических образований рассматриваются как поля пространственных переменных, для которых известны не только значения изучаемого свойства, но и

его пространственные координаты. При изучении свойств одного геологического объекта поля, как правило, бывают непрерывными, а при изучении закономерностей, характеризующих взаимное расположение нескольких объектов в пространстве, — дискретными.

Геометрические и аналитические методы моделирования геологических полей способствуют объективному выделению и количественному описанию тенденций, наблюдаемых в изменении

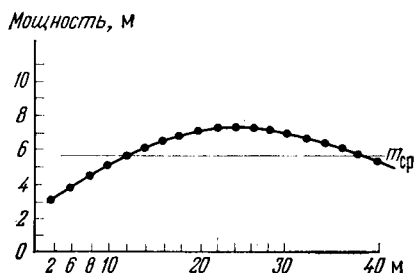
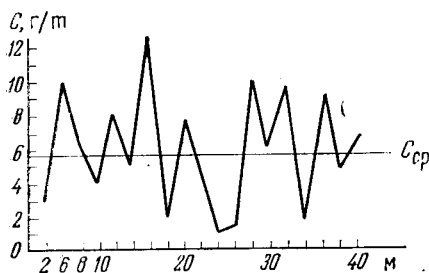


Рис. 22. Пример случайного изменения содержания золота по штреку:
 $C_{ср}$ — среднее содержание золота

Рис. 23. Пример неслучайного изменения мощности залежи:
 $m_{ср}$ — среднее значение мощности залежи

свойств геологических образований, а в ряде случаев позволяют выявить новые, не известные ранее закономерности. Для целей моделирования при этом используются результаты геологического картирования, геохимических и шлиховых съемок, геофизических наблюдений, геологоразведочных работ и т. д. Выявляемые закономерности используются для решения многочисленных проблем и задач. Пространственные закономерности изменения гранулометрического и минерального состава терригенных отложений позволяют установить направление сноса обломочного материала и восстановить палеогеографическую обстановку периода формирования толщ осадочных пород. По закономерным изменениям концентрации различных минералов выявляется зональность интрузивных массивов и месторождений полезных ископаемых. Пространственные закономерности изменения геофизических полей широко используются при геологическом картировании и поисках полезных ископаемых. Выявление закономерностей в изменении параметров рудных тел на ранних стадиях изучения позволяет более обоснованно оценивать не затронутые разведкой фланги и глубокие горизонты месторождений, а также способствует правильному выбору размеров сети наблюдений при продолжении разведочных работ. Математическое моделирование геохимических и геофизических полей позволяет более надежно выяв-

лять аномалии, перспективные на обнаружение месторождений полезных ископаемых.

С помощью моделирования дискретных геологических полей устанавливаются закономерности расположения месторождений и рудопоявлений определенного генетического типа относительно интрузий, тектонических нарушений или других элементов геологического строения.

§ 2. ГОРНО-ГЕОМЕТРИЧЕСКИЕ МОДЕЛИ

Основоположником методов горно-геометрического моделирования является выдающийся отечественный ученый П. К. Соболевский. В основу геометрической модели П. К. Соболевского положено представление о функциональных связях наблюдаемых свойств с пространственными координатами. Им сформулированы принципы горно-геометрического моделирования тел полезных ископаемых:

1) если значения изучаемого признака в каждой точке тела полезного ископаемого выразить соответствующим числом a , то этот параметр может рассматриваться как функция координат пространства $a=f(x, y, z)$, а тело полезного ископаемого — как некоторое геохимическое поле, выраженное множеством точек пространства;

2) функция $a=f(x, y, z)$ должна удовлетворять условиям конечности, однозначности, непрерывности и плавности.

Из существа геометрической модели П. К. Соболевского следует, что если известна функция $a=f(X, Y, Z)$, то можно установить значения параметра a в любой наперед заданной точке тела полезного ископаемого или любого другого геологического тела.

Геометрическое моделирование числовых значений любого свойства производится графическим путем с помощью изолиний. На планах и разрезах изолинии образуют топографические поверхности и могут рассматриваться как графические выражения функций типа $a=f(x, y)$. Аналитическое выражение таких функций практически невозможно вследствие их исключительной сложности. Поэтому П. К. Соболевским разработан оригинальный и логически стройный математический аппарат различных операций с топографическими поверхностями (топофункциями), с помощью которых можно производить с ними любые арифметические и алгебраические действия.

Графические модели самых различных свойств природных геологических тел широко используются в структурной геологии, в геологии полезных ископаемых, в рудничной геологии и в методике поисков и разведки полезных ископаемых. Методы горно-геометрического моделирования изучаются в курсе геометризации недр. На принципах П. К. Соболевского были разработаны и различные аналитические методы описания измен-

чивости, 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 содержания полезных компонентов в рудных телах (рис. 24), карты геохимических и геофизических полей

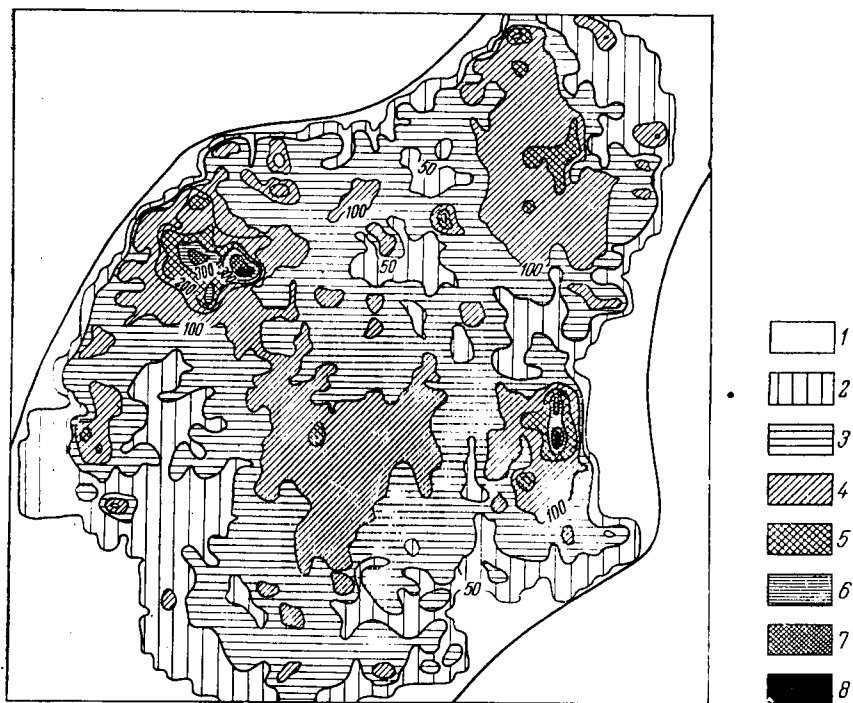


Рис. 24. План изосодержаний молибдена в условных единицах по уступам карьера:

1 — <10; 2 — 10—50; 3 — 50—100; 4 — 100—200; 5 — 200—300; 6 — 300—400;
7 — 400—500; 8 — >500

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 выдержанными в пространстве свойствами. К таким объектам относятся пласты осадочных пород, границы интрузивных образований, рудные тела с простой морфологией и относительно равномерным характером оруденения и т. п.

§ 3. МОДЕЛЬ П. Л. КАЛЛИСТОВА

Методы, описывающие закономерности пространственного изменения признака, так же как и статистические методы, не обеспечивают возможности объективной оценки изменчивости наблюдаемых свойств. Если статистические методы не учитывают влияния плавных, закономерных изменений, то геометрические методы игнорируют влияние многочисленных случайных отклонений по отдельным пунктам наблюдений. Поэтому в 1956 г. П. Л. Каллистовым было предложено различать две составляющих наблюдаемой изменчивости свойств — случайную и закономерную [24].

При случайной изменчивости значения признака, наблюдаемые в различных, в том числе и 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.

Примером почти чистой случайной изменчивости могут служить беспорядочные колебания содержания золота по данным опробования одного из жильных месторождений (см. рис. 22).

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.

Неслучайная изменчивость свойств реальных геологических тел полезных ископаемых обычно настолько сложна, что выразить ее аналитически как функцию координат пространства не удастся. 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:

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).

Примером неслучайной изменчивости может служить изменение мощности линзообразной залежи полезного ископаемого (см. рис. 23). 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 вида, а для характеристики случайной изменчивости используется коэффициент вариации, вычисленный через отклонения каждого частного значения от скользящей средней, т. е. от уровня неслучайной изменчивости (рис. 25).

Методика П. Л. Каллистова может применяться для решения аналогичной задачи не по профилям, а по площади. Для

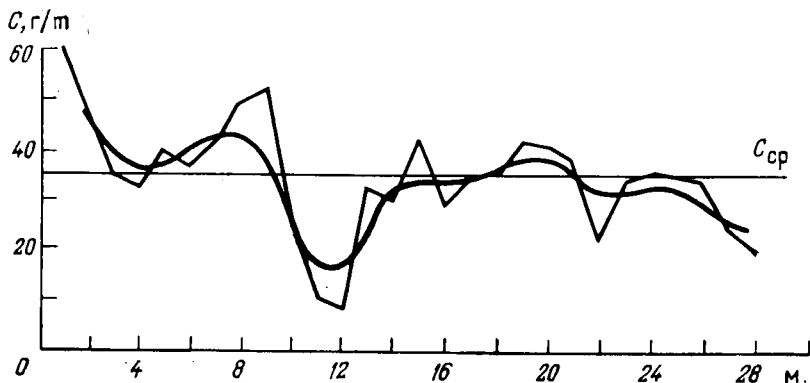


Рис. 25. Изменение содержания золота по штреку. По П. Л. Каллистову

этого по точкам среднеарифметических значений признака, рассчитанным для центров разведочных ячеек путем двух- или трехкратного сглаживания, проводятся изолинии (рис. 26), 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 скорее описательный прием, так как получаемая скользящая средняя маскирует и сдвигает наблюдаемые точки максимальных и минимальных значений признака. При сглаживании ряда в него всегда вносится некоторая ложная взаимозависимость между отдельными членами, обусловленная техникой процесса сглаживания. Кроме того, характер скользящей средней зависит от

количества проб в скользящем окне, а также от количества последовательных сглаживаний, причем число проб, участвующих в образовании каждого сглаженного значения, растет по мере увеличения числа сглаживаний. Значения коэффициентов вариации случайной составляющей, вычисленные при последовательном сглаживании от скользящих средних, зависят не только от свойства исследуемого ряда, но также от способа сгла-

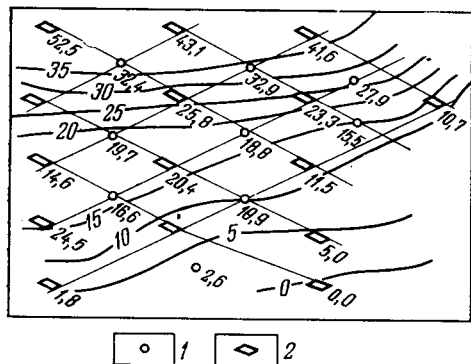


Рис. 26. Сглаживание значений признака по площади.

По А. А. Петрову.

1 — скважины; 2 — шурфы

живания и не могут поэтому использоваться в качестве объективных критериев оценки доли случайной составляющей.

§ 4. ТРЕНД-АНАЛИЗ

Идеи, заложенные в моделях П. К. Соболевского и П. Л. Каллистова, получили дальнейшее развитие в так называемых методах тренд-анализа. В этих методах свойство геологического объекта описывается случайным полем непрерывной скалярной переменной u , причем

$$u = f(x, y) + \varepsilon, \quad (IV.1)$$

где $f(x, y)$ — функция координат пространства (тренд);
 ε — случайная переменная.

С помощью тренд-анализа решаются задачи трех типов:

- проверяются гипотезы о наличии каких-либо закономерностей в пространственной изменчивости свойств геологических объектов;

- выделяются и количественно описываются региональные закономерности геологических полей (поверхности тренда), отражающие наиболее общие особенности геологического строения изучаемого объекта;

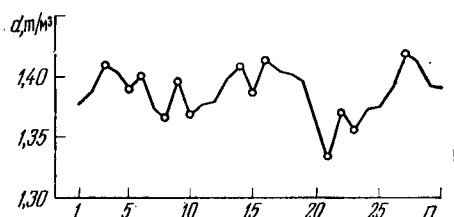
- выделяются локальные «аномальные» участки на фоне общих (фоновых) закономерностей.

Решение задач первого типа основано на сравнении свойств наблюдаемых полей или упорядоченных последовательностей

со свойствами полей или последовательностей, в которых закономерная составляющая (тренд) заведомо отсутствует. Для проверки гипотезы о наличии тренда в упорядоченных последовательностях, например, в результатах замеров определенного свойства геологического объекта по профилю, используются способы смены знака и количества скачков.

Точкой смены знака в упорядоченной последовательности называется такой элемент последовательности, в котором знак приращения изменяется на противоположный — убывание на возрастание или наоборот (рис. 27).

Рис. 27. Изменение объемной массы угля по простиранию пласта в пределах шахтного поля. Кружком обозначены точки смены знака.
 n — расстояние между скважинами



Число точек смены знака в случайных последовательностях зависит только от общего количества элементов последовательности N . Если $N > 10$, статистическое распределение числа точек смены знака близко к нормальному с математическим ожиданием

$$M(t) = \frac{2N - 4}{3} \quad (IV.2)$$

и дисперсией

$$\sigma^2(t) = \frac{16N - 29}{90} \quad (IV.3)$$

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 IV.2. Поскольку в таблицах приводятся значения нормированной функции нормального распределения, разницу между фактическим и теоретическим числом точек смены знака следует разделить на $\sqrt{\sigma^2(t)}$ для получения вероятностного критерия

$$Z = \frac{t - M(t)}{\sqrt{\sigma^2(t)}} \quad (IV.4)$$

В случайной последовательности значения t и $M(t)$ не должны существенно отличаться, следовательно, вероятность больших по абсолютной величине значений критерия Z будет ма-

ла. Так как значения критерия Z распределены нормально с параметрами 0 и 1, то по его величине с помощью таблиц нормированного нормального распределения [3, 52] можно оценить вероятность полученного по исследуемому ряду отклонения фактического числа точек смены знака от теоретического при условии, что ряд случаен (прил. 4, 5). Если эта вероятность мала (например, меньше 0,05), то гипотеза о случайном 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. 27). Математическое ожидание числа точек смены знака для $N=30$ будет

$$M(t) = \frac{2 \cdot 30 - 4}{3} \approx 18,7,$$

а дисперсия $\sigma^2(t) = \frac{16 \cdot 30 - 29}{90} \approx 0,5$. Фактическое число точек смены знака по профилю 13, следовательно, значение критерия Z составит

$$Z = \frac{13 - 18,7}{5} \approx -2,55,$$

что соответствует вероятности 0,0054 (прил. 4). Вероятность полученного отклонения фактического числа точек смены знака от теоретического для случайного ряда настолько мала, что гипотезу о случайном 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 последовательность состоит из двух типов элементов, которые условно можно обозначить знаками $+$ и $-$. 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

$$M(u) = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad (IV.5)$$

и дисперсией

$$\sigma(u) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}. \quad (IV.6)$$

Как и в предыдущем способе, фактическое значение числа скачков (u) сравнивается с теоретическим по исследуемому ряду $M(u)$ по критерию

$$Z = \frac{u - M(u)}{\sqrt{\sigma^2(u)}}. \quad (\text{IV.7})$$

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 игнимбригов андезитового состава по профилю вкрест простирания потока вулкана (рис. 28). 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 скачков $u = 6$.

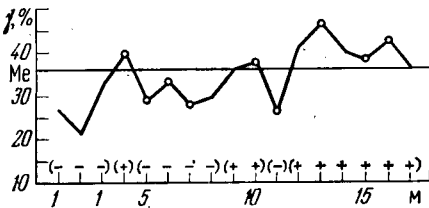


Рис. 28. Изменение общей пористости (γ) четвертичных игнимбригов вкрест простирания потока вулкана.

Вычислим математическое ожидание и дисперсию количества скачков для случайного ряда при $n_1 = 9$ и $n_2 = 8$:

$$M(u) = \frac{2 \cdot 9 \cdot 8}{9 + 8} + 1 \approx 9,5;$$

$$\sigma(u) = \frac{2 \cdot 9 \cdot 8 (2 \cdot 9 \cdot 8 - 9 - 8)}{(9 + 8)^2 (9 + 8 - 1)} \approx 4.$$

Значение критерия Z при этом составит

$$Z = \frac{6 - 9,5}{4} = 1,75,$$

что соответствует вероятности 0,04. Таким образом, при доверительной вероятности 0,95 гипотезу о случайном 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 «скачков» 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.

Необходимость выделения региональных закономерностей геологических полей с помощью математических методов возникает в тех случаях, когда в изменчивости исследуемого объекта наблюдаются значительные по амплитуде локальные флуктуации, затушевывающие общий характер закономерных изменений.

Способы выделения закономерной составляющей (поверхности тренда) основаны на замене исходных данных производными от них величинами либо на аппроксимации эмпирических данных единой функцией координат пространства.

Простейшим видом преобразования исходных данных является рассмотренное выше невзвешенное осреднение (способ скользящего окна). Для сглаживания используют также линейные преобразования вида

$$f(x, y) = v \sum k(\xi, \eta) u(\xi, \eta), \quad (IV.8)$$

где x, y — координаты центра площадки осреднения;

v — нормировочный множитель;

$k(\xi, \eta)$ — коэффициенты, зависящие от координат точки задания сглаживаемого поля в пределах площадки трансформации;

$u(\xi, \eta)$ — значения исследуемой величины в пределах площадки трансформации.

Наибольшее распространение получили сглаживающие преобразования Вулхауса, Спенсера, Шеппарда. В формуле Вулхауса используется 15 последовательных наблюдений, по семи наблюдений в каждую сторону от величины u_0 , которую необходимо сгладить

$$u'_0 = \frac{1}{125} [25u_0 + 24(u_1 + u_{-1}) + 21(u_2 + u_{-2}) + 7(u_3 + u_{-3}) + 3(u_4 + u_{-4}) - 2(u_6 + u_{-6}) - 3(u_7 + u_{-7})]. \quad (IV.9)$$

В формуле Спенсера используется по 10 значений с каждой стороны

$$u_0' = \frac{1}{350} [60u_0 + 57(u_1 + u_{-1}) + 47(u_2 + u_{-2}) + 33(u_3 + u_{-3}) + \\ + 18(u_4 + u_{-4}) + 6(u_5 + u_{-5}) - 2(u_6 + u_{-6}) - 5(u_7 + u_{-7}) - \\ - 5(u_8 + u_{-8}) - 3(u_9 + u_{-9}) - (u_{10} + u_{-10})]. \quad (\text{IV.10})$$

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 [6], а графики или поля изменения коэффициентов корреляции между различными физическими свойствами помогают интерпретировать результаты геофизических работ при геологическом картировании.

Пример. Молибденовое оруденение скарнового и грейзенового типа приурочено к лейкократовым гранитам, над которыми отмечается незначительное повышение удельного электрического сопротивления (ρ_k) и понижение уровня гравитационного поля (Δg). Однако выделение рудомещающего интрузива по какому-либо одному признаку оказывается затруднительным (рис. 29). Между значениями ρ_k и Δg были вычислены коэффициенты кор-

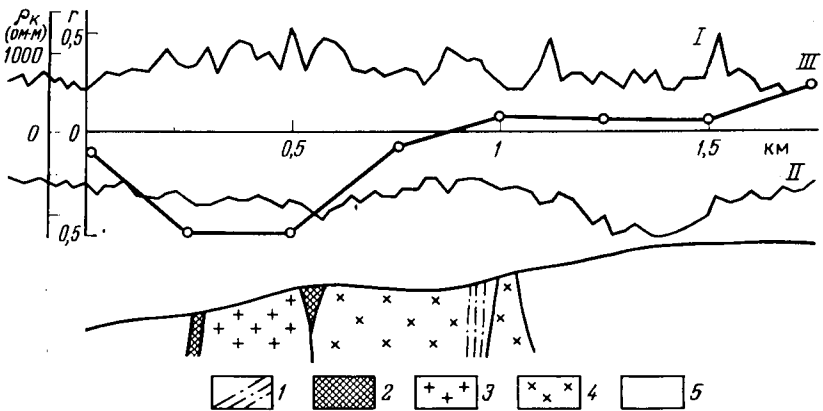


Рис. 29. График геофизических наблюдений и коэффициенты корреляции над рудопроявлением Чемуртай. По Л. Л. Ляхову и Г. А. Соловьеву:
 I — ρ_k (удельное электрическое сопротивление); II — Δg (уровень гравитационного поля); III — коэффициент корреляции. Прочие обозначения: 1 — зона разлома; 2 — скарны; 3 — лейкократовые граниты; 4 — микродиориты и габро-диориты; 5 — песчаники и известняки

реляции по участкам профиля протяженностью 500 м. По графику изменения коэффициентов корреляции рудоносный интрузив четко выделяется пониженными значениями r [33].

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 обуславливает корреляцию между соседними значениями [15];

— 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.

Выделение региональных закономерностей путем аппроксимации эмпирических данных функцией координат пространства связано с довольно сложными вычислениями, обычно требующими применения ЭЦВМ. 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 исходные данные записываются в виде матрицы порядка $r \times c$, где r — количество профилей наблюдений, а c — число наблюдений в каждом профиле. Функция $f(x, y)$, описывающая тренд, выражается через ортогональные полиномы в следующем виде:

$$f(x, y) = \beta_{00} + \beta_{01} \xi_1^1(x) + \beta_{01} \xi_1^1(y) + \dots + \beta_{pq} \xi_p^1(x) \xi_q^1(y), \quad (\text{IV.11})$$

где β — полиномиальные коэффициенты;

$\xi_p^1(x)$ — значения ортогональных полиномов степени p , соответствующие направлению X ;

$\xi_q^1(y)$ — аналогичные значения для степени q направления Y ;

β_{00} — среднее арифметическое изучаемого признака;

β_{01} и β_{10} — линейные коэффициенты направления X и Y ;

β_{02} и β_{20} — аналоги квадратов двух аргументов.

При построении поверхностей тренда при неравномерном расположении результатов наблюдений используются методы

матричной алгебры. Вручную могут быть рассчитаны только полиномы первого порядка, а для построения аппроксимирующих поверхностей более высокого порядка необходимо применение ЭЦВМ. Поэтому в геологической практике неравномерная сеть наблюдений часто трансформируется в равномерную путем осреднения результатов наблюдений в пределах элементарных площадок либо путем отнесения к узлам равномерной сети ближайших замеров.

Полиномы первой степени описывают общую для всего исследуемого участка тенденцию к возрастанию или убыванию изучаемого признака по определенному направлению (линейный тренд), а полиномы более высоких степеней отражают закономерности более высоких порядков.

- *Пример.* При изучении изменения средних размеров зерен циркона во фракциях тяжелых минералов из верхней части галечникового слоя Ашдаун (юго-восток Англии) [27] выяснилось следующее. По характеру поверхности тренда первого порядка (рис. 30, б) можно заключить, что циркон

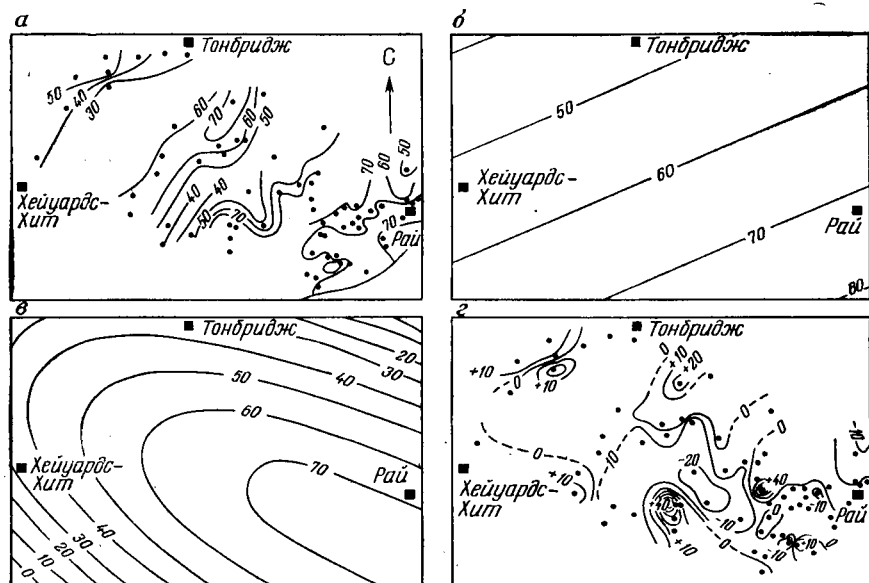


Рис. 30. Изменение размеров зерен циркона (в микронах) в галечниковом слое. По У. Крамбейну:

а — карта результатов наблюдений; б — плоскость тренда; в — поверхность тренда второго порядка; г — карта отклонений от поверхности второго порядка

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 (рис. 30, в).

Выбор степени аппроксимирующего полинома и оценка значимости выявленных закономерностей могут осуществляться с помощью дисперсионного анализа (см. гл. I). 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. 27.

Т а б л и ц а 27

Дисперсионный анализ тренда размеров зерен циркона

Тип изменчивости	Сумма квадратов	Число степеней свободы	Средний квадрат	Критерий Фишера
Линейная модель	4 982,01	2	2 491,00	11,89
Отклонения от линейной модели	13 403,93	64	209,44	
Квадратичная модель	1 266,19	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		

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.

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 изучае-

мых 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 выявлении 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 (рис. 30, 2). Подобные карты могут быть построены и для случайной составляющей после сглаживания исходных данных скользящим окном или с помощью преобразования какого-либо другого вида.

В тех случаях, 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 характера изменчивости предложен целый ряд производных характеристик. Для оценки степени «изрезанности» поля используют коэффициент «аномальности» [6], который определяется по относительной разнице между длиной графика изучаемого свойства на участке профиля постоянной длины и длиной этого участка. На рис. 31 повышенные значения коэффициента аномальности магнитного поля четко фиксируют крупные разрывные структуры.

При прослеживании геологических границ по геофизическим данным применяются горизонтальные градиенты или энтропия поля [6]. 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 изоли-

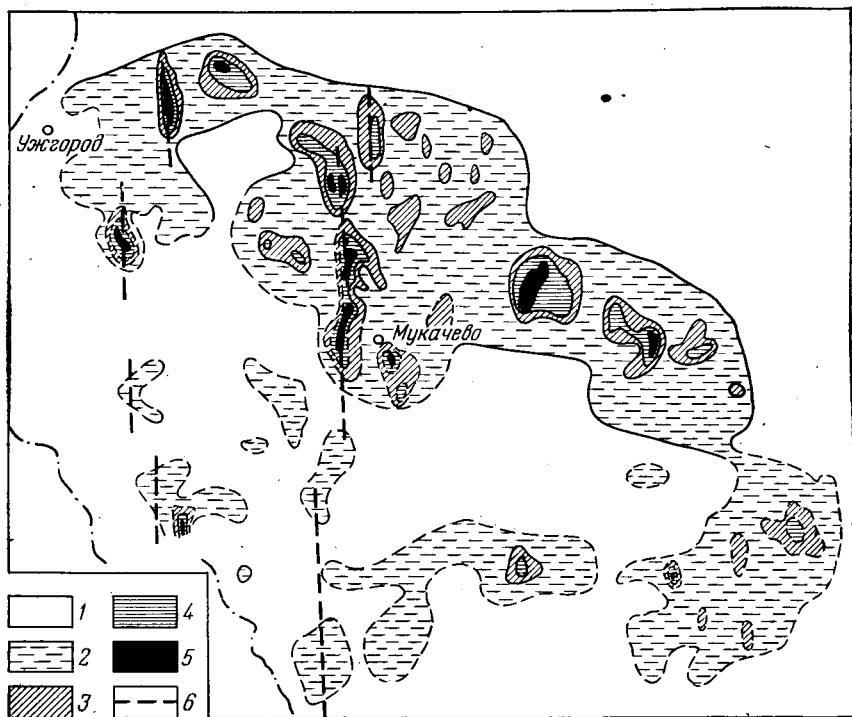


Рис. 31. Схема изменчивости аномалийного магнитного поля Закарпатья.
По Г. И. Кругляковой.

1 — невозмущенное поле; возмущенность поля (в %): 2 — 0—20, 3 — 20—40, 4 — 40—60, 5 — >60; 6 — разломы древнего заложения

ниями, попадающими в пределы элементарной площадки трансформации.

На рис. 32 видна приуроченность ореолов свинца в дельювии к областям повышенных значений горизонтальных градиентов и энтропии гравитационного поля, а наиболее крупный из ореолов приурочен к пересечению двух зон повышенной энтропии.

§ 5. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ПОЛЯ

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 пренебрежимо малы по сравнению с исследуемой территорией.

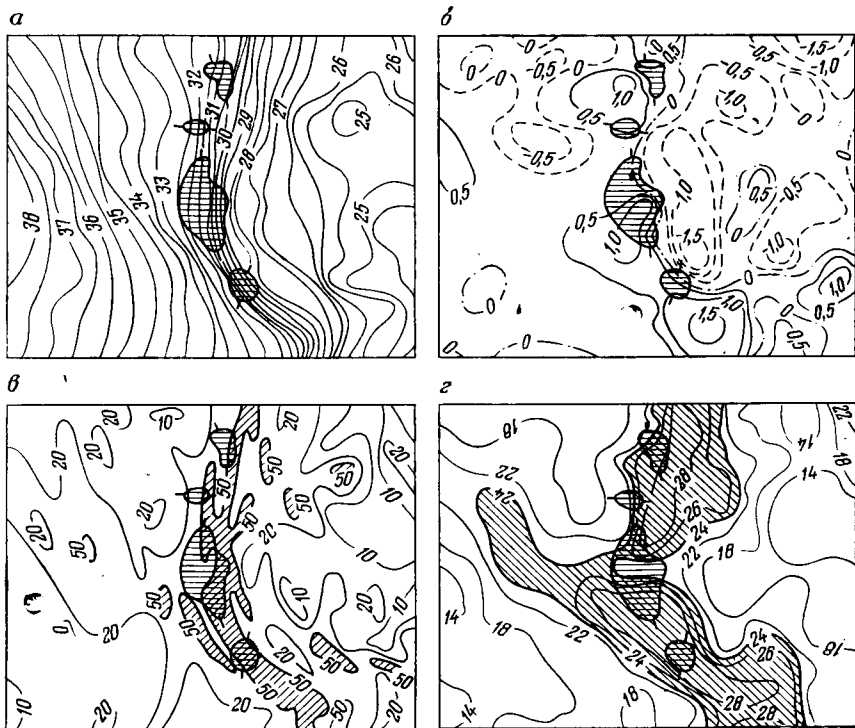


Рис. 32. Различные способы описания гравитационного поля. По Н. Н. Боровко.

а — гравитационное поле в изолиниях; б — локальные аномалии: в — горизонтальные градиенты; г — карта энтропии.
Косая штриховка — аномальные значения горизонтальных градиентов и энтропии, горизонтальная — ореолы свинца в элкви

Выделяют общую задачу, когда проверяется гипотеза о случайном расположении исследуемых объектов, и локальную, когда требуется выделить области относительного сгущения и разряжения точечных объектов [6].

Для решения общей задачи при изучении распределения точечных объектов на плоскости исследуемая территория разбивается на квадратные ячейки постоянного размера. При этом часть ячеек (p) 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 случайном расположении точечных объектов вероятность того, что новый квадрат окажется пустым, будет равна

$$p_N = (1 - p)^N. \quad (\text{IV.12})$$

Перебираются все возможные варианты объединения исходных ячеек в квадраты по $N=4, 9, 16$ и т. д. штук и находятся доли пустых ячеек p'_N для каждого N . Повышенная доля p'_N пустых ячеек по сравнению с теоретической вероятностью p_N указывает на группировку «роение» объектов на отдельных участках, а пониженная — на регулярность их расположения на плоскости.

Пример. Проверяется гипотеза о случайном расположении медных месторождений в западной части штата Аризона [6]. Для этого исследуемая площадь была разделена на исходные квадраты площадью 250 км^2 , составляющие 0,25 часть от квадратных ячеек, изображенных на рис. 33. График изменения величины для случайного расположения точек в зависимости от размеров ячеек приведен на рис. 34. Фактическое значение числа

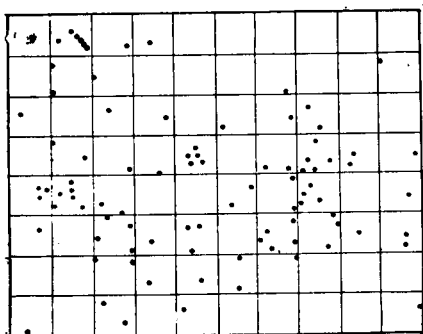


Рис. 33. Схема расположения медных месторождений на территории западной части штата Аризона

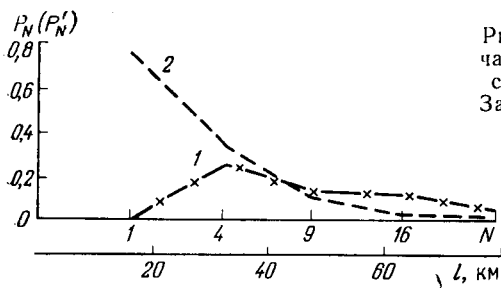


Рис. 34. Проверка гипотезы о случайном расположении медных месторождений в штате Аризона. Зависимость доли пустых ячеек от длины стороны квадрата: 1 — фактическая, 2 — теоретическая

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

$$v = \frac{m}{p} - n, \quad (\text{IV.13})$$

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

$$p_1 = \sum_{i=m}^n C_n^i p^i (1-p)^{n-i}. \quad (\text{IV.14})$$

Вероятность случайного попадания не более чем m объектов при тех же условиях может быть определена по формуле

$$p_2 = \sum_{i=0}^m C_n^i p^i (1-p)^{n-i}. \quad (\text{IV.15})$$

Величины p_1 и p_2 могут быть найдены для заданных n , m и p по специальным таблицам [58]. 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. 35). Для расчета избыточной плотности аномалий использовалась квадратная палетка с соотношением площадей квадратов $p = \frac{1}{9}$ (см. рис. 35). Расчет осуществлялся по формуле (IV. 13). Например, для точки, помеченной на рис. 35 крестом, при $m=2$, $n=3$ и $p = \frac{1}{9}$

$$v = 9 \cdot 2 - 3 = 15.$$

С помощью формулы (IV. 14) были установлены и околтурены участки сгущения аномалий, значимо отличающиеся от площадей со случайным расположением точечных объектов при доверительной вероятности 0,95 (см. рис. 35). Сопоставление изолиний избыточной плотности аномалий со схемой ландшафтного районирования данной территории (рис. 36) позволяет сделать следующие выводы. Площади с избыточной плотностью радиометрических аномалий 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.

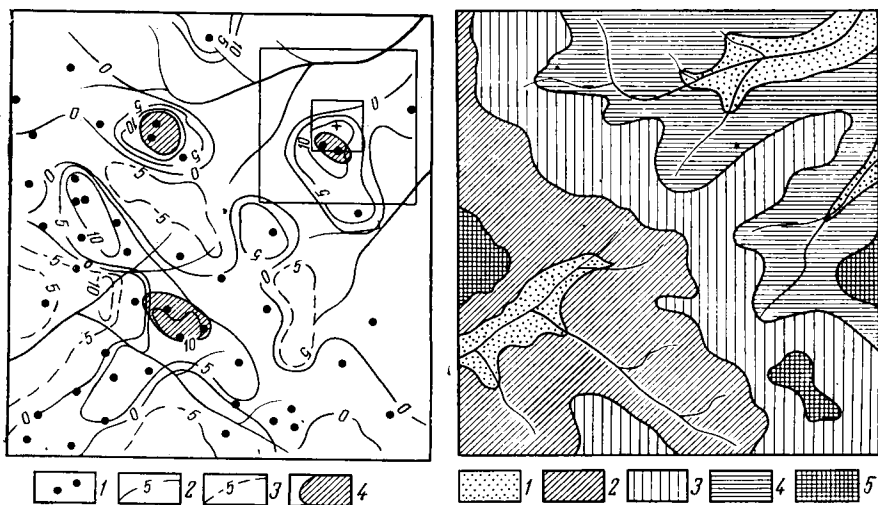


Рис. 35. Расположение радиометрических аномалий в горно-таежном районе: 1 — радиометрические аномалии; 2 — изолинии избыточной плотности аномалий; 3 — изолинии недостаточной плотности аномалий; 4 — участки сгущения, значительно отличающиеся от площадей со случайным расположением аномалий

Рис. 36. Схема морфологически-ландшафтного районирования горно-таежного района. Рельеф:

1 — аллювиальный; 2 — крутых склонов; 3 — склонов средней крутизны; 4 — водоразделов; 5 — аккумулятивный

О зависимости расположения n точечных объектов относительно объектов с конечной суммарной площадью, составляющей часть p от всей исследуемой площади, можно судить по нарушению равенства

$$\frac{m}{n} = p, \quad (\text{IV.16})$$

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 (IV.14).

Пример. В одном из районов Южного Урала медноколчеданные месторождения и рудопроявления приурочены к площадям развития кислых вулканитов (рис. 37). Эта тенденция настолько очевидна, что не нуждается в статистической проверке. Однако на некоторых участках развития благоприятных пород рудопроявления отсутствуют. Высказано предположение [20], что в пределах площадей развития кислых вулканитов размещение проявлений медноколчеданной минерализации определяется локальными аномалиями тектонической напряженности. Тектоническая напряженность оценивалась по произведению количества трещин на их суммарную длину в пределах скользящего окна площадью 4 км². Для проверки этой гипотезы примем за исследуемую площадь общую площадь развития кислых вулканитов, а в качестве площадных объектов — участки аномалий тектонической напряженности в пределах этих пород. Отношение суммарной площади

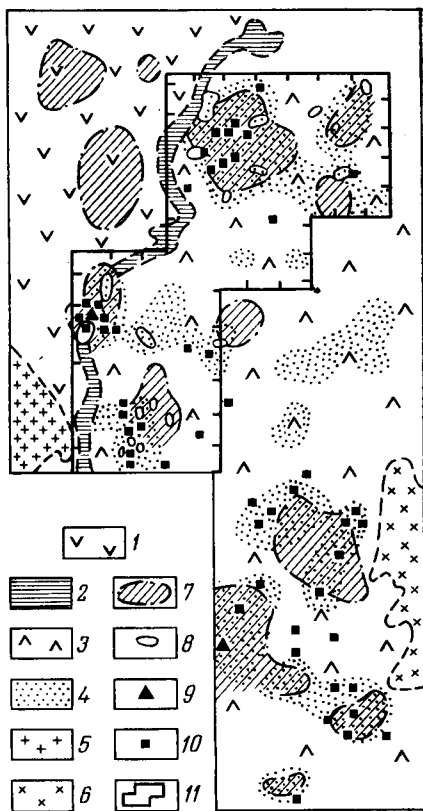
аномалий в пределах кислых вулканитов к общей площади этих пород (p) составляет 0,46, а отношение количества рудопроявлений и месторождений в кислых вулканитах в пределах аномалий (m) к общему их числу в данных породах (n) равно

$$\frac{m}{n} = \frac{24}{44} \approx 0,55.$$

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 случайном расположении точек минерализации относительно аномалий, определенная по формуле (IV. 14), составляет 0,36. Это не позволяет считать связь медноколчеданных рудопроявлений и месторождений с аномалиями тектонической напряженности доказанной.

Рис. 37. Расположение медноколчеданных 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 субвулканические тела; 9 — месторождения; 10 — рудопроявления; 11 — площадь детальных геофизических исследований

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 [6], У. Крамбейна и Ф. Грейбила [27]. Способы проверки гипотез о наличии тренда подробно описаны в книге Р. Л. Миллера и Д. С. Кана [36].

ГЛАВА V

МОДЕЛИ ТИПА СЛУЧАЙНЫХ ФУНКЦИЙ

§ 1. ПРИНЦИПЫ МОДЕЛИРОВАНИЯ СВОЙСТВ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ С ПОМОЩЬЮ СЛУЧАЙНЫХ ФУНКЦИЙ

При решении многих геологических задач возникает необходимость количественной оценки пространственной изменчивости изучаемых признаков как функции их природной изменчивости и расстояний между смежными пунктами наблюдений. Методы статистики случайных величин для этих целей мало пригодны, 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 прибрежно-морских 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, описывающих пространственную изменчивость самых различных геологических признаков в одномерном, двухмерном 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. 38). Каждая реализация отражает общие тенденции в изменении изучаемого свойства, но они несколько отличаются друг от друга. В данном примере это обусловлено аппаратур-

ными погрешностями, некоторым несовпадением точек замеров и другими случайными факторами.

Совокупность значений случайной функции $X(l)$ при любом фиксированном значении аргумента l называется *сечением* слу-

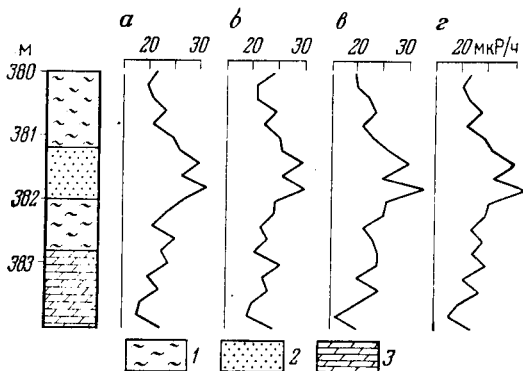


Рис. 38. Результаты гамма-каротажа (в мкР/ч) интервала по скважине:
 а — основной каротаж; б, в — контрольный; г — усредненная каротажная кривая
 1 — алевролиты; 2 — песчаники; 3 — известняки

чайной функции и является обычной случайной величиной (рис. 39).

Таким образом, случайная функция совмещает в себе черты случайной величины и функции. Если зафиксировать значение аргумента, она превращается в обычную случайную вели-

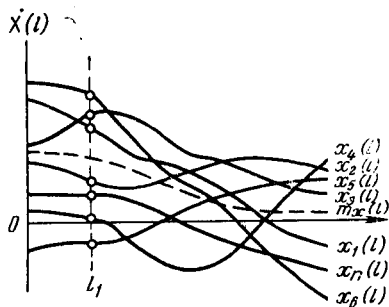


Рис. 39. Пример случайной функции

чину, а отдельная реализация может рассматриваться как обычная (неслучайная) функция.

Модель типа случайной функции так же, как и статистическая модель, 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 согла-

сия. При оперировании со случайными функциями широко используются основные понятия математической статистики о функциях и моментах распределения, уровнях значимости и доверительных вероятностях событий.

Основными характеристиками случайной функции являются ее математическое ожидание $M_x(l)$, дисперсия $D_x(l)$ и корреляционная функция $K_x(l, l')$. В отличие от числовых характеристик случайных величин, представляющих собой определенные числа, характеристики случайных функций представляют собой в общем случае не числа, а функции.

Математическим ожиданием случайной функции $X(l)$ называется неслучайная функция $M_x(l)$, которая при каждом значении аргумента l равна математическому ожиданию соответствующего сечения случайной функции. Оценка математического ожидания по ансамблю из N реализаций рассчитывается по формуле

$$\overline{x(l)} = \frac{\sum_{k=1}^N x_k(l)}{N} \quad (V.1)$$

и является геометрическим местом точек средних значений изучаемого свойства по сечениям. В приведенном выше примере оценкой математического ожидания является усредненная каротажная диаграмма, рассчитанная по трем реализациям (рис. 38, з).

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

$$D_x(l) = \frac{\sum_{k=1}^N [x_k(l) - \overline{x(l)}]^2}{N - 1} \quad (V.2)$$

Для описания основных свойств случайной функции математического ожидания и дисперсии недостаточно. Ни одна из этих характеристик не отражает особенностей внутренних связей между отдельными сечениями случайной функции. С этой целью используется специальная характеристика — корреляционная функция, которая характеризует 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.

$$K_x(l, l') = M[\overset{\circ}{X}(l') \overset{\circ}{X}(l'')]. \quad (V.3)$$

В этом выражении $\overset{\circ}{X}(l')$ и $\overset{\circ}{X}(l'')$ — значения центрированной случайной функции при значениях аргумента l' и l'' . Центрирование случайной функции осуществляется путем вычитания из значений случайной функции ее математического ожидания

$$\overset{\circ}{X}(l) = X(l) - M_x(l). \quad (V.4)$$

В частном случае, когда оба аргумента случайной функции совпадают, т. е. $l' = l''$, корреляционная функция $K_x(l', l'')$ обращается в дисперсию случайной функции

$$K_x(l', l'') = M\{[X(l')]^2\} = D_x(l'). \quad (V.5)$$

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, которая изменяется относительно некоторого ее среднего значения. Изменения математических ожиданий и дисперсий этих случайных величин можно описать неслучайными функциями координат центров элементарных участков и рассматривать как математическое ожидание и дисперсию случайной функции.

Упорядоченный ряд точечных значений признака, полученных путем опробования по заданному направлению 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.

Использование аппарата теории случайных функций для решения геологических задач сопряжено с рядом сложностей и ограничений. Главное из них заключается в том, что случайная функция подобно случайной величине, является совокупностью частных реализаций, в то время как результаты геологических наблюдений чаще всего представляют собой только 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 наблюдаемую 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 наблюдаемых значений признака.

§ 2. ИСПОЛЬЗОВАНИЕ МОДЕЛИ СТАЦИОНАРНОЙ СЛУЧАЙНОЙ ФУНКЦИИ 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, корреляционная 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

$$K_x(r) = \frac{1}{L-r} \int_0^{L-r} [f(x) - M_x][f(x+r) - M_x] dx, \quad (V.6)$$

где L — длина исследуемого профиля (участка);

r — расстояние между точками наблюдений, выраженное числом интервалов между ними;

$f(x)$ — переменная величина (например, значения содержания металла в пробах в ряду, который начинается значением x_1 и заканчивается значением $x_n - r$);

$f(x+r)$ — переменная величина (например, значение содержания металла в пробах в ряду, который начинается значением $x+r$ и заканчивается значением x_n);

M_x — среднее значение переменной величины $f(x)$ в интервале от 0 до L .

На практике часто пользуются корреляционной функцией, нормированной по дисперсии

$$\rho_x(r) = \frac{K_x(r)}{D_x}. \quad (V.7)$$

Функция $\rho_x(r)$ представляет собой коэффициент корреляции между сечениями случайной функции, разделенными интервалами r . Очевидно, что $\rho_x(0)$ равна единице. Другой, весьма удобной для расчетов характеристикой является структурная функция стационарной случайной функции, основанная на вычислении средних квадратов разностей значений признака в точках, отстоящих друг от друга на расстоянии r . Она определяется выражением

$$\gamma_x(r) = \frac{1}{L-r} \int_0^{L-r} [f(x+r) - f(x)]^2 dx. \quad (V.8)$$

Между структурной и корреляционной функциями существует почти строгая зависимость при r , заведомо меньшем, чем L :

$$\gamma_x(r) = 2[K_x(0) - K_x(r)]. \quad (V.9)$$

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. 40). Ее структурная функция возрастает в том же интервале значений r , а при расстоянии, большем предела корреляции, равна удвоенному значению статистической дисперсии исследуемого свойства (рис. 41).

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

$$K_x(r) = \frac{\sum_{i=1}^{n-r} (x_i - \bar{x})(x_{i+r} - \bar{x})}{(n-1-r)}; \quad (\text{V.10})$$

$$\gamma_x(r) = \frac{\sum_{i=1}^{n-r} (x_i - x_{i+r})}{(n-1-r)}, \quad (\text{V.11})$$

где i — порядковый номер замера в исследуемом ряду значений X ;

n — общее число замеров;

r — последовательно принимает значения 1, 2, 3, ..., $(n-1)$.

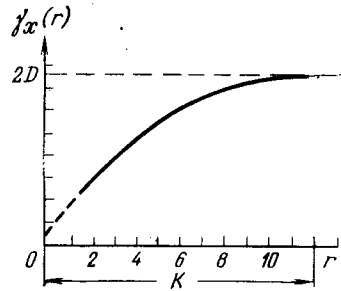
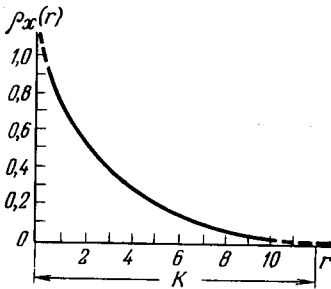


Рис. 40. Графическое изображение корреляционной функции

Рис. 41. Графическое изображение структурной функции

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. 38) приведен в табл. 28.

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 (рис. 42).

Каждая реализация эргодической случайной функции $X_1(l)$ обладает примерно одинаковыми средним значением, средней

Расчет первой точки нормированной корреляционной функции значений интенсивности гамма-излучения по скважине

i	x_i	$x_i - \bar{x}$	x_{i+1}	$x_{i+1} - \bar{x}$	$(x_i - \bar{x}) \times$ $\times (x_{i+1} - \bar{x})$	$(x_i - \bar{x})^2$
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	6,25
6	24,5	1,0	26,0	2,5	2,50	1,00
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	20,0	-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

$$\bar{x} = \frac{\sum \bar{x}_i}{n} = \frac{490,5}{21} = 23,5; S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{243,50}{20} = 12,175;$$

$$\rho_x(1) = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{(n-1+1) S^2} = \frac{135,25}{19 \cdot 12,175} = 0,585.$$

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

двух 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.) [32]. 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. 40. Наличие в пределах исследуемого поля крупных

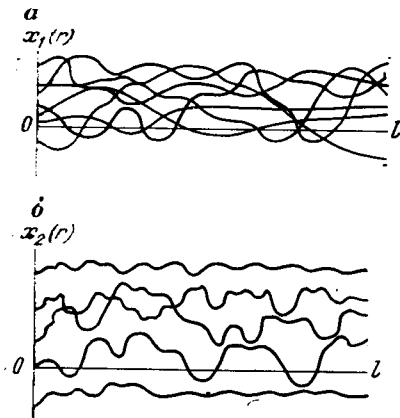


Рис. 42. Случайные функции:
а — эргодическая; б — неэргодическая

элементов неоднородности приводит 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 целью рассматриваются первые две-три точки нормированной корреляционной функции изучаемой последовательности и оценивается статистическая значимость их отличия от нуля по формуле (III.10).

Этот способ является более надежным по сравнению с рассмотренными в главе IV статистическими способами проверки гипотезы о наличии тренда, так как позволяет выявлять слабопроявленные закономерности.

Пример. Проверим гипотезу о наличии закономерности в изменении интенсивности гамма-излучения по скважине (см. табл. 28)

$$t = \frac{\rho_x(1)}{\sqrt{1 - \rho_x^2(1)}} \sqrt{n-2} = \frac{0,585}{\sqrt{1 - 0,585^2}} \sqrt{20-2} \approx 3,06.$$

Расчетное значение критерия Стьюдента превышает табличное для доверительной вероятности и числа степеней свободы 18 [3], следовательно, гипотеза о наличии закономерности в изучаемой последовательности может быть принята.

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 [32]. Для участков молодой океанической коры характерна упорядоченная линейная система гребней или разломов срединно-океанических хребтов, обуславливающая закономерные, периодические колебания АМП, которые отчетливо видны на корреляционных функциях (рис. 43). Вдали же от рифтов упорядоченность топографии океанического дна вуалируется 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. 44).

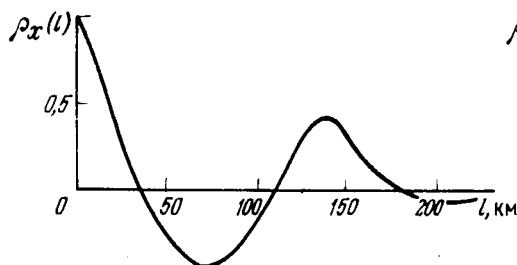


Рис. 43. Корреляционная функция аномального магнитного поля по профилю 23. По В. Н. Луговенко [32]

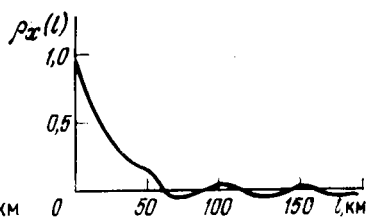


Рис. 44. Корреляционная функция аномального магнитного поля по профилю «Анабарский». По В. Н. Луговенко [32]

Эталонами и основой для классификации промышленных типов месторождений важнейших видов минерального сырья могут служить типизированные графики характеристик случайных функций важнейших свойств залежей полезных ископае-

мых, составленные по обобщенным экспериментальным данным.

По виду корреляционных функций, рассчитанных для различных интервалов разрезов толщ метаморфических пород, можно обоснованно проводить их расчленение на более мелкие стратиграфические подразделения [44].

3. Оценка предельных расстояний между точками наблюдений, при которых сохраняются корреляционные связи между значениями исследуемых свойств, 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 должны рассматриваться как совокупности случайных, независимых величин и 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 м и 135 м (рис. 45). Следовательно, показатель анизотропии данного свойства в плоскости залежи будет равен

$$A = \frac{K_2}{K_1} = \frac{135}{75} = 1,8.$$

Величина коэффициента анизотропии может быть использована в качестве критерия оптимальности выбранной геометрии сети наблюдений. Так,

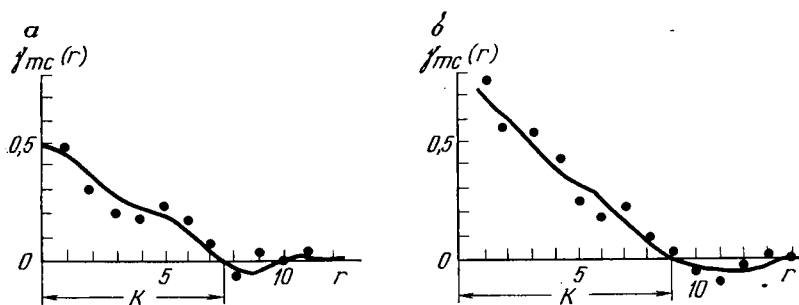


Рис. 45. Корреляционные функции линейных запасов металла рудной залежи: а — по простиранию ($r=10$ м); б — по падению ($r=15$ м)

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 75×135 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

$$\rho_x(r) = e^{-\alpha r}. \quad (V.12)$$

В этом случае для определения дисперсии случайной составляющей ($\sigma_{\text{случ}}^2$) можно использовать формулу [41]

$$\sigma_{\text{случ}}^2 = \sigma_0^2 [1 - \rho_x^2(r)], \quad (V.13)$$

где σ_0^2 — общая статистическая дисперсия изучаемого признака;

$\rho_x^2(r)$ — значение нормированной корреляционной функции для расстояния между точками наблюдения, равного r .

Дисперсия закономерной составляющей находится из равенства

$$\sigma_{\text{зак}}^2 = \sigma_0^2 - \sigma_{\text{случ}}^2. \quad (V.14)$$

При наличии в изменчивости изучаемого свойства закономерной составляющей в статистические формулы для расчета интервальных оценок целесообразно вводить поправки за связь. Для этого в формулу (I.34) вместо общей дисперсии подставляется дисперсия случайной составляющей, определенная по формуле (V.13).

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. 45, определим долю случайной составляющей в изменчивости линейного запаса по падению рудной 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 пере-сечениями, 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 3 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 останется 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 наблюдаемой изменчивости могут служить также показателями степени разведанности 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.

§ 3. МОДЕЛЬ ТИПА ПОЛИГАРМОНИЧЕСКОЙ СЛУЧАЙНОЙ ФУНКЦИИ

Многие геологические объекты отличаются периодическим характером изменчивости свойств. Периодические колебания установлены в изменчивости пористости нефтегазоносных осадочных толщ [11], минеральном составе метаморфических пород [44]. Определенная ритмичность отмечается также в расположении тектонических нарушений как в региональном, так и в локальном масштабе [32, 2]. При этом, однако, закономерные периодические колебания обычно осложняются нерегулярными случайными флуктуациями.

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

$$M_x(l)_e = A_0 + \sum_{k=1}^{k=v} A_k \cos(\omega_k l + \varphi_k), \quad (V.15)$$

где v — количество гармоник;
 A_k, ω_k, φ_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

$$X(r) = M_x(r) + h_x(r), \quad (V.16)$$

где $h_x(r)$ — 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 наблюдений $X(r)$, позволяющие усилить в преобразованном ряде $X'(r)$ роль периодической составляющей [45]. Наиболее приемлемым из них является метод, основанный на оценке спектральной плотности дисперсии $S_x(\omega)$, получаемой в результате разложения в ряд Фурье корреляционной функции

$$S_x(\omega) = \frac{2}{\pi} \int_0^{\infty} K_x(r) \cos \omega r dr. \quad (V.17)$$

При изучении изменений параметра дискретной сетью наблюдений спектральная плотность $S_x(\omega)$ заменяется линейчатым спектром амплитуд гармоник. Каждое значение линейчатого спектра вычисляется по формуле

$$d_k = \frac{1}{2} A_k^2. \quad (V.18)$$

Линейчатый спектр показывает, каким образом общая дисперсия признака распределяется между отдельными гармониками.

Спектр амплитуд случайной полигармонической последовательности вида (V.16), рассчитанный по ограниченному интервалу наблюдений, представляет собой результат наложения резких пиков, соответствующих определенным частотам гармоник, на относительно постоянные для всех частот значения d_k , обусловленные случайной составляющей (рис. 46, в). Количество таких пиков соответствует числу гармоник полигармонической составляющей, а их высота равна половине квадрата амплитуды гармоники данной частоты.

Величина периодов закономерных колебаний может быть определена по формуле

$$T_k = \frac{(N-1)r}{\omega_k}, \quad (V.19)$$

где N — общее количество замеров по профилю;

r — расстояние между соседними пунктами наблюдений.

Значения d_k , обусловленные случайной составляющей, имеют бета-распределение Фишера [45], плотность которого задается формулой

$$P(x) = (N - 1)(1 - x)^{N-2}, \quad (\text{V.20})$$

где N — общее число значений спектра.

Математическое ожидание и стандартное отклонение этого распределения соответственно равны

$$d_k' = \frac{1}{N}; \quad (\text{V.21})$$

$$\sigma_d = \sqrt{\frac{N-1}{N^2(N+1)}}. \quad (\text{V.22})$$

Используя эти формулы, можно с наперед заданной доверительной вероятностью проверить гипотезу о принадлежности тех или иных пиков спектра к случайным флуктуациям.

Пики, превышающие предельное значение (например, $d_k' + 3\sigma_d = d_k''$ при доверительной вероятности 99%), объясняются наличием гармоник и могут быть использованы для оценки

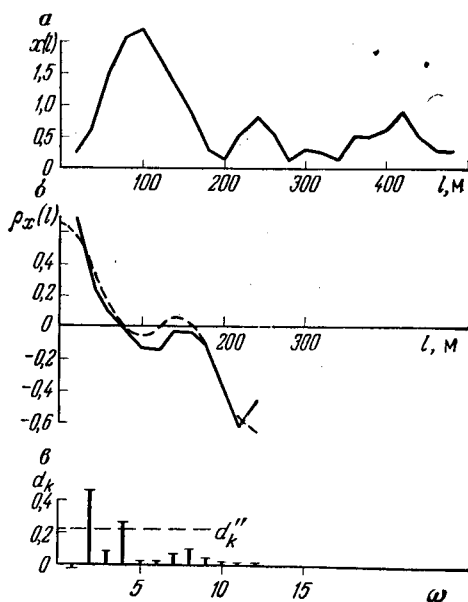


Рис. 46. Характеристики изменчивости линейного запаса железа вкострости простирания прибрежно-морской россыпи ильменит-магнетитовых песков:

a — изменение линейного запаса по профилю; b — корреляционная функция; c — спектр амплитуд

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

$$d_k' (\text{испр.}) = \frac{1 - \sum_1^m d_k (\text{ан})}{N - m}, \quad (\text{V.23})$$

где $\sum_1^m d_k(\text{ан})$ — сумма аномальных значений спектра;

m — количество значений спектра.

Периоды и амплитуды аномальных значений спектра служат основой для аппроксимации корреляционной функции и исходного ряда значений тригонометрическими полиномами.

Для расчета эмпирических спектров амплитуд на ЭВМ рекомендуется приближенная формула

$$d_k = \frac{2\rho(0)}{n} + \frac{4}{n} \sum_{r=1}^{\frac{n}{2}-1} \rho(r) \cos \frac{2\pi}{n} kr + \frac{2\rho(r=n/2)}{n} (-1)^k, \quad (\text{V.24})$$

где n — общее число равномерно размещенных наблюдений по всей реализации;

k — номер гармоники;

$\rho(r)$ — значения нормированной корреляционной функции.

Значения амплитуд для k , равного 0 и $n/2$, при этом делаются пополам.

Пример. Прибрежно-морская россыпь ильменит-магнетитовых песков (рис. 46, а) разведана скважинами и шурфами по сети 200×20 м [25]. По данным опробования шурфов и скважин были вычислены эмпирическая корреляционная функция (рис. 46, б) и спектр амплитуд (рис. 46, в) линейных запасов железа (произведений содержания металла на мощность песков). Фоновое значение спектра амплитуд равно

$$d_k' = \frac{1}{N} = \frac{1}{12} \approx 0,083,$$

а минимальное аномальное при доверительной вероятности 0,95 составит

$$d_k'' = 0,083 + 2 \sqrt{\frac{11}{12^2(12+1)}} \approx 0,235.$$

Таким образом, аномальными следует считать значения спектра на частотах 1 и 3, которые соответствуют колебаниям значений линейного запаса с периодами 480 и 160 м.

Исправленное значение фона будет равно

$$d_k'(\text{исп}) = \frac{1 - (0,460 + 0,260)}{12 - 2} = 0,028,$$

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 (рис. 46, б) удовлетворительно аппроксимируется полиномом

$$\rho(l) = 0,432 \cos \frac{2\pi l}{480} + 0,232 \cos \frac{2\pi l}{160}.$$

Однако модель, основанная на разделении наблюдаемой изменчивости на полигармоническую составляющую и случай-

ную 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 в общем случае член $h_x(r)$ в выражении (V.16) представляет собой не случайную величину, а реализацию стационарной случайной функции.

Приближенная оценка корреляционной функции составляющей $h_x(r)$ может быть получена путем последовательного вычитания из эмпирической корреляционной функции гармонических составляющих.

Для приведенного выше примера остаточный график корреляционной функции члена $h_x(r)$, построенный путем последовательного вычитания гармонических колебаний с частотами 1 и 3, характеризуется незначительными отклонениями от нуля почти для всех значений l (рис. 47). Исключение составляют

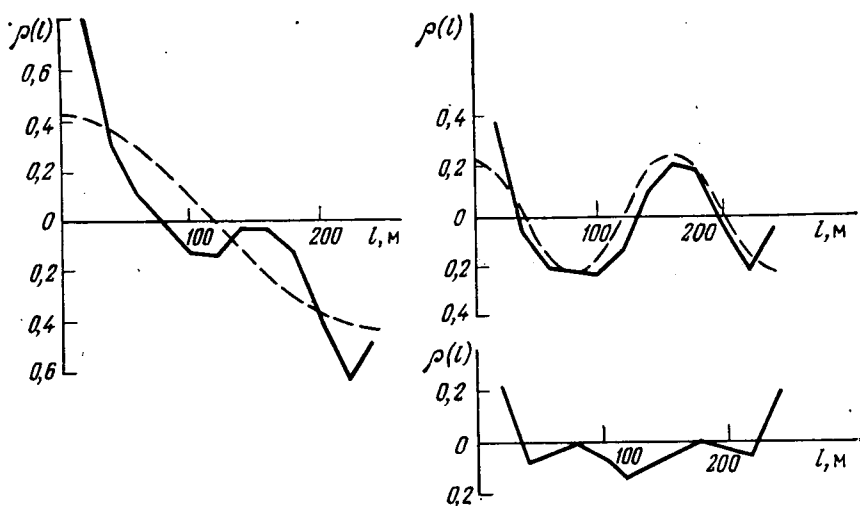


Рис. 47. Последовательное вычитание из эмпирической корреляционной функции отдельных гармоник спектра

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 составляющую, которая обусловлена закономерными на всем протяжении изучаемого участка изме-

нениями свойства и описывается с помощью тригонометрического полинома;

— коррелированную составляющую, описывающую закономерные в локальной области изменения свойства с помощью корреляционной функции, зависящей только от расстояния между точками наблюдений;

— случайную составляющую, описывающую незакономерные случайные колебания свойства, которые могут рассматриваться как реализация случайной величины («белый шум»).

Координированная закономерная составляющая, описывая общие закономерности пространственного изменения свойств геологических объектов, отражает главные особенности их формы и строения. С ее помощью можно:

— выделять элементы неоднородности в строении объектов исследования на различных стадиях их изучения;

— предсказывать наиболее вероятные значения свойства в любой точке исследуемого участка;

— 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 размер и 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 флуктуации параметров и должна учитываться при расчете точности средних значений параметров.

Способ выражения закономерной (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.

Пример. На основании изучения закономерностей распределения трещин в некоторых интрузиях В. А. Апродовым было высказано предположение о волновой природе явлений трещиноватости [2]. Им было подмечено, что изменение густоты трещин от центра интрузии к ее контакту происходит волнообразно с определенным периодом (рис. 48). Это явление

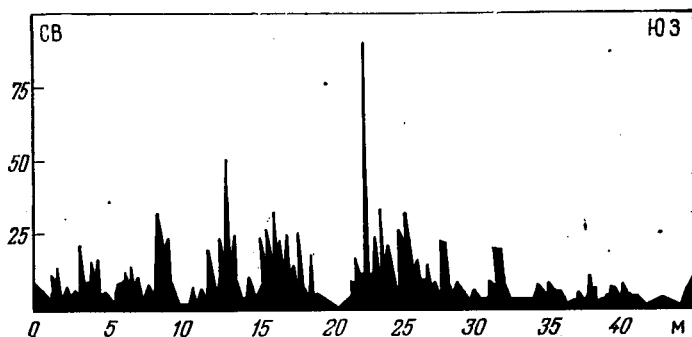


Рис. 48. Гармоническое распределение трещин в интрузии. По В. А. Апродову.

По горизонтали отложены расстояния от контакта интрузии, по вертикали — количество трещин, проходящихся на 1 пог. м обнажения

В. А. Апродов объяснил отражением упругих волн от контактовых 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. 49) 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 периодическим колебаниям с периодом T , равным

$$T = \frac{L}{\omega} = \frac{50 \text{ м}}{13} \approx 4 \text{ м.}$$

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 формулам (V.21) и (V.22). В данном примере фоновое значение спектра амплитуд равно

$$d'_k = \frac{1}{74} \approx 0,0135,$$

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 anomальное значение при доверительной вероятности 0,99 составит

$$d_k'' = d_k' + 3\sigma_d = 0,0135 + 0,0399 = 0,0534.$$

Значение спектра на частоте 4 немного превышает минимальное anomальное, что позволяет признать реальным существование периодических колебаний плотности трещин с периодом 11 м. Спектр амплитуд позволил также установить, что ведущая роль в изменении 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 м выражены слабее.

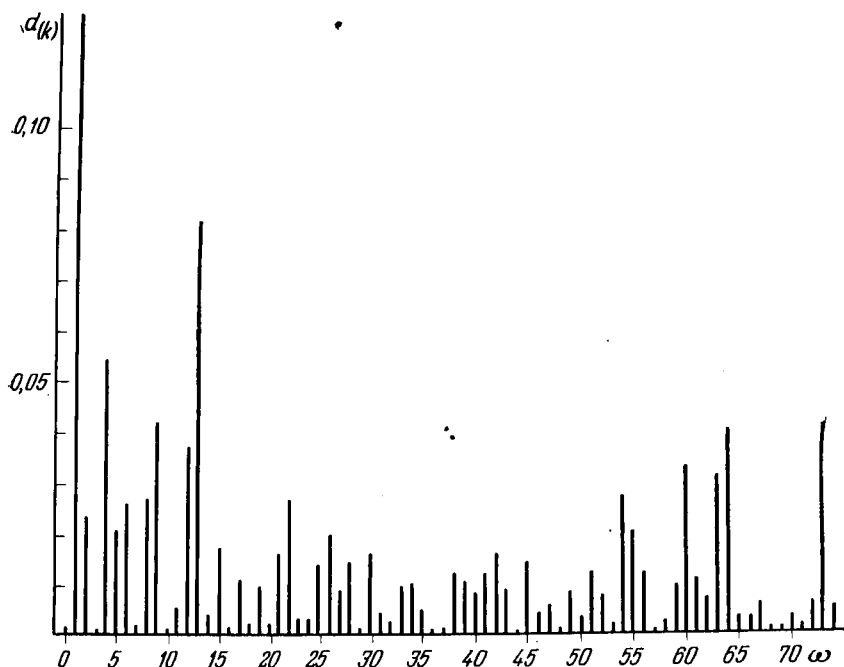


Рис. 49. Спектр амплитуд распределения трещин в интрузии

Эмпирические спектры амплитуд позволяют определять 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, доминирующим на том или ином уровне строения месторождения.

Пример. Флюоритовая жила разведана редкой сетью скважин (100×100 м) и на интервале длиной 320 м прослежена по простираанию штольней. На штольневом горизонте жила опробовалась на полную мощность через 10 м. По результатам опробования (рис. 50) штольневый горизонт

были построены корреляционная функция и спектр амплитуд значений линейного запаса (рис. 51). В спектре амплитуд выделяются аномальные значения на частотах 2 и 4, соответствующие колебаниям с периодами 160 и 80 м, на которые приходится около 45% от общей дисперсии линейного запаса. Следовательно, принятая на месторождении сеть наблюдений является запрещенной, так как расстояние между соседними скважинами по простиранию жилы близко к периоду закономерных колебаний. Данная сеть не позволяет надежно выявить даже наиболее крупные закономерные изменения линейных запасов с периодом 160 м. Для этого расстояние между

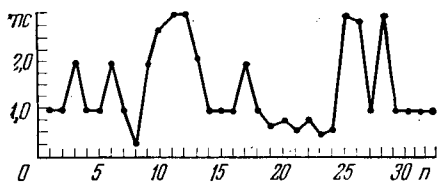


Рис. 50. Изменение линейного запаса по простиранию флюоритовой жилы

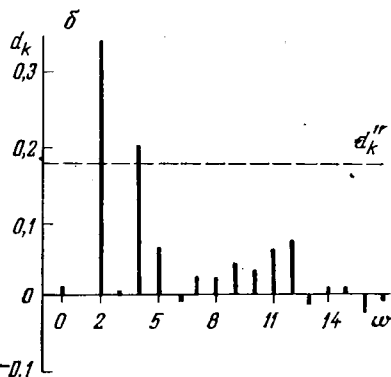
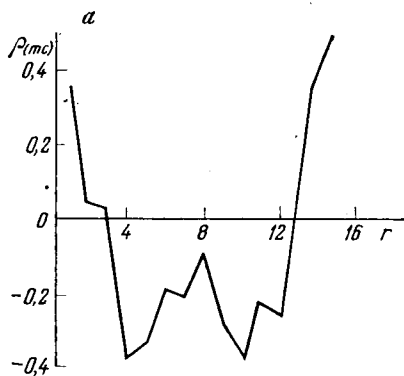


Рис. 51. Изменения линейного запаса по простиранию флюоритовой жилы:

a — корреляционная функция; *б* — изменения линейного запаса по простиранию

разведочными пересечениями должно быть уменьшено до 40 м, а для выявления неоднородностей более высокого порядка — до 20 м.

При расчете ошибок вычисления среднего значения линейного запаса необходимо учитывать закономерную составляющую. Возможные ошибки определения среднего значения в зависимости от количества пересечений в подсчетном блоке при расстоянии между соседними пересечениями меньше 40 м можно оценить по графику, изображенному на рис. 52.

Для более глубокого ознакомления с математическим аппаратом теории случайных функций может быть рекомендована книга М. Дж. Кендалла и А. Стюарта [26]; математический аппарат гармонического анализа подробно рассмотрен в книге М., Г. Серебренникова и А. А. Первозванского [45]. Примеры практического использования случайных функций при геологи-

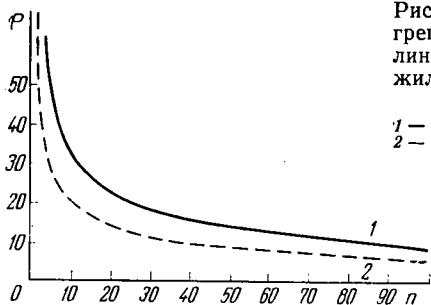


Рис. 52. Зависимость величины погрешности расчета среднего значения линейного запаса по флюоритовой жиле от количества разведочных пересечений:

1 — без учета закономерной составляющей;
2 — с учетом закономерной составляющей

ческом картировании приведены в работе Б. М. Роненсона, В. М. Ройзенмана и Н. А. Ортенберга [44]. Опыт использования данной модели при решении задач разведки месторождений обобщен в работе А. Б. Каждана [23].

ФАКТОРЫ, ОПРЕДЕЛЯЮЩИЕ ВЫБОР
И ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ
МАТЕМАТИЧЕСКИХ МЕТОДОВ В ГЕОЛОГИИ

В практике геолого-математического моделирования возможны случаи, когда

— при изучении свойств различных геологических объектов приемлемой окажется одна и та же математическая модель;

— на различных стадиях изучения какого-либо свойства одного и того же объекта потребуется применение различных математических моделей;

— для изучения изменчивости различных свойств одного и того же объекта на данной стадии потребуется применение различных моделей;

— при решении различных геологических задач на определенной стадии изучения геологического объекта для описания одного и того же свойства могут использоваться различные математические модели;

— изменение методики изучения какого-либо свойства геологического объекта, например, изменение геометрии проб, может привести к изменению характера наблюдаемой изменчивости этого свойства и тем самым вызвать необходимость применения новой математической модели.

В каждом конкретном случае выбор математической модели зависит от цели и задач проводимых геологических исследований, особенностей изучаемых объектов и методов их изучения.

**§ 1. ВЛИЯНИЕ ТИПА ГЕОЛОГИЧЕСКОЙ ЗАДАЧИ
НА ВЫБОР МАТЕМАТИЧЕСКОЙ МОДЕЛИ**

При использовании математических методов для решения различных геологических задач обычно возникает одна из трех нижеперечисленных ситуаций.

1. Геологическая задача удовлетворительно решается по совокупности полученных качественных характеристик. Мера и число необходимы только в качестве дополнительных характеристик изучаемых объектов (например, камеральная обработка результатов многих геологических съемок, описание и увязка геологических разрезов, изучение минерального состава руд и многие другие геологические операции). Однако даже в этих случаях мера и число могут оказать заметное, а иногда и ре-

шающее влияние на окончательный результат. Академик А. В. Сидоренко [46] приводит следующий пример. Всегда считалось, что в докембрии резко преобладали магматические процессы, а в последующие геологические эры — осадочные. Однако простые количественные подсчеты площадей распространения магматических и осадочно-метаморфических пород в докембрии показали, что магматические процессы и в этот период имели такое же подчиненное значение, как и в последующие.

2. Поставленная геологическая задача не может быть полностью решена только на основании качественных характеристик, а совокупность количественных характеристик не позволяет вывести объективное решение из-за того, что разброс наблюдаемых фактических значений признаков вуалирует истинную картину. Оно может быть получено на основе статистической обработки исходных данных. Такая ситуация может возникнуть при классификации горных пород по данным химических, минералогических или гранулометрических анализов, при изучении влияния свойств вмещающих пород на интенсивность процессов рудоотложения, при изучении изменчивости состава изверженных пород в эндоконтактной зоне или зональности метасоматических образований и т. д.

3. Геологические задачи исследований включают выявление достоверных количественных характеристик, которые могут быть получены только с помощью математических методов обработки исходных данных. К таким задачам относятся, например, вычисления кларков или геохимических фонов различных элементов, вычисление средних содержаний элементов, химических соединений или минералов в породах или рудах, вычисление абсолютного возраста геологических образований по данным изотопных анализов, а также задачи, решение которых основано на совместном изучении и оценке важнейших составляющих изменчивости для выявления направленных пространственных изменений признаков и более объективной оценки статистических характеристик, обусловленных случайным разбросом фактических данных.

Таким образом, применение математических методов оказывает геологам реальную помощь при:

— статистической обработке экспериментальных геологических данных,

— изучении изменчивости свойств геологических объектов,

— прогнозировании неизвестных свойств геологических объектов по совокупности известных свойств.

Статистическая обработка экспериментальных геологических данных. К числу важнейших задач статистической обработки экспериментальных геологических данных относятся.

1. Оценка средних значений признаков и характеристик их изменчивости с целью классификации объектов изучения или

выявления систематических изменений признака в пространстве. Например:

1) статистическая обработка минералогических, химических или гранулометрических анализов пород для целей их классификации;

2) вычисление средних размеров рудных скоплений и содержаний металлов в них при оценке запасов полезных ископаемых;

3) обработка результатов лабораторных анализов проб;

4) анализ статистических связей рудопроявлений и месторождений с конкретными геологическими структурами.

2. Выявление существенных геологических различий между изучаемыми объектами. Например:

1) обработка результатов анализа физико-механических свойств вмещающих рудоносных пород для решения вопроса об их влиянии на интенсивность рудного процесса;

2) обработка результатов определения абсолютного возраста пород (например, К-аргоновым методом) для решения вопроса о том, одна или несколько возрастных границ могут быть установлены по всей совокупности аналитических данных.

3. Оценка корреляционных связей между двумя признаками для предсказания значений одного признака по другому. Например:

1) изучение корреляционных связей породообразующих элементов для суждения об условиях формирования магматических пород;

2) изучение корреляционных связей между рудообразующими элементами для оценки содержаний одного элемента по данным анализов на другой элемент (например, расчет содержания серебра в рудах по содержанию в них свинца);

3) оценка силы поискового критерия (например, содержания пиропы в шлихе при поисках алмазов).

Изучение изменчивости свойств геологических объектов в пространстве или во времени. Количественные характеристики изменчивости используются при решении следующих задач.

1. Изучение общих тенденций в изменении свойств геологических объектов для суждения о процессах их формирования. Например:

1) определение областей сноса обломочного материала при формировании терригенных отложений по изменениям гранулометрического и минерального состава обломочной фракции;

2) определение глубины формирования интрузивных массивов по изменениям химического и минерального состава или физико-механических свойств слагающих их пород;

3) выявление текстурных особенностей (линейчатости, полосчатости) при микроскопическом изучении пород;

4) выявление рудоконтролирующих факторов путем изуче-

ния пространственного положения геохимических аномалий и точек минерализации относительно различных геологических объектов.

2. Определение густоты сети наблюдений для изучения параметров геологических объектов с заданной детальностью. Например:

1) определение расстояний между профилями и точками замеров при проведении геофизических работ на основе разделения аномалий геофизических полей на региональную и локальную части;

2) определение густоты разведочной сети для оценки запасов месторождения на основе выделения закономерной и случайной составляющих изменчивости параметров рудных тел.

Прогнозирование неизвестных значений свойств геологических объектов по совокупности значений известных признаков. Например:

1) оценка глубины эрозионного среза месторождения по параметрам геохимических ореолов рассеяния элементов-спутников на поверхности;

2) оценка перспектив рудоносности отдельных площадей по совокупности геологических факторов;

3) распознавание геологических объектов по совокупности геофизических данных.

Из перечисленных примеров видно, что мера и число способствуют более полному изучению природных явлений при решении практически любых геологических задач, а во многих случаях математические методы могут рассматриваться как одно из мощных средств познания недр. В совокупности с традиционными геологическими методами математические методы позволяют глубже проникать в сущность геологических процессов и извлекать максимальную информацию из выборочных экспериментальных данных.

Выбор математической модели для решения конкретной геологической задачи является наиболее ответственной и не всегда простой операцией. В настоящее время численные методы решения разработаны только для наиболее простых задач. Это часто вызывает необходимость упрощения сложных геологических задач или их разделения на более простые частные задачи.

В одних случаях характер задачи предопределяет вид математической модели, необходимый для ее решения, в других — одна и та же задача может быть решена различным путем на основе разных математических моделей.

Наиболее просто выбор модели осуществляется при сравнении геологических объектов с целью их классификации. К задачам такого типа относится решение вопроса о комагматичности интрузивных и эффузивных пород, выделение видов ископаемых организмов, разбраковка геохимических и геофизи-

ческих аномалий и т. п. При решении этих задач геологические объекты полагаются внутренне однородными и характеризуются набором средних значений каких-либо свойств, что позволяет использовать статистические модели. Вид статистической модели (одномерная или многомерная) определяется в данном случае количеством изученных свойств.

Задачи, связанные с предсказанием по одним признакам значений других признаков, также не вызывают трудностей при выборе математической модели. При их решении также постулируется однородность изучаемого объекта и используются статистические модели. Однако эффективность решения этих задач снижается, если объект оказывается внутренне неоднородным. Поэтому предположение об однородности объекта, основанное на геологическом изучении, необходимо проверять количественными методами (см. гл. I).

Наиболее сложен выбор математической модели при решении задач, основанных на математическом описании изменчивости геологических объектов. К этому типу задач относятся: определение интервальных оценок средних значений свойств геологических объектов, изучение влияния различных факторов на характер их изменчивости, анализ сети наблюдений и т. п.

При решении этих задач свойства геологических объектов в общем случае должны рассматриваться как пространственные переменные, а в их изменчивости необходимо выделять случайную и закономерную составляющие. Это вызывает необходимость использования горно-математических моделей, методов тренд-анализа, моделей типа случайных функций. Вместе с тем при отсутствии закономерностей в пространственной изменчивости геологических объектов эти задачи могут успешно решаться с помощью статистических моделей. Математические методы изучения пространственных переменных весьма разнообразны, поэтому выбор наиболее эффективной модели должен базироваться на анализе свойств исследуемых объектов и методики их изучения.

§ 2. СВОЙСТВА ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ, ВЛИЯЮЩИЕ НА ВЫБОР И ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Для характеристики объектов геолого-математического моделирования У. Крамбейном [27] введено понятие «геологической совокупности». Геологическая совокупность — это множество объектов, событий или чисел, которые являются предметами геологических исследований. В качестве геологических совокупностей могут выступать геохимические поля, совокупности интрузий определенного возраста, совокупности шлифов определенной породы, совокупности зерен минералов в шлиховой пробе и т. п. Иногда совокупность может быть разделена

по какому-либо признаку на множество самостоятельных совокупностей. Так, например, совокупность всех эффузивных пород Камчатки может быть разделена на самостоятельные совокупности по возрасту, химическому составу или по приуроченности к определенным вулканическим аппаратам. Образование одного вулканического аппарата, в свою очередь, могут рассматриваться как совокупность отдельных лавовых потоков. Определенный лавовый поток при петрографическом изучении может рассматриваться как совокупность отобранных из него образцов, каждый образец может быть использован для изготовления нескольких шлифов, каждый шлиф представляет собой совокупность минералов.

При геолого-математическом моделировании необходимо четко определить границы исследуемой геологической совокупности, ее элементарные составляющие, а также выбрать свойства элементарных составляющих, подлежащих измерению. У. Крамбейн [27] выделяет два вида геологических совокупностей:

1) совокупности, образованные первичными индивидами или объектами, по каждому из которых измеряется одно или несколько свойств. К ним относятся: совокупности ископаемых организмов определенного вида, совокупности минералов в шлифе или шлихе и т. п.;

2) совокупности, в которых каждый объект представляет собой набор исходных объектов, свойства которых обычно не измеряются, а каждый составной объект характеризуется средним значением определенных свойств либо количеством исходных объектов. К таким совокупностям относятся совокупности образцов или проб, по которым определяются физико-механические свойства или химический состав пород и руд, гранулометрический состав обломочного материала и т. п. Основной особенностью таких совокупностей является то, что свойства полученных по ним числовых совокупностей зависят от размеров составных объектов, т. е. проб.

Определение границ геологической совокупности и составляющих ее элементов производится исходя из конкретных задач. Так, например, при изучении россыпного месторождения золота задачи, связанные с его промышленной оценкой, решаются по совокупности проб, содержание золота в которых будет больше определенного значения (бортового содержания). Одновременно с этим для определения областей сноса золота и его коренных источников проводится изучение формы и размеров отдельных золотин в пробах, отобранных в разных участках россыпи и за ее пределами. При этом изменяется не только объект измерения, но и границы исследуемой геологической совокупности.

Результаты статистических исследований можно распространять только на исследуемую совокупность. Правомочность

перенесения полученных выводов на более широкий круг объектов или явлений должна обосновываться, исходя из чисто геологических соображений.

Переход от геологической совокупности к совокупности чисел, отражающих какое-либо свойство изучаемых объектов, также иногда вызывает трудности. Прежде всего необходимо убедиться в том, что измеренная переменная величина действительно хорошо отражает изменения интересующего нас свойства. 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 разные величины — частное от деления радиуса кривизны самого острого конца гальки на ее средний радиус, отношение среднего радиуса максимальных окружностей, описывающих вершины всех углов частицы в ее проекции на плоскости, к радиусу наибольшего круга, вписанного в эту проекцию [56], и т. д.

Существенное влияние на выбор математической модели оказывает характер измерения свойств объектов. У. Крамбейн [27] выделяет следующие шкалы измерений, употребляемые в геологии.

Номинальная шкала используется для классификации объектов по признаку равенства их свойств. Число в данном случае используется в качестве кода качественных признаков объектов, например, цвета пород: красный — 1, черный — 2 и т. д. Круг арифметических действий с числами номинальной шкалы ограничен: подсчет числа объектов, принадлежащих каждому классу, выявление наиболее распространенного класса и т. п. Номинальная шкала измерений используется для обозначения факторов при дисперсионном анализе, а также для фиксации качественных признаков при решении задач распознавания образов.

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 полуколичественного спектрального анализа проб и т. п. Измерения порядковой шкалы могут использоваться для сравнения геологических объектов, выявления корреляционной связи между их свойствами и проверки некоторых других геологических гипотез статистическими методами. Однако при этом могут использоваться только непараметрические статистические критерии согласия.

Относительная шкала используется для измерения свойств геологических объектов, когда измерение данного свойства можно однозначно оценить количественно. К числам относительной шкалы относятся замеры мощности тел полезных ископаемых, результаты химических анализов пород и руд, замеры элементов залегания различных геологических объектов и т. д. Числа данной шкалы могут использоваться для создания всех рассмотренных выше математических моделей.

При изучении свойств геологических объектов как пространственных переменных выбор математической модели зависит от размерности пространства. Сложность математического моделирования и вычислительных операций значительно увеличивается при переходе от одномерного пространства (случайные последовательности) к двумерному (случайные поля) и тем более к трехмерному. Поэтому в процессе моделирования необходимо стремиться к снижению размерности пространства за счет перехода к усредненным числовым характеристикам или разделения общей задачи на несколько частных. Так, например, при изучении характера распределения содержания полезного компонента в рудном теле обычно выделяют две самостоятельные задачи. Изменчивость содержания по мощности рудного тела изучается по отдельным интервальным или секционным пробам, причем результаты опробования по каждому разведочному пересечению, ориентированному по направлению мощности рудного тела, рассматриваются как отдельные реализации случайной последовательности. Для изучения распределения в плоскости рудного тела результаты опробования по каждому разведочному пересечению осредняются, а средние содержания по сквозным пересечениям рассматриваются как одна реализация двумерного случайного поля.

Аналогичная проблема возникает при изучении объектов, характеризующихся несколькими свойствами, каждое из которых условно может быть принято в качестве координаты пространства. Сведение многомерных случайных величин к комбинации трех-, а еще лучше двумерных моделей позволяет кроме аналитических методов широко использовать более наглядные и простые графические методы исследования. Для этого иногда простые числовые характеристики свойств объекта заменяются комбинированными, представляющими собой произведения, отношения или суммы нескольких исходных числовых показателей свойств объекта. Так, например, при изучении рудных тел часто в качестве пространственной переменной исследуется произведение мощности рудного тела на содержание полезного компонента (продуктивность или линейный запас). При изучении геохимических аномалий иногда вместо содержаний отдельных элементов изучаются отношения содержаний определенных элементов или групп элементов друг к другу и т. д.

При выборе математической модели необходимо также учитывать область существования исследуемых числовых характеристик, так как каждый тип математических моделей имеет модификации, разработанные специально для непрерывных или дискретных величин.

§ 3. ВЛИЯНИЕ СПОСОБА ИЗУЧЕНИЯ ОБЪЕКТА НА ВЫБОР МАТЕМАТИЧЕСКОЙ МОДЕЛИ

В геологической практике геологические совокупности в силу своих больших размеров и ряда других причин обычно не могут быть исследованы полностью. О свойствах всей геологической совокупности геолог судит по какой-то ее части, доступной для наблюдения и опробования, которую М. Розенфельд [57] 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 и объема отбираемых проб.

Выделяют [27, 35] три основных системы расположения точек наблюдения: равномерное, случайное и многостадийное опробование (рис. 53).

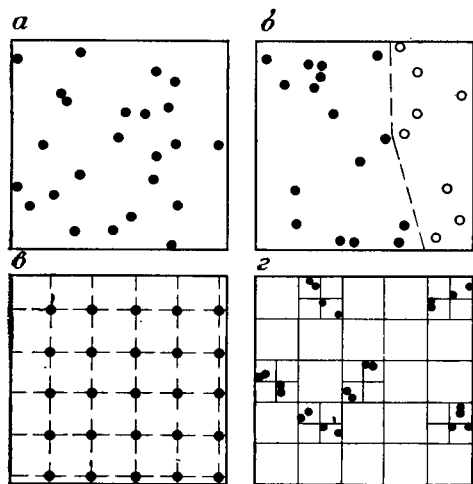


Рис. 53. Системы расположения точек наблюдений. По У. Крамбейну:
а — случайное расположение проб; *б* — случайное стратифицированное опробование (объект разделен на две подсовокупности); *в* — равномерное опробование; *г* — многостадийное опробование

Наиболее распространенным и в большинстве случаев предпочтительным является равномерное опробование, когда точки наблюдений в плоскости изучаемого объекта располагаются по

правильной геометрической сети (см. рис. 53, в). Равномерное опробование позволяет с одинаковой детальностью изучить все части изучаемого объекта и является наиболее эффективным для выявления средних значений изучаемых свойств при наличии в их изменчивости закономерной составляющей. Исключение представляет только довольно редко встречающийся случай, когда в изменчивости свойства присутствуют периодические колебания с периодом, равным или кратным размерам ячейки сети наблюдений. Кроме того, равномерное расположение точек наблюдений существенно упрощает вычислительные операции при тренд-анализе, расчете основных числовых характеристик моделей типа случайных функций и построении горно-геометрических моделей.

При случайном опробовании места точек наблюдений или отбора проб выбираются случайно с помощью таблиц случайных чисел. Часто к случайному близко размещение наблюдений по естественным обнажениям. Случайное опробование обычно применяется в тех случаях, когда исследователя не интересуют закономерности изменения изучаемого свойства в пространстве или априорно известно, что таких закономерностей нет. Случайный способ отбора может, например, использоваться для составления выборок проб для контрольных анализов.

Многостадийное опробование применяется для изучения изменчивости свойств геологических объектов на разных уровнях строения. Для этого исследуемый объект разделяется на приблизительно равные по величине участки, которые, в свою очередь, разделяются на участки меньшего масштаба, и т.д. В пределах каждого участка опробуется только определенная часть участков более высокого порядка. За счет этого общее количество наблюдений при многостадийном опробовании существенно сокращается по сравнению с равномерным. Многостадийное опробование применяется при проведении поисково-съёмочных работ, особенно с применением аэродесантных операций.

Детальность выявления структуры изучаемого объекта зависит от густоты сети дискретных наблюдений. Чем гуще эта сеть, 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 элемента неоднородности, то между значениями изучаемого признака в смежных пунктах наблюдений возникнут более или менее отчетливые автокорреляционные связи, а в характеристике наблюдаемой изменчивости признака отчетливо проявится неслучайная ее составляющая. Если же проводить наблюдения по более редкой сети, так, чтобы смежные

пункты приходились на различные элементы неоднородности, корреляционные связи между значениями изучаемого признака в них не проявятся, а в характеристике изменчивости признака будет доминировать случайная ее составляющая. Таким образом, изменчивость свойств одного и того же природного геологического объекта может быть оценена как случайная при редкой сети и как неслучайная после сгущения этой сети (рис. 54).

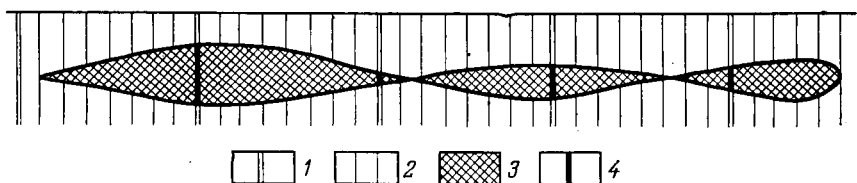


Рис. 54. Зависимость характера изменчивости мощности залежи от густоты сети наблюдений:

1 — по редкой сети наблюдений преобладает случайная составляющая; 2 — по густой сети наблюдений преобладает закономерная составляющая изменчивости; 3 — рудная залежь; 4 — рудный интервал по редкой сети наблюдений

Еще более заметное влияние на представления об изменчивости изучаемых свойств природных геологических образований оказывают размеры отбираемых проб. Так, например, если опробовать золоторудную зону пробами микроскопических размеров, изменчивость содержания золота по таким пробам окажется исключительно высокой, так как отдельные редкие пробы будут состоять из самородков золота, некоторые пробы — из золотосодержащих сульфидов или кварца, а остальные пробы — из минеральных зерен вмещающих пород. Если же увеличить размеры проб на порядок по сравнению с размерами золотосодержащих минеральных агрегатов, то изменчивость содержаний по пробам заметно снизится, хотя содержание в отдельных пробах с самородками золота еще будет резко отличаться от содержаний в остальных пробах. Продолжая увеличивать размеры проб, можно настолько снизить наблюдаемую изменчивость содержаний в них, что оно будет представляться практически равномерным (например, среднее содержание золота в загруженных рудой товарных вагонах). Выравнивание средних содержаний происходит в данном примере по закону «больших чисел» за счет резкого увеличения общего количества входящих в пробу элементов неоднородности высокого порядка (рис. 55).

Зависимость частных значений изучаемых признаков от густоты сети наблюдений и от размеров отбираемых проб свидетельствует о том, что задача полноценного и всестороннего количественного описания природной изменчивости свойств гео-

логических образований практически невыполнима. Природная изменчивость существует объективно, отражая комплекс геологических и физико-химических условий формирования изучаемого объекта, и представляет собой явление весьма сложного и многопланового характера. При решении конкретных геологических задач из всех возможных проявлений природной изменчивости оцениваются только те, знание которых необходимо для решения поставленных задач и только на тех уров-

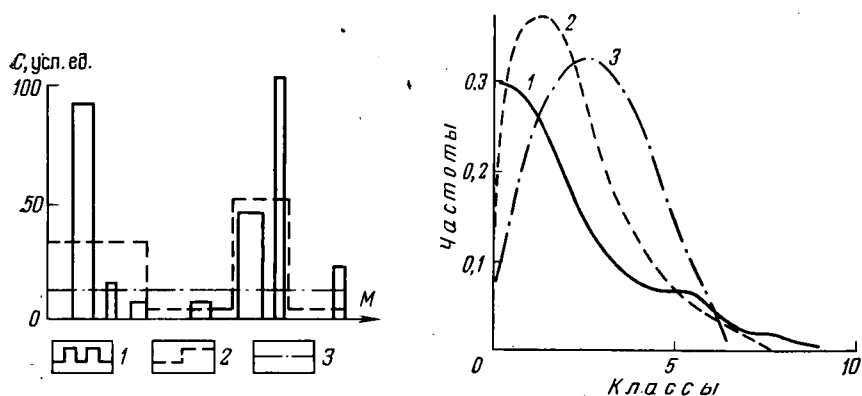


Рис. 55. Связь представлений о степени однородности размещения золота в зависимости от размера проб:

1 — по малым секционным пробам; 2 — по большим секционным пробам; 3 — по сквозной пробе

Рис. 56. 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

нях строения, которые выявляются при данном масштабе проводимых исследований.

Следовательно, понятие наблюдаемой изменчивости принципиально отличается от широко распространенного понятия природной изменчивости свойств геологических образований.

Наблюдаемая изменчивость отражает детальность наших представлений об изменчивости реально существующего геологического образования в зависимости не только от его природной сложности, но также и от условий экспериментальных наблюдений, положенных в основу ее характеристики.

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. Так, например,

Д. А. Зенковым [22] показано, что гиперболовидное распределение золота по секционным борздовым пробам сменяется асимметричным логнормальным для «сквозных» проб и почти симметричным для «площадных» проб (рис. 56). Поэтому далеко идущие генетические выводы по виду статистических характеристик делать нельзя.

При использовании методов корреляционного и регрессионного анализов в геологии и разведке также необходимо учитывать влияние геометрии проб — их объемов, весов, размеров и ориентировки на количественные характеристики выборочных оценок связи. Так, например, при выявлении характера и силы связи между содержаниями двух элементов в рудах какого-либо месторождения следует пользоваться данными анализов таких проб, геометрия которых отвечает условиям 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 покусковой сортировки), то для целей корреляционного анализа должны использоваться данные о содержаниях элементов в штучных пробах.

На рис. 57 приводится график зависимости выборочных значений коэффициента корреляции между содержаниями основ-

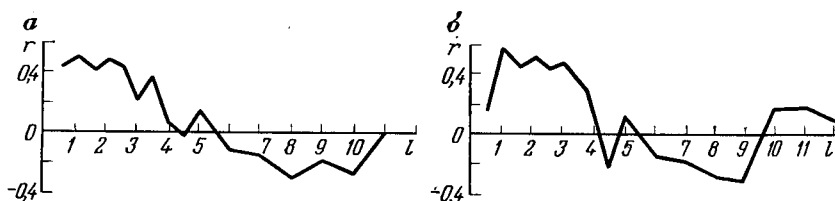


Рис. 57. Изменения выборочных значений коэффициента корреляции между содержаниями основного и сопутствующего компонента в зависимости от длин борздовых проб:

а — для логарифмов содержаний полезных компонентов; б — для логарифмов метрпроцентов полезных компонентов

ного и сопутствующего элемента в рудах одного из редкометальных месторождений в зависимости от геометрических характеристик анализируемых проб. Из рассмотрения графика видно, что значение выборочного коэффициента корреляции изменяется в широком диапазоне — от статистически 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 [47].

§ 4. РОЛЬ ГЕОЛОГИЧЕСКОГО АНАЛИЗА ПРИ ВЫБОРЕ ГЕОЛОГО-МАТЕМАТИЧЕСКОЙ МОДЕЛИ

В зависимости от морфологии и внутреннего строения объекта, размеров проб и расстояний между смежными пунктами наблюдений экспериментальные данные могут представлять собой совокупность случайных взаимонезависимых величин, совокупность случайных автокоррелированных величин или совокупность случайных величин с отчетливо проявленной периодичностью, функционально связанной с координатами пространства.

В каждом из перечисленных случаев выбор математической модели будет определяться степенью соответствия ее свойств наблюдаемым свойствам объекта, а также целями и задачами проводимых исследований. В первом случае с позиций соответствия математических и типовых геологических моделей должен использоваться аппарат вариационной статистики случайных величин, во втором случае — аппарат теории стационарных случайных функций, а в третьем случае — аппарат гармонического анализа случайных функций или тренд-анализа. При изучении одновременно двух и более случайных величин используются статистические модели двумерного и многомерного статистического анализа. Таким образом, по мере накопления информации и детализации представлений о природе объекта более простые модели закономерно сменяются более сложными.

Статистические модели используются в тех случаях, когда вследствие природных особенностей пространственного размещения изучаемого геологоразведочного параметра, недостаточного числа наблюдений или редкой разведочной сети наблюдаемая изменчивость моделируемого свойства проявляется только как случайная, а неслучайную ее составляющую выделить не удается.

Если по совокупности экспериментальных данных в наблюдаемой изменчивости моделируемого признака выделяются как случайная, так и неслучайная ее составляющие, применяется одна из трех вышеупомянутых математических моделей.

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 трудно переоценить ведущую роль геологического анализа на всех стадиях использования математических методов.

Геологический анализ лежит в основе сбора исходной цифровой информации, поскольку в процессе геологических исследований размещение пунктов наблюдений и проб 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 в первую очередь с геологической точки зрения. Все изучаемые объекты — массивы горных пород, скопления полезных ископаемых, их участки или блоки — должны быть проверены с точки зрения их геологической однородности, а для совокупностей, не однородных в геологическом отношении, должны быть установлены границы. Для каждой выделяемой совокупности цифровых данных должна быть проверена 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 отдельно 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 — также и 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.

Геологический анализ должен оказывать влияние не только на оценку и систематизацию исходных цифровых данных, но и на обоснование возможности применения той или иной конкретной математической модели. С его помощью, в зависимости

от полноты геологической изученности объекта, должны определяться те задачи, которые могут быть решены на данной стадии исследования с применением математических методов. Так, например, по данным сплошного геохимического опробования керна колонковых скважин могут быть получены исходные цифровые данные для количественной оценки эндогенных ореолов рассеяния редких металлов, которые можно использовать для поисков редкометального оруденения, связанного с гидротермально-измененными гранитоидными породами. Если среди большого количества ореольных проб с повышенными (аномальными) содержаниями встретятся единичные рудные пробы с богатыми содержаниями редких металлов, то с чисто математической точки зрения вся совокупность данных опробования будет представлять собой однородный статистический коллектив. Однако, если учесть, что рудные пробы встречаются только среди метасоматитов определенного состава и свидетельствуют о наличии уже не ореолов рассеяния, а промышленно-ценных рудных скоплений, то с геологических позиций правильнее выделить их в качестве самостоятельной совокупности. Эту весьма ограниченную совокупность единичных рудных проб нецелесообразно подвергать какой-либо математической обработке. Однако результаты статистической обработки совокупности ореольных проб могут оказаться существенно иными вследствие исключения рудных проб.

В дальнейшем, по мере сгущения сети скважин, может быть накоплен статистический материал, достаточный для получения количественной характеристики изменчивости редкометального оруденения в пределах выявленных рудных залежей. При относительно редкой разведочной сети значения наблюдаемых свойств залежи будут считаться взаимонезависимыми, если в смежных пересечениях не проявится заметных корреляционных связей между ними. Для выяснения пространственных закономерностей изучаемых свойств подобная цифровая информация непригодна, а для ее математической обработки наиболее целесообразно использование статистической модели.

При дальнейшем сгущении сети наблюдений рано или поздно появится необходимость изменения математической модели, а также иной группировки исходных цифровых данных. Между значениями наблюдаемых свойств в смежных пунктах наблюдения начнут появляться корреляционные связи и, следовательно, всю совокупность цифровых данных уже нельзя будет рассматривать, как взаимонезависимую. В этих новых условиях совокупность цифровых данных будет нести в себе информацию не только об уровне изменчивости изучаемого признака, но и о закономерностях его пространственного размещения. В результате появится возможность применить более сложную математическую модель, позволяющую учитывать

раздельно влияние случайной и неслучайной составляющей изменчивости.

Более густая сеть наблюдений обеспечит получение значительно более полной геологической информации. Поэтому внутри отдельных участков, которые из-за недостатка фактических данных считались ранее условно однородными, появится возможность выделения новых, более дробных, геологически самостоятельных единиц. Однако при выделении таких геологически самостоятельных объектов возникнет необходимость в соответствующей перегруппировке исходной цифровой информации с целью ее раздельной обработки.

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, 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.

Более подробно проблема выбора математических моделей при решении различных задач геологии и разведки рассмотрена в работах У. Крамбейна и Ф. Грейбилла [27], Н. Н. Боровко [6] и А. Б. Каждана [23].

1. *Андерсен Т.* Введение в многомерный статистический анализ. М., Физматгиз, 1963. 500 с.
2. *Апродов В. А.* Геологическое картирование. М., Госгеолыздат, 1952. 371 с.
3. *Большаев Л. Н., Смирнов Н. В.* Таблицы математической статистики. М., ВЦ АН СССР, 1968. 474 с.
4. *Бондаренко В. Н.* Статистические решения некоторых задач геологии. М., Недра, 1970. 246 с.
5. *Бондаренко В. Н., Верховская Л. А., Сорокина Е. П.* Применение многомерных статистических методов для изучения структуры геохимических полей.— В кн.: Математические методы при геохимических исследованиях. Свердловск, 1973, с. 23—25.
6. *Боровко Н. Н.* Статистический анализ пространственных геологических закономерностей. Л., Недра, 1971. 173 с.
- ✓ 7. *Бугаец А. Н., Дуденко Л. Н.* Математические методы при прогнозировании месторождений полезных ископаемых. Л., Недра, 1976. 270 с.
8. *Ван дер Варден Б. Л.* Математическая статистика. М., Изд-во иностр. лит., 1960. 434 с.
9. *Викентьев В. А., Гуськов О. И., Шумилин М. В.* О методах установления связи между мощностью рудного тела и содержанием полезного компонента.— Изв. вузов. Сер. Геология и разведка, 1968, № 1, с. 37—42.
10. *Викентьев В. А., Шумилин М. В.* Оценка точности коэффициента рудоносности при подсчете запасов.— Изв. вузов. Сер. геология и разведка, 1971, № 3, с. 83—91.
11. *Вистелиус А. Б.* Ритмы пористости в нижнеказанских отложениях Южной Татарии.—Тр. Ленингр. о-ва естествоисп. Т. 68, вып. 2, 1951, с. 150—167.
12. *Вистелиус А. Б.* Задачи геохимии и информационные меры.— Сов. геология, 1964, № 12, с. 5—26.
13. *Геология и математика* / Ю. А. Воронин, Б. К. Алабин, С. В. Гольдин и др. Новосибирск, Наука, 1967. 254 с.
14. *Геология и математика.* Под ред. Э. Э. Фотиади. Новосибирск, Наука, 1967. 254 с.
15. *Гуськов О. И., Шумилин М. В.* О применении статистических методов для анализа плотности разведочной сети.— Изв. вузов. Сер. геология и разведка, 1968, № 6, с. 69—77.
16. *Гуськов О. И., Шумилин М. В.* Способы уточнения оценок средних параметров при подсчете запасов.— Разведка и охрана недр, 1976, № 5, с. 28—31.
17. *Дубов Р. И.* К теории оценивания информативности и эффективности методов исследований.— В кн.: Ежегодник—1969. Иркутск, 1970, с. 297—301.
18. *Дубов Р. И.* О применении принципа максимального правдоподобия для классификации геохимических аномалий.— В кн.: Ежегодник—1970. Иркутск, 1971, с. 362—366.
19. *Езекиэл М., Фокс К. А.* Методы корреляций и регрессий линейных и криволинейных. М., Статистика, 1966. 558 с.
20. *Емельяненко В. У., Левит Н. В.* Опыт применения аэрофотометода для выявления перспективных площадей с медноколчеданным оруденением в районах Южного Урала.— Разведка и охрана недр, 1972, № 1, с. 12—14.
21. *Загоскин В. А., Шиманский А. А.* Строение первичных ореолов рас-

- сеяния редкометалльных пегматитов натро-литиевого типа. — В кн.: Геохимия пегматитов Восточной Сибири. М., Наука, 1971, с. 195—206.
22. *Зенков Д. А.* Эволюция статистических показателей содержания компонентов рудных месторождений в связи с методами исследования при разведке. — В кн.: Вопросы методики опробования рудных месторождений при разведке и эксплуатации. М., Госгеолтехиздат, 1962, с. 35—44.
23. *Каждан А. Б.* Методологические основы разведки полезных ископаемых. М., Недра, 1974. 271 с.
24. *Каллистов П. Л.* Изменчивость оруденения и плотность наблюдений при разведке и опробовании. — Сов. геология, 1956, сб. 53, с. 118—151.
25. *Кащеев Л. П., Кушнарев П. И.* О выборе разведочной сети при разведке прибрежно-морских россыпных месторождений. — Изв. вузов. Сер. геология и разведка, 1975, № 3, с. 115—119.
26. *Кендалл М. Дж., Стьюарт А.* Многомерный статистический анализ и временные ряды. М., Наука, 1976. 736 с.
27. *Крамбейн У., Грейбилл Ф.* Статистические модели в геологии. М., Мир, 1969. 398 с.
28. *Крамбейн У., Кауфмен М., Мак-Кеммон Р.* Модели геологических процессов. М., Мир, 1973. 150 с.
29. *Криге Д. Г.* Роль математической статистики в методах уточнения оценки промышленного оруденения на рудниках Южной Африки. — В кн.: Вопросы математической геологии. Л., Наука, 1968, с. 252—272.
30. *Кутолин В. А.* Статистическое изучение химизма базальтов. М., Наука, 1969. 142 с.
31. *Лоули Д., Максвелл А.* Факторный анализ как статистический метод. М., Мир, 1967. 144 с.
32. *Луговенко В. Н.* Статистический анализ аномального магнитного поля территории СССР. М., Наука, 1974. 199 с.
33. *Ляхов Л. Л., Соловьев Г. А.* Корреляционный способ комплексной интерпретации материалов геофизических исследований при поисках редкометаллических месторождений. — Изв. вузов. Сер. геология и разведка, 1969, № 8, с. 75—79.
34. *Математическая обработка данных в поисковой геохимии.* Под ред. Р. И. Дубова. Новосибирск, Наука, 1976. 208 с.
35. *Матерон Ж.* Основы прикладной геостатистики. М., Мир, 1968. 408 с.
36. *Миллер Р. Л., Кан Дж. С.* Статистический анализ в геологических науках. М., Мир, 1965. 482 с.
37. *Налимов В. В.* Теория эксперимента. М., Наука, 1971. 207 с.
38. *Опыт применения множественной корреляции для решения некоторых задач поисков и разведок / М. К. Косыгин, Г. В. Росляков, В. А. Загоскин и др.* — Изв. вузов. Сер. геология и разведка, 1966, № 9, с. 3—10.
39. *Пирс Дж.* Символы, сигналы, шумы. М., Мир, 1967. 334 с.
40. *Рао С. Р.* Линейные статистические методы и их применение. М., Наука, 1968. 548 с.
41. *Рац М. В.* Неоднородности горных пород и их физических свойств. М., Наука, 1968. 108 с.
42. *Родионов Д. А.* Функции распределения содержания элементов и минералов в изверженных горных породах. М., Наука, 1964. 101 с.
43. *Родионов Д. А.* Статистические методы разграничения геологических объектов по комплексу признаков. М., Недра, 1967. 158 с.
44. *Роненсон Б. М., Ройзенман В. М., Ортенберг Н. А.* Методы расчленения метаморфических комплексов. Л., Недра, 1976. 191 с.
45. *Серебренников М. Г., Первозванский А. А.* Выявление скрытых периодичностей. М., Наука, 1965. 244 с.
46. *Сидоренко А. В.* Геология — наука будущего. М., Знание, 1964. 64 с.
47. *Сидорков Е. А., Шумилини М. В., Тихонов В. Н.* О влиянии размера проб при использовании корреляционных методов подсчета запасов попутных компонентов. — Изв. вузов. Сер. геология и разведка, 1974, № 2, с. 85—89.
48. *Факторный анализ в нефтяной геологии (обзор) / М. Д. Белонин, И. В. Татаринев, О. М. Калинин и др.* М., Изд. ВИЭМС, 1971. 56 с.

49. Харбух Дж., Бонэм-Картер Г. Моделирование на ЭВМ в геологии. М., Мир, 1974. 318 с.
50. Харман Г. Современный факторный анализ. М., Статистика, 1972. 486 с.
51. Шарапов И. П. Применение математической статистики в геологии. М., Недра, 1965. 260 с.
52. Янко Я. Математико-статистические таблицы. М., Госстатиздат, 1961. 243 с.
53. Aitchison J., Brown J. A. C. The lognormal distribution. Cambridge Univ. Press, London, 1957. 176 p.
54. Ferguson T. S. Proceeding of the 4-th Berkeley symposium on mathematical statistics and probability. 1, 1961. 253 p.
55. Krige D. G. Effective pay limits for selective mining.— Journ. S.—Afr. Inst. Min. Met., 62, 1961—1962, p. 345—363.
56. Olson J. S., Potter P. E. Variance components of cross-bedding direction in some basal Pennsylvanian sandstones of the eastern interior basin—statistical methods. J. Geol., 62, 1954, p. 26—49.
57. Rosenfeld M. A. Petrographic variation in the Oriskany sandstone.— Bull. Geol. Soc. Am., 1954, 65, p. 1298—1299.
58. Tables of the cumulative binomial probabilities. Ordnance corps. USA, Wash., 1952.

Приложение 1

Допустимые значения критерия Стьюдента
при данном объеме выборки N и уровне значимости P

N	P												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,16	0,33	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66	636,62
2	14	29	45	62	0,82	1,06	1,39	1,89	2,92	4,30	6,97	9,93	31,60
3	14	28	42	58	77	0,98	1,25	1,64	2,35	3,18	4,54	5,84	12,94
4	13	27	41	57	74	94	1,19	1,53	2,13	2,78	3,75	4,60	8,61
5	13	27	41	56	73	92	1,16	1,48	2,02	2,57	3,37	4,03	6,86
6	13	27	40	55	72	91	1,13	1,44	1,94	2,45	3,14	3,71	5,96
7	13	26	40	55	71	90	1,12	1,42	1,90	2,37	3,00	3,50	5,41
8	13	26	40	55	71	89	1,11	1,40	1,86	2,31	2,90	3,36	5,04
9	13	26	40	54	70	88	1,10	1,38	1,83	2,26	2,82	3,25	4,78
10	13	26	40	54	70	88	1,09	1,37	1,81	2,23	2,76	3,17	4,59
11	13	26	40	54	70	88	1,09	1,36	1,80	2,20	2,72	3,11	4,44
12	13	26	40	54	70	87	1,08	1,36	1,78	2,18	2,68	3,06	4,32
13	13	26	39	54	69	87	1,08	1,35	1,77	2,16	2,65	3,01	4,22
14	13	26	39	54	69	87	1,08	1,35	1,76	2,15	2,62	2,98	4,14
15	13	26	39	54	69	87	1,07	1,34	1,75	2,13	2,60	2,95	4,07
16	13	26	39	54	69	87	1,07	1,34	1,75	2,12	2,58	2,92	4,02
17	13	26	39	53	69	86	1,07	1,33	1,74	2,11	2,57	2,90	3,97
18	13	26	39	53	69	86	1,07	1,33	1,73	2,10	2,55	2,88	3,92
19	13	26	39	53	69	86	1,07	1,33	1,73	2,09	2,54	2,86	3,88
20	13	26	39	53	69	86	1,06	1,33	1,73	2,09	2,53	2,85	3,85
21	13	26	39	53	69	86	1,06	1,32	1,72	2,08	2,52	2,83	3,82
22	13	26	39	53	69	86	1,06	1,32	1,72	2,07	2,51	2,82	3,79
23	13	26	39	53	69	86	1,06	1,32	1,71	2,07	2,50	2,81	3,77
24	13	26	39	53	69	86	1,06	1,32	1,71	2,06	2,49	2,80	3,75
25	13	26	39	53	68	86	1,06	1,32	1,71	2,06	2,49	2,79	3,73
26	13	26	39	53	68	86	1,06	1,32	1,71	2,06	2,48	2,78	3,71
27	13	26	39	53	68	86	1,06	1,31	1,70	2,05	2,47	2,77	3,69
28	13	26	39	53	68	86	1,06	1,31	1,70	2,05	2,47	2,76	3,67
29	13	26	39	53	68	85	1,06	1,31	1,70	2,05	2,46	2,76	3,66
30	13	26	39	53	68	85	1,06	1,31	1,70	2,04	2,46	2,75	3,65
40	13	26	39	53	68	85	1,05	1,30	1,68	2,02	2,42	2,70	3,55
60	13	25	39	53	68	85	1,05	1,30	1,67	2,00	2,39	2,66	3,46
120	13	25	39	53	68	85	1,04	1,29	1,66	1,98	2,36	2,62	3,37
∞	13	25	39	52	67	84	1,04	1,28	1,65	1,96	2,33	2,58	3,29

Приложение 2

Значения критерия Фишера при уровне значимости
 $P=0,05$ для числа степеней свободы k_1 и k_2

k_2	k_1												
	1	2	3	4	5	6	8	12	16	24	50	∞	
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	246,5	249,0	251,8	254,3	
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,43	19,45	19,47	19,50	
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,69	8,64	8,58	8,53	
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,84	5,77	5,70	5,63	

k_2	k_1											
	1	2	3	4	5	6	8	12	16	24	50	∞
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,60	4,53	4,44	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,92	3,84	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,49	3,41	3,32	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,20	3,12	3,03	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,98	2,90	2,80	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,82	2,74	2,64	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,70	2,61	2,50	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,60	2,50	2,40	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,51	2,42	2,32	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,44	2,35	2,24	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,39	2,29	2,18	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,33	2,24	2,13	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,29	2,19	2,08	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,25	2,15	2,04	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,21	2,11	2,00	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,18	2,08	1,96	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,15	2,05	1,93	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,13	2,03	1,91	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,11	2,00	1,88	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	2,09	1,98	1,86	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	2,07	1,96	1,84	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	2,05	1,95	1,82	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	2,03	1,93	1,80	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	2,02	1,91	1,78	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	2,00	1,90	1,77	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,99	1,89	1,76	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,94	1,83	1,70	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,90	1,79	1,66	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,87	1,76	1,63	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,85	1,74	1,60	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,81	1,70	1,56	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,79	1,67	1,53	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,77	1,65	1,51	1,32
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,76	1,64	1,49	1,30
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,75	1,63	1,48	1,28
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,72	1,60	1,45	1,25
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,71	1,59	1,44	1,22
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,69	1,57	1,42	1,19
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,68	1,55	1,39	1,15
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,67	1,54	1,38	1,13
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,66	1,54	1,38	1,11
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,65	1,53	1,36	1,08
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,64	1,52	1,35	1,00

Значения критерия Фишера при уровне значимости $P=0,01$ для числа степеней свободы k_1 и k_2

k_2	k_1											
	1	2	3	4	5	6	8	12	16	24	50	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6169	6234	6302	6366
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,42	99,44	99,46	99,48	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,83	26,60	26,35	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	14,15	13,93	13,69	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,29	9,89	9,68	9,47	9,24	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,52	7,31	7,09	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,27	6,07	5,85	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,48	5,28	5,06	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,92	4,73	4,51	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,52	4,33	4,12	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,21	4,02	3,80	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,98	3,78	3,56	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,78	3,59	3,37	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,62	3,43	3,21	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,48	3,29	3,07	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,37	3,18	2,96	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,27	3,08	2,86	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,20	3,00	2,79	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	3,12	2,92	2,70	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	3,05	2,86	2,63	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,99	2,80	2,58	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,94	2,75	2,53	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,89	2,70	2,48	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,85	2,66	2,44	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,81	2,62	2,40	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,78	2,58	2,36	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,74	2,55	2,33	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,71	2,52	2,30	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,68	2,49	2,27	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,66	2,47	2,24	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,07	2,74	2,56	2,37	2,13	1,90
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,48	2,29	2,05	1,80
45	7,23	5,11	4,25	3,77	3,45	3,23	2,94	2,61	2,43	2,23	1,99	1,75
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,56	2,38	2,18	1,94	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,32	2,12	1,87	1,60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,78	2,45	2,28	2,07	1,82	1,53
80	6,96	4,88	4,04	3,56	3,26	3,04	2,74	2,42	2,24	2,03	1,78	1,49
90	6,92	4,85	4,01	3,53	3,23	3,01	2,72	2,39	2,21	2,00	1,75	1,45
100	6,90	4,82	3,98	3,51	3,21	2,99	2,69	2,37	2,19	1,98	1,73	1,43
125	6,84	4,78	3,94	3,47	3,17	2,95	2,66	2,33	2,15	1,94	1,69	1,37
150	6,81	4,75	3,91	3,45	3,14	2,92	2,63	2,31	2,13	1,92	1,66	1,33
200	6,76	4,71	3,88	3,41	3,11	2,89	2,60	2,28	2,09	1,88	1,62	1,28
300	6,72	4,68	3,85	3,38	3,08	2,86	2,57	2,24	2,06	1,85	1,59	1,22
400	6,70	4,66	3,83	3,37	3,06	2,85	2,56	2,23	2,04	1,84	1,57	1,19
500	6,69	4,65	3,82	3,36	3,05	2,84	2,55	2,22	2,03	1,83	1,56	1,16
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,53	2,20	2,01	1,81	1,54	1,11
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,99	1,79	1,52	1,00

Приложение 4

Функция нормального распределения с параметрами 0;1
(для отрицательных значений z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
-,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
-,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
-,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
-,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
-,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
-,7	2420	2389	2358	2327	2297	2266	2236	2206	2177	2148
-,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
-,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	0985
-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	0594	0582	0570	0559
-1,6	0548	0537	0526	0515	0505	0495	0485	0475	0465	0455
-1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	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	0139	0135	0132	0129	0125	0122	0119	0116	0113	0110
-2,3	0107	0104	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	0022	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
-3,2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
-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	0002	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

Приложение 5

Функция нормального распределения с параметрами 0;1
(для положительных значений z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,0	0,5000	0,5040	0,5080	0,5120	0,5136	0,5199	0,5239	0,5279	0,5319	0,5359
1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141

Продолжение прилож. 5

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8925	8944	8962	8980	9015
1,3	9032	9049	9066	9082	9100	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9278	9292	9306	9319
1,5	9331	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9648	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9874	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9924	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	9986	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1	9990	9991	9991	9991	9991	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9996	9995	9996	9996	9996	9996	9996	9996
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9997	9998
3,5	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,6	9998	9998	9998	9998	9999	9999	9999	9999	9999	9999

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Адекватность геологических и математических моделей** 154
Амплитуда гармонических колебаний 132
Анализ гармонический 131
 — дисперсионный 43
 — корреляционный 63
 — поверхностей тренда 113
 — регрессионный 64
Аномальное значение 40, 111
Вектор 70
Вектор-столбец 70
Вектор-строка 70
Вероятность доверительная 33
 — достоверного события 18
 — случайного события 18
Выборка 16
Гипотезы статистические (проверка) 25
 — об однородности изучаемого объекта 39
 — о линейности корреляционной связи 56, 65
 — о наличии корреляционной связи 50
 — о наличии тренда 103, 127
 — о равенстве выборочных оценок 33
 — о равенстве дисперсий 38
 — о соответствии эмпирических распределений теоретическим 25
Диагональные элементы матрицы
Дискретные случайные поля 112
Дискриминантная функция 79
Дискриминантный анализ 80
Дисперсионный анализ 43, 110
 — двухфакторный 44
 — однофакторный 43
Дисперсия 20
 — условная 47
Доверительный интервал 29
Закон распределения 18
 — биномиальный 23
 — логнормальный 22
 — нормальный 22
Изменчивости уровень 96
 — характер 96
Изменчивость геологических объектов 96
 — закономерная 100
 — наблюдаемая 15, 96, 152
 — случайная 100
Ковариация (корреляционный момент) 49
Корреляционная зависимость 46
Корреляционное отношение 50
 — поле точек 52
Корреляция 49
 — линейная 55
 — нелинейная 55
 — обратная 49
 — прямая 49
Коэффициент асимметрии 20
 — аномальности 111
 — вариации 20
 — корреляции 49
 — корреляции информационный (А. Б. Вистелиуса) 60
 — корреляции множественный 73
 — корреляции парный 73
 — ранговый Спирмена 53
Коэффициенты линейной регрессии 48
Критерии согласия 32
 — непараметрические 34, 36
 — параметрические 34
Критерий Вилкоксона (Манна—Уитни) 36
 — Махаланобиса 81
 — Родионова 35
 — Смирнова 40
 — Фергюссона 40
 — χ^2 — Пирсона 25
 — F — критерий Фишера 39, 43, 65, 110
 — t — критерий Стьюдента 30, 34
 — X — критерий Ван дер Вардена 36
Критерия мощность 33
Линия регрессии 47
Математическое ожидание 19
Матрица, определение 70
 — ковариационная 71
 — корреляционная 71
 — наблюдений 70
 — ортогональных полиномов 88
Матриц транспонирование 70
Матрицы порядок 70
 — элемент 70
Медиана 19
Метод главных компонент 86

- максимального правдоподобия 83
- множественной регрессии 71
- Мода** 19
- Модели геологические** 9, 154
 - геолого-математические 15
 - горно-геометрические 12, 98
 - графические 12
 - двумерные статистические 46
 - многомерные статистические 69
 - объемно-макетные 11
 - одномерные статистические 16
 - типа случайных функций 119
 - физические 12
- Моделирования объекты** 12
 - принципы 11
- Моменты статистические** 20
- Опробование равномерное** 150
 - многостадийное 149
 - случайное 149
 - стратифицированное 149
- «Остатки» от тренда** 111
- Оценка выборочная** 17
 - дисперсии 26, 28
 - интервальная 27
 - максимально-правдоподобная 28
 - максимально эффективная 27
 - состоятельная 27
 - точечная 27
- Ошибка второго рода** 33
 - первого рода 33
- Переменная пространственная** 96
- Плотность вероятности**
 - логнормального распределения 23
 - нормального распределения 22
- Полиномы ортогональные** 108
 - тригонометрические 110
- Предел корреляции** 124
- Разделение дисперсии на составляющие** 130
- Разделение асимметричное** 21
 - биномиальное 22
 - выборочного среднего 29
 - логнормальное 22
 - нормальное 22
 - симметричное 20
 - Стьюдента 51
 - условное 47
 - Фишера 132
- Сглаживание исходных данных** 101
 - преобразованием Вулхауза 106
 - преобразованием Спенсера 107
 - скользящей корреляцией 107
 - скользящим окном 101
- Системы плохо организованные** 11
 - хорошо организованные 11
- Скаляр** 70
- Случайная величина** 18, 122
 - последовательность 103, 119
 - функция 119
- Случайная функция полигармоническая** 131
 - стационарная 123, 127
- Случайная функция эргодическая** 123, 125
- Случайное поле** 119, 122
- Случайной функции дисперсия** 121
 - корреляционная функция 121
 - математическое ожидание 121
 - реализация 119
 - сечение 120
- Совокупность генеральная** 16
 - геологическая 145
 - опробуемая 149
- Составляющие наблюдаемой изменчивости**
 - координированная 135
 - коррелированная 136
 - случайная 136
- Спектральная плотность дисперсии** 132
- Спектры амплитуд** 132
- Среднее арифметическое** 27
 - гармоническое 27
 - геометрическое 27
- Стандарт (среднее квадратичное отклонение)** 20
- Статистика, определение** 16
- Статистические гипотезы** 18, 32
- Топофункции** 98
- Тренд, анализ** 102
 - второго порядка 109
 - линейный 109
- Уравнение линейной регрессии** 47
- Уровень значимости** 33
- Уровень строения** 13
- Формализация геологических понятий** 7, 9
- Функция корреляционная (автокорреляционная)** 123
 - нормированная Лапласа 22
 - распределения дифференциальная 19
 - распределения интегральная 18
 - структурная 124
- Характеристики рассеяния** 19, 20
- Частота** 18
- Частота наблюдаемая** 18
- Шкала измерения** 147
 - номинальная 147
 - относительная 148
 - порядковая 147
- Элементы неоднородности** 13, 136
- Экссесс** 21

ОГЛАВЛЕНИЕ

	Стр.
Предисловие	3
Введение	5
§ 1. «Математика — новое оружие геолога»	5
§ 2. Особенности использования математических методов в геологии и разведке	6
§ 3. Краткая историческая справка об использовании математических методов в геологии	9
§ 4. Принципы и методы геолого-математического моделирования	11

Глава I

Одномерные статистические модели

§ 1. Сущность и условия применения одномерной статистической модели	16
§ 2. Основные понятия теории вероятностей	18
§ 3. Законы статистических распределений	21
§ 4. Выбор вида статистической модели	24
§ 5. Оценка числовых характеристик свойств геологических объектов	26
§ 6. Статистическая проверка геологических гипотез	32

Глава II

Двумерные статистические модели

§ 1. Сущность и основные числовые характеристики двумерной статистической модели	46
§ 2. Проверка гипотез о наличии корреляционной связи	50
§ 3. Использование корреляционных связей для предсказания свойств геологических объектов	63

Глава III

Многомерные статистические модели

§ 1. Метод множественной регрессии	71
§ 2. Метод линейной дискриминантной функции	79
§ 3. Метод главных компонент	86

Глава IV

Моделирование пространственных переменных

§ 1. Геологические объекты как поля пространственных переменных	96
§ 2. Горно-геометрические модели	98
§ 3. Модель П. Л. Каллистова	100
§ 4. Тренд-анализ	102
§ 5. Дискретные случайные поля	112

Глава V

Модели типа случайных функций

§ 1. Принципы моделирования свойств геологических объектов с помощью случайных функций	118
§ 2. Использование модели стационарной случайной функции для решения геологических задач	122
§ 3. Модель типа полигармонической случайной функции	130

Глава VI

Факторы, определяющие выбор

и эффективность использования математических методов в геологии

§ 1. Влияние типа геологической задачи на выбор математической модели	140
§ 2. Свойства геологических объектов, влияющие на выбор и эффективность использования математических моделей	144
§ 3. Влияние способа изучения объекта на выбор математической модели	148
§ 4. Роль геологического анализа при выборе геолого-математической модели	153

Список литературы	157
-------------------	-----

Приложения

1. Допустимые значения критерия Стьюдента при данном объеме выборки N и уровне значимости P	160
2. Значения критерия Фишера при уровне значимости $P=0,05$ для числа степеней свободы k_1 и k_2	160
3. Значения критерия Фишера при уровне значимости $P=0,01$ для числа степеней свободы k_1 и k_2	162
4. Функция нормального распределения с параметрами 0; 1 (для отрицательных значений z)	163
5. Функция нормального распределения с параметрами 0; 1 (для положительных значений Z)	163

Предметный указатель	165
----------------------	-----

ИБ № 2507

Алексей Борисович Каждан
Олег Иванович Гуськов
Альфред Антонович Шиманский

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

В ГЕОЛОГИИ И РАЗВЕДКЕ ПОЛЕЗНЫХ ИСКОПАЕМЫХ

Редактор издательства А. М. Поспелова

Обложка художника А. Е. Чучканова Художественный редактор Е. Л. Юрковская
Технические редакторы Б. А. Илясова, О. А. Болтунова Корректор И. Н. Таранева

Сдано в набор 01.08.78 Подписано в печать 15.03.79 Т-06033
Формат 60×90^{1/16}. Бумага № 3. Гарнитура литер. Печать высокая. Печ. л. 10,5
Уч.-изд. л. 11,0 Тираж 6500 экз. Заказ 1797/7148-14. Цена 40 коп.

Издательство «Недра». 103633, Москва, К-12, Третьяковский проезд, 1/19.

Московская типография № 6 Союзполиграфпрома
при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли.
109088, Москва, Ж-88, Южнопортовая ул., 24.