

Ю.А. ТКАЧЕВ

ТЕОРИЯ
И МЕТОДЫ ПОДСЧЕТА
ЗАПАСОВ
ПОЛЕЗНЫХ
ИСКОПАЕМЫХ

АКАДЕМИЯ НАУК СССР

КОМИ ФИЛИАЛ
ИНСТИТУТ ГЕОЛОГИИ

550 8.: 519.

Ю. А. ТКАЧЕВ

ТЕОРИЯ
И МЕТОДЫ ПОДСЧЕТА
ЗАПАСОВ
ПОЛЕЗНЫХ
ИСКОПАЕМЫХ

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

доктор геол.-минер. наук В. А. ДЕДЕЕВ



ЛЕНИНГРАД
«НАУКА»
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ
1988



5029

Ю. А. Ткачев. Теория и методы подсчета запасов полезных ископаемых. — Л.: Наука, 1988. — 224 с.

Книга посвящена применению математических методов и ЭВМ при обработке геологоразведочных данных, подсчете запасов и оптимизации плотности разведочной сети. В ней изложены корреляционно-регрессионные и другие методы обработки данных, направленные на использование максимума информации и повышение точности оценивания. Большое внимание уделено метрологии измерений и анализов, проблеме оптимальной и надежной обработки проб. Проанализированы категории промышленных запасов по точности, степени изученности и производственному назначению. Выявлены неизвестные ранее аспекты оптимизации геологоразведочных работ по этапам и стадиям, сформулированы критерии оптимальности и изложены практические методики, реализующие их. Каждый метод сопровождается одним или несколькими численными примерами, заимствованными из практики разведки месторождений нефти и рудных полезных ископаемых. Имеются ссылки на разработанные в Институте геологии Коми филиала АН СССР программы на языке ФОРТРАН-IV.

Работа адресована специалистам поисково-разведочного профиля и специалистам по применению математических методов и ЭВМ в геологии.

Библ. 230 назв. Ил. 47. Табл. 63.

Рецензенты:

Л. В. ГРИГОРЬЕВА, А. А. ПРЕДОВСКИЙ

T $\frac{1904050000-572}{042(02)-88}$ 213-87—II

ISBN 5-02-024430-9

© Издательство «Наука», 1988 г.

ПРЕДИСЛОВИЕ

Подсчет запасов месторождений полезных ископаемых в широком смысле слова представляет собой тот фокус, в котором пересекается пучок самых различных отраслей знаний. Качественно и надежно его можно провести только при хорошем знании геологического строения месторождения, квалифицированном применении методов математической обработки данных на ЭВМ, соблюдении экономических принципов рационального освоения месторождений. В современных условиях научно-технического прогресса руководитель подсчета запасов должен обладать знаниями в области геологии, экономики, математики, информатики и метрологии.

Передача месторождения из разведки в разработку — самый решающий момент в практическом освоении недр, связанный с известным риском, полностью устранить который можно только ценой огромных дополнительных затрат. Нет одинаковых месторождений, поэтому подсчет запасов всегда является творческим актом, который по существу приравнивается к научной геологической работе. В небольшом объеме, явно недостаточном для поставленной задачи, автор попытался рассмотреть различные аспекты разведки на предмет повышения ее достоверности с минимальными затратами. Для этого потребовалось касаться анализа погрешностей измерений, обработки проб и их анализов, погрешностей усреднения данных по отдельным горным выработкам, блокам и месторождению в целом, анализа категорий запасов и их точности, стадийности геологоразведочных работ и различных сторон оптимизации детальности разведки на различных стадиях. Хотя все эти вопросы находятся (и, как надеется автор, изложены) в тесной связи друг с другом, автор не хотел бы создать у читателя представления, что изложением охвачены все составные части затронутых вопросов и все методы их решения. Напротив, их выбор продиктован интересами или возможностями автора.

Прежде автор работал в области подсчета запасов твердых полезных ископаемых, а в последние 10 лет — преимущественно нефти, благодаря чему столкнулся с двумя недостаточно пересекающимися потоками литературы — нефтяной и «твердо-рудной». Поэтому в книге при выборе численных примеров, терминологии и ссылок на источники сознательно не делалось предпочтения какому-либо типу полезного ископаемого.

Работа выполнена в Институте геологии Коми филиала АН СССР при энергичном содействии дирекции и является общей частью разрабатываемой проблемы автоматизированной обработки геологоразведочных данных и подсчета запасов нефти в условиях региона. В процессе выполнения работы и написания этой книги автор постоянно пользовался консультациями и советами М. Д. Белонина (ВНИГРИ), А. Б. Каждана (МГРИ), Ю. В. Рошина (ВНИИХТ), В. П. Якуцени (ИГ Коми филиала АН СССР, ВНИГРИ), Ф. А. Ципорина (ВНИГРИ). Значительную работу по редактированию рукописи провел В. А. Дедеев (ИГ Коми филиала АН СССР). Большая помощь в подготовке рукописи и реализации численных примеров была оказана Т. И. Ивановой, а в оформлении работы — В. А. Носковым и С. Ю. Перетягиной (ИГ Коми филиала АН СССР). Всем перечисленным лицам автор выражает большую благодарность.

Глава 1

ВВЕДЕНИЕ

Разведка месторождений полезных ископаемых требует значительных материальных ресурсов и денежных средств. Отдача от них начинается спустя много лет после начала изысканий. Большую часть этих средств поглощает детальная разведка, заканчивающаяся подсчетом запасов и передачей месторождения в промышленное освоение. Повышение надежности разведки, снижение погрешностей оценки запасов и других геологических и горно-технических параметров месторождений имеют важнейшее народнохозяйственное значение.

Как и в других отраслях, в геологической разведке увеличение положительного эффекта можно достигнуть либо экстенсивным путем, увеличивая объем работ, либо интенсивным, извлекая максимум информации из геологоразведочных данных. Эта книга посвящена геолого-статистическим методам увеличения точности оценивания основных параметров месторождений и в конечном итоге повышения достоверности разведки. Тема эта ни в коей степени не является новой: она существует с тех пор, как планомерно ведется разведка и разработка полезных ископаемых. Но каждый исторический период наполняет ее своим содержанием. По нашему мнению, три момента определяют новый взгляд на тему: разработка и широкое распространение математических идей и методов оптимизации производства, применение во всех сферах деятельности вычислительной техники, курс на интенсивные методы ведения хозяйства.

В книге мы не пытались систематически охватить все относящиеся к этой теме вопросы. В ней излагается ряд разработок, тесно связанных друг с другом и расположенных приблизительно в той последовательности, в какой геологоразведчик может использовать их при обработке геологоразведочных данных и подсчете запасов. Для понимания материала книги от читателя не требуется специальной математической подготовки, превышающей программу геологоразведочного вуза: внимательное прослеживание выкладок вполне достаточно. Знакомясь с обширной геолого-математической литературой и с тем, как практические геологи применяют изложенные в ней результаты, мы пришли к следующему выводу: самым слабым местом применения математических методов в геологии является стык содержательного анализа решаемого вопроса с подбором адекватных математических средств и их специализация. Поэтому при изложении материала основное внимание мы уделяли анализу причин, по которым тот или иной метод может или не может быть применен.

Схема, которой мы придерживались в работе, состояла в следующем: а) постановка задачи в геологических терминах (изложение геологической задачи), б) математическая формулировка задачи, в) рассуждения и выкладки для получения решения, г) составление основной программы, реализующей решение на ЭВМ или микрокалькуляторе, д) составление подсобных программ, моделирующих изучаемую ситуацию (с известным по модели ответом), е) обработка смоделированных данных по основной программе и сравнение полученных

результатов с ответом, ж) обработка по основной программе практического примера, как правило, упрощенного.

Когда среди известных нам математических методов не находилось подходящих для решения поставленной задачи, мы моделировали ситуацию и находили приближенное решение имитацией многократного выборочного исследования модели. Для решения многих сложных задач геологической разведки такой метод нам представляется единственно возможным. Он, в частности, был реализован при оптимизации обработки геологических проб для изучения изменения контрастности руды при ее дроблении. Этот же метод лежит в основе предлагаемого алгоритма оптимизации плотности детальной разведки.

Мы считаем необходимым заранее высказать свое отношение к геостатистике. Хотя в работе ее методы в явном виде не были использованы и книгу в целом можно назвать «догеостатистической», идеи геостатистики использовались в упрощенных вариантах постоянно. Какой бы вопрос ни решался, по мере движения от абстрактных схем к реальности все острее нами ощущалась необходимость в применении геостатистических методов. Последние представляют особую тему, достойную многих книг, недостаток которых остро ощущается в советской литературе.

Пути повышения эффективности геологической разведки условно можно подразделить на технические, методические и организационно-экономические. В книге рассматриваются только две последние. Методический путь заключается в разработке и применении таких методов обработки геологоразведочных данных, которые позволяют извлечь из них максимум информации и повысить точность подсчета запасов без увеличения объема геологоразведочных работ. Организационно-экономический путь заключается в выборе такой степени разведанности, которая дает с учетом последующей разработки максимальный народнохозяйственный эффект. В соответствии с этим книгу также можно подразделить на две части.

В первой части разработан комплекс методов повышения достоверности оценки параметров при разведке месторождений нефти. Он впервые в едином плане учитывает большую часть источников погрешностей начиная от измерений и анализов до геометризации и итоговых расчетов. На основе новой прикладной концепции уравнения регрессии разработаны корреляционные методы объединения геологоразведочной информации разной достоверности и по несовпадающим объектам измерений с учетом цифровой и качественной априорной информации о зависимостях. Впервые поставлен и решен вопрос об оптимизации схем обработки геологических проб, обеспечивающих минимум затрат энергии при заданной точности. Предложены принцип доукомплектования геологоразведочных выборок и методика для его реализации на основе корреляционных методов, разработаны методы статистической обработки таких выборок (сфинкс-выборки).

Применительно к геологии и разведке месторождений разработаны и широко иллюстрированы примерами методы оценивания сложных объектов и их погрешностей с помощью выборок сложной структуры и разной представительности. Разработаны методы оценивания параметров, являющихся функциями коррелирующих величин.

Во второй части впервые выявлены и детально проанализированы различные аспекты оптимизации разведки. Для каждого из аспектов сформулированы критерии оптимизации, разработаны методики и составлены реализующие их алгоритмы. Практически для каждого аспекта оптимизации приведен практический численный пример.

Все изложенные методы повышения достоверности оценки запасов и оптимизации месторождений использованы как составные части автоматизированного комплекса обработки геологоразведочных данных и подсчета запасов нефти для условий Тимано-Печорского региона.

Глава 2

КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЕ МЕТОДЫ ПОВЫШЕНИЯ ДОСТОВЕРНОСТИ ОЦЕНКИ ПАРАМЕТРОВ МЕСТОРОЖДЕНИЙ

Среди многочисленных статистических методов корреляционно-регрессионный анализ занимает в геологии и разведке месторождений важное место, что связано с существом самого метода и с особенностями геологии как отрасли знания. Геологическое исследование состоит прежде всего в установлении зависимостей между различными явлениями и объектами. Эти зависимости затушеваны влиянием посторонних факторов, и для их распознавания необходимо применение статистических методов. Поэтому изучение корреляций и регрессий дает важную информацию для построения моделей геологических процессов и месторождений.

Статистические зависимости являются инструментом дополнительной, косвенной оценки и прогноза. Даже в отсутствие правдоподобных гипотез о явлениях наблюдаемые зависимости между изучаемыми величинами позволяют использовать эмпирические уравнения регрессии после соответствующей их обработки в качестве инструмента косвенного «измерения», когда прямые измерения экономически невыгодны, сопряжены с большими техническими трудностями, требуют большого времени ожидания (например, отработки месторождения) или недоступны.

Корреляционно-регрессионные методы оценки особенно важны в геологии и разведке месторождений, т. е. тогда, когда в распоряжении исследователя имеется большой набор величин (характеризующих то или иное геологическое свойство), измеренных различными методами с различной точностью. Согласование их друг с другом, получение комплексных оценок, повышающих точность подсчетных параметров и надежность выводов без дополнительных затрат на разведку, представляют собой специфические слабо разработанные вопросы практического применения математической статистики.

Оценка параметров (например, запасов нефти), являющихся функцией нескольких переменных, в условиях, когда не в каждой опробованной точке исследуемого пространства измерены или вычислены значения всех требующихся величин, также представляет собой важный, но слабо разработанный вопрос.

Наконец, важным для геологии и разведки месторождений является анализ изменения коэффициентов корреляции и параметров уравнений регрессии при переходе от одних выборочных элементов к другим, например при изменении массы отбираемых проб, при изменении объема части геологического тела, участвующего в геофизическом измерении, при переходе от значений в единичных пробах к средним значениям из n проб и т. д.

Теория оценок по регрессии и их погрешностей в математической статистике разработана. Однако она применима лишь в ограниченном числе наиболее простых случаев. Большинство практических задач геологии и разведки месторождений требует различных ее доработок и адаптации к конкретным условиям. В этом видит автор свою основную цель.

§ 1. Анализ и классификация ситуаций, возникающих при изучении и использовании статистических зависимостей в геологии и разведке месторождений

Использование уравнений регрессии в оценочных целях, т. е. для оценки одной случайной величины по значению другой (оценки по регрессии, как принято писать в литературе по математической статистике) — один из основных аспектов применения корреляционно-регрессионного анализа в геологии и других естественных науках. Несмотря на это, в литературе по математической статистике почти не уделяется внимания анализу ситуаций его применения. Некоторые из них в математической статистике вообще не рассмотрены, и рекомендаций для них не дано (см., например, У. Кокрен, 1976, с. 208).

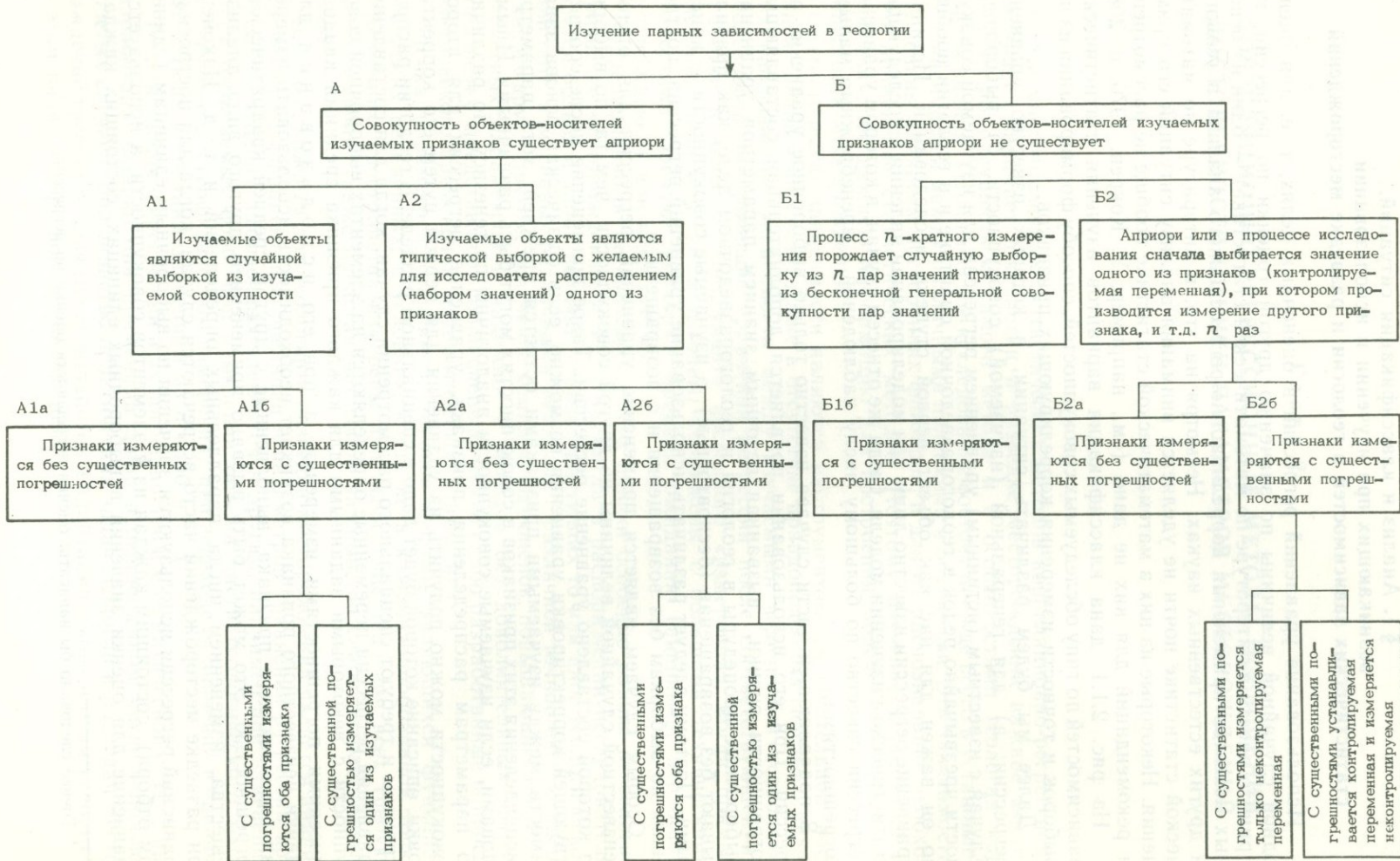
На рис. 2.1.1 дана классификация вариантов изучения статистических зависимостей по типу обследуемых совокупностей, способу формирования из них выборок и точности измерения коррелирующих признаков.

Далее мы будем различать ситуации, в которых известно уравнение регрессии: а) для генеральной (изучаемой) совокупности, б) выборочное. Случай с известным (истинным) уравнением регрессии для изучаемой совокупности чрезвычайно редок в геологоразведочной практике и в геологии вообще, но он важен для нас как объект первой степени исследований. Истинное уравнение регрессии известно лишь при моделировании явления и последующем статистическом изучении модели. Сюда же отнесем случай, в котором уравнение регрессии известно по большому числу наблюдений с пренебрежимо малыми погрешностями.

В подавляющей части случаев известно только выборочное уравнение регрессии. При его использовании появляется дополнительная составляющая погрешности оценки,¹ вызванная неточным знанием параметров уравнения. Выборочные процедуры в геологии и геологоразведочном деле, как правило, бывают без возвращения (бесповторные), а изучаемые совокупности — конечными. Поэтому следует различать использование уравнений регрессии, построенных по выборкам без возвращения и с возвращением.

Особым случаем является применение уравнений регрессии для оценки неизвестной случайной величины из другой совокупности, чем та, по выборке из которой составлено уравнение регрессии. Такое применение после соответствующей корректировки уравнения возможно, если физическая основа зависимости между изучаемыми признаками остается прежней, хотя параметры распределения этих признаков в совокупностях могут быть различными. Иными словами, если изучаемые совокупности аналогичны по генезису, но различны по параметрам распределения признака, уравнение регрессии для второй совокупности можно получить из уравнения для первой путем его корректировки, описание которой будет дано в специальном разделе. В геологии распространены и требуют специального рассмотрения случаи, когда для составления уравнения регрессии переменные определяются на элементах выборочной совокупности с некоторыми заданными для каждого признака средними квадратическими погрешностями измерения, а при его использовании для оценки неизвестного признака возникает необходимость использовать измерения оцениваемого признака, выполненные с другой средней квадратической погрешностью. Это может быть вызвано применением другого вида анализа вещества, изменением числа параллельных определений и т. д. Наконец, при разведке месторождений часто встречаются случаи, когда для построения уравнений регрессии используются значения по выборочным единицам (единицам отбора), состоящим каждая из k элементов совокупности, а используется уравнение для оценки значений в выборочных единицах, состоящих каждая

¹ Точнее следовало бы написать: оценки погрешности оценки по регрессии.



из m элементов. Так в нефтяной геологии и геохимии используют «сборные пробы», состоящие из смеси индивидуальных.

Многие из перечисленных выше случаев использования уравнений регрессии в математической статистике еще не описаны и требуют разработки соответствующих рекомендаций. Обстоятельством, влияющим на применение тех или иных формул для расчета погрешностей оценок, является также цель, с которой производится оценка по регрессии. В простейшем случае оцененное значение используется только для нужд, связанных с данным элементом изучаемой совокупности: данной пробой, данным керновым образцом. В более сложном случае оцененные по уравнению регрессии значения дополняют выборочную совокупность величин, полученных непосредственными измерениями, и служат наравне с ними различным целям — главным образом для оценки параметров изучаемой совокупности: ее среднего значения, дисперсии, коэффициента корреляции и погрешностей оценки средних значений.

Таким образом, мы рассматриваем уравнение регрессии в качестве инструмента оценивания наравне с другими измерительными приборами, наделенного определенной погрешностью измерения в зависимости от тесноты зависимости между оцениваемой и оценивающей величинами и от различных перечисленных выше обстоятельств. Далее будет показано с помощью выкладок и подтверждено экспериментами, что, несмотря на большую в среднем погрешность оценки по регрессии по сравнению с прямыми измерениями, она (оценка по регрессии) существенно уточняет изучаемые параметры при разведке месторождений.

В теории выборочного метода известны так называемые оценки по отношению и оценки по регрессии. В геологии такие оценки фактически применялись уже давно, например при подсчете запасов элементов-примесей в рудах. Однако это носило эпизодический характер, не опиралось на какую-либо теорию, и потенциальные возможности метода не использовались.

§ 2. Оценка геологических переменных по истинному уравнению регрессии

Под истинным уравнением регрессии мы подразумеваем уравнение с параметрами, подсчитанными по всей генеральной (изучаемой) совокупности. Его построение требует измерения обеих коррелирующих величин в каждом элементе изучаемой совокупности. В ней не остается элементов, в которых мы могли бы с пользой для дела произвести оценку интересующего нас признака по регрессии. Тем не менее рассмотреть ее важно по следующим причинам: а) она наиболее проста и способствует анализу более сложных случаев оценки; б) нередки случаи использования (с коррективами, которые будут рассмотрены в дальнейшем) истинного уравнения регрессии на аналогичные геологические совокупности; в) при экспериментальной работе с математическими моделями месторождений уравнение регрессии известно как вытекающее из принятой модели.

Рис. 2.1.1. Основные ситуации, возникающие при изучении парных корреляционных зависимостей в геологии и разведке месторождений.

Ситуация *A1* наиболее типична при выборочных исследованиях в геологии. Ситуация *A2* — при систематическом опробовании, при формировании специальных совокупностей, например «хорошо опробованных интервалов» или «хорошо проанализированных образцов» для сопоставления керновых и геофизических данных по пористости при подсчете запасов нефтяных месторождений, для эталонирования геофизических методов. Ситуация *B1* соответствует изучению коррелированных погрешностей измерений двух величин или изучению зависимости между величинами в нерегулируемом эксперименте. Ситуация *B2* характерна в аналитической практике при построении градуировочных графиков и при изучении физических зависимостей в контролируемых экспериментах.

Пусть известно истинное уравнение регрессии

$$M(y/x) = \alpha + \beta x. \quad (2.2.1)$$

Требуется найти среднюю квадратическую погрешность оценки по регрессии переменной y . В дальнейшем будем различать среднюю квадратическую погрешность оценки y в заданном элементе совокупности ($\sigma_{\hat{y}/x_0}$) с известным значением $x=x_0$ и среднее значение средней квадратической погрешности оценки, вычисленное по всем элементам изучаемой совокупности ($\sigma_{\hat{y}/x}$).

Для нормально распределенных коррелированных величин имеет место равенство $\sigma_{\hat{y}/x_0} = \sigma_{\hat{y}/x}$. Оно становится очевидным, если учесть, что дисперсия условного распределения случайной величины y в случае так называемой нормальной корреляции не зависит от значения x .

Величина $\sigma_{\text{ост}}^2 = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$ (где N — объем изучаемой совокупности,

y_i — значение величины y в i -м элементе совокупности, \hat{y}_i — оценка этого значения по регрессии) есть остаточная дисперсия, которая и определяет искомую погрешность

$$\sigma_{\hat{y}/x_0} = \sigma_{\hat{y}/x} = \sigma_{\text{ост}} = \sigma_y \sqrt{1 - \rho_{xy}^2}, \quad (2.2.2)$$

где σ_y^2 — дисперсия величины y в изучаемой совокупности, ρ_{xy} — коэффициент корреляции. Иногда формулу (2.2.2) удобнее использовать в виде

$$\sigma_{\hat{y}/x} = \beta_{yx} \sigma_x \sqrt{\frac{1}{\rho_{xy}} - 1}, \quad (2.2.3)$$

где β_{yx} — угловой коэффициент уравнения регрессии y на x .

На рис. 2.2.1 дана иллюстрация изложенных положений. Они известны из математической статистики, но в геологии применяются не всегда корректно. В частности, вывод о том, что средняя квадратическая погрешность оценки по регрессии постоянна при всех значениях оцениваемой величины, верный только для истинного уравнения регрессии, обычно распространяется на все другие случаи.

Рассмотрим влияние погрешностей измерений (анализов) на изменение коэффициента корреляции между наблюдаемыми значениями изучаемых величин и параметров уравнений регрессии. Математическая сторона этого вопроса обсуждалась М. С. Бартлетом (Bartlett, 1948) и Д. В. Линдли (Lindley, 1953). В литературе по геохимии формула для коэффициента корреляции между истинными содержаниями компонентов (если бы они были определены без погрешностей) впервые приведена у А. И. Гавришина и Г. А. Вострокнутова (1971). Подробный анализ явления на геологических примерах дан Ю. А. Ткачевым и Я. Э. Юдовичем (1975). Оно заключается в уменьшении коэффициента корреляции $\rho_{\text{погр}}$ между измеренными значениями коррелированных величин с увеличением погрешностей измерения $\sigma_{\text{ан } x}$, $\sigma_{\text{ан } y}$:

$$\rho = \rho_{\text{погр}} \frac{\sqrt{\sigma_x^2 + \sigma_{\text{ан } x}^2}}{\sigma_x} \cdot \frac{\sqrt{\sigma_y^2 + \sigma_{\text{ан } y}^2}}{\sigma_y}, \quad (2.2.4)$$

$$\rho_{\text{погр}} = \rho \frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_{\text{ан } x}^2}} \cdot \frac{\sigma_y}{\sqrt{\sigma_y^2 + \sigma_{\text{ан } y}^2}}, \quad (2.2.5)$$

в уменьшении угловых коэффициентов прямых регрессии и увеличении угла между этими прямыми. Погрешность оценки по регрессии при этом увеличивается. Однако специально заботиться об учете погрешностей измерений приходится лишь тогда, когда в элементах выборки, примененной для составления уравнения регрессии, измерения проводились с погрешностью $\sigma_{\text{ан } x}$, а при

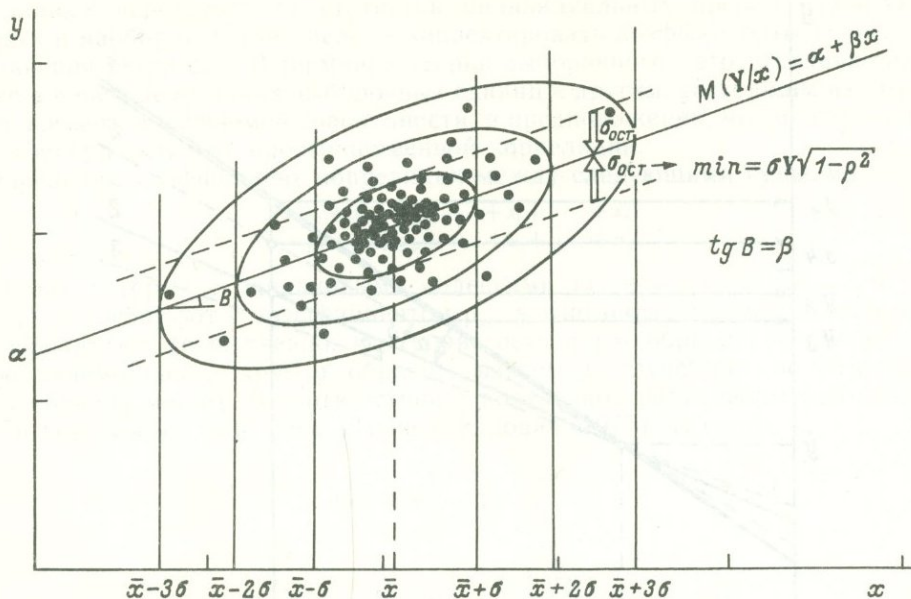


Рис. 2.2.1. Эллипсы рассеивания при нормальной корреляции и остаточное среднее квадратическое отклонение $\sigma_{ост}$ относительно прямой регрессии.

его использовании для оценки другой переменной мы располагаем результатами измерения величины x с другой погрешностью $\sigma_{ан2, x}$, т. е. $\sigma_{ан1, x} \neq \sigma_{ан2, x}$. Уравнение регрессии здесь должно быть скорректировано: угловой его коэффициент β уменьшается, если $\sigma_{ан2} > \sigma_{ан1}$, и увеличивается, если $\sigma_{ан2} < \sigma_{ан1}$. Новое его значение β_2 будет равно

$$\beta_2 = \beta_1 \frac{\sigma_x^2 + \sigma_{ан1, x}^2}{\sigma_x^2 + \sigma_{ан2, x}^2}, \quad (2.2.6)$$

где σ_x^2 — дисперсия величины x , если бы последняя была измерена без погрешностей.

Соответственно изменится и погрешность оценки: увеличится, если $\sigma_{ан2, x} > \sigma_{ан1, x}$ и уменьшится в противоположном случае:

$$\sigma_{\hat{y}/x}^2 = \sigma_y^2 \sqrt{1 - \rho^2 \left(\frac{\sigma_x^2 + \sigma_{ан1, x}^2}{\sigma_x^2 + \sigma_{ан2, x}^2} \right)}. \quad (2.2.7)$$

Эти соотношения показаны на рис. 2.2.2, из которого видно, что оценка \hat{y}/x будет тем больше сдвигаться к среднему значению \bar{y} в выборке, по которой составлено уравнение регрессии, чем больше погрешность анализа оценочной величины x . Напомним, что сейчас мы рассматриваем случай, когда выборка достаточно велика для того, чтобы пренебречь погрешностями в оценке параметров уравнения регрессии.

Частным случаем описанного является использование n -кратных измерений x как средства повышения точности оценки по регрессии. При этом $\sigma_{ан2} = \sigma_{ан1} / \sqrt{n}$ и эффективность повышения точности целиком зависит от того, какую долю общей наблюдаемой дисперсии величины x составляет дисперсия анализа. Пусть известны истинное уравнение регрессии и коэффициент корре-

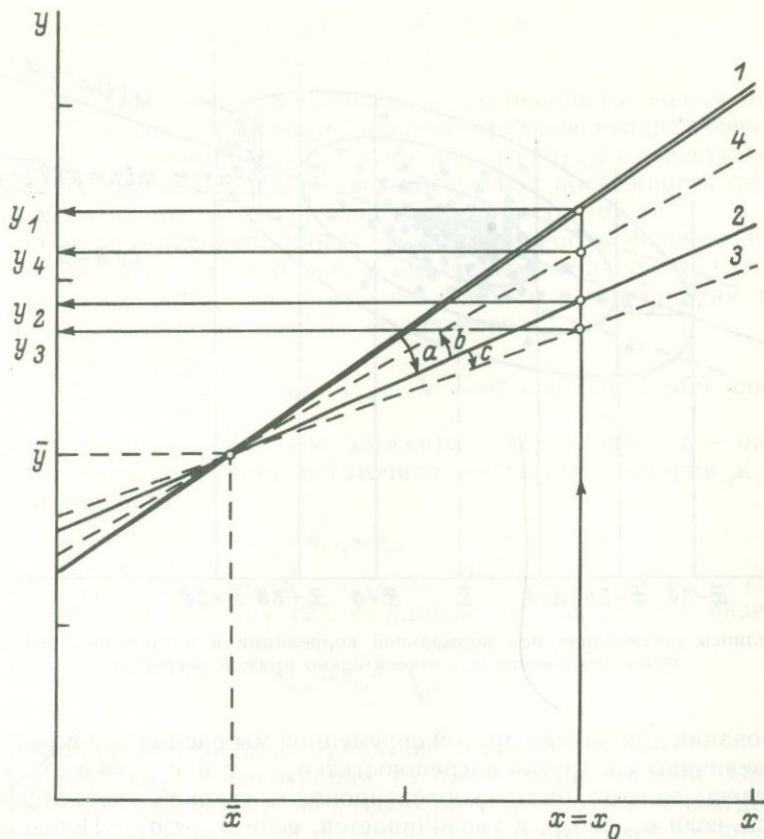


Рис. 2.2.2. Корректировка уравнения регрессии при изменении погрешностей анализа величины, по которой производится оценка.

1 — уравнение регрессии y на x при условии, что величина x измерена без погрешностей; 2 — уравнение регрессии, если погрешность измерения x составляет $\sigma_{ан1}$; 3 — скорректированное уравнение регрессии, если $\sigma_{ан2} > \sigma_{ан1}$; 4 — то же, если $\sigma_{ан2} < \sigma_{ан1}$; y_1, y_2, y_3, y_4 — различные оценки y при заданном значении x .

ляции между наблюдаемыми величинами (ρ). Поставим следующий вопрос: изменится ли погрешность оценки по регрессии, если нам станут известны погрешности, с которыми измерялись изучаемые величины? Точнее: всегда ли расчетная погрешность оценки по регрессии (формула (2.2.2)) соответствует фактической? Оказывается, не всегда, а информация о погрешностях измерений x и y позволяет привести их в соответствие. Для этого формула (2.2.2) преобразуется следующим образом:

$$\sigma_{y/x}^2 = \sigma_y' \sqrt{1 - \rho'^2}. \quad (2.2.8)$$

где $\sigma_y' = \sqrt{\sigma_y^2 - \sigma_{анy}^2}$, $\rho' = \rho \frac{\sigma_y}{\sigma_y'}$ (σ_y^2 — дисперсия измеренной с погрешностью $\sigma_{анy}$ оцениваемой случайной величины y). Поскольку реально мы располагаем только тем, о чем нам известно, т. е. тем значением точности, которое мы рассчитали по формуле (2.2.2) или (2.2.8), то приходим к выводу, что знание погрешностей измерений приводит к повышению точности оценки по регрессии.

В геохимической практике и в геологоразведочном деле часто пользуются объединенными или «сборными» пробами, включающими от двух до нескольких десятков индивидуальных проб. Отсюда возникает необходимость при сравне-

нии данных переходить от статистик индивидуальных проб к статистикам сборных и наоборот, в том числе — корректировать коэффициенты корреляции и уравнения регрессии. В терминах теории выборочного метода задача заключается в переходе от одних выборочных единиц к другим, состоящим из другого числа элементов изучаемой совокупности, в предположении, что между элементами совокупности нет пространственной корреляции.

Определим коэффициент корреляции между следующими суммами:

$$z = x_1 + x_2 + x_3 + \dots + x_i + \dots + x_n,$$

$$u = y_1 + y_2 + y_3 + \dots + y_i + \dots + y_n.$$

слагаемые которых с одинаковыми индексами (x_1 и y_1 , x_2 и y_2 , ..., x_i и y_i , x_n и y_n) коррелируют с коэффициентом ρ_{xy} , а слагаемые с разными индексами не коррелируют (естественно, что пористость в i -м образце не коррелирует с пористостью в другом образце, взятом в случайном по отношению к i -му образцу месте). В соответствии с полученной ранее формулой коэффициента корреляции для сумм (Ткачев, Юдович, 1975) получим

$$\rho_{zu} = \rho_{xy} \sum_{i=1}^n \frac{\sigma_{x_i}}{\sigma_z} \cdot \frac{\sigma_{y_i}}{\sigma_u}.$$

Но

$$\sigma_z = \sqrt{\sum \sigma_{x_i}^2} = \sigma_x \sqrt{n}, \quad \sigma_u = \sqrt{\sum \sigma_{y_i}^2} = \sigma_y \sqrt{n},$$

поэтому

$$\rho_{zu} = \rho_{xy} \sum_{i=1}^n \frac{\sigma_x}{\sigma_x \sqrt{n}} \cdot \frac{\sigma_y}{\sigma_y \sqrt{n}} = \rho_{xy}. \quad (2.2.9)$$

Угловой коэффициент уравнения регрессии также останется прежним:

$$\beta_{zu} = \frac{\sigma_z}{\sigma_u} \rho_{zu} = \frac{\sigma_y / \sqrt{n}}{\sigma_x / \sqrt{n}} \rho_{xy} = \beta_{yx}. \quad (2.2.10)$$

Этот результат означает, что коэффициент корреляции между средними значениями из однопипных слагаемых равен коэффициенту корреляции между самими переменными. Сравним теперь два пути оценки по регрессии среднего значения \bar{y}_n в группе из n проб: а) n раз по уравнению регрессии y на x в индивидуальных пробах с последующим усреднением результатов; б) один раз по уравнению регрессии, составленному по сборным пробам. Используя первый путь, получим следующую погрешность оценки отдельного значения y : $\sigma_{\hat{y}/x} = \sigma_y \sqrt{1 - \rho_{xy}^2}$ и

следующую погрешность оценки среднего по группе из n проб:

$$\sigma_{\bar{y}_n/x} = \sigma_{y/x}/n = \sigma_y \sqrt{(1 - \rho_{xy}^2)/n}. \quad (2.2.11)$$

Используя второй путь, получим такую же погрешность оценки среднего \bar{y}_n :

$$\sigma_{\bar{y}_n/x} = \sigma_{\bar{y}} \sqrt{(1 - \rho_{\bar{y}_n \bar{y}}^2)} = \sigma_y \sqrt{(1 - \rho_{xy}^2)/n}. \quad (2.2.12)$$

При подсчете запасов сопутствующих компонентов среди практиков часто возникает вопрос: что использовать — сборные или индивидуальные пробы? Изложенное выше показывает, что оба варианта равноценны, если используемое общее число элементов совокупности единичных проб остается прежним.

В процессе разведки месторождений не всегда обеспечиваются идентичность выборочных единиц и неизменность погрешностей измерений. Поэтому в каждом случае применения уравнений регрессии для оценок необходимо проводить тщательный анализ ситуации. Лишь полное сохранение неизменными всех усло-

вий, имевших место при построении уравнения регрессии, обеспечивает корректное его применение при оценке. Любое изменение условий требует корректировки параметров уравнений, т. е. формул оценок переменных и их погрешностей.

§ 3. Оценка геологических переменных по выборочному уравнению регрессии

Трудно предположить, что в распоряжении геологоразведчика будет иметься требующееся для него уравнение регрессии, построенное по данным всей изучаемой совокупности. Нельзя исчерпать пробами все месторождение, а если это и было бы возможно, то применение уравнения регрессии потеряло бы смысл. Основной, наиболее распространенный в практике случай — оценка изучаемой переменной по **выборочному** уравнению регрессии, построенному, как правило, по выборке, составляющей незначительную долю совокупности.

При использовании выборочного уравнения регрессии требуется знание а) погрешностей¹ оценок параметров уравнения — углового коэффициента прямой регрессии, называемого также коэффициентом регрессии ($s_{b_{yx}}$), и свободного члена ($s_{a_{yx}}$); б) погрешности оценки среднего значения оцениваемой величины \bar{y}/x при заданном значении оценивающей величины x , обозначаемой $s_{\hat{y}/x}$; в) погрешности оценки величины, указанной в (б), но не при заданном значении x , а в среднем по изучаемой совокупности $\bar{s}_{\hat{y}/x}$; г) погрешности оценки индивидуального значения y при заданном значении x ($s_{\hat{y}/x}$); д) погрешности оценки величины, указанной в (г), но не при заданном значении оценивающей величины x , а в среднем по изучаемой совокупности ($s_{\hat{y}/x}$).

В математической статистике основное внимание уделяется анализу погрешностей положения прямой регрессии, т. е. пункту (а). Напротив, при обработке геологоразведочных данных и подсчете запасов используются главным образом погрешности $s_{\hat{y}/x}$ и $s_{\hat{y}/x}$. Упомянутые виды погрешностей в прикладной геологической литературе часто путают, поэтому в процессе изложения мы уделили особое внимание содержательному анализу их смысла и области применения.

Остаточная дисперсия ($s_{\text{ост } y}^2$). Эта величина характеризует разброс y относительно выборочной прямой регрессии. По смыслу она соответствует величине $s_{\text{ост } y}^2$ из предыдущего параграфа и определяется по формуле

$$s_{\text{ост } y}^2 = s_y^2 (1 - r^2) \frac{n-1}{n-2}, \quad (2.3.1)$$

где s_y^2 и r — оценки дисперсии y и коэффициента корреляции между x и y , определенные из выборки объема n .

Погрешность коэффициента регрессии ($s_{b_{yx}}$) и свободного члена ($s_{a_{yx}}$).

Из математической статистики известно, что

$$s_{b_{yx}} = \frac{s_{\text{ост } y}}{s_x \sqrt{n-1}} = \frac{s_y \sqrt{1-r^2}}{s_x \sqrt{n-2}}, \quad (2.3.2)$$

$$s_{a_{yx}} = s_{\text{ост } y} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_x^2 (n-1)}}. \quad (2.3.3)$$

Для оценок погрешностей подсчетных параметров месторождений величины $s_{b_{yx}}$ и $s_{a_{yx}}$ непосредственно не употребляются, но могут быть полезны, например

¹ Здесь и далее подразумеваются среднеквадратические погрешности, если не указаны какие-либо иные.

при разделении пород-коллекторов нефтяных месторождений на литотипы. Предполагается, что для каждого литотипа коллектора характерны свои параметры зависимости, например между пористостью и проницаемостью. Проверку гипотезы о равенстве этих параметров проводят по известному критерию t , а для этого должны быть известны величины s_b и s_a .

Погрешность оценки индивидуального значения y при заданном значении x_i ($s_{\hat{y}/x_i}$). Эта погрешность наиболее важна при оценке параметров и подсчете запасов. В специальных руководствах по математической статистике формула для $s_{\hat{y}/x_i}$ в явном виде не приводится, а в прикладных руководствах для биологов, геологов и экономистов она иногда встречается, но везде записана по-разному. Так, у Л. Зака (1976) приведена следующая формула:

$$s_{\hat{y}/x_i} = s_y \sqrt{1-r^2} \sqrt{\frac{n-1}{n-2}} \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n-1)s_x^2}} = s_{\text{ост } y} \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n-1)s_x^2}}. \quad (2.3.4)$$

Погрешность индивидуальной оценки по регрессии состоит из трех слагаемых. Первое из них есть остаточное среднее квадратическое отклонение относительно выборочной прямой регрессии. В некоторых руководствах для геологов (Карлье, 1966) в этом слагаемом отсутствует множитель $\sqrt{(n-1)/(n-2)}$. При $n=2$ всегда выполняется равенство $|r|=1$, если исключить указанный множитель, то $s_{\hat{y}/x_i}=0$, что неверно. Таким образом, этот множитель необходим. Второе слагаемое отражает погрешность положения прямой регрессии в направлении оси y и равно $s_{\text{ост } y}/\sqrt{n}$. В некоторых руководствах приведено $s_{\text{ост } y}/\sqrt{n-1}$. Третье слагаемое отражает влияние погрешности угла наклона прямой регрессии

$$s_{\text{ост } y} \sqrt{\frac{(x_i - \bar{x})^2}{(n-1)s_x^2}} = s_{\text{ост } y} \frac{x_i - \bar{x}}{s_x} \cdot \frac{1}{\sqrt{n-1}}.$$

У Э. Карлье (1966) и в некоторых других руководствах вместо $\sqrt{n-1}$ указано \sqrt{n} .

Величина третьего слагаемого сильно зависит от значения x в том элементе совокупности, для которого оценивается y . Она возрастает прямо пропорционально отклонению x от среднего, т. е. для элементов совокупности, сильно отклоняющихся (по оцениваемой величине x) от среднего, оценка по регрессии мало надежна.

Средняя по совокупности погрешность оценки индивидуального значения ($s_{\hat{y}/.}$). Эта величина также находит широкое применение в геологии. Когда говорят о погрешности оценки по зависимости, не уточняя, обычно подразумевают эту величину. Формулу для нее можно получить, имея в виду, что $s_{\hat{y}/.}^2 = M\{s_{\hat{y}/x}^2\}$, т. е.

$$s_{\hat{y}/.}^2 = s_{\text{ост } y}^2 \left[1 + \frac{1}{n} + \frac{1}{(n-1)s_x^2} \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x) dx \right], \quad (2.3.5)$$

где $f(x)$ — плотность вероятности величины x . Интеграл в (2.3.5) есть дисперсия величины x в совокупности относительно выборочного среднего. Она равна $\sigma_x^2/n + \sigma_x^2 = \sigma_x^2(n+1)/n$. Заменяя σ_x^2 ее несмещенной оценкой s_x^2 , получим

$$s_{\hat{y}/.} = s_{\text{ост } y} \sqrt{1 + \frac{1}{n} + \frac{n+1}{n(n-1)}} = s_{\text{ост } y} \sqrt{\frac{n+1}{n-1}}. \quad (2.3.6)$$

Подставляя в (2.3.6) выражение для $s_{\text{ост } y}$, получаем

$$s_{\hat{y}/.} = s_y \sqrt{1-r^2} \sqrt{\frac{n+1}{n-2}}. \quad (2.3.7)$$

Сравнивая эту погрешность с соответствующей ей погрешностью оценки по истинному уравнению регрессии, замечаем, что она не намного больше даже для очень малых выборок. По формуле (2.3.7) для $n=3$ погрешность оценки по выборочному уравнению регрессии всего в 2 раза превышает погрешность оценки по истинному уравнению, для $n=5$ — немногим более чем в 1.4 раза, для $n=10$ — только в 1.17 раза.

Для проверки столь, на наш взгляд, неожиданных и оптимистичных для любителей оценок по регрессии результатов была проведена на ЭВМ обширная серия экспериментов «Модель». Она заключалась в моделировании совокупности ($N=200$) нормально распределенных коррелированных величин $\{x_i, y_i\}$ с заданными $\mu_x, \sigma_x, \mu_y, \sigma_y, \rho_{xy}, \alpha_{xy}, \beta_{xy}, \alpha_{yx}, \beta_{yx}$, где α и β — коэффициенты заданного по модели уравнения регрессии. В экспериментах мы варьировали всеми этими величинами. Из полученных совокупностей с помощью случайного механизма формировались выборки численностью от 3 до 200 элементов. По этим выборкам вычислялись необходимые статистики и составлялись выборочные уравнения регрессии y на x и x на y . В автоматическом же режиме вычислялись оценки по регрессии (обеих переменных) и их погрешности по формулам

$$s_{\hat{y}/x}^2 = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}/x_i - y_i)^2}, \quad (2.3.8)$$

$$s_{\hat{x}/y}^2 = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}/y_i - x_i)^2}.$$

Оценка по регрессии проводилась только для тех элементов совокупности, которые не использовались при составлении уравнений регрессии. Это максимально приближало условия эксперимента к обычным в геологии. Всего в экспериментах «Модель» было сделано 186 204 оценки по регрессии. Для сопоставления результатов всех экспериментов на последнем этапе обработки погрешности нормировались следующим образом:

$$A = s_{\hat{y}/x}^2 / (s_y \sqrt{1-r^2}),$$

$$B = s_{\hat{x}/y}^2 / (s_x \sqrt{1-r^2}), \quad (2.3.9)$$

так что нормированные погрешности (величины A и B) были инвариантны относительно изменения дисперсий случайных величин и коэффициентов корреляции между ними.

Окончательные результаты приведены в табл. 2.3.1. Помещенные в нее величины A и B являются по существу коэффициентами, на которые необходимо умножить величину $s_y \sqrt{1-r^2}$ (или $s_x \sqrt{1-r^2}$), чтобы получить среднюю погрешность оценки по уравнению регрессии, составленному по выборке из n элементов. Если формула (2.3.7) верна, то таким коэффициентом должно быть выражение $\sqrt{(n+1)/(n-2)}$. В табл. 2.3.2 и на рис. 2.3.1 дано сравнение экспериментальных и расчетных значений нормированной погрешности оценки по регрессии. Хорошее их совпадение наблюдается при $n \geq 20$, т. е. приведенные выше формулы для оценки погрешностей (2.3.4) и (2.3.7) могут быть рекомендованы только для не очень малых n . Для малых выборок нами вычислены

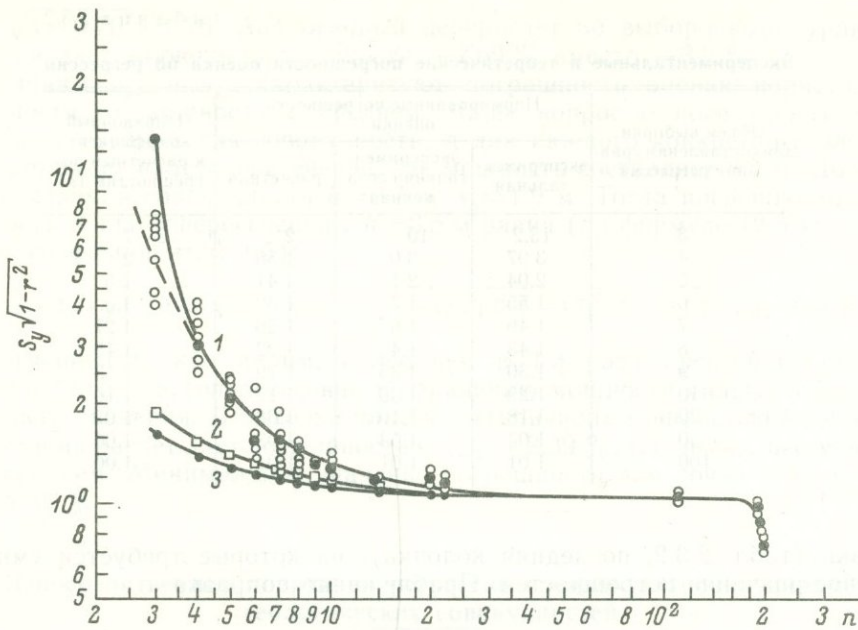


Рис. 2.3.1. Экспериментальные (кривая 1) и расчетные (кривая 2 — по формуле (2.3.7), кривая 3 — по формуле Э. Карлье) нормированные погрешности оценки по регрессии в зависимости от объема выборки n .

Таблица 2.3.1

5029
 Результаты имитационных экспериментов на ЭВМ по моделированию корреляции и изучению погрешностей оценки по выборочному уравнению регрессии

Объем выборок n	$K=11$		$K=11$		$K=50$		$K=23$		$K=23$		Среднее взвешенное по числу модельных выборок		Среднее из А и В
	$\rho_{xy}=0.6695$		$\rho_{xy}=0.6541$		$\rho_{xy}=0.4700$		$\rho_{xy}=0.213$		$\rho_{xy}=0.8818$				
	А	В	А	В	А	В	А	В	А	В	А	В	
3	7.65	7.43	6.65	6.98	23.7	23.4	5.52	5.44	4.022	4.325	13.24	13.16	13.2
4	3.88	3.48	2.60	2.68	2.98	3.08	—	—	—	—	3.06	3.08	3.07
5	2.38	2.19	2.22	2.40	1.96	1.91	1.90	1.93	1.934	2.356	2.01	2.07	2.04
6	1.36	1.32	2.23	1.82	1.52	1.46	—	—	—	—	1.60	1.49	1.55
7	1.33	1.51	1.23	1.30	1.58	1.54	1.34	1.57	1.414	1.386	1.44	1.49	1.46
8	1.35	1.26	1.31	1.34	1.48	1.53	—	—	—	—	1.43	1.46	1.45
9	—	—	1.28	1.34	1.28	1.32	—	—	—	—	1.28	1.32	1.30
10	—	—	1.24	1.19	1.31	1.24	1.37	1.27	1.316	1.319	1.32	1.26	1.29
14	—	—	1.125	1.106	1.19	1.20	—	—	—	—	1.18	1.18	1.18
20	—	—	—	—	—	—	1.13	1.17	1.232	1.207	1.18	1.19	1.18
22	—	—	1.079	1.084	1.12	1.13	—	—	—	—	1.11	1.12	1.11
50	—	—	—	—	1.03	1.05	—	—	—	—	1.03	1.05	1.04
70	—	—	—	—	1.01	1.00	—	—	—	—	1.01	1.00	1.00
109	—	—	0.991	1.05	—	—	—	—	—	—	1.00	1.05	1.05
130	—	—	—	—	1.01	1.00	—	—	—	—	1.01	1.00	1.00
150	—	—	—	—	0.94	1.03	—	—	—	—	0.94	1.03	0.99
180	—	—	—	—	0.97	1.03	—	—	—	—	0.97	1.03	1.00
190	—	—	—	—	0.92	0.993	—	—	—	—	0.96	—	0.96
193	—	—	—	—	0.93	0.955	—	—	—	—	0.94	—	0.94
195	—	—	—	—	0.80	0.92	—	—	—	—	0.86	—	0.86
197	—	—	—	—	0.82	0.86	—	—	—	—	0.84	—	0.84
199	—	—	—	—	0.77	0.69	—	—	—	—	0.73	—	0.73

Примечание. K — число выборок.

Экспериментальные и теоретические погрешности оценки по регрессии

Объем выборки для составления уравнения регрессии	Нормированные погрешности оценки			Поправочный коэффициент к расчетным погрешностям (λ)
	экспериментальная	экспериментальная сглаженная	расчетная	
3	13.2	10	2	6—7
4	3.07	3.0	1.58	2
5	2.04	2.1	1.41	1.5
6	1.55	1.7	1.32	1.3
7	1.46	1.6	1.26	1.27
8	1.45	1.4	1.22	1.15
9	1.30	1.35	1.19	1.13
10	1.29	1.30	1.17	1.10
20	1.18	1.10	1.08	1.02
50	1.04	1.03	1.03	1.00
100	1.01	1.01	1.01	1.00

поправки (табл. 2.3.2, последняя колонка), на которые требуется умножить расчетное значение погрешностей. Приблизленно поправка

$$\lambda = \sqrt{(n+1)/(n-1)}. \quad (2.3.10)$$

Пример. Пусть задача состоит в оценке пористости m нефтенасыщенного пласта песчаника по зависимости ее от мощности h . Представим, что в результате предыдущей разведки достаточно точно по очень большому числу данных известно уравнение регрессии пористости пласта на его мощность в данном пластопересечении. Коэффициент корреляции $\rho_{mh} = 0.70$, среднее квадратическое отклонение $\sigma_m = 2.1$. Средняя дисперсия оценки по (2.2.2) $\sigma_{\hat{y}/x} = 2.1 \sqrt{1 - 0.7^2} \approx 1.5$ абс. %. Видоизменим пример и будем считать, что параметры регрессии и другие статистики оценены по выборке объемом пять пластопересечений (определений). Применяя (2.3.7) с поправочным коэффициентом из табл. 2.3.2, получим

$$s_{\hat{y}/x} = 2.1 \sqrt{(1 - 0.7^2)(5+1)/(5-2)} \cdot 1.5 \approx 3.2 \text{ абс. \%}$$

Последняя величина ($\pm 3.2\%$) существенно превышает первую ($\pm 1.5\%$) из-за недостаточной надежности параметров уравнения регрессии ($n=5$).

Проведенный расчет является по существу недостаточно полным, неверным. Необходимо выяснить, какова погрешность оценки пористости по пластопересечениям, использованным при построении уравнения регрессии. Пусть она равна ± 0.8 абс. % (это значение нельзя считать завышенным, так как оно включает и погрешность анализа пористости в образцах, и погрешность усреднения в пластопересечении, которая при большой вертикальной изменчивости и малом числе образцов может быть и более значительной). В соответствии с (2.2.5) корректируем значения σ_m и ρ_{mh} :

$$\sigma'_m = \sqrt{\sigma_m^2 - \sigma_{ан м}^2} = \sqrt{2.1^2 - 0.8^2} = 1.94,$$

$$\rho'_{mh} = \rho_{mh} \sigma_m / \sigma'_m = 0.7 \cdot 2.1 / 1.94 = 0.76.$$

Теперь расчет по (2.2.2) дает существенно меньшее значение $\sigma_{\hat{y}/x} =$

$= 1.94 \sqrt{1 - 0.76^2} = 1.26$. Аналогичный перерасчет по выборочному уравнению регрессии дает значение погрешности $\pm 2.69\%$ вместо $\pm 3.2\%$.

Величины $\sigma_{\hat{y}_j}$ и $s_{\hat{y}_j}$ характеризуют погрешность оценки пористости по зависимости от мощности в среднем. Чаще вопрос о погрешности оценки встает не в среднем для совокупности, а для каждого конкретного элемента. Пусть это будет пластопересечение, где мощность составляет 8.5 м. Допустим, что в изучаемой совокупности $\bar{h} = 3.7$ м, $s_h = 1.9$ м. Тогда погрешность оценки пористости в пластопересечении с $h = 8.5$ м равна (по формуле (2.3.4) с поправочным коэффициентом 1.5):

$$s_{\hat{y}_{x=8.5}} = 1.94 \sqrt{1 - 0.76^2} \sqrt{(5-1)/(5-2)} \sqrt{1 + 1/5 + (8.5 - 3.7)^2 / (5-1)} \cdot 1.9^2 \cdot 1.5 = 3.63\%.$$

Эта погрешность, как видим, более чем в 1.3 раза превышает среднюю, ввиду того что пластопересечение по значению мощности сильно отклоняется от среднего значения. Для точки пласта с «типичной» мощностью $h = 4$ м соответствующий расчет дает погрешность $s_{\hat{y}_{x=4.0}} = 2.39\%$, т. е. несколько меньшую, чем в среднем. Минимальное значение погрешности для точки с $h = \bar{h} = 3.7$ м равно 2.38% .

§ 4. Непосредственная оценка по регрессии средних значений параметров геологических совокупностей

Возможны два принципиально различающихся способа использования уравнения регрессии при оценивании параметров изучаемых совокупностей. Один из них, более простой, заключается в непосредственном расчете по уравнению регрессии среднего значения изучаемой совокупности:

$$\hat{Y}_N = \bar{y}_n + b(\bar{X}_N - \bar{x}_n), \quad (2.4.1)$$

где \hat{Y}_N — оценка по регрессии среднего значения в совокупности объема N ; \bar{y}_n — среднее значение оцениваемой величины y в выборке объема n ; \bar{X}_N — среднее значение оцениваемой величины x в изучаемой совокупности; \bar{x}_n — ее среднее значение в выборке объема n . Второй способ заключается в оценке по регрессии значения y в каждом элементе совокупности с последующим расчетом средних значений. Он будет рассмотрен в § 7 этой главы.

Первый способ заключается по существу в корректировке выборочной оценки \bar{y}_n в большую или меньшую сторону в зависимости от среднего значения другой величины \bar{X}_N в совокупности в выборке. Этот способ недостаточно известен в геологии, хотя и применяется при обработке геологоразведочных данных, особенно в рудничной геологии. Геолог рассуждает при этом следующим образом: если разведочные данные оказались заниженными для величины x (например, мощности), то таковыми же они, вероятно, окажутся и для величины y (например, содержания), коррелирующей с x .

У. Кокрен (1976) приводит следующую формулу для погрешности оценки по регрессии средних значений в ограниченных совокупностях:

$$\sigma_{\hat{Y}_N} = \sigma_y \sqrt{1 - \rho_{xy}^2} \sqrt{(1-f)/n} = \sigma_{\text{ост } y} \sqrt{(1-f)/n}, \quad (2.4.2)$$

где f — коэффициент отбора ($f = n/N$); ρ_{xy} — коэффициент корреляции, известный для всей изучаемой совокупности. Формула (2.4.2) применима, если коэффициент регрессии b_{yx} также известен. Обычно ни ρ_{xy} , ни b_{yx} в изучаемой совокупности неизвестны, а оцениваются по выборке. Тогда формулу (2.4.2) необходимо изменить таким образом, чтобы вместо остаточной дисперсии $\sigma_{\text{ост } y}^2$ в ней стояла бы величина, аналогичная ей по смыслу. В (2.4.2) остаточная

дисперсия отражает среднюю погрешность оценки величины y по истинному уравнению регрессии. Других составляющих этой погрешности нет. При оценке по выборочному уравнению эта погрешность равна

$$s_{\hat{y}} \approx s_y \sqrt{(1-r^2)(n+1)(n-2)} \approx s_y \sqrt{(1-r^2)(n+1)^2/(n-1)(n-2)}.$$

Эта величина и должна быть подставлена в (2.4.2) вместо $\sigma_{\text{ост } y}$, что дает

$$s_{\hat{Y}_N} \approx s_y \sqrt{\frac{1-f}{n} (1-r^2)(n+1)^2/(n-1)(n-2)}. \quad (2.4.3)$$

В данной формуле n означает численность выборки, в элементах которой измерены значения x , y и по которой рассчитаны параметры уравнения регрессии и значение s_y .

Обращаем внимание (специальным индексом N) на то, что при расчете по этой формуле предполагается, что значение \bar{X}_N , входящее в формулу (2.4.1), точно известно для всей изучаемой совокупности. В разведке месторождений такие случаи иногда встречаются. Например, x может быть мощностью пласта, среднее значение которой достаточно точно известно, так как определено, допустим, по плотной сети скважин, пробуренных без отбора керна, в то время как y — содержание, определяемое по редкой сети скважин с отбором керна. Однако это скорее исключение, чем правило. Обычно величина x определяется, хотя и в большей по объему k ($k \geq n$) выборке, чем y , но далеко не исчерпывающей изучаемую совокупность. Покажем, что погрешность оценки по регрессии (s_y) в этом случае будет равна

$$s_{\hat{y}} = s_y \sqrt{\left(1 - \frac{k}{N}\right) \frac{1}{k} + \left(1 - \frac{n}{k}\right) \frac{1}{n} (1-r^2) \cdot \frac{(n+1)^2}{(n-1)(n-2)}}, \quad (2.4.4)$$

где N — объем изучаемой совокупности; k — объем выборки, в элементах которой измерены значения величины x ; n — объем подвыборки из нее, в которой измерены значения x , и y . Дисперсия оценки величины \bar{Y}_N состоит из двух слагаемых: дисперсии оценки величины \bar{Y}_k (среднего в выборке из k элементов) и дисперсии распространения этой оценки на всю совокупность. Первое слагаемое есть не что иное, как погрешность оценки по регрессии, для которой применима формула (2.4.2). Она равна $s_1^2 = (1-n/k)1/ns_y^2(1-r^2)(n+1)^2/(n-1)(n-2)$. Выражение для второго слагаемого очевидно: $s_2^2 = (1-k/N) \times \times 1/ks_y^2$. Таким образом, равенство (2.4.4) доказано.

Если в (2.4.4) положить $k=N$, что соответствует рассмотренному в самом начале случаю, то получаем

$$s_{\hat{Y}_N} = s_y \sqrt{\left(1 - \frac{n}{N}\right) \frac{1}{N} (1-r^2) \frac{(n+1)^2}{(n-1)(n-2)}},$$

что совпадает с известным результатом. Если положить $k=n$ (ни в одном элементе совокупности, кроме подвыборки объемом n , значение x не измерено), то $s_{\hat{Y}_N} = s_y \sqrt{(1-n/N)/n}$, т. е. оценка по регрессии исчезает, и мы имеем обычную оценку «по среднему на единицу». При $N > k > n$ и $r=0$, а также n , достаточно большим, чтобы пренебречь множителем $(n+1)^2/(n-1)(n-2)$, получаем

$$s_{\hat{Y}_n}^2 = \left(1 - \frac{k}{N}\right) \frac{s_y^2}{k} + \left(1 - \frac{n}{k}\right) \frac{s_y^2}{n} = s_y^2 \left(\frac{1}{k} - \frac{1}{N} + \frac{1}{n} - \frac{1}{k}\right) = s_y^2 \frac{1}{n} \left(1 - \frac{n}{N}\right),$$

т. е. оценка по регрессии также по существу исчезает, хотя формально в $k-n$ элементах уравнение регрессии использовалось для оценки. Прямая

регрессии при этом параллельна оси x , и для всех x мы получаем оценку $y = \bar{y}_n$. Можно поставить вопрос: существует ли некоторое минимальное значение $|\rho|$, ниже которого применение оценки по регрессии бесполезно или вредно? Для ответа на этот вопрос найти такое значение ρ , при котором выполняется условие

$$\left(1 - \frac{k}{N}\right) \frac{1}{k} + \left(1 - \frac{n}{k}\right) \frac{1}{n} (1 - r^2) \leq \left(1 - \frac{n}{N}\right) \frac{1}{N},$$

т. е.

$$1 - \rho^2 \leq \left(\frac{1}{n} - \frac{1}{N} - \frac{1}{k} + \frac{1}{N}\right) / \left(\frac{1}{n} - \frac{1}{k}\right) = 1,$$

откуда $\rho^2 \geq 0$. Таким образом, вопреки убеждению большинства практиков-геологов оценка по регрессии никогда не увеличивает погрешность, а при $|\rho| > 0$ уменьшает ее. Добавим при этом: если ρ и параметры уравнения регрессии установлены без погрешностей. В противном случае предыдущее условие следует заменить на следующее:

$$\left(1 - \frac{k}{N}\right) \frac{1}{k} + \left(1 - \frac{n}{k}\right) \frac{1}{n} (1 - r^2) \frac{(n+1)^2}{(n-1)(n-2)} \leq \left(1 - \frac{n}{N}\right) \frac{1}{N},$$

что дает $|r| \geq \sqrt{1 - (n-1)(n-2)/(n+1)^2}$, т. е. действительно существует минимальное значение модуля выборочного коэффициента корреляции, зависящее от n :

n	3	4	5	6	7	8	9	10	20	50	100
$ r $	0.935	0.872	0.816	0.769	0.730	0.694	0.653	0.636	0.470	0.310	0.221

Любопытно, что эти значения почти точно соответствуют критическим значениям r для уровня значимости $\alpha = 0.02$.

С учетом важности оценок по регрессии в геологии был проведен обширный ряд экспериментов. На ЭВМ с помощью датчика случайных чисел моделировалась совокупность объема $N=200$. Из нее случайным образом по схеме выборочной процедуры без возвращения формировались выборки различного объема n (от 20 до 180). Число выборок с заданным значением n обычно составляло 8.

На первом этапе эксперимента по каждой выборке проводился расчет \bar{y}_n , а по всем 8 выборкам — экспериментальная погрешность оценки значения Y_N с помощью выборочного среднего. Теоретические значения погрешностей определялись по формуле для оценки «по среднему на единицу»:

$$s_{\bar{y}_N} = s_y \sqrt{\left(1 - \frac{n}{N}\right) \frac{1}{n}}. \quad (2.4.5)$$

Результаты экспериментов приведены в табл. 2.4.1, а также на рис. 2.4.1. Они показали хорошее совпадение теоретических и наблюдаемых погрешностей оценок средних в ограниченных совокупностях. Результат этого эксперимента важен для геологической аудитории тем, что формула (2.4.5) вызывает в ней недоверие. Для нас же он важен прежде всего тем, что подтвердил корректность схемы программного комплекса самого эксперимента, который в дальнейшем используется для апробации не столь очевидных формул, как эта.

На втором этапе экспериментов «Имопр» моделировалась оценка по регрессии, когда параметры зависимости вычислялись по выборке объема n , а значения оцениваемой величины x «измерялись» во всех остальных элементах

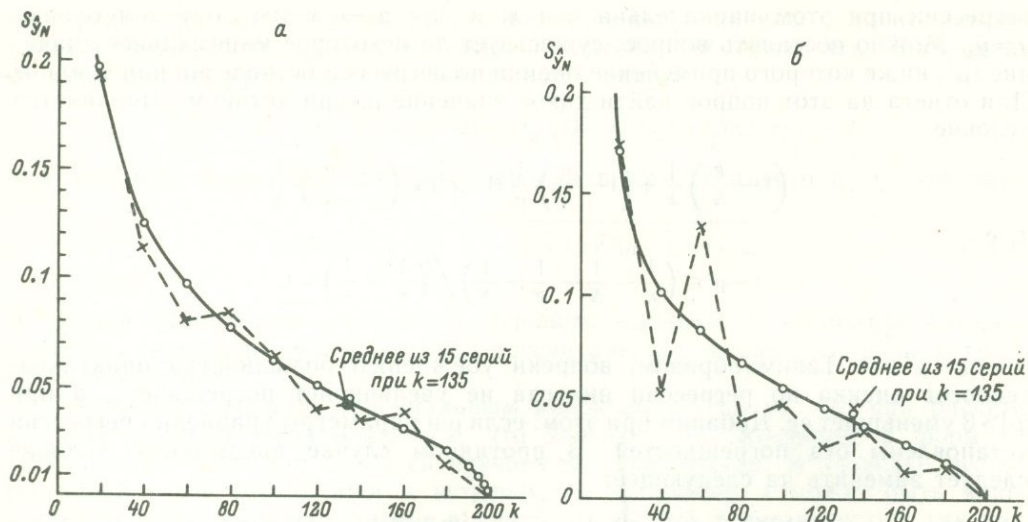


Рис. 2.4.1. Теоретическая (сплошная) и экспериментальная (штриховая) кривые погрешности оценки среднего в совокупности $N=200$ с помощью выборочных средних при различных ее объемах выборки k . Отбор без возвращения.

а — основной эксперимент, б — дополнительный эксперимент.

Таблица 2.4.1

Теоретические и экспериментальные погрешности оценок среднего значения в совокупности объемом 200 с помощью выборок объема n

Номер серии эксперимента	Число выборок t	Объем выборки n	Среднее квадратическое отклонение в совокупности	Средние квадратические погрешности среднего		
				экспериментальная	теоретическая	знак отклонения
1	8	135	0.8842	0.0491	0.0434	—
2	8	135	0.8741	0.0319	0.0429	+
3	8	135	0.8270	0.0589	0.0406	—
4	8	135	0.7751	0.0285	0.0381	+
5	8	135	0.9091	0.0390	0.0443	+
6	8	135	0.9180	0.0615	0.0449	—
7	8	135	0.9347	0.0504	0.0467	—
8	8	135	0.8977	0.0449	0.0444	—
9	8	135	0.8834	0.0349	0.0438	+
10	8	135	0.8591	0.0378	0.0416	+
11	8	135	0.8437	0.0256	0.0419	+
12	8	135	0.8709	0.0390	0.0425	+
13	8	135	0.8375	0.0410	0.0410	=
14	8	135	0.8711	0.0377	0.0421	+
15	8	135	0.8946	0.0402	0.0420	+
Среднее по 15 сериям	8	135	0.8720	0.0414	0.0427	+
1	8	20	0.8859	0.1931	0.1916	—
2	8	40	0.8859	0.1143	0.1239	+
3	8	60	0.8859	0.0823	0.0972	+
4	8	80	0.8859	0.0845	0.0770	—
5	8	100	0.8859	0.0632	0.0632	=
6	8	120	0.8859	0.0400	0.0515	+
7	8	140	0.8859	0.0339	0.0411	+
8	8	160	0.8859	0.0394	0.0320	—
9	8	180	0.8859	0.0156	0.0213	+

совокупности ($N-n$). Теоретические погрешности оценки по регрессии вычислялись по формуле (2.4.3), экспериментальные — по формуле $s_{\hat{Y}_N}^2 = \frac{1}{(m-1)} \sum_{i=1}^m (\hat{Y}_i - \bar{Y}_N)^2$, где m — число выборок для данного значения n , \hat{Y}_i — оценка по регрессии (по формуле (2.4.1)) в i -й выборке. Результат одной из серий приведен в табл. 2.4.2 и на рис. 2.4.2. Отклонение экспериментальных данных от теоретических небольшое.

Таблица 2.4.2

Теоретические и экспериментальные погрешности оценки по регрессии среднего значения в совокупности $N=200$. Модельный коэффициент корреляции $\rho=0.8$

Объем выборки n	Нормированные погрешности оценки по регрессии $s_{\hat{Y}_N}/s_y$		Знак отклонения	Объем выборки n	Нормированные погрешности оценки по регрессии $s_{\hat{Y}_N}/s_y$		Знак отклонения
	теоретическая	экспериментальная			теоретическая	экспериментальная	
5	0.459	0.456	+	120	0.035	0.036	+
10	0.240	0.230	—	140	0.028	0.020	—
20	0.145	0.148	+	160	0.021	0.020	—
40	0.088	0.084	—	180	0.014	0.012	—
60	0.066	0.061	—	190	0.010	0.009	—
80	0.053	0.053	=	200	0.000	0.000	=
100	0.043	0.046	+				

Рисунок 2.4.2 наглядно иллюстрирует возможность значительного снижения погрешностей оценок по регрессии в сравнении с обычными оценками («по среднему на единицу»). При коэффициенте корреляции $\rho=0.8$ погрешность уменьшается в 1.55—1.60 раза, а такой коэффициент между геологоразведочными величинами на месторождениях — обычное явление. Лишь при $n=5$ погрешность оценки по регрессии выше обычной «по среднему на единицу». Это вызвано тем, что значение r (0.800) меньше критического для $n=5$ ($r_{кр}=0.816$). На рис. 2.4.3 показаны результаты экспериментального исследования погрешностей оценки по регрессии в зависимости от изменчивости оцениваемой переменной в совокупности.

На третьем этапе эксперимента «Имопр» имитировалась ситуация, максимально приближенная к реальной разведке или опробованию месторождений. Из совокупности объемом $N=200$ формировались выборки объемом k ($k=10, 20, 40, \dots, 200$), а из каждой из них — подвыборки объемом n ($n < k$, $n=5, 10, 20, 40, \dots, 180, 190$). В элементах подвыборок были «измерены» значения x и y , и по этим данным строились уравнения регрессии. В элементах выборок были известны значения x . В остальной части совокупности численностью $N-k$ значения x (разумеется, и $y!$) не были известны. По каждой выборке вычислялись оценки по регрессии и их погрешности $s_{\hat{y}}$.

Результаты этого эксперимента представлены в табл. 2.4.3. Теоретические значения погрешностей рассчитывались по формуле

$$s_{\hat{Y}_N} = s_y \sqrt{\left(\frac{1}{k} - \frac{1}{N}\right) + \left(\frac{1}{n} - \frac{1}{k}\right) (1-r^2) \frac{(n+1)^2}{(n-1)(n-2)}}. \quad (2.4.6)$$

Для этой цели приводим программу для м/к «Электроника» БЗ-34 (табл. 2.4.4). Как видно из табл. 2.4.3, экспериментальные погрешности незначительно отклоняются от теоретических. Систематических расхождений не установлено. Погрешности увеличиваются с уменьшением n и k , причем влияние n существеннее. При малых n сильно увеличивается вклад в погрешность, зависящий от неточности положения прямой регрессии, поэтому в верхних строках таблицы уменьшение погрешности с увеличением k сказывается мало.

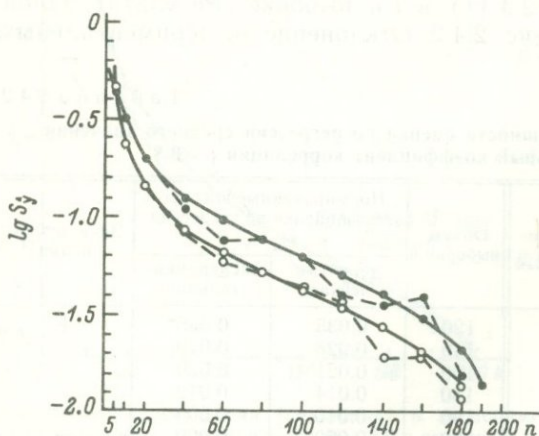


Рис. 2.4.2. Сопоставление эмпирических (штриховые кривые) и расчетных (сплошные) погрешностей оценки средних $S_{\bar{x}}$ по обычной и гибридной выборке в зависимости от числа оцененных элементов k . Обычная выборка объемом k элементов (темные кружки) и «гибридная» выборка объемом 200 элементов, из которых $200-k$ оценено по уравнению регрессии (светлые кружки). Модельный коэффициент корреляции между переменными в генеральной совокупности $\rho=0.8$.

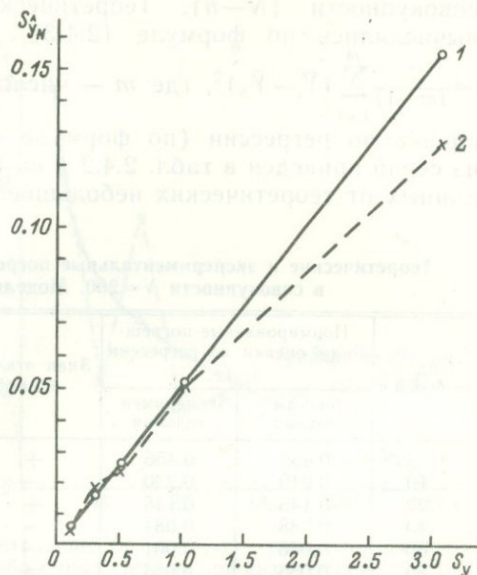


Рис. 2.4.3. Теоретические и экспериментальные погрешности средних, полученных с помощью оценки по регрессии, для различной изменчивости оцениваемой переменной.

1 — теоретические значения погрешностей, 2 — экспериментальные. Объем изучаемой совокупности $N=200$, уравнение регрессии построено по 135 элементам, значение y оценено в 65 элементах, модельный коэффициент корреляции $\rho=0.8$.

Обращаем внимание на то, что в первой строке табл. 2.4.3 погрешность с увеличением k не уменьшается, а увеличивается. Это вызвано тем, что при $n=5$ значение $r=0.8$ не является достаточным для полезного применения оценок по регрессии: они становятся хуже обычных. При этом, естественно, они становятся тем хуже, чем большая доля элементов совокупности получает оценку по регрессии.

Пример. На месторождении нефти отобрано и проанализировано на пористость 371 образец керна. В 22 из них определена проницаемость. В качестве одного из вариантов применена оценка пористости и проницаемости залежи методом среднего арифметического. Пористость распределена по нормальному закону с параметрами $\bar{K}_n=23.85\%$, $s_{K_n}=2.64\%$. Проницаемость распределена по логнормальному закону с $\ln K_{np}=5.78$, $s \ln K_{np}=2.07$, между логарифмом проницаемости и пористостью установлена корреляция $r=0.801$. Уравнение регрессии логарифма проницаемости на пористость $\ln K_{np}=-7.34+0.55K_p$. Считаем, что изучаемая совокупность (залежь) является практически неограниченной по сравнению с выборкой в 371 образец, что выборку можно считать случайной и что автокорреляция пористости по площади отсутствует на расстояниях, равных расстоянию между скважинами, а по вертикали (в скважинах) — образцов друг от друга. Считаем также, что погрешностью анализов можно пренебречь.

Тогда погрешность оценки пористости можно посчитать по формуле (2.4.4) (с использованием программы в табл. 2.4.4), положив $r=0.8$, $N=10^6$, $n=k=371$, а погрешность обычной оценки проницаемости по 22 образцам, приняв $r=0.8$, $N=10^6$, $n=k=22$. Погрешность оценки средней проницаемости по ре-

Теоретические и экспериментальные погрешности оценки по регрессии для случая, когда значение оцениваемой величины известно в k элементах ($k < N$, $N = 200$, модельный коэффициент корреляции $\rho = 0.8$)

Объем выборки n	Теоретическая (сверху) и экспериментальная (снизу) нормированная погрешность для различных значений k												
	5	10	20	40	60	80	100	120	140	160	180	190	200
5	—	0.451	0.455	0.457	0.458	0.458	0.458	0.459	0.459	0.459	0.459	0.459	0.459
		0.459	0.466	0.454	0.439	0.459	0.463	0.452	0.472	0.464	0.438	0.457	0.456
10	—	—	0.274	0.256	0.249	0.246	0.244	0.242	0.241	0.241	0.241	0.240	0.240
			0.277	0.245	0.254	0.243	0.231	0.253	0.239	0.248	0.237	0.236	0.230
20	—	—	—	0.178	0.165	0.158	0.154	0.151	0.148	0.147	0.146	0.145	0.145
				0.187	0.126	0.149	0.154	0.147	0.170	0.159	0.130	0.147	0.148
40	—	—	—	—	0.123	0.112	0.105	0.100	0.097	0.094	0.092	0.091	0.090
					0.128	0.121	0.100	0.101	0.083	0.088	0.094	0.086	0.084
60	—	—	—	—	—	0.095	0.087	0.081	0.076	0.072	0.069	0.067	0.066
						0.086	0.091	0.080	0.079	0.068	0.068	0.066	0.061
80	—	—	—	—	—	—	0.077	0.070	0.064	0.060	0.056	0.054	0.053
							0.082	0.077	0.062	0.062	0.055	0.056	0.053
100	—	—	—	—	—	—	—	0.063	0.057	0.051	0.047	0.045	0.043
								0.060	0.058	0.053	0.044	0.046	0.046
120	—	—	—	—	—	—	—	—	0.051	0.045	0.040	0.037	0.035
									0.052	0.049	0.044	0.038	0.036
140	—	—	—	—	—	—	—	—	—	0.040	0.034	0.031	0.028
										0.045	0.032	0.031	0.020
160	—	—	—	—	—	—	—	—	—	—	0.028	0.025	0.021
											0.027	0.028	0.020
180	—	—	—	—	—	—	—	—	—	—	—	0.019	0.014
												0.014	0.012
190	—	—	—	—	—	—	—	—	—	—	—	—	0.010
													0.009

грессии вычисляем по той же программе, положив $r=0.8$, $N=10^6$, $n=371$, $k=22$.

Результаты расчетов приведены в табл. 2.4.5. Они убедительно свидетельствуют в пользу применения оценок по регрессии. Если известна погрешность измерения проницаемости, то параметры уравнения регрессии и r должны быть откорректированы в соответствии с формулами § 2 этой главы, в частности

Программа для м/к «Электроника» БЗ-34 расчета погрешности оценки по регрессии

00.F1/X 01.П4 02.ИП1 03.FX² 04.1 05.XY 06. — 07.П5 08.ИП2 09.F1/X 10.П6 11.ИП3 12.F1/X 13.П7
 14.ИП6 15. — 16.П6 17. ИП4 18.ИП7 19. — 20.ИП5 21. × 22.П8 23.ИП4 24.F1/X 25.П9 26.1 27. + 28.ПА
 29.ИП9 30.2 31. — 32.ИПА 33.FX² 34.XY 35. ÷ 36.ПВ 37.ИП9 38.1 39. — 40.ИПВ 41.XY 42. ÷ 43.
 ИП8 44. × 45.П8 46.ИП6 47. + 48.F√X 49. С/П. Инструкция: ввести r в П1 — 0,8, N в П2 — 200,
 k в П3 — 10, n в регистр x — 5. Начать счет В/О, С/П. Ответ: 0.451.

Результаты сопоставления точности оценивания коллекторских свойств с использованием и без использования уравнения регрессии

Оцениваемый параметр и способ оценки	Число проанализированных образцов	Число образцов, участвующих в оценке	Погрешность оценки $s_{\bar{y}}$	Двухсигмовый доверительный интервал для логарифма	Двухсигмовый доверительный интервал
Пористость, %	371	371	0.14	—	23.56—24.13
Проницаемость, мД	22	22	0.44	4.90—6.66	134—781
Проницаемость по регрессии, мД	22	371	0.31 0.26 *	5.16—6.40 5.26—6.3 *	174—602 192—544 *

* С учетом погрешности измерения проницаемости в образцах выборки, использованной для построения уравнения регрессии.

значение r увеличится. В нашем примере логарифмическая погрешность анализа образцов на проницаемость составляет 0.5. Тогда $s'_{\ln k_{np}} = 1.83$, $r' = 0.85$, погрешность оценки средней проницаемости снизится до 0.26, т. е. станет почти в два раза меньше, чем обычная оценка по 22 образцам.

Изложенные результаты естественным образом обобщаются на случай множественной регрессии, при этом r заменяется на R . Заканчивая рассмотрение оценок по регрессии, заметим, что таковыми являются по существу оценки коллекторских свойств по данным геофизических исследований скважин. Действительно, такая оценка производится по схеме: измерение некоторых геофизических сигналов, преобразование их в требуемый геофизический (аналитический) параметр, установление уравнения регрессии этого параметра на оцениваемое свойство, например пористость, на некотором множестве пластопересечений коллекторов (пористость в них определена непосредственными измерениями в образцах), оценка среднего значения геофизического параметра по залежи, использование его для оценки по регрессии требуемого среднего значения пористости.

§ 5. Некоторые свойства реальных геологоразведочных выборок

Изложенный выше первый способ оценивания по регрессии имеет в геологии ограниченное применение, так как геологическое опробование месторождений и других геологических тел редко укладывается в схему простого случайного отбора. Различные процедуры взвешивания, расчет оценок с предварительной геометризацией параметров, сложная иерархическая структура системы разведочных выработок и отбора образцов — все это практически исключает первый способ оценивания по регрессии. Методика разведки и схема опробования в геологии таковы, что каждая точка измерения, каждый элемент геологоразведочной выборки уникальны. Как правило, различными в них будут погрешности измерения, представительность, набор измеренных величин. Может оказаться, что элемент выборки, для которого индивидуальная погрешность

оценки по регрессии оказалась особенно высокой, является единственным или главным источником информации для большого блока месторождения. Фактически это приведет к высокой погрешности оценки среднего по месторождению. С помощью первого способа оценивания по регрессии это обстоятельство учесть невозможно.

Второй способ оценивания по регрессии, или, говоря точнее, использования зависимостей для повышения точности оценок, заключается в индивидуальном оценивании значения интересующей нас величины y в тех элементах совокупности, в которых она не измерена. Уравнение регрессии выступает здесь в роли некоторого аналога измерительного прибора. В результате его применения образуется выборка, в каждом элементе которой значение y будет все-таки определено пусть даже со значительной погрешностью.

Для математика-статистика и изучаемая совокупность, и выборка являются прежде всего множествами значений изучаемой величины. Для геолога выборка — это множество образцов, проб, скважин, пластов и т. д., т. е. физических объектов — носителей этой величины. Они, как правило, имеют различные форму, массу, ориентировку в пространстве и другие характеристики. Одной из таких существенных характеристик является способ, с помощью которого определяется изучаемая величина в данном элементе совокупности — непосредственным измерением или оцениванием по регрессии. Вряд ли найдется в геологии случай, чтобы во всех пробах непосредственными измерениями (анализами) были определены все характеристики и в таблицах исходных данных не было бы пустых клеток. Этого невозможно добиться по экономическим, техническим, а в некоторых случаях — по принципиальным причинам, когда измерение одного свойства приводит к уничтожению пробы или разрушению образца.

Таким образом, выборки объема k , в n элементах которых изучаемое свойство непосредственно измерено, а в $k-n$ элементах оценено по регрессии, являются в геологии широко распространенным явлением и требуют специального названия и специальных методов статистической обработки. Наиболее подходит для этого случая термин *сфинкс-выборка*. Оценивание по регрессии с образованием сфинкс-выборки можно рассматривать как доукмплектование исходных данных. Доукмплектование как один из методов повышения достоверности оценок параметров особенно необходимо при вычислении параметра, являющегося функцией двух или нескольких характеристик, потому что точки, в которых отсутствует хотя бы одна из них, не могут участвовать в оценивании, снижая его точность. В связи с использованием сфинкс-выборок возникают следующие вопросы.

1. Имеют ли систематическое смещение среднее \bar{y}_k , дисперсия $s_{y_k}^2$ и коэффициент корреляции r_{xy} , вычисленные по сфинкс-выборке и рассматриваемые в качестве оценок соответствующих параметров изучаемой совокупности.

2. Является ли индивидуальная (для каждого элемента выборки) погрешность оценки по регрессии аналогом погрешности измерения и какие поправки в нее следует внести, чтобы она стала таковой.

3. Какова погрешность оценки параметров изучаемой совокупности сфинкс-выборками в сравнении с обычными выборками.

Рассмотрение свойств сфинкс-выборок начнем со среднего значения оцениваемой переменной в выборке. Уравнение регрессии дает несмещенную оценку значения y в каждом ее элементе, поэтому и выборочное среднее $\bar{y}_k = \frac{1}{k} \times$

$\times \left(\sum_{i=1}^n y_i + \sum_{i=n+1}^k \hat{y}/x_i \right)$, где y_i — измеренные значения, \hat{y}/x_i — значения, оценен-

ные по уравнению регрессии, является несмещенной оценкой среднего значения в выборке (\bar{Y}_k) независимо от доли оцененных элементов $(k-n)/k$, а также независимо от числа элементов n , использованных при построении уравнения регрессии.

Покажем, что дисперсия величины y в сфинкс-выборке ниже дисперсии в ее обычной выборке. Действительно, дисперсия оцененной величины y/x меньше дисперсии соответствующей величины y_k на величину остаточной дисперсии $s_y^2(1-r^2)$. В сфинкс-выборке объема k с подвыборкой объема n она будет равна

$$s_y^{sf2} = \frac{s_y^2 n + [s_y^2 - s_y^2(1-r^2)](k-n)}{k} = s_y^2 \left[1 - \left(1 - \frac{n}{k}\right)(1-r^2) \right]. \quad (2.5.1)$$

При $r=1$ получим $s_y^{sf2} = s_y^2$, что вполне согласуется с интуитивными представлениями, а при $r=0$ и $n=0$ получим $(s_y^{sf})^2 = 0$, т. е. дисперсия оцененных по уравнению регрессии величин превратится в ноль, так как они в этом случае все будут равны среднему значению. При $n=0$ и $r \neq 0$ будем иметь $(s_y^{sf})^2 = s_y^2 \cdot r^2$, при $n=k$ $(s_y^{sf})^2 = s_y^2$. Из (2.5.1) следует, что вычисленную обычным способом по сфинкс-выборке дисперсию следует исправить по формуле

$$s_y^2 = (s_y^{sf})^2 / (1 - (1 - n/k)(1 - r^2)). \quad (2.5.2)$$

В некоторых задачах требуется восстановить не только природную дисперсию s_y^2 , но и то ее значение, которое наблюдалось бы в сфинкс-выборке, если бы мы в $k-n$ элементах ее значение y оценивали не по уравнению регрессии, а с помощью некоторого измерительного прибора с погрешностью, соответствующей погрешности уравнения регрессии ($s_{наб\ y}^2$). Так как $s_{наб\ y}^2 = s_y^2 + s_{ан\ y}^2$, а $s_{ан\ y}^2 = s_y^2 - (s_y^{sf})^2$, то

$$s_{наб\ y}^2 = s_y^2 + s_y^2 - (s_y^{sf})^2 = 2s_y^2 - (s_y^{sf})^2 = (s_y^{sf})^2 \left[\frac{2}{1 - \left(1 - \frac{n}{k}\right)(1 - r^2)} - 1 \right]. \quad (2.5.3)$$

Для практической апробации формулы и иллюстрации изложенного использовались материалы эксперимента «Имопр». В табл. 2.5.1, 2.5.2 и на рис. 2.5.1, 2.5.2 приведены результаты экспериментов, в которых значение коэффициента корреляции изменялось от 0.2 до 0.95, а значение n — от 20 до 180 при заданном $k=200$. Каждая цифра является средней из 8 случайных выборок с заданным n . Схождение между значениями $(s_y^{sf})^2$ по формуле (2.5.1) и фактически полученными очень хорошее, что позволяет рекомендовать формулу (2.5.2) для статистической обработки данных сфинкс-выборки и использовать полученное по ней значение в качестве более надежной оценки дисперсии признака в изучаемой совокупности, чем ее значение по обычной выборке объема n .

Весьма любопытно изменение коэффициента корреляции между x и y в сфинкс-выборке. Очевидно, что он будет значительно завышен, так как в части выборки величина y оценивалась как линейная функция x , и в этой части выборки коэффициент корреляции между переменными равен единице. Коэффициент корреляции в сфинкс-выборке (r^{sf}) будет средневзвешенным значением между коэффициентами корреляции в составных ее частях: единица — в оцененной части объемом $k-n$, r_n — в подвыборке объема n . Весовыми коэффициентами служат произведения относительных объемов этих частей сфинкс-выборки на стандартные отклонения:

$$r^{sf} = 1 \cdot (k-n)/k \cdot s_{k-n}/s_y^{sf} + r \cdot n/k \cdot s_y/s_y^{sf}.$$

Стандартное отклонение величины y в оцененной части сфинкс-выборки численностью $k-n$ (s_{k-n}) определяется по формуле (2.5.2) при значении n , равном нулю, $s_{k-n} = s_y \sqrt{1 - (1 - r^2)} = s_y r$, откуда

$$r^{sf} = r_n \frac{s_y}{s_y^{sf}} \left(\frac{k-n}{k} + \frac{n}{k} \right) = r_n \frac{s_n}{s_y^{sf}} = r_n \frac{1}{\sqrt{1 - \left(1 - \frac{n}{k}\right)(1 - r_n^2)}}. \quad (2.5.4)$$

Таблица 2.5.1

Средние квадратические отклонения оцениваемой величины в сфинкс-выборках в зависимости от числа оцененных по уравнению регрессии значений $k-n$ (объем сфинкс-выборки $k=200$, модельный коэффициент корреляции $\rho=0.8$)

Наименование показателя	Среднеквадратические отклонения при различных $k-n$								
	180	160	140	120	100	80	60	40	20
1 - я серия экспериментов									
Коэффициент корреляции r в подвыборке объема n	0.824	0.801	0.810	0.807	0.812	0.811	0.805	—	—
Стандартное отклонение s_y в подвыборке объема n	2.138	2.080	2.106	2.087	2.084	2.060	2.062	—	—
Эмпирическое стандартное отклонение в сфинкс-выборке	1.646	1.656	1.729	1.787	1.841	1.874	1.929	—	—
Теоретическое стандартное отклонение в сфинкс-выборке s_y^{sf}	1.803	1.757	1.835	1.856	1.898	1.913	1.950	—	—
2 - я серия экспериментов									
Коэффициент корреляции r в подвыборке объема n	0.776	0.742	0.751	0.754	0.756	0.756	0.759	0.807	0.806
Стандартное отклонение s_y в подвыборке объема n	2.288	2.151	2.237	2.251	2.261	2.275	2.245	2.261	2.263
Эмпирическое стандартное отклонение в сфинкс-выборке	1.953	1.863	1.928	2.002	2.040	2.094	2.113	2.168	2.110
Теоретическое стандартное отклонение в сфинкс-выборке s_y^{sf}	1.833	1.721	1.864	1.938	2.002	2.070	2.094	2.181	2.223

Таблица 2.5.2

Средние квадратические отклонения оцениваемой величины в сфинкс-выборках в зависимости от коэффициента корреляции (объем сфинкс-выборки $k=200$, число оцененных значений $k-n=65$)

Наименование показателя	Средние квадратические отклонения при различных модельных коэффициентах корреляции ρ					
	0.2	0.4	0.6	0.8	0.9	0.95
1 - я серия экспериментов						
Коэффициент корреляции r в подвыборке	—	0.013	0.513	0.709	0.884	0.943
Среднее квадратическое отклонение s_y в подвыборке объема $n=135$	—	4.306	2.803	2.075	1.843	1.776
Эмпирическое среднее квадратическое отклонение в сфинкс-выборке	—	3.676	2.443	1.924	1.784	1.739
Теоретическое среднее квадратическое отклонение (s_y^{sf}) в сфинкс-выборке	—	3.672	2.444	1.900	1.776	1.744
2 - я серия экспериментов						
Коэффициент корреляции r в подвыборке	0.147	0.461	0.660	0.805	0.888	0.939
Среднее квадратическое отклонение s_y в подвыборке объема $n=135$	9.113	4.679	3.104	2.284	1.994	1.750
Эмпирическое среднее квадратическое отклонение в сфинкс-выборке	7.534	4.040	2.813	2.135	1.908	1.795
Теоретическое среднее квадратическое отклонение s_y^{sf} в сфинкс-выборке	7.526	4.036	2.805	2.149	1.924	1.716

Таким образом, в качестве выборочной оценки коэффициента корреляции в совокупности по сфинкс-выборке следует принимать величину

$$r = r^{sf} \sqrt{1 - \left(1 - \frac{n}{k}\right) (1 - r_n^2)} = r_n \quad (2.5.5)$$

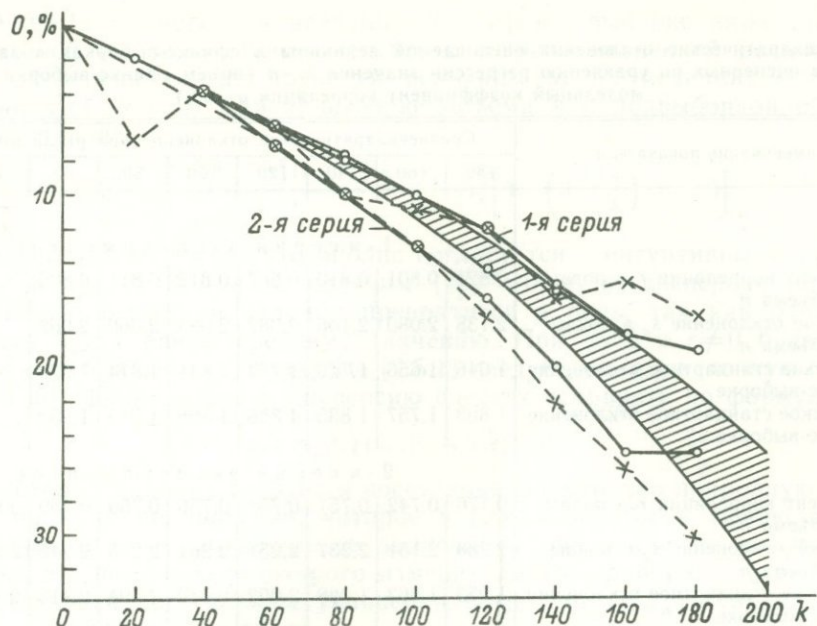


Рис. 2.5.1. Изменение эмпирического среднего квадратического отклонения величины y в сфинкс-выборке (в % от S_y , ось ординат) в зависимости от числа оцененных по уравнению регрессии элементов $n-k$ (ось абсцисс).

Сплошная линия — теоретические значения, штриховая — эмпирические по двум сериям. Заштрихована область теоретических значений при ρ от 0.75 до 0.80. Объем сфинкс-выборки $k=200$, средний эмпирический коэффициент корреляции $r=0.8$.

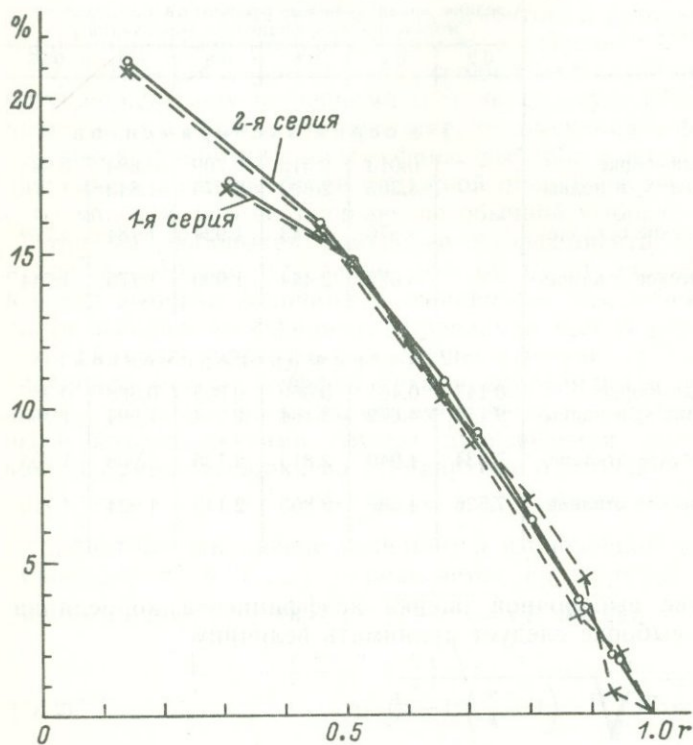


Рис. 2.5.2. Изменение среднего квадратического отклонения величины y в сфинкс-выборке (в % от S_y , ось ординат) с увеличением коэффициента корреляции (ось абсцисс) при постоянном числе оцененных элементов в совокупности $k-n=65$, объем сфинкс-выборки $k=200$.

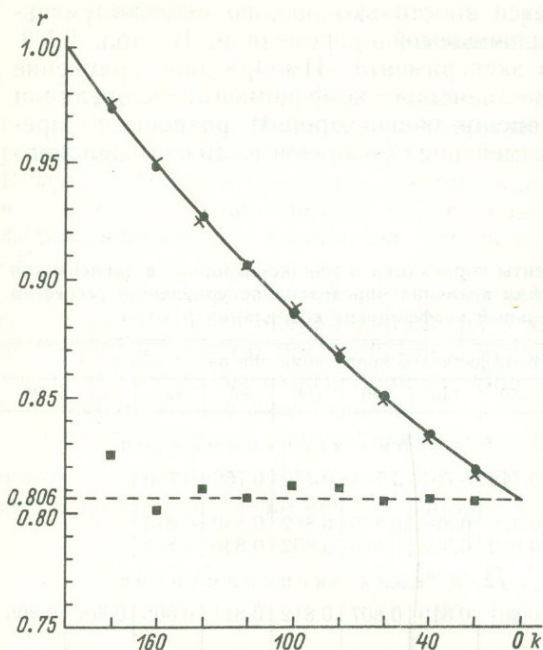


Рис. 2.5.3. Изменение коэффициента корреляции в сфинкс-выборке (ось ординат) в зависимости от числа оцененных значений $k-n$ в выборке (ось абсцисс). 2-я серия экспериментов.

Темные кружки и соединяющая их сплошная — теоретические значения, крестики — эмпирические значения, квадратики — коэффициент корреляции в подвыборке объема n .

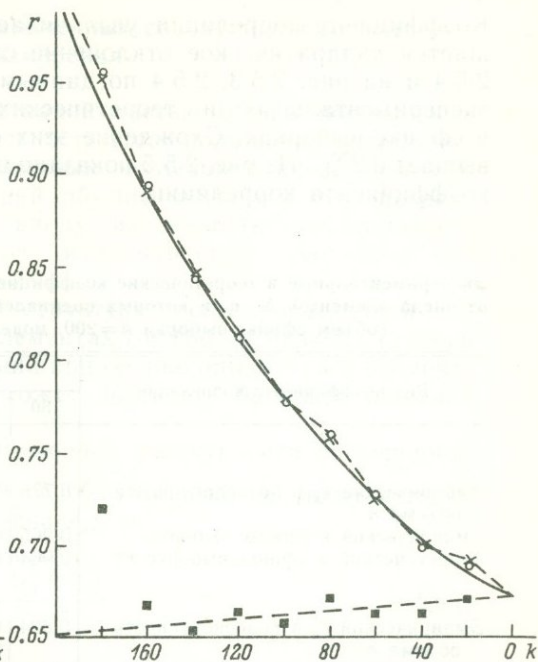


Рис. 2.5.4. Изменение коэффициента корреляции r в сфинкс-выборке в зависимости от числа $k-n$, оцененных по уравнению регрессии значений. 3-я серия экспериментов.

Кружки и соединяющая их сплошная — теоретические значения, крестики — эмпирические значения, квадратики — конкретная реализация модельного коэффициента корреляции в подсовкупности объема n , штриховая прямая — значение модельного коэффициента корреляции в экспериментах.

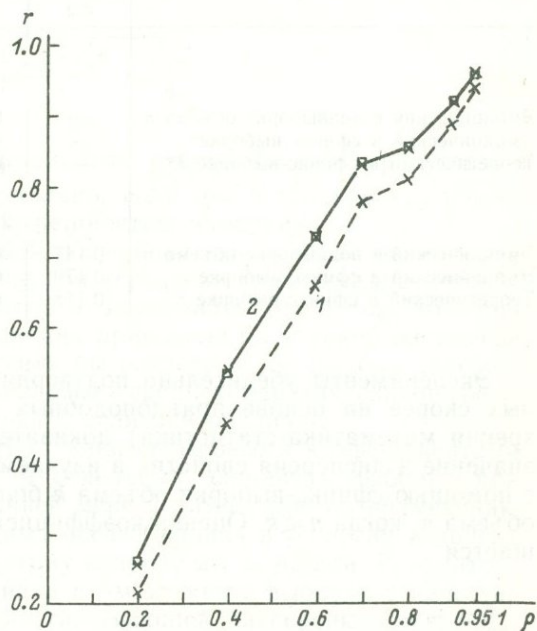


Рис. 2.5.5. Изменение коэффициента корреляции в сфинкс-выборке r в зависимости от модельного коэффициента корреляции ρ .

Объем сфинкс-выборки $k=200$, число элементов с оценкой по регрессии $k-n=65$. 1 — эмпирический коэффициент корреляции в обычной подвыборке объемом $n=135$, 2 — эмпирический коэффициент корреляции в сфинкс-выборке, квадратики — теоретический.

Коэффициент корреляции увеличивается во столько раз, во сколько уменьшается квадратическое отклонение оцениваемой переменной. В табл. 2.5.3, 2.5.4 и на рис. 2.5.3, 2.5.4 по данным эксперимента «Имопр» дано сравнение экспериментальных и теоретических значений коэффициента корреляции в сфинкс-выборках. Схождение этих оценок очень хорошее: разность не превышает 0.2 %. На рис. 2.5.5 показано изменение r^{sf} в зависимости от модельного коэффициента корреляции.

Таблица 2.5.3

Экспериментальные и теоретические коэффициенты корреляции в сфинкс-выборках в зависимости от числа элементов $k-n$, в которых оцениваемая величина определена по уравнению регрессии (объем сфинкс-выборки $k=200$, модельный коэффициент корреляции $\rho=0.8$)

Вид коэффициента корреляции	Коэффициенты корреляции при различных $k-n$								
	180	160	140	120	100	80	60	40	20
1 - я серия экспериментов									
Эмпирический r_n в подсовкупности объема n	0.776	0.742	0.751	0.754	0.756	0.756	0.759	—	—
Эмпирический в сфинкс-выборке	0.965	0.926	0.898	0.875	0.852	0.830	0.813	—	—
Теоретический в сфинкс-выборке r^{sf}	0.964	0.932	0.903	0.876	0.852	0.830	0.810	—	—
2 - я серия экспериментов									
Эмпирический r_n в подсовкупности объема n	0.824	0.801	0.810	0.807	0.812	0.811	0.805	0.807	0.806
Эмпирический в сфинкс-выборке	0.977	0.951	0.927	0.908	0.889	0.870	0.850	0.835	0.818
Теоретический в сфинкс-выборке r^{sf}	0.974	0.950	0.928	0.907	0.888	0.869	0.852	0.835	0.820

Таблица 2.5.4

Экспериментальные и теоретические коэффициенты корреляции в сфинкс-выборках в зависимости от модельного коэффициента корреляции (объем сфинкс-выборки $k=200$, число оцененных значений $k-n=65$)

Вид коэффициента корреляции	Коэффициенты корреляции при различных модельных коэффициентах корреляции ρ					
	0.2	0.4	0.6	0.8	0.9	0.95
1 - я серия экспериментов						
Эмпирический в подвыборке объема n	—	0.313	0.513	0.709	0.884	0.943
Эмпирический в сфинкс-выборке	—	0.374	0.583	0.813	0.916	0.960
Теоретический в сфинкс-выборке r^{sf}	—	0.372	0.588	0.774	0.917	0.960
2 - я серия экспериментов						
Эмпирический в подвыборке объема n	0.147	0.461	0.660	0.805	0.888	0.939
Эмпирический в сфинкс-выборке	0.179	0.534	0.733	0.854	0.919	0.957
Теоретический в сфинкс-выборке r^{sf}	0.178	0.534	0.730	0.855	0.920	0.958

Эксперименты убедительно подтвердили справедливость формул, полученных скорее на основе правдоподобных рассуждений, чем строгих (с точки зрения математика-статистика) доказательств. Они показывают, что среднее значение и дисперсия свойства в изучаемой совокупности могут быть оценены с помощью сфинкс-выборки объема k более надежно, чем по обычной выборке объема n , когда $n < k$. Оценка коэффициента корреляции, естественно, не улучшается.

§ 6. Особенности оценки средних в ограниченных совокупностях с учетом погрешности измерения

Погрешность измерения является составной частью общей погрешности оценки параметра. Несмотря на общеизвестность этого положения, удовлетворительной методики ее учета при выборочных исследованиях не имеется. В геологии и при разведке месторождений обычно считается, что погрешностями измерений можно пренебречь ввиду их малости по сравнению с погрешностью, вызванной природной изменчивостью геологического объекта.

При многоступенчатой иерархической структуре геологоразведочной выборки погрешность оценки параметра в элементах i -й (более низкой) ступени выступает в качестве погрешности измерения при оценке параметра в элементе $(i+1)$ -й (более высокой) ступени. Она может достигать больших значений и пренебрегать ею уже нельзя.

При выборочном исследовании неограниченной совокупности погрешность оценки $s_{\bar{y}}$ среднего, как известно, равна $s_{\bar{y}} = s_y / \sqrt{n}$. Если погрешность измерения в каждом из n элементов выборки равна $s_{ан}$, то

$$s_y = \frac{s_{наб\ y}}{\sqrt{n}} = \sqrt{\frac{(s_y^2 + s_{ан}^2)}{n}}, \quad (2.6.1)$$

где s_y^2 — природная дисперсия величины y . Это равенство удобнее записать в таком виде:

$$s_y = \sqrt{\frac{s_y^2}{n} + \frac{s_{ан}^2}{n}} = \sqrt{\frac{(s_{наб}^2 - s_{ан}^2)}{n} + \frac{s_{ан}^2}{n}}, \quad (2.6.2)$$

так как непосредственному расчету поддается прежде всего наблюдаемая дисперсия измеренных величин $s_{наб}^2$. Соответствующее равенство для ограниченной совокупности примет вид

$$s_{\bar{y}} = \sqrt{(1-f) \frac{(s_{наб}^2 - s_{ан}^2)}{n} + \frac{s_{ан}^2}{n}}. \quad (2.6.3)$$

Существенно, что множитель с коэффициентом отбора $f = n/N$ присутствует только у первого слагаемого. Действительно, если $n = N$, то $s_{\bar{y}} = s_{ан} / \sqrt{n}$, т. е. остается только член, определяемый погрешностью измерения.

При различной среднеквадратической погрешности измерения величины y в выборке необходимо найти выражение для «средней» среднеквадратической погрешности измерения $s_{ан}^2$, смысл которой заключался бы в том, что при подстановке в (2.6.2) и (2.6.3) эта величина приводила бы к такой же оценке погрешности среднего, какую мы получили бы расчетами по другому способу, позволяющему использовать «индивидуальные» среднеквадратические погрешности измерения $s_{ан_i}$ в элементе выборки i , если они различны.

Кроме того, необходимо определить, с какими весовыми коэффициентами следует брать результаты измерения в отдельных элементах выборки при оценке среднего совокупности. По крайней мере в литературе по применению математической статистики в выборочных исследованиях в геологии и других естественных науках рекомендаций по этому вопросу мы не нашли. Рассматривая значение интересующего нас признака в i -м элементе выборки в качестве независимой оценки среднего совокупности, имеющей дисперсию $s_{\bar{y}}^2 = s_y^2 + s_{ан_i}^2$,

и применяя известную формулу $\bar{y} = \sum_{i=1}^n y_i / s_{\bar{y}}^2 / \sum_{i=1}^n 1/s_{y_i}^2$, к рассматриваемому случаю, получаем¹

$$\bar{y} = \sum_{i=1}^n \frac{y_i}{s_y^2 + s_{ан_i}^2} / \sum_{i=1}^n \frac{1}{s_y^2 + s_{ан_i}^2}, \quad (2.6.4)$$

Рекомендации взвешивать именно таким способом, при котором весовые коэффициенты зависят не столько от погрешностей измерения, сколько от их отношения (дробь) к изменчивости признака в совокупности, мы не встречали. Если в (2.6.4) положим $s_y^2 = 0$, то получим обычную формулу для наиболее эффективной оценки величины, несколько измерений которой выполнены с различными погрешностями. При значениях $s_{ан}$, которыми по сравнению с s_y можно пренебречь, получим обычное среднее арифметическое. При одном и том же наборе погрешностей $s_{ан_i}$ весовые коэффициенты будут различаться тем сильнее, чем меньше s_y . Средняя квадратическая погрешность оценки среднего по (2.6.4) равна

$$s_{\bar{y}} = \sqrt{\frac{1}{\sum_{i=1}^n 1/(s_y^2 + s_{ан_i}^2)}}. \quad (2.6.5)$$

Если бы все $s_{ан_i}^2$ были равны некоторому среднему $\bar{s}_{ан}^2$, то

$s_{\bar{y}} = \sqrt{(s_y^2 + \bar{s}_{ан}^2)/n}$. Приравнивая правую часть этого равенства к правой части (2.6.5), получаем выражение для искомой величины $s_{ан}^2$:

$$s_{ан}^2 = \frac{n}{\sum 1/s_y^2 + s_{ан_i}^2} - s_y^2. \quad (2.6.6)$$

Однако определить $s_{ан}^2$ непосредственно по (2.6.6) невозможно, так как входящая туда величина s_y^2 — природная дисперсия нам неизвестна. По результатам обработки данных нам известна общая (наблюдаемая) дисперсия $s_{наб}^2$, являющаяся суммой

$$s_{наб}^2 = s_y^2 + s_{ан}^2. \quad (2.6.7)$$

Подставляя это в (2.6.6), получим $s_{наб}^2 = \frac{n}{\sum 1/(s_{наб}^2 - \bar{s}_{ан}^2 + s_{ан_i}^2)}$. Таким образом, средняя дисперсия анализа $\bar{s}_{ан}^2$ определяется путем решения уравнения

$$\frac{n}{\sum_{i=1}^n 1/s_{наб}^2 - \bar{s}_{ан}^2 + s_{ан_i}^2} - s_{наб}^2 = 0. \quad (2.6.8)$$

Разрешить его относительно $\bar{s}_{ан}^2$ практически невозможно, особенно при больших n , однако его можно легко решить, например методом последовательных приближений. Для этого в табл. 2.6.1 мы приводим программу для м/к «Электроника» БЗ-34 и схему решения. Выбрав в качестве первого приближения значение $\bar{s}_{ан}^2 = 1/n \sum s_{ан_i}^2$, путем 5—6 итераций определяют $\bar{s}_{ан}^2$ с необходимой степенью точности. Получив $\bar{s}_{ан}^2$, из (2.6.7) определяют s_y^2 . Эти величины

¹ Равенство справедливо, если «физическая» представительность всех элементов выборки (проб) одинакова.

Программа для расчета средней погрешности измерения при неравноточных измерениях

00.FX² 01.ИП2 02.+03.ИП5 04.— 05.F1/X 06.ИП3 07.+08.ПЗ 09.С/П 10.F1/X 11.ИП1 12.×13.С/П 14.ИП2 15.—16.С/П. Инструкция: 1) Занести n в рег. П1, $s_{\text{наб}}^2$ в рег. П2. 2) занести O в рег. ПЗ, $s_{\text{ан}}^2$ в рег. П5. 3) занести $s_{\text{ан}}$ в рег. X, В/О, С/П. 4) все ли $s_{\text{ан}}$ занесены? Если да, перейти к п. 5, если нет — к п. 2.5) С/П, в рег. X появится промежуточное значение, С/П. 6) Результат отрицательный? Если да, то уменьшить $s_{\text{ан}}^2$ и перейти к п. 2, если нет, то увеличить $s_{\text{ан}}^2$ и перейти к п. 2. Если результат близок к нулю, перейти к п. 7. 7). Принять введенное $s_{\text{ан}}^2$ в качестве решения.

подставляются в (2.6.2) и (2.6.3) при расчете погрешностей оценки среднего с учетом погрешностей измерений.

Рассмотрим следующий пример, являющийся фрагментом автоматизированного подсчета запасов нефти на одной из залежей Харьгинского месторождения Тимано-Печорской провинции. Исходные данные и результаты расчетов приведены в табл. 2.6.2. Обращает на себя внимание незначительная вариация среднескважинных значений пористости ($s_{\text{наб}}^2 = 1.562$). Следовательно, заранее можно предсказать, что результаты существенно не изменятся, какие бы способы расчета не применялись.

Таблица 2.6.2

Среднескважинные значения пористости и погрешности, их оценки. Получение средней пористости и ее погрешности по залежи

Исходные данные			Обработка данных		
номер скважины	среднескважинная пористость (y_j), %	средняя квадратическая погрешность оценки среднескважинной пористости ($s_{\text{ан}j}$)	$S_{\text{ан}j}^2$	весовой коэффициент $\frac{1}{s_y^2 + s_{\text{ан}j}^2}$	$\frac{y_j}{s_y^2 + s_{\text{ан}j}^2}$
43	22.8	0.573	0.328	0.715	16.30
53	23.7	0.196	0.038	0.903	21.40
61	23.3	0.510	0.260	0.716	16.68
64	21.5	6.270	39.310	0.025	0.54
65	22.2	0.412	0.170	0.806	17.89
Сумма	113.5	—	39.609	3.16	72.8

$$s_{\text{наб}}^2 = 1.562, \bar{y} = 22.7, s_y^* = \sqrt{1.562/n} \approx 0.56, \bar{s}_{\text{ан}}^2 = 0.49, s_y^2 = s_{\text{наб}}^2 - s_{\text{ан}}^2 = 1.07,$$

$$\hat{y} = 72.8/3.16 = 23.04, \hat{s}_y = \sqrt{1/3.16} = 0.563 \approx 0.56, \hat{y} = 23.0 \pm 0.5 \%$$

Расчетами по программе (из табл. 2.6.1) определяем среднюю «аналитическую» дисперсию $s_{\text{ан}}^2 = 0.49$, что соответствует средней квадратической погрешности оценки среднескважинных $\sqrt{0.49} = 0.7$ %. Природная дисперсия лишь вдвое превышает аналитическую, поэтому последней пренебречь нельзя. В табл. 2.6.2 вычислены весовые коэффициенты для среднескважинных. Они мало различаются друг от друга за исключением скв. 64, в которой пористость оценена с очень большой погрешностью. Эта скважина практически выпадает из оценки средней пористости. Погрешность оценки среднего по залежи (точнее, по рассматриваемой части залежи) $s_{\bar{y}} = 0.56$ мы могли бы получить в данном случае и более простым способом: $s_{\bar{y}} = \sqrt{s_{\text{наб}}^2/n} \approx 0.56$.

Чтобы понять необходимость определения $\bar{s}_{\text{ан}}^2$, разовьем наш пример следующим образом. Пусть месторождение разбито на 8 геологических блоков, пять из которых изучены. Пусть третья колонка табл. 2.6.2 отражает погрешность оценки пористости в соответствующем блоке месторождения. Тогда, рассматривая полученную оценку пористости в качестве средней для всех 8 блоков, в соответствии с (2.6.2) будем иметь $s_{\bar{y}} = \sqrt{(1-5/8)(1.56-0.49)/5 +$

$\pm 0.49/\sqrt{5} = \sqrt{0.08+0.1} = 0.42$. Погрешность оценки по месторождению не станет меньше, чем $\sqrt{0.4a/5} \approx 0.3$, что будет достигнуто при коэффициенте отбора $f=1$. Формально применяя формулу $s_{\bar{y}} = \sqrt{(1-f)s_y^2/n}$, при $f=1$ мы получили бы абсурдный результат $s_{\bar{y}}=0$, так как погрешность оценки в отдельных скважинах (блоках) в данном случае весьма существенна. Подобная ошибка встречается очень часто.

§ 7. Оценка среднего в совокупности через индивидуальную оценку y по зависимости в каждом элементе сфинкс-выборки

Теперь, после того как установлена зависимость между погрешностью оценки среднего с учетом погрешностей измерения в ограниченных совокупностях по выборке, полученной отбором без возвращения, мы можем закончить обсуждение второго способа оценки среднего по регрессии. Он, как было изложено, заключается в индивидуальной оценке значения y по зависимости в тех элементах выборки, где нет прямых измерений этой величины. Было бы весьма заманчиво так откорректировать некоторые характеристики полученной сфинкс-выборки, чтобы в дальнейшем обрабатывать ее стандартными методами, учитывая погрешность оценки y в каждом элементе выборки, но не обращая каждый раз внимания на то, как в нем произведена оценка — непосредственным измерением или по зависимости от вспомогательной величины x . Обычная выборка выступала бы тогда как частный случай сфинкс-выборки с $k=n$, т. е. с числом оцененных по зависимости значений y , равным $k-n=0$. При этом возникают следующие вопросы: 1) какую погрешность оценки по зависимости следует принимать в расчет как аналог погрешности измерения, 2) с какими весовыми коэффициентами следует брать значения величины y при оценивании среднего в совокупности.

Будь в сфинкс-выборке все погрешности «измерения» одинаковыми, а совокупность — неограниченной, можно было бы погрешность оценки среднего определять по формуле (2.6.1).

Напомним, что в § 5 получены формулы для корректировки наблюдаемой дисперсии $s_{\text{наб } y}^2$ признака в сфинкс-выборке (формула (2.5.3)). Однако главные проблемы связаны с определением погрешности индивидуальной оценки y по зависимости в конкретном элементе выборки, а именно той погрешности, которую необходимо принимать во внимание при вычислении погрешности оценки среднего в совокупности. Приведем простые рассуждения, показывающие, что эта погрешность не равна обычной, детально обсужденной в § 3 погрешности индивидуальной оценки по уравнению регрессии. Пусть известно, что $\rho_{xy}=0$. Прямая регрессии в этом случае горизонтальна; все оцененные значения равны среднему по подвыборке \bar{y}_n , т. е. такая оценка по регрессии фактически не должна улучшить оценки среднего в совокупности. Погрешность (если изучаемая совокупность неограниченна) останется равной $s_{\bar{y}} = s_y \sqrt{1/n}$. Но, рассуждая формально, мы имеем сфинкс-выборку объема k , в $k-n$ элементах которой погрешность «измерения» равна $s_{y|x} = s_y \sqrt{1-\rho_{\bar{y}=0}^2} s_y$, а в n элементах — нулю. Погрешность оценки среднего в совокупности, полученной по выборке объема k , будет равна

$$s_{\bar{y}_k} = \sqrt{\frac{1}{\sum_{i=n+1}^k \frac{1}{s_y^2 + s_{\text{ан}}^2} + \sum_{i=1}^n \frac{1}{s_y^2 + 0}}} = \sqrt{\frac{1}{\sum_{i=n+1}^k \frac{1}{s_y^2 + s_y^2} + \sum_{i=1}^n \frac{1}{s_y^2}}} =$$

$$= s_y \sqrt{2/(k+n)} = s_y \sqrt{2/(m+n+n)},$$

где m — число оцененных по уравнению регрессии элементов выборки. Если $m=0$, то получаем правильный результат: $s_{\bar{y}_k} = s_{\bar{y}_n} = s_y \sqrt{1/n}$, однако при $m > 0$ дробь $2/(m+2n)$, а вместе с ней и погрешность уменьшатся, что заведомо неверно. Так, при $m=2n$ получится $s_{\bar{y}} = s_y / \sqrt{2n}$, тогда как должно оставаться $s_{\bar{y}} = s_y / \sqrt{n}$.

Кажущееся повышение точности происходит оттого, что оценка по регрессии складывается из двух источников информации: первый — значение величины x в оцениваемом элементе, второй — среднее значение y в совокупности объема n , т. е. \bar{y}_n . Оба источника информации полезны, если оценка используется в локальных целях. Однако, когда она используется далее для оценки среднего в совокупности, из которой образована подвыборка, этот источник дублирует значение \bar{y}_n , притом столько раз, сколько элементов выборки оценено по регрессии.

Концепция, что уравнение регрессии есть способ оценивания по двум источникам информации, позволяет определить отдельно долю точности, вносимую каждым из этих источников. Для простоты рассмотрим случай с известным точным уравнением регрессии (или построенному по большому числу данных: $n \geq 20$). Тогда

$$s_{y|x}^2 = s_y^2 (1-r^2) = \frac{1}{1/s_y^2 + 1/s_{зав}^2} = \frac{s_y^2 s_{зав}^2}{s_y^2 + s_{зав}^2},$$

где $s_{зав}$ — погрешность оценки по первому источнику информации — «по зависимости», s_y — по второму, т. е. погрешность оценки среднего с помощью измерения величины y в одном элементе выборки. Точнее, величина s_y выступает здесь в роли погрешности оценки значения y в элементе совокупности по среднему \bar{y}_n . Определим отсюда $s_{зав}^2 = (1-r^2) s_{зав}^2 + (1-r^2) s_y^2$, или

$$s_{зав} = s_y \sqrt{\frac{1-r^2}{r^2}} = \frac{s_y}{\sqrt{1/(1-r^2)-1}}. \quad (2.7.1)$$

Как в содержательном смысле можно интерпретировать величину $s_{зав}^2$. Предположим, что зависимость установлена по выборке из совокупности № 1, а применять ее для оценки y предполагается в совокупности № 2, в отношении которой известно, что установленная зависимость в ней сохраняется, но о параметрах совокупности, в частности \bar{y} , в ней ничего неизвестно. Второго источника информации не существует, и погрешность оценки по зависимости должна быть в этом случае большей, а именно $s_{зав}$ (так как $r^2 \leq 1$, то $s_{зав} > s_{y/x}$).

Рассмотрим оценку по уравнению регрессии как средневзвешенное из двух величин:

$$y|x = \frac{\bar{y}_n \cdot 1/s_y^2 + y_{зав} \cdot 1/s_{зав}^2}{1/s_y^2 + 1/s_{зав}^2}.$$

Отсюда определим $y_{зав}$, подставляя вместо $s_{зав}$ его выражение из (2.7.1): $y_{зав} = y_{пер} \cdot 1/r^2 - \bar{y}_n [(1-r^2)/r^2]$. Подставив в это равенство выражение для $y_{пер}$:

$$y_{зав} = \frac{a - \bar{y}_n}{r^2} + \bar{y}_n + \frac{b_{yx}}{r^2} x, \quad (2.7.2)$$

обнаруживаем, что полученное есть сопряженное уравнение регрессии к уравнению регрессии y на x .

Следовательно, можно ожидать, что, оценивая среднее в совокупности как средневзвешенное с учетом погрешностей «измерений» по (2.6.4), мы получим оценку, в точности равную оценке по регрессии. Формулу (2.6.4) для данного случая можно записать в следующем виде:

$$\bar{y} = \frac{\sum_{i=1}^n y_i \cdot \frac{1}{s_y^2} + \sum_{i=n+1}^k y_{зав_i} \cdot \frac{1}{s_y^2 + s_{зав}^2}}{\sum_{i=1}^n \frac{1}{s_y^2} + \sum_{i=n+1}^k \frac{1}{s_n^2 + s_{зав}^2}} = \frac{n\bar{y}_n + r^2 \sum_{i=n+1}^k y_{зав_i}}{n + (k-n)r^2}. \quad (2.7.3)$$

В табл. 2.7.1 и на рис. 2.7.1 приведена модельная совокупность двух коррелированных величин, которую мы использовали для иллюстрации и практического подтверждения изложенных положений. Из нее с помощью таблицы случайных чисел было получено 13 выборок объемом $n=20$. Параметры выборок приведены в табл. 2.7.2, а в табл. 2.7.3 дано сравнение простых выборочных оценок (по среднему на единицу), оценок по регрессии, оценок по (2.7.3) и их погрешностей. Прежде всего обратим внимание на то, что простая выборочная оценка (графа 2) почти в два раза менее точна, чем оценки с использованием зависимостей. Оценка \bar{y}_n как средневзвешенное значение во всех k элементах выборки с весами $1/s_y^2$ для подвыборки объема n и с весами $1/(s_y^2 + s_{зав}^2)$ для тех ее элементов, которые были оценены по сопряженному уравнению регрессии, во всех выборках оказалась практически идентичной оценке по регрессии. Среднее значение из всех средних (3.1008) практически равно истинному среднему в совокупности (3.10). Среднеквадратическое отклонение выборочных средних (0.059) оказалось почти в точности равным среднему по 13 выборкам расчетному среднеквадратическому отклонению по формуле (2.6.5) — 0.058.

Таким образом, концепция подтвердилась на практическом примере, и мы можем сделать следующие выводы: 1) для того чтобы образовать сфинкс-выборку, недостающие значения y необходимо оценивать по сопряженному уравнению регрессии; 2) в качестве погрешности «анализа» в этом случае выступает величина $s_{зав}$; 3) дальнейшая обработка такой сфинкс-выборки может проводиться обычными методами, в частности оценка среднего в совокупности и ее погрешность рассчитываются по формулам (2.6.4) и (2.6.5).

Если n недостаточно велико (≤ 20), то следует брать не по (2.7.1), а по формуле

$$s_{зав} = s_y \sqrt{\frac{1-r^2}{r^2} \frac{(n-1)(n+1)}{(n-2)n}}. \quad (2.7.4)$$

В качестве входящей в формулы s_y можно брать выборочное значение s_{y_n} , однако более эффективно использовать величину

$$s_y = \sqrt{[n \cdot s_{y_n}^2 + (k-n) b_{зав}^2 s_{x_k}^2 r^2] / K}. \quad (2.7.5)$$

В нашем примере, если во всех элементах совокупности определить y по сопряженному уравнению регрессии, то стандартное отклонение этой величины равно 0.618, т. е. $\sigma_x \cdot b_{зав} = 0.601 \cdot 1.028 = 0.618$, откуда $s_y = 0.618 / 0.82 = 0.5066$. Фактически по модели было $\sigma_y = 0.507$.

Рассмотрим погрешность оценивания среднего неограниченной совокупности сфинкс-выборкой объема k с подвыборкой объема n при условии, что n

Таблица 2.7.1

Модельная совокупность двух коррелированных величин

№ п/п	x	y	№ п/п	x	y	№ п/п	x	y
0	1.25	1.81	33	2.26	2.95	66	2.26	3.47
1	1.74	2.20	34	2.32	2.97	67	2.72	3.37
2	1.40	2.10	35	2.12	3.03	68	2.77	3.36
3	1.54	2.10	36	2.16	3.11	69	2.86	3.24
4	1.28	2.28	37	2.28	3.10	70	2.95	3.18
5	1.71	2.27	38	2.38	3.03	71	3.27	3.08
6	1.97	2.34	39	2.50	2.94	72	2.86	3.33
7	1.76	2.39	40	2.49	2.98	73	2.92	3.35
8	1.58	2.44	41	2.62	2.89	74	2.52	3.60
9	1.40	2.56	42	2.66	2.84	75	2.66	3.51
10	1.85	2.49	43	2.88	2.73	76	2.78	3.54
11	1.97	2.46	44	1.97	3.28	77	2.82	3.56
12	2.14	2.47	45	1.47	3.07	78	2.93	3.47
13	2.34	2.45	46	2.60	3.02	79	3.02	3.38
14	2.59	2.41	47	2.64	3.00	80	2.99	3.33
15	2.03	2.56	48	2.75	2.90	81	2.99	3.51
16	1.87	2.65	49	2.88	2.94	82	2.56	3.70
17	1.42	2.71	50	2.55	3.08	83	3.19	3.47
18	2.04	2.65	51	2.44	3.13	84	3.20	3.58
19	1.95	2.76	52	2.35	3.23	85	2.97	3.68
20	1.72	2.80	53	2.43	3.19	86	2.84	3.77
21	2.20	2.72	54	2.53	3.16	87	3.52	3.42
22	2.34	2.66	55	2.58	3.13	88	3.83	3.27
23	2.11	2.81	56	2.73	3.09	89	2.85	3.96
24	2.36	2.71	57	2.86	3.10	90	3.13	3.84
25	2.47	2.73	58	2.65	3.18	91	3.42	3.69
26	2.51	2.71	59	2.72	3.17	92	3.15	4.06
27	2.05	2.88	60	2.44	3.25	93	3.44	3.97
28	1.72	3.04	61	2.52	3.24	94	3.75	3.77
29	1.95	3.07	62	2.60	3.24	95	3.23	4.26
30	2.27	2.82	63	2.57	3.32	96	3.56	4.20
31	2.40	2.55	64	2.69	3.28	97	3.82	4.00
32	2.19	2.95	65	2.48	3.38	98	4.04	4.00
						99	4.00	4.36

$N=100$, $\bar{x}=2.52$, $\sigma_x=0.601$, $\bar{y}=3.10$, $\sigma_y=0.507$, $r_{xy}=0.82$, $y=1.36+0.692x$.

Таблица 2.7.2

Статистические характеристики выборок из модельной совокупности табл. 2.7.1

Номер выборки	\bar{x}	σ_x	\bar{y}	σ_y	r_{xy}	b_{yx}	a_{yx}	$b_{зав}$	$a_{зав}$
1	2.50	0.546	3.17	0.471	0.794	0.685	1.46	1.086	0.455
2	2.56	0.709	3.15	0.578	0.884	0.721	1.30	0.922	0.790
3	2.66	0.762	3.14	0.613	0.920	0.740	1.17	0.874	0.814
4	2.51	0.633	3.03	0.484	0.748	0.572	1.59	1.022	0.465
5	2.60	0.663	3.18	0.590	0.848	0.755	1.22	1.049	0.453
6	2.32	0.539	3.00	0.434	0.685	0.552	1.72	1.175	0.274
7	2.55	0.514	3.13	0.419	0.788	0.642	1.49	1.034	0.493
8	2.52	0.690	3.10	0.517	0.909	0.681	1.38	0.824	1.024
9	2.28	0.505	2.90	0.420	0.756	0.629	1.47	1.100	0.392
10	2.35	0.623	2.95	0.434	0.856	0.596	1.55	0.814	1.037
11	2.52	0.386	3.12	0.415	0.705	0.758	1.21	1.530	-0.736
12	2.64	0.741	3.16	0.532	0.862	0.619	1.53	0.833	0.961
13	2.26	0.541	2.93	0.432	0.756	0.604	1.56	1.056	0.543
Средние	2.48	0.613	3.07	0.492	0.808	0.658			
Средние квадратические отклонения средних	0.130	0.125	0.095	0.069	—	0.070	—	—	—

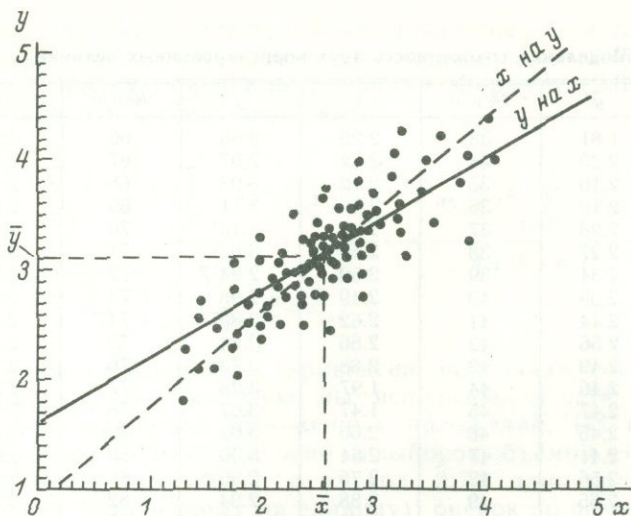


Рис. 2.7.1. Точечная диаграмма двумерной совокупности случайных величин, полученной моделированием с использованием таблицы случайных чисел.

Объем совокупности $N=100$, $\bar{x}=2.52$, $\sigma_x=0.601$, $\bar{y}=3.10$, $\sigma_y=0.507$. Сплошная линия — прямая регрессии y на x ($y=1.36+0.692x$), штриховая — x на y ($x=-0.496+0.972y$), она же — прямая функциональной зависимости y от x ($y=0.51+1.028x$).

достаточно велико и погрешность измерения y в элементах подвыборки равна нулю. Она равна

$$s_{\bar{y}}^2 = \frac{1}{\sum_{i=1}^n 1/s_y^2 + \sum_{i=n+1}^k 1/(s_y^2 + s_{ан}^2)}$$

Таблица 2.7.3

Сравнение оценок по регрессии с оценками «по зависимости»

Номер выборки	Выборочное среднее \bar{y}_n	Оценка \bar{y} по регрессии	Среднеквадратическая погрешность оценки	Оценка \bar{y} «по зависимости»	Среднеквадратическая погрешность оценки
1	3.17	3.19	0.057	3.19	0.057
2	3.15	3.12	0.054	3.12	0.065
3	3.14	3.03	0.048	3.02	0.067
4	3.03	3.03	0.064	3.04	0.061
5	3.18	3.12	0.063	3.10	0.068
6	3.00	3.11	0.063	3.19	0.058
7	3.13	3.11	0.052	3.10	0.051
8	3.10	3.10	0.043	3.10	0.057
9	2.90	3.06	0.055	3.12	0.053
10	2.95	3.05	0.045	3.08	0.050
11	3.12	3.12	0.059	3.12	0.052
12	3.16	3.09	0.054	3.07	0.061
13	2.93	3.08	0.057	3.16	0.054
Средние	3.07	3.093	0.054	3.1008	0.058
	$s_{\bar{y}}=0.099$	$s_{\bar{y}}=0.044$		$s_{\bar{y}}=0.059$	

Примечание: Оценка «по зависимости» включает индивидуальную оценку y в $k-n$ элементах по уравнению зависимости (сопряженному уравнению регрессии) и расчет средневзвешенной оценки из всех элементов сфинкс-выборки. $s_{зав}$ бралась по формуле (2.7.4).

так как имеется n независимых оценок \bar{y} , каждая с дисперсией s_y^2 (элементы подвыборки), и $k-n$ (независимых друг от друга и от n предыдущих) оценок, каждая с дисперсией $s_y^2 + s_{ан}^2$. Подставляя в качестве $s_{ан}$ в предыдущее равенство величину $s_{зав}$ (по формуле (2.7.1)), получим

$$s_{y_k}^2 = \frac{s_y^2}{kr^2 + n(1-r^2)}. \quad (2.7.6)$$

При $r=1$ получаем верный результат $s_{y_k} = s_y/\sqrt{k}$, при $r=0$ получаем также верный результат $s_{y_k} = s_y/\sqrt{n}$. Тем самым убеждаемся в том, что в рассматриваемом случае в качестве аналитической погрешности следует брать $s_{зав}$ по (2.7.1).

Из (2.7.6) можно получить выражение средней аналитической погрешности для рассматриваемого случая:

$$s_{\bar{y}}^2 = \frac{\bar{s}_{ан}^2 + s_y^2}{k} = \frac{s_y^2}{kr^2 + n(1-r^2)},$$

откуда

$$\bar{s}_{ан}^2 = s_y^2 \frac{(1-r^2) \left(1 - \frac{n}{k}\right)}{r^2 + \frac{n}{k}(1-r^2)}. \quad (2.7.7)$$

Если $n=k$, то при любом r получим $\bar{s}_{ан}^2=0$, что и ожидалось по смыслу. При $r=1$ также $\bar{s}_{ан}^2=0$. Если $r=0$, то $\bar{s}_{ан}^2 = (k/n-1) s_y^2$, причем если $k \gg n$, то погрешность оценки среднего в сфинкс-выборке равна $s_{ан}/\sqrt{k} = s_y\sqrt{k/n}/\sqrt{k} = s_y/\sqrt{n}$.

Геолога в такой же или в большей степени интересуют ситуации, когда нельзя пренебрегать погрешностью измерений $s_{анy}$ также и в подвыборке объема n . Схема оценки погрешности при этом сохраняется:

$$s_{\bar{y}}^2 = \frac{1}{\sum_{i=1}^n \frac{1}{s_y^2 + s_{анy}^2} + \sum_{i=n}^k \frac{1}{s_y^2 + s_{зав}^2}} = \frac{1}{\frac{n}{s_y^2 + s_{ан}^2} + \frac{k-n}{s_y^2 + s_{зав}^2}}. \quad (2.7.8)$$

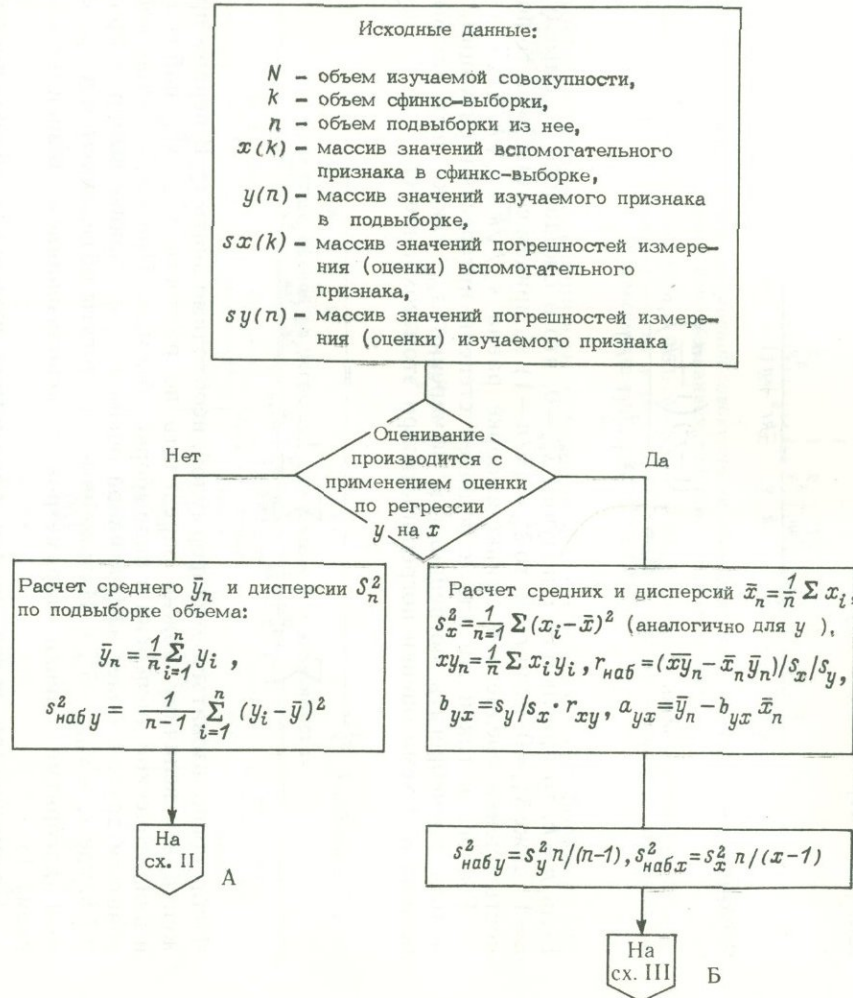
Чтобы воспользоваться этой формулой, необходимо знать $s_{анy}^2$. В первом приближении эту величину можно определить по разности $s_{наб}^2 - s_{анy}^2$ наблюдаемой и аналитической дисперсий в подвыборке объема n . При $k \gg n$ лучше воспользоваться другой, более эффективной оценкой s_y по сфинкс-выборке (формула 2.7.5, где $s_{y_n}^2 = s_{наб_{y_n}}^2 - \bar{s}_{анy}^2$). Изложенные результаты объединяются в предлагаемом алгоритме оценки параметров с использованием зависимостей (см. схему 1).

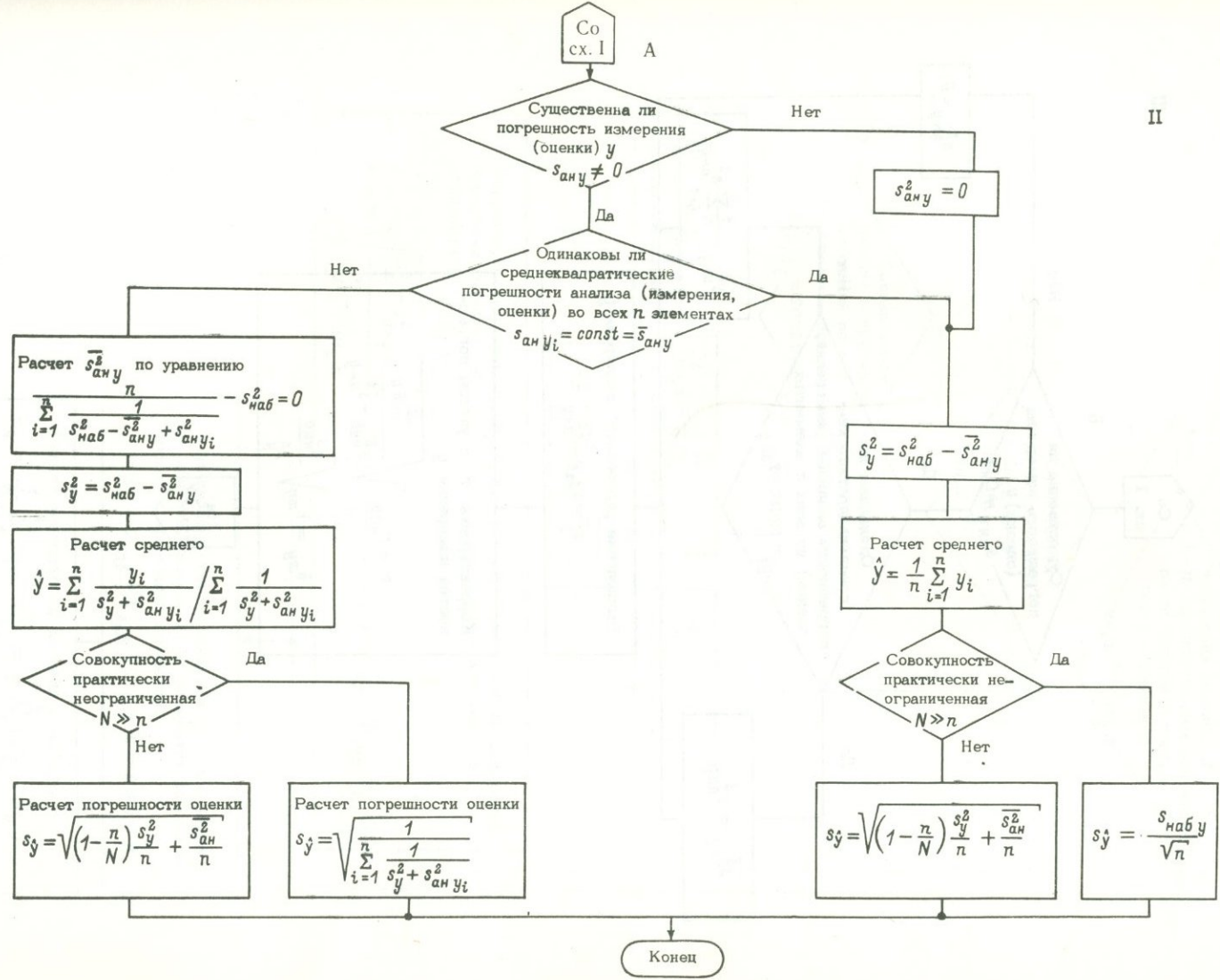
Преимущества такой оценки параметров изучаемых совокупностей становятся особенно очевидными, если в качестве вспомогательной используется одна из нескольких величин. В большом количестве образцов, проб или других элементов геологоразведочных выборок бывает известен целый ряд характеристик, коррелирующих с изучаемой. Поэтому сфинкс-выборку можно сформировать, используя уравнения регрессии на ту из вспомогательных величин, которая измерена в данном элементе выборки. При обычном оценивании «по регрессии» (первый способ оценивания по регрессии) таких возможностей не предусматривается.

Определенный резерв повышения точности оценивания кроется в том, что можно использовать сразу несколько вспомогательных величин. Нам не ясно,

Алгоритм оценивания среднего величины y и ее погрешности в изучаемой неавтокоррелированной совокупности

I





Расчет $\bar{s}_{ан y}^2$ по уравнению

$$\frac{1}{\sum_{i=1}^n \frac{1}{s_{наб}^2 - s_{ан y}^2 + s_{ан y_i}^2}} - s_{наб}^2 = 0$$

$$s_y^2 = s_{наб}^2 - \bar{s}_{ан y}^2$$

Расчет среднего

$$\hat{y} = \frac{\sum_{i=1}^n \frac{y_i}{s_y^2 + s_{ан y_i}^2}}{\sum_{i=1}^n \frac{1}{s_y^2 + s_{ан y_i}^2}}$$

Совокупность практически неограниченная
 $N \gg n$

Расчет погрешности оценки

$$s_y = \sqrt{\left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} + \frac{s_{ан}^2}{n}}$$

Расчет погрешности оценки

$$s_y = \sqrt{\frac{1}{\sum_{i=1}^n \frac{1}{s_y^2 + s_{ан y_i}^2}}}$$

$$s_y^2 = s_{наб}^2 - \bar{s}_{ан y}^2$$

Расчет среднего

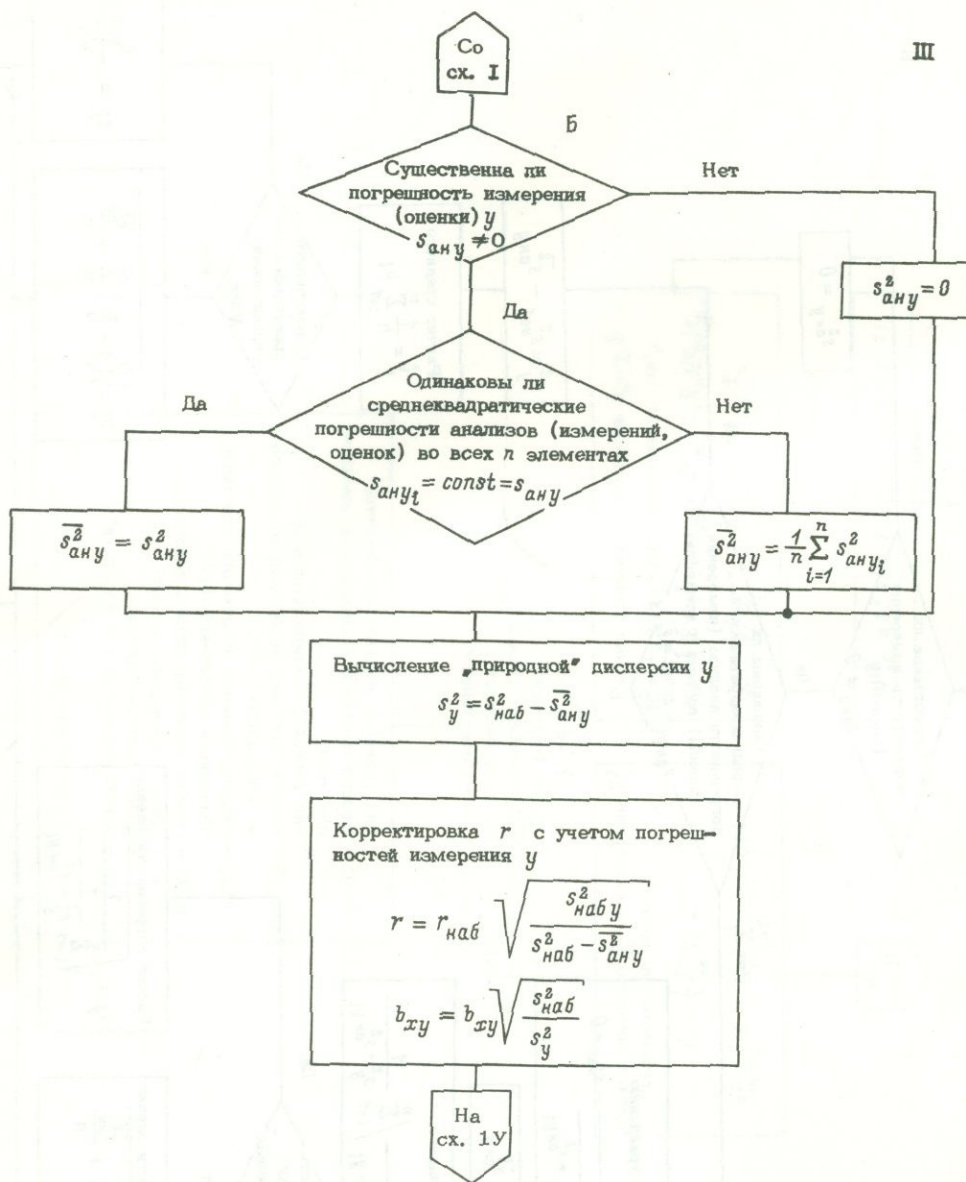
$$\hat{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

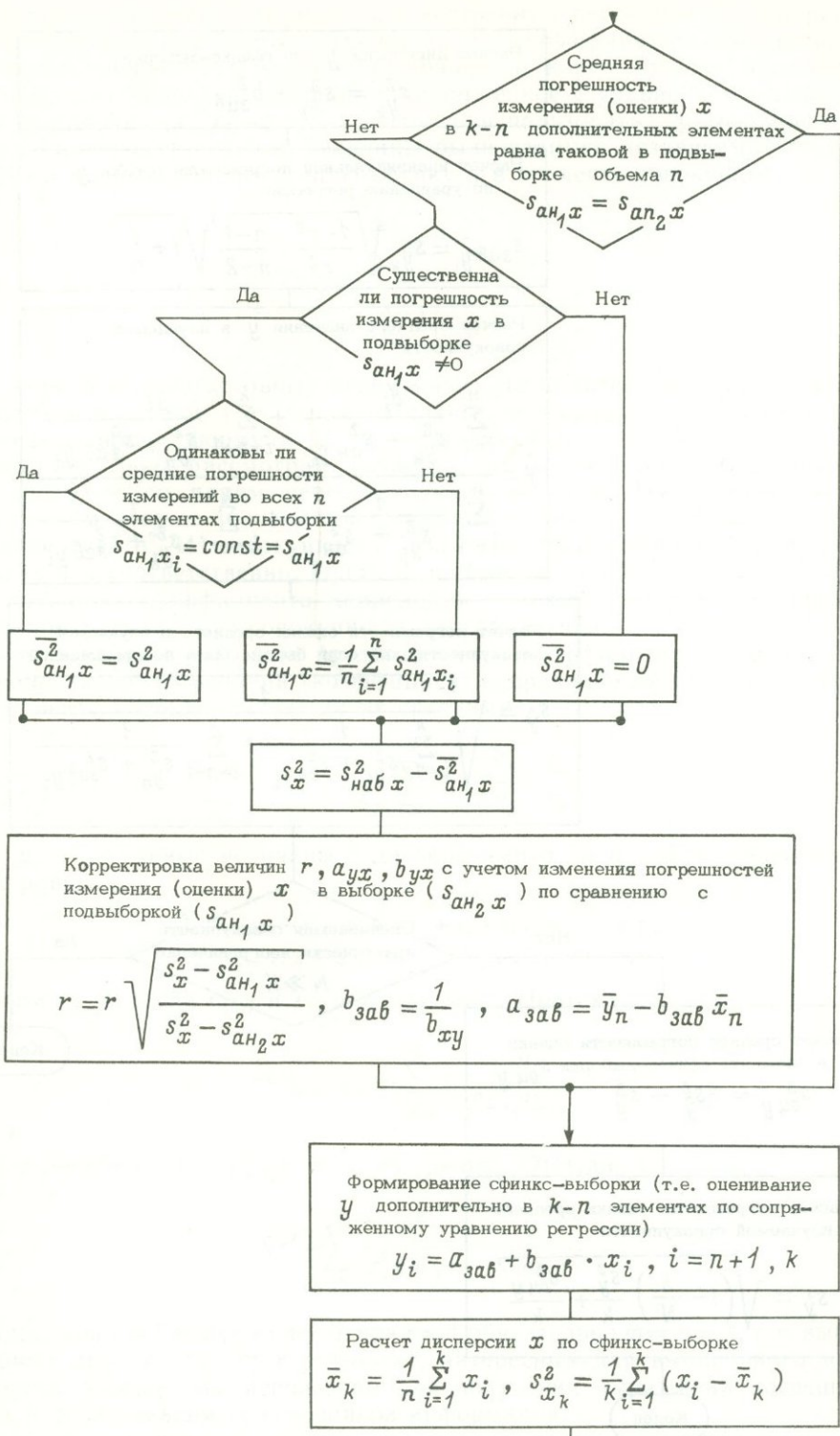
Совокупность практически неограниченная
 $N \gg n$

$$s_y = \sqrt{\left(1 - \frac{n}{N}\right) \frac{s_y^2}{n} + \frac{s_{ан}^2}{n}}$$

$$s_y = \frac{s_{наб} y}{\sqrt{n}}$$

Конец





Расчет дисперсии y по сфинкс-выборке

$$s_{y_k}^2 = s_{x_k}^2 \cdot b_{зав}^2$$

Расчет индивидуальной погрешности оценки y по уравнению регрессии

$$s_{зав y_i} = s_{y_k} \sqrt{\frac{1-r^2}{r^2} \cdot \frac{n-1}{n-2}} \sqrt{1 + \frac{1}{n}}$$

Расчет среднего значения y в изучаемой совокупности

$$\hat{y} = \frac{\sum_{i=1}^n \frac{y_i}{s_{y_k}^2 + s_{ан y_i}^2} + \sum_{i=n+1}^k \frac{y_i}{s_{y_n}^2 + s_{зав y_i}^2}}{\sum_{i=1}^n \frac{1}{s_{y_k}^2 + s_{ан y_i}^2} + \sum_{i=n+1}^k \frac{1}{s_{y_n}^2 + s_{зав y_i}^2}}$$

Расчет погрешности оценки среднего в изучаемой совокупности, как если бы она была неограниченной

$$s_{\hat{y}} = \sqrt{\frac{1}{\sum_{i=1}^k \frac{1}{s_{y_n}^2 + s_{ан y_i}^2} + \sum_{i=n+1}^k \frac{1}{s_{y_n}^2 + s_{зав y_i}^2}}}}$$

Нет

Оцениваемая совокупность
практически неограничена
 $N \gg K$

Да

Расчет средней погрешности оценки
 y в элементе сфинкс-выборки $s_{оц y}$

$$s_{оц y}^2 = k s_{\hat{y}}^2 - s_y^2$$

Расчет погрешности оценки среднего
в изучаемой совокупности

$$s_{\hat{y}} = \sqrt{\left(1 - \frac{k}{N}\right) \frac{s_y^2}{k} + \frac{s_{оц y}^2}{k}}$$

Конец

Конец

можно ли использовать для этого множественную регрессию при сохранении всей изложенной выше схемы с заменой парного коэффициента множественным R . Остается неясным, в частности, вопрос, в каком смысле понимать (и существует ли) уравнение регрессии, сопряженное с уравнением множественной регрессии y на $(x_1, x_2, x_3, \dots, x_n)$. Проще использовать парные зависимости.

Если бы вспомогательные величины были взаимно независимы, то можно было бы использовать следующую оценку y в элементе совокупности:

$$y = \frac{\sum_{i=1}^l y_i \cdot 1/s_{3ав_i}^2}{\sum_{i=1}^l 1/s_{3ав_i}^2}, \quad (2.7.9)$$

где l — число вспомогательных переменных. На практике распространен противоположный случай, т. е. вспомогательные переменные — зависимые источники информации и формула (2.7.9) не применима.

Предварительно рассмотрим коэффициент корреляции между выборочными средними по парам пересекающихся выборок. Пусть имеется совокупность элементов со значениями x_i , распределенными нормально со средним μ_x и дисперсией σ_x^2 . Из совокупности отбирается пара выборок A_i и B_i численностью m и n соответственно. Пусть A_i и B_i имеют k общих элементов. Определим значение коэффициента корреляции r между выборочными средними \bar{x}_A и \bar{x}_B при условии, что совокупность неограниченна.

Воспользуемся следующим представлением коэффициента корреляции для сумм коррелируемых величин (коэффициент корреляции между суммами равен коэффициенту корреляции между их средними):

$$r = \sum_{i=1}^m \sum_{j=1}^n r_{x_i y_j} \frac{\sigma_{x_i}}{\sigma_x} \cdot \frac{\sigma_{y_j}}{\sigma_y},$$

где x_i, y_j — случайные величины с дисперсиями σ_{x_i} и σ_{y_j} . Для нашего случая суммы равны:

$$\begin{aligned} \sum x_A &= x_1 + x_2 + x_3 + \dots + x_k + x_{k+1} + \dots + x_m, \\ \sum x_B &= x_1 + x_2 + x_3 + \dots + x_k + x_{m+1} + \dots + x_n. \end{aligned}$$

При этом $r_{x_i x_j} = 1$ для $i = j$ и $r_{x_i x_j} = 0$ для $i \neq j$. Поэтому

$$r = r_{(\sum x_A, \sum x_B)} = \sum_{i=1}^k 1 \cdot \frac{\sigma_{x_i} \sigma_{x_i}}{\sigma_{\sum x_A} \sigma_{\sum x_B}},$$

где $\sigma_{x_i} = \text{const} = \sigma_x$. Но $\sigma_{\sum x_A}^2 = m\sigma_x^2$, $\sigma_{\sum x_B}^2 = n\sigma_x^2$. Отсюда

$$\sqrt{r^2} = \sum_{i=1}^k \frac{\sigma_x^2}{\sigma_x^2 \sqrt{m} \sqrt{n}} = \frac{k}{\sqrt{m} \sqrt{n}}.$$

Очевидно, при $k=0$ получим $n=0$, при $r=1$ обязательно $m=n=k$, т. е. выборки повторяют друг друга. Этот результат, интересный для выборочного метода в геологии вообще, мы используем для получения выражения погрешности оценки и двух зависимых источников информации.

Пусть y_1 — оценка параметра из первого источника информации, s_1^2 — дисперсия этой оценки, y_2, s_2^2 — то же, из другого источника информации,

r — коэффициент корреляции между оценками y_1 и y_2 . Требуется определить дисперсию комплексной оценки \bar{y} :

$$\bar{x} = (y_1/\omega_1 + y_2/\omega_2) / (1/\omega_1 + 1/\omega_2).$$

где ω_1 и ω_2 — весовые коэффициенты, причем $\frac{1}{\omega_1 + \omega_2}$ есть дисперсия оценки y из двух источников. Воспользуемся для получения требуемой дисперсии частным случаем, заключающимся в следующем: \bar{y}_1 — оценка среднего по выборке m элементов, x — то же, по выборке n элементов. Пусть выборки содержат k общих (повторяющихся) элементов. В соответствии с результатами, изложенными выше, коэффициент корреляции между y_1 и y_2 равен $\sqrt{r^2} = k/\sqrt{m} \sqrt{n}$. Точность комплексной оценки мы в данном случае знаем: она равна $\frac{1}{s_y^2(m+n-k)}$. Будем также исходить из того, что весовые коэффициенты пропорциональны объемам выборок и имеют вид $\omega_1 = m/s_y^2 + m\Delta$, $\omega_2 = n/s_y^2 + n\Delta$, поэтому

$$s_y^2 = s_y^2/(m+n-k) = \frac{1}{m/s_y^2 + m/s_y^2 + (m+n)\Delta}. \quad (2.7.10)$$

Теперь требуется определить Δ :

$$\begin{aligned} (m+n-k)/s_y^2 &= m/s_y^2 + n/s_y^2 + (m+n)\Delta, \\ (m+n)\Delta &= (m+n-k)/s_y^2 - m/s_y^2 - n/s_y^2 = -k/s_y^2, \\ \Delta &= -k/s_y^2(m+n). \end{aligned}$$

Но $k = r \sqrt{m} \sqrt{n}$. Подставляя это, получаем $\Delta = -\sqrt{r^2} \sqrt{m} \sqrt{n} / (s_x^2(m+n))$. Подставляя это в (2.7.10) и преобразовывая, получаем

$$s_y^2 = \frac{1}{\left(\frac{1}{s_1^2} + \frac{1}{s_2^2} - \sqrt{r^2} \frac{1}{s_1 s_2}\right)}, \quad (2.7.11)$$

где s_1 — погрешность оценки y из первого источника информации, s_2 — из второго, r — коэффициент корреляции между этими источниками. Проведенный вывод не является доказательством, поскольку опирается на частный случай, но результат заслуживает некоторого доверия. Действительно, при $r=0$ это дает известную формулу для независимых источников информации. Когда $r^2=1$ обязательно равенство $s_1^2 = s_2^2$ и

$$s_y^2 = s_{\text{св}}^2 = \frac{1}{1/s_1^2} s_1^2 = s_1^2,$$

так как источники информации дублируют друг друга.

Обобщение этого результата на несколько источников информации сложно, однако при наличии ЭВМ, позволяющей пренебрегать вычислительными трудностями, можно избежать затруднений, вызванных зависимостью между вспомогательными переменными следующим образом. Первая вспомогательная величина x_1 используется как обычно. Затем строится уравнение регрессии второй вспомогательной величины x_2 на первую: $x_2 = a_{21} + b_{21}x_1$ и вычисляются так называемые остаточные величины

$$\Delta_{2i} = x_{2i} - (a_{21} + b_{21}x_{1i}),$$

которые и используются в качестве независимого от x_1 источника информации для оценки y . Вместо третьей вспомогательной величины используются остатки

$\Delta_3 = x_3 - f(x_1, x_2)$, где $f(x_1, x_2)$ — множественное уравнение регрессии x_3 на x_1 и x_2 .

В результате будет иметься столько независимых источников информации, сколько имеется (неважно, зависимых или нет) вспомогательных переменных: $x_1, \Delta_2, \Delta_3, \dots, \Delta_l$. Комплексную оценку y получим как среднее взвешенное по формуле

$$y = \frac{\sum_{i=1}^l y_i / s_{зав_i}^2}{\sum_{i=1}^l 1 / s_{зав_i}^2}, \quad (2.7.12)$$

а погрешность оценки — по формуле

$$s_{y_e} = \sqrt{\frac{1}{\sum_{i=1}^l 1 / s_{зав_i}^2}},$$

где y_i — оценка y по сопряженному уравнению регрессии с уравнением регрессии y на i -ю вспомогательную величину (если $i \neq 1$, то на соответствующий остаток).

s_{y_e} всегда не больше, чем погрешность оценки зависимости от любой из использованных вспомогательных величин, что и определяет полезность такой вычислительной процедуры, направленной на полную «утилизацию» всей информации. Величины y и s_y , напомним, — оценка изучаемой переменной в элементе сфинкс-выборки и ее погрешность, которые обычным порядком используются в оценке среднего совокупности по второму способу.

Глава 3

ОЦЕНИВАНИЕ СРЕДНЕГО ЗНАЧЕНИЯ ПРОИЗВЕДЕНИЯ НЕСКОЛЬКИХ ВЕЛИЧИН С ПРИМЕНЕНИЕМ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА

§ 1. Общие замечания

При разведке месторождений интерес представляют не столько отдельные величины, сколько функции от нескольких величин. Частным, но самым важным случаем является произведение двух или нескольких величин, широко используемое при подсчете запасов. Это произведение мощности на содержание (так называемый «метропроцент»), мощности на содержание и на объемный вес полезного ископаемого (линейный запас). Запасы нефти Q — это интеграл от произведения:

$$Q = \int_S h K_n K_{np} \rho dS, \quad (3.1.1)$$

где h — мощность залежи, K_n , K_{np} — коэффициенты пористости и проницаемости, ρ — плотность нефти в пластовых условиях. Довольно часто он заменяется произведением средних значений: $Q = \bar{h} \bar{K}_n \bar{K}_{np} \bar{\rho}_n \int_S dS = \bar{h} \bar{S} \bar{K}_{np} \bar{\rho}_n$, где \bar{S} —

площадь залежи. В этом и во всех подобных случаях возникают вопросы, правомерна ли такая замена и в каких случаях ее можно осуществлять без риска ввести существенную систематическую погрешность. Они интенсивно обсуждались в геологоразведочной литературе, в частности в рамках непрекращающейся дискуссии о лучшем среднем. В последнее время эти вопросы отошли на задний план в связи с широким внедрением методов аппроксимации на ЭВМ значений свойств по площади месторождения трендами, сплайнами, с помощью процедур крайгинга (в отечественной практике — крайне редко). При этом оценивание среднего значения произведения легко осуществляется интегрированием произведения соответствующих «поверхностей», как говорят геологи. Однако проблема по существу осталась нерешенной: неясно, в каких случаях оценка (3.1.1) эффективнее (т. е. имеет меньшую погрешность), чем произведение средних? Каким образом поступать, если данных по какому-либо свойству явно недостаточно для сколько-нибудь уверенной аппроксимации?

Наконец, автор (Ткачев, 1973, 1975) поставил следующие вопросы. Во-первых, как изменится погрешность оценки среднего значения произведения геологоразведочных параметров, если в тех точках наблюдений, где не измерено то или иное требуемое свойство, его оценивать по уравнению регрессии на измеренные свойства (т. е. с образованием сфинкс-выборки)? Во-вторых, нельзя ли получить с использованием корреляций и регрессий несмещенную и наиболее эффективную оценку непосредственно, без оценивания недостающей

характеристики в каждой точке наблюдения, т. е. в каждом элементе геолого-разведочной выборки? В-третьих, останутся ли верными тождества, связывающие среднее значение произведения с произведением средних для сфинкс-выборки?

§ 2. Тождества, связывающие среднее значение произведения нескольких величин с произведением их средних

Из определения ковариации двух случайных величин $\text{cov}_{xy} = M\{(x-\bar{x})(y-\bar{y})\}$ легко получить следующее равенство: $\bar{x}\bar{y} = M(xy) = \bar{x}\bar{y} + \text{cov}_{xy}$, или $\bar{x}\bar{y} = \bar{x}\bar{y} + \rho_{xy}\sigma_x\sigma_y$. Оно остается верным и для выборочных значений соответствующих величин:

$$\bar{x}\bar{y}_n = \bar{x}_n\bar{y}_n + r_{xy}S_{x_n}S_{y_n}. \quad (3.2.1)$$

В дискуссиях о лучшем среднем это тождество по существу подразумевалось, но впервые в геологии в связи с этим вопросом было записано и применено В. И. Кузьминым (1966 и др.). В последующие годы оно встречается и в работах других авторов (Ткачев, 1973; Демин, Конторович, 1973, и др.). Значительно позже оно появляется у Д. Марсала (Марсал, 1978), который полагает, что «его до сих пор явно не рассматривали» (с. 100).

Для трех переменных требуется рассмотреть величину, которая в теории вероятностей и математической статистике практически не употребляется: $\text{cov}_{xyz} = M\{(x-\bar{x})(y-\bar{y})(z-\bar{z})\}$. Раскрывая скобки, получим $M\{(x-\bar{x})(y-\bar{y})(z-\bar{z})\} = M(xyz) - \bar{y}M(xz) - \bar{x}M(yz) - \bar{z}M(xy) + 2\bar{x}\bar{y}\bar{z} = M(xyz) - \bar{y}(\bar{x}\bar{z} + \text{cov}_{xz}) - \bar{x}(\bar{y}\bar{z} + \text{cov}_{yz}) - \bar{z}(\bar{x}\bar{y} + \text{cov}_{xy}) + 2\bar{x}\bar{y}\bar{z} = M(xyz) - \bar{y}\bar{x}\bar{z} - \bar{z}\text{cov}_{xz} - \bar{x}\bar{y}\bar{z} - \bar{x}\text{cov}_{yz} - \bar{z}\bar{y}\bar{x} - \bar{z}\text{cov}_{yx} + 2\bar{x}\bar{y}\bar{z} = M(xyz) - \bar{x}\bar{y}\bar{z} - \bar{z}\text{cov}_{xy} - \bar{y}\text{cov}_{xz} - \bar{x}\text{cov}_{yz}$, т. е. $\text{cov}_{xyz} = \bar{x}\bar{y}\bar{z} - \bar{x}\bar{y}\bar{z} - \bar{z}\text{cov}_{xy} - \bar{y}\text{cov}_{xz} - \bar{x}\text{cov}_{yz}$, откуда $\bar{x}\bar{y}\bar{z} = \bar{x}\bar{y}\bar{z} + \text{cov}_{xyz} + \bar{x}\text{cov}_{yz} + \bar{y}\text{cov}_{xz} + \bar{z}\text{cov}_{xy}$, или

$$\bar{x}\bar{y}\bar{z} = \bar{x}\bar{y}\bar{z} + \text{cov}_{xyz} + \bar{x}\rho_{yz}\sigma_y\sigma_z + \bar{y}\rho_{xz}\sigma_x\sigma_z + \bar{z}\rho_{xy}\sigma_x\sigma_y. \quad (3.2.2)$$

Попытку получить равенство, по смыслу аналогичное (3.2.2), предприняли В. И. Демин и А. Э. Конторович (1973). При выводе они по существу исходили из того, что оценивание по регрессии может проводиться для y и z по уравнению регрессии только на x . В результате получено выражение, несимметричное относительно всех трех величин и имеющее вследствие этого более частный характер, но весьма полезное для практики. Д. Марсал (1978) приводит для трех переменных приближенное выражение, отличающееся от (3.2.2) отсутствием слагаемого cov_{xyz} .

Для четырех переменных аналогичные выкладки дают следующее тождество:

$$\begin{aligned} \bar{x}\bar{y}\bar{z}\bar{u} = & \bar{x}\bar{y}\bar{z}\bar{u} + \text{cov}_{xyz} + \bar{u}\text{cov}_{xyz} + \bar{z}\text{cov}_{uxy} + \\ & + \bar{y}\text{cov}_{xuz} + \bar{x}\text{cov}_{yzu} + \bar{z}\bar{u}\text{cov}_{xy} + \bar{y}\bar{u}\text{cov}_{xz} + \bar{y}\bar{z}\text{cov}_{xy} + \\ & + \bar{x}\bar{u}\text{cov}_{yz} + \bar{x}\bar{z}\text{cov}_{yu} + \bar{x}\bar{y}\text{cov}_{uz}. \end{aligned} \quad (3.2.3)$$

Обобщая по аналогии на n величин, можно записать

$$\begin{aligned} \prod_{i=1}^n x_i = & \prod_{i=1}^n \bar{x}_i \left[1 + \frac{\text{cov}_{i_1, i_2, i_3, \dots, i_n}}{\prod_{i=1}^n \bar{x}_i} \right] + \sum_{j=1}^n \frac{\text{cov}_{i_1, i_2, i_3, \dots, i_n (i \neq j)}}{\prod_{i=1, i \neq j}^n \bar{x}_i} + \\ & + \sum_{j_1=1}^n \sum_{j_2=j_1+1}^n \frac{\text{cov}_{i_1, i_2, i_3, \dots, i_n (i \neq j_1, i \neq j_2)}}{\prod_{i=1, i \neq j_1, j_2}^n \bar{x}_i} + \dots + \sum_{j_1=1}^n \sum_{j_2=j_1+1}^n \frac{\text{cov}_{j_1, j_2}}{x_{j_2} x_{j_1}}. \end{aligned} \quad (3.2.4)$$

У Д. Марсала в тождествах для трех и более переменных присутствуют только слагаемые типа последнего в (3.2.4), что сильно искажает их правую часть.

Тождества (3.2.1) — (3.2.4) сами по себе бесполезны для оценивания, так как значительно легче непосредственно вычислить их левую часть, чем правую. Они лишь делают очевидным то, что произведение средних не равно среднему произведению.

§ 3. Анализ условий, при которых произведение средних предпочтительнее среднего значения произведения

Если оцениваемой является величина $\int_S mcdS$ или $\frac{1}{N} \sum_{i=1}^N m_i c_i = \bar{m} \bar{c}_N$, где

m — мощность пласта в точке i , c — содержание полезного ископаемого в ней, N — объем изучаемой совокупности, то, спрашивается, зачем искать другую оценку этой величины, в частности заменять ее величиной $\bar{m}_N \bar{c}_N$. Следует, однако, принять во внимание, что в распоряжении геологоразведчика имеется только выборка ограниченного объема $n < N$. В этих условиях, вероятно, имеет смысл вопрос: не является ли величина $\bar{m}_n \bar{c}_n$ лучшей оценкой величины $\bar{m} \bar{c}_N$, чем $\bar{m} \bar{c}_n$, несмотря на наличие в ней смещения? Действительно, давно установлен геологический факт, что при положительной корреляции между m и c коэффициент вариации среднего значения произведения mc по ряду выборок ($V_{(mc)}$) больше коэффициента вариации $V_{(\bar{m}\bar{c})}$ по этим же выборкам (см. например, табл. 3.3.3). Погрешность оценки параметра $\bar{m} \bar{c}_N$ статистикой $\bar{m} \bar{c}_n$ по n данным (S_{mc}/\sqrt{n}) при этом также больше, чем погрешность оценки параметра $\bar{m}_n \bar{c}_n$ статистикой $\bar{m}_n \bar{c}_n$ ($S_{(\bar{m}\bar{c})}$). Может ли получиться, что при некоторых n и r разность этих погрешностей превысит смещение и величина $\bar{m}_n \bar{c}_n$ будет более предпочтительной оценкой $\bar{m} \bar{c}_N$, чем $\bar{m} \bar{c}_n$? Здесь возможны по крайней мере три варианта постановки задачи: а) выбрать такую оценку, которая в среднем (точнее — в среднем квадратическом) имела бы меньшую погрешность, чем альтернативная; б) выбрать оценку, которая была бы чаще ($P > 0.5$) более близкой к истинной, чем альтернативная; в) выбрать оценку, которая минимизировала бы максимальное отклонение, т. е. гарантировала от может быть маловероятных, но грубых промахов. Выяснение этого вопроса во втором варианте постановки посвятил многочисленные работы В. И. Кузьмин (1966, 1967, 1968, 1972а, 1972б, 1972в, 1976, и др.). Анализ его результатов необходимо предпослать одно замечание. Обсуждаемое В. И. Кузьминым соотношение средневзвешенных $X_{взв}$ и среднеарифметических \bar{X}_a оценок эквивалентно соотношению рассматриваемых нами оценок $\bar{X}\bar{Y}$ и $\bar{X}\bar{Y}$, так как доля обе части равенства на \bar{Y} :

$$\bar{X}\bar{Y}/\bar{Y} = \bar{X}\bar{Y}/\bar{Y} + r_{xy} S_x S_y / \bar{Y},$$

получаем $\bar{X}_{взв} = \bar{X}_{ар} (1 + r_{xy} V_x V_y)$.

Приведем одну из таблиц, в которую В. И. Кузьмин свел результаты своих теоретических и экспериментальных исследований (табл. 3.3.1). В этой таблице $\Delta X_b = \bar{X}_{взв} - \bar{X}_{ист}$, $\Delta X_a = \bar{X}_a - \bar{X}_{ист}$, $\bar{X}_{взв} = \bar{X} Y_n / \bar{Y}_n$, $\bar{X}_a = \bar{X} \bar{Y}_n / \bar{Y}_n$. К сожалению, таблицей невозможно воспользоваться практически, так как никогда не бывает известно заранее, с каким из четырех случаев мы имеем дело. Чтобы узнать это, надо иметь $x_{ист}$, тем самым образуется замкнутый круг и рекомендации «повисают в воздухе». Приведем характерное высказывание В. И. Кузьмина (1967, с. 140): «Операция взвешивания (т. е. по существу расчет $(\bar{m}\bar{c})_n$, — Ю. Т.) при подсчете запасов является трудоемкой и нежелательной. С другой стороны, при малых значениях $V_x \cdot V_y$ и r_{xy} взвешивание не дает заметного уточнения и даже может приводить к некоторому ухудшению

Таблица 3.3.1

Случай	Условия, при которых выполняется взвешивание		Результат, к которому приводит применение формул средневзвешенного
	$r > 0$	$r < 0$	
1	$\Delta X_b > 0,$ $\Delta X_a < 0$	$\Delta X_b < 0,$ $\Delta X_a > 0$	Точность средневзвешенного повышается при некотором $n > a$
2	$\Delta X_b < 0,$ $\Delta X_a < 0$	$\Delta X_b > 0,$ $\Delta X_a > 0$	Взвешивание всегда дает лучший результат независимо от числа наблюдений
3	$\Delta X_b > 0,$ $\Delta X_a > 0$	$\Delta X_b < 0,$ $\Delta X_a < 0$	Взвешивание всегда дает худший результат
4	$\Delta X_b < 0,$ $\Delta X_a > 0$	$\Delta X_b > 0,$ $\Delta X_a < 0$	Случай невозможен

оценки. Не удовлетворяясь полученными результатами, В. И. Кузьмин вновь и вновь возвращается к анализу этого вопроса, проделывает длинные выкладки, но не получает убедительных выводов (Кузьмин, 1981, и др.). В конце концов автор рекомендует не аналитическое решение вопроса, а номограммы, полученные моделированием на ЭВМ двух коррелирующих величин. Распределения их подбирались так, чтобы охватить наиболее распространенные в геологии типы. Особенностью экспериментов было то, что по их результатам не высчитывались непосредственно погрешности сравниваемых оценок, а определялась некоторая промежуточная статистика, по которой уже аналитическим путем рассчитывались погрешности. Из этих номограмм мы сделали необходимую практическую выборку (табл. 3.3.2). В принятой В. И. Кузьминым постановке задачи, по-видимому, единственный путь решения заключается в моделировании совокупностей, выборок из них и сравнении погрешностей выборочных оценок. Практически это не представляется удобным для пользователя.

Таблица 3.3.2

Таблица предпочтения среднеарифметической оценки. Составлена автором по данным П. И. Кудряшова и В. И. Кузьмина (1981)

Модельный коэффициент корреляции ρ	Число проб в выборке, начиная с которого средневзвешенная оценка с вероятностью более 0.5 ближе к истинной, чем среднеарифметическая		
	Распределение X гиперболическое, Y — правоасимметричное		Распределения X и Y левоасимметричные $V_X=1.0, V_Y=0.6$
	$V_X=1.42,$ $V_Y=0.85$	$V_X=3.74;$ $V_Y=0.85$	
0.05	32	80	80
0.10	18	54	36
0.15	15	40	22
0.20	10	32	17
0.25	9	28	15
0.30	9	—	10
0.35	8	—	8
—0.05	7	78	18
—0.10	3	53	10
—0.15	2	40	8
—0.20	1	33	5
—0.25	1	27	3
—0.30	—	—	—

Часто более широкий взгляд на проблему упрощает ее решение. В более общем виде рассматриваемая задача заключается в следующем. Имеются смещенная оценка $\hat{\theta}$ (т. е. $\bar{m}\bar{c}$) и оценка смещения δ (т. е. cov_{mc}), полученные с некоторой существенной случайной погрешностью σ_δ . Очевидно, что в первом варианте постановки задачи (минимизировать среднеквадратическую погрешность оценки параметра) поправку на смещение следует вводить только тогда, когда оценка смещения больше, чем погрешность, связанная с применением поправки на смещение. Действительно, дисперсия оценивания смещенной оценкой (без применения поправки) равна

$$\sigma_{\hat{\theta}}^2 = \sigma_{\hat{\theta}^*}^2 + \delta^2, \quad (3.3.1)$$

где $\sigma_{\hat{\theta}}^2$ — дисперсия смещенной оценки относительно ее математического ожидания, равного $\hat{\theta} - \delta$; δ — смещение. Дисперсия оценивания этого же параметра несмещенной оценкой (с применением поправки) равна

$$\sigma_{\hat{\theta}}^2 = \sigma^2 (\hat{\theta}^* + \delta) = \sigma_{\hat{\theta}^*}^2 + \sigma_\delta^2 + 2 \text{cov}_{\hat{\theta}^*\delta}. \quad (3.3.2)$$

Из сопоставления последних равенств легко устанавливается условие, при котором внесение поправки на смещение не целесообразно, а именно: правая часть (3.3.1) должна быть меньше правой части (3.3.2): $\sigma_{\hat{\theta}^*}^2 + \delta^2 < \sigma_{\hat{\theta}^*}^2 + \sigma_\delta^2 + 2 \text{cov}_{\hat{\theta}^*\delta}$, или

$$\delta^2 < \sigma_\delta^2 + 2 \text{cov}_{\hat{\theta}^*\delta}. \quad (3.3.3)$$

В аналитической практике смещение обычно некоррелировано с оценкой параметра (например, содержания, поэтому достаточно простого условия $\delta^2 < \sigma_\delta^2$, т. е. чтобы смещение было меньше среднеквадратической погрешности его оценки. Применим этот результат к нашему случаю. Рассмотрим тождество $\bar{m}\bar{c}_n = \bar{m}_n\bar{c}_n + \text{cov}_{m_n c_n}$. Оно позволяет нам рассматривать оценку $\bar{m}\bar{c}_n$ как сумму смещенной оценки $\bar{m}_n\bar{c}_n$ и смещения $\text{cov}_{m_n c_n}$. Тогда условие предпочтительности средней арифметической в соответствии с (3.3.3) можно записать следующим образом:

$$s_{\bar{m}_n\bar{c}_n}^2 + \text{cov}_{m_n c_n}^2 < s_{\bar{m}_n\bar{c}_n}^2 + s_{\text{cov}_{m_n c_n}}^2 + 2 \text{cov} \{ \bar{m}_n\bar{c}_n, \text{cov}_{m_n c_n} \}, \quad (3.3.4)$$

или

$$r_{mc}^2 s_m^2 s_c^2 < s_{\text{cov}_{m_n c_n}}^2 + r \{ \bar{m}_n, \bar{c}_n, \text{cov}_{m_n c_n} \} s_{\bar{m}_n\bar{c}_n} s_{\text{cov}_{m_n c_n}}. \quad (3.3.5)$$

При решении этого неравенства относительно n мы воспользуемся тремя экспериментальными фактами: 1) $r_{s_m s_c} \approx |r_{mc}|$; 2) $r \{ r_{mc}, s_m \cdot s_c \} \approx |r_{mc}|$; 3) $r \{ \bar{m}\bar{c}, \text{cov}_{mc} \} \approx |r_{mc}|$. Найдем выражение для $s_{\bar{m}_n\bar{c}_n}^2$ как для дисперсии произведения коррелирующих величин:

$$\begin{aligned} s_{\bar{m}_n\bar{c}_n}^2 &= s_m^2 \cdot \bar{c}^2 + s_c^2 \bar{m}^2 + 2 r_{mc} s_m s_c \bar{m}\bar{c} = \\ &= 1/n (s_m^2 \bar{c}^2 + s_c^2 \bar{m}^2 + 2 r_{mc} s_m s_c \bar{m}\bar{c}). \end{aligned} \quad (3.3.6)$$

Дисперсию ковариации определим как дисперсию произведения трех зависимых сомножителей: $s_{\text{cov}_{m_n c_n}}^2 = s_r^2 \cdot s_m s_c$. Найдем сначала $s^2 \{ s_m s_c \}$:

$$s^2 \{ s_m s_c \} = s_m^2 s_c^2 [(s_{s_m}/s_m)^2 + (s_{s_c}/s_c)^2] + 2r \{ s_m, s_c \} (s_{s_m}/s_m) \cdot (s_{s_c}/s_c).$$

Используя экспериментальный факт, что $r \{ s_m s_c \} = |r_{mc}|$, а также, что $s_{s_m} = s_m / \sqrt{2n}$, $s_{s_c} = s_c / \sqrt{2n}$, получим:

$$s^2 \{ s_m s_c \} = s_m^2 s_c^2 (1/2n + 1/2n + 2\sqrt{r_{mc}^2} \cdot 1/\sqrt{2n} \cdot 1/\sqrt{2n}) = \frac{s_m^2 s_c^2}{n} (1 + \sqrt{r_{mc}^2}). \quad (3.3.7)$$

Обозначая $s_m \cdot s_c = A$, запишем

$$s_{r_{mc} \cdot s_m \cdot s_c}^2 = s_{r_{mc} \cdot A}^2 = r_{mc}^2 A^2 [(s_d/r_{md})^2 + (s_A/A)^2 + 2r_{mc} A \{s_d/r_{md} \cdot s_A/A\}].$$

Используя, что $r_{mc} \{s_d/r_{md}\} \approx |r_{mc}|$ и что $s_r = (1-r^2)/\sqrt{n-1} \approx (1-r^2)/\sqrt{n}$, получим:

$$s_{cov_{mc}}^2 = r_{mc}^2 s_x^2 s_y^2 \left[\frac{(1-r_{mc}^2)^2}{nr_{mc}^2} + \frac{s_m^2 s_c^2 (1+|r_{mc}|)}{ns_m^2 s_c^2} + 2r_{mc} \frac{(1-r_{mc}^2)}{\sqrt{n} r_{mc}} \cdot \frac{s_m s_c \sqrt{1+|r_{mc}|}}{\sqrt{n} s_m s_c} \right] = \frac{s_m^2 s_c^2}{n} [(1-r^2)^2 + r^2(1+|r|) + 2(1-r^2)|r|\sqrt{1+|r|}] = \frac{s_m^2 s_c^2}{n} B. \quad (3.3.8)$$

Хотя формула (3.3.8) приближенная, в численном эксперименте, который будет описан ниже, она дала величину, совпадающую с эмпирической дисперсией (0.009 и 0.0088 соответственно). Подставим полученное в (3.3.5):

$$r^2 s_m^2 s_c^2 < \frac{s_m^2 s_c^2}{n} B + 2r \frac{s_m s_c}{\sqrt{n}} \sqrt{B} \frac{s_m s_c}{\sqrt{n}} \sqrt{\frac{1}{V_m^2} + \frac{1}{V_c^2} + 2r \frac{1}{V_m V_c}},$$

откуда

$$r^2 < 1/n [B + 2r \sqrt{B} \sqrt{C}], \quad (3.3.9)$$

где выражение для B см. в (3.3.8), а $C = \sqrt{1/V_m^2 + 1/V_c^2 + 2r/V_m V_c}$. Таким образом, среднее арифметическое лучше в том случае, когда

$$n < B + 2r_{mc} \sqrt{B} \sqrt{C} / r_{mc}^2. \quad (3.3.10)$$

Для апробации полученных результатов был проведен специальный расширенный численный эксперимент. Совокупность коррелирующих величин для него заимствована из табл. 2.7.2. Из нее с помощью таблицы случайных чисел (П. Мюллер и др., 1982) было получено 13 выборок объемом по 20 элементов. По каждой выборке вычислялись \bar{x} , \bar{y} , s_x , s_y , r_{xy} , cov_{xy} , $\bar{x} \cdot \bar{y}$ и \overline{xy} (табл. 3.3.3). Для изучения влияния на оценки отрицательной зависимости (при прочих равных условиях) была образована величина $y^* = 2\mu_y - y$. Таким образом обеспечивалось: $\bar{y}^* = \bar{y}$, $s_y = s_y$, $r_{xy} = -r_{xy^*}$, $cov_{xy} = -cov_{xy^*}$. По серии из 13 выборок определялись эмпирические значения погрешностей оценок следующих величин: $\bar{x} \cdot \bar{y}$, cov_{xy} , \overline{xy} , \overline{xy}^* , cov_{xy^*} , \overline{xy}^* , которые сопоставлялись с теоретическими. Теоретическое значение величины $s_{\overline{xy}}$ вычислялось по формуле (3.3.6), величины $s_{cov_{xy}}$ — по формуле (3.3.8). Схождение результатов хорошее.

Значение	$s_{\overline{xy}}$	$s_{cov_{xy}}$	$s_{\overline{xy}^*}$	$s_{cov_{xy^*}}$
Теоретическое	0.662	0.095	0.22	0.095
Эмпирическое	0.614	0.094	0.24	0.094

Затем проводился расчет теоретической погрешности оценки \overline{XY}_N (и \overline{XY}_N^*) величиной $\bar{X}_n \bar{Y}_n$, т. е. $S\hat{\theta}$ (3-я и 5-я колонки табл. 3.3.4; левая часть неравенства (3.3.4)). Расчет проводился по программе из табл. 3.3.5, и заинтересованному читателю легко повторить его. Теоретические значения не отклоняются от эмпирических более чем на стандартное отклонение погрешности:

Значение	$s\hat{\theta}$ при $r=0.85$	$s\hat{\theta}$ при $r=-0.82$
Теоретическое	0.724	0.364
Эмпирическое	0.746 ± 0.15	0.290 ± 0.06
	0.746	0.290
Стандартное отклонение эмпирической погрешности	$\sqrt{2 \cdot 13} = 0.15$	$\sqrt{2 \cdot 13} = 0.06$

Сопоставление произведения выборочных ($\bar{x}\bar{y}$) средних и выборочного среднего значения произведения коррелированных величин (модельный пример), способ получения исходных данных — см. пример к табл. 2.7.2

Номер выборки (по 20 элементов в выборке)	\bar{x}	\bar{y}	$\bar{y}^* = 2\bar{y}_N - \bar{y}_n$	Модель $\rho_{xy} = +0.82$			Модель $\rho_{xy^*} = -0.82$			Примечания
				$\bar{x}_n \bar{y}_n$	$\text{cov}_{xy} = r \times s_x \cdot s_y$	$\bar{x}\bar{y}$	$\bar{x}\bar{y}^*$	$\text{cov}_{xy^*} = r \times s_x \cdot s_{y^*}$	$\bar{x}\bar{y}^*$	
1	2.50	3.17	3.03	7.92	0.205	8.13	7.58	-0.204	7.38	$r_{xy} = +0.911$
2	2.56	3.15	3.05	8.06	0.362	8.43	7.81	-0.362	7.45	
3	2.66	3.14	3.06	8.35	0.430	8.78	8.14	-0.430	7.71	$\mu_x \mu_y = 7.81$
4	2.51	3.03	3.17	7.61	0.229	7.83	7.96	-0.229	7.73	
5	2.60	3.18	3.02	8.27	0.332	8.60	7.85	-0.332	7.52	$\mu_{xy} = 7.81 + 0.250 = 8.06$
6	2.32	3.00	3.20	6.96	0.160	7.12	7.42	-0.160	7.26	$r_{xy^*} = -0.911$
7	2.55	3.13	3.07	7.98	0.170	8.15	7.83	-0.170	7.66	$\mu_x \mu_y = 7.81$
8	2.52	3.10	3.10	7.81	0.324	8.14	7.81	-0.324	7.49	$\mu_{xy} = 7.81 - 0.250 = 7.56$
9	2.28	2.90	3.30	6.61	0.160	6.77	7.52	-0.160	7.36	$r_{\bar{x}\bar{y}}, \text{cov}_{xy} = +0.63$
10	2.35	2.95	3.25	6.93	0.231	7.16	7.64	-0.231	7.41	
11	2.52	3.12	3.08	7.86	0.113	7.98	7.76	-0.113	7.65	
12	2.64	3.16	3.04	8.34	0.340	8.68	8.03	-0.340	7.69	
13	2.26	2.93	3.27	6.62	0.177	6.80	7.39	-0.177	7.21	$r_{\bar{x}\bar{y}^*}, \text{cov}_{xy^*} = +0.70$
Среднее	2.48	3.07	3.13	7.64	0.249	7.89	7.75	-0.249	7.50	
Эмпирическое среднее квадратичное отклонение	0.130	0.095	0.095	0.614	0.094	0.706	0.220	0.094	0.176	
Эмпирическая дисперсия	0.0169	0.03×10^{-3}	$9.03 \cdot 10^{-3}$	0.38	8.84×10^{-3}	0.498	0.048	8.84×10^{-3}	0.031	
Теоретическое среднее квадратичное отклонение				0.662 ¹	0.095 ²		0.24 ¹	0.095 ²		
Квадратическая ошибка с учетом смещения относительно величины $\bar{X}\bar{Y}$ во всей совокупности				0.746	—	0.726	0.29	—	0.186	

¹ Теоретическое среднее квадратическое отклонение, вычисленное по формуле

$$s_{\bar{x}\bar{y}} = \sqrt{1/n (s_x^2 \bar{y}^2 + s_y^2 \bar{x}^2 + 2r s_x s_y \bar{x}\bar{y})}$$

² Теоретическое среднее квадратическое отклонение, вычисленное по формуле

$$s_{\text{cov}} = \sqrt{1/n s_x^2 s_y^2 \cdot B}$$

Сопоставление теоретических и эмпирических погрешностей оценок \bar{xy} по 13 выборкам объема $n=20$

Номер выборки	Оценка без поправки, т. е. $\bar{x}_n\bar{y}_n$				Оценка с поправкой, т. е. \bar{xy}_n			
	$\rho = +0.82$		$\rho = -0.82$		$\rho = +0.82$		$\rho = -0.82$	
	$s_{\bar{xy}}^2$	$s_{\bar{xy}}$	$s_{\bar{xy}}^2$ *	$s_{\bar{xy}}^*$	$s_{\bar{xy}}^2$	$s_{\bar{xy}}$	$s_{\bar{xy}}^2$	s_{xy}
1	0.483	0.650	0.099	0.315	0.468	0.684	0.0328	0.181
2	0.782	0.884	0.198	0.445	0.855	0.925	0.0240	0.155
3	0.963	0.981	0.245	0.495	1.04	1.02	0.0151	0.123
4	0.485	0.696	0.136	0.369	0.538	0.733	0.0507	0.225
5	0.724	0.851	0.176	0.419	0.797	0.893	0.0269	0.164
6	0.319	0.564	0.0956	0.309	0.354	0.595	0.0485	0.220
7	0.351	0.592	0.0798	0.283	0.388	0.623	0.0312	0.177
8	0.672	0.820	0.165	0.407	0.736	0.858	0.0224	0.150
9	0.285	0.534	0.0728	0.270	0.316	0.562	0.0294	0.172
10	0.435	0.659	0.114	0.338	0.480	0.693	0.0315	0.177
11	0.229	0.478	0.0512	0.226	0.252	0.502	0.0269	0.164
12	0.772	0.878	0.205	0.453	0.849	0.921	0.0402	0.200
13	0.322	0.567	0.0875	0.296	0.357	0.597	0.0351	0.187
Теоретическая среднеквадрати- ческая ошибка оценки		0.724		0.364		0.756		0.179
Эмпирическая среднеквадрати- ческая ошибка оценки		0.746		0.290		0.726		0.186

Таблица 3.3.5

Программа для м/к «Электроника» БЗ-34 для расчета погрешности оценки среднего значения произведения двух величин (метропроцента) по выборке объема n

00.ИП2 01.ИП1 02. — 03.П7 04.ИП4 05.ИП3 06. — 07.П8 08.ФХ² 09.ИП7 10.ФХ²11. + 12.П9 13.ИП5 14.2 15×16.ИП7 17.×18.ИП8 19.×20.ИП9 21.×22.П9 23.ИП1 24.ФХ² 25.ИП3 26.ФХ² 27.×28.ИП6 29. — 30.ИП9 31.×32.П9 33.ИП5 34.ФХ² 35.ИП2 36.ФХ² 37.×38.ИП4 39.ФХ² 40.×41.ИП9 42+43.С/П 44.ФХ 45.С/П. Инструкция: ввести в регистр памяти: x — в П1, σ — в П2, y — в П3, σ_y — в П4, r_{xy} — в П5, n — в П6. На индикаторе — дисперсия оценки, С/П — средняя квадратическая погрешность оценки.

Наконец, рассчитывались теоретические значения погрешностей оценок величин \bar{XY}_n и \bar{XY}_n^* статистикой $\bar{XY}_n = \bar{X}\bar{Y} + cov_{xy}$ (правая часть неравенства (3.3.4) с использованием выражений (3.3.6) и (3.3.8)). Программа для расчета приведена в табл. 3.3.6. Их отклонения от эмпирических также очень невелики:

Значение	$s_{\bar{xy}}$ при $r=0.85$	$s_{\bar{xy}}$ при $r=-0.82$
Теоретическое	0.756	0.179
Эмпирическое	0.726 ± 0.14	0.186 ± 0.04
Стандартное отклонение эмпирической погрешности	$\frac{0.726}{\sqrt{2 \cdot 13}} = 0.14$	$\frac{0.186}{\sqrt{2 \cdot 13}} = 0.04$

Таблица 3.3.6

Программа для м/к «Электроника» БЗ-34 вычисления погрешности оценки произведения двух величин (погрешности среднего значения метропроцента)

00.ФХ² 01.Ф√Х 02.П6 03.1 04.+05.Ф√Х 06.ИП6 07.×08.П7 09.ИП5 10.ФХ² 11.1 12.=13. — 14.П8 15.ИП7 16.×17.2 18.×19.П7 20.ИП6 21.1 22.+23.ИП5 24.ФХ² 25.×26.ИП7 27.+28.П7 29.ИП8 30.ФХ² 31.ИП7 32.+33.П7 34.ИП2 35.ИП1 36.— 37.П8 38.ИП4 39.ИП3 40.— 41.П9 42.ИП8 43.×44.ИП5 45.×46.2 47.×48.ИП8 49.ФХ² 50.+51.ИП9 52.ФХ² 53.+54.П0 55.ИП7 56.ИП8 57.ФХ² 58.×59.ИП9 60.ФХ² 61.×62.П4 63.Ф√Х 64.ИП0 65.Ф√Х 66.×67.ИП6 68.×69.2 70.×71.ИП4 72.+73.ИП0 74.+75.ИП1 76.ФХ² 77.×78.ИП3 79.ФХ² 80.×81.ИП8 82.— 83.С/П 84.Ф√Х 85.С/П. Инструкция: ввести x в П1, s_x в П2, y в П3, s_y в П4, r_{xy} в П5, n в П6, s_{xy}^2 — на индикаторе, С/П s_{xy} на индикаторе.

Таким образом, изложенная приближенная теория погрешностей рассматриваемых оценок подтверждается по крайней мере этими экспериментами, и мы можем на ее основании приступить к заключительному этапу их анализа.

В табл. 3.3.7 и 3.3.8 приведены максимальные значения объема выборки n в зависимости от r , V_x и V_y , при которых средние арифметические оценки предпочтительнее средневзвешенных. Таблицы рассчитаны по формуле (3.3.10) с применением программы из табл. 3.3.9. При положительной и тесной зависимости между величинами и высоких коэффициентах вариации область применения среднего арифметического ограничена очень малыми и малыми выборками. При едва обнаруживающихся зависимостях ($r=0.1 \div 0.2$), а также при малых коэффициентах вариации ($<20\%$) максимальный объем выборки достигает нескольких десятков, а при сочетании этих условий — нескольких сотен. Вообще, чем выше смещение (табл. 3.3.10), тем меньше объем выборки, при котором среднее арифметическое предпочтительнее.

Таблица 3.3.7

Предельный объем выборки, при котором среднее арифметическое и \bar{x}_y предпочтительнее среднего взвешенного и \bar{x}_y в зависимости от r_{xy} , V_x и V_y

V_x, V_y	$r=0.9$	$r=0.8$	$r=0.7$	$r=0.6$	$r=0.5$	$r=0.4$	$r=0.3$	$r=0.2$	$r=0.1$
2	6	7	8	10	12	17	24	44	136
1.5	7	8	10	11	14	18	26	47	142
1.2	8	9	11	13	16	20	29	50	147
1	9	10	12	14	17	22	31	53	152
0.8	11	12	13	16	20	25	35	58	161
0.6	13	15	17	20	24	30	40	66	174
0.5	15	17	19	23	27	33	45	72	185
0.4	18	21	23	27	31	39	51	81	202
0.3	24	26	30	34	39	48	63	96	229
0.2	34	38*	42	48	55	66	85	126	283
0.1	65	72	80	89	102	121	152	218	445
0.05	127	140	154	173	197	231	287	400	770

* Обсуждаемый пример.

Таблица 3.3.8

Предельный объем выборки, при котором среднее арифметическое ($n \bar{x}_y$) предпочтительнее среднего взвешенного (и \bar{x}_y) в зависимости от r_{xy} , V_x и V_y ($r_{xy} < 0$)

V_x, V_y	$r=-0.9$	$r=-0.8$	$r=-0.7$	$r=-0.6$	$r=-0.5$	$r=-0.4$	$r=-0.3$	$r=-0.2$	$r=-0.1$
2	2	3	3	4	5	8	13	27	105
1.5	2	2	2	3	4	6	11	25	100
1.2	2	2	2	2	3	5	10	23	95
1	0	0	0	2	2	4	8	20	90
0.8	0	0	0	0	0	2	6	16	84
0.6	0	0	0	0	0	0	0	10	71
0.5	0	0	0	0	0	0	0	5	61
0.4	0	0	0	0	0	0	0	0	47
0.3	0	0	0	0	0	0	0	0	22
0.2	0	0*	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0
0.05	0	0	0	0	0	0	0	0	0

* Обсуждаемый пример.

Для нашего примера ($r \approx 0.8 \sqrt{V_x V_y} = 0.2$) $n_{кр} < 38$. Поскольку $n = 20 < n_{кр}$, то следует сделать заключение, что среднее арифметическое здесь предпочтительнее. Действительно, средняя квадратическая погрешность (с учетом возникающего смещения) средней арифметической равна 0.724, взвешенной — 0.756

Программа для м/к «Электроника» БЗ-34 расчета предельного объема выборки l , при котором xy предпочтительнее, чем xy , в зависимости от r , V_x , V_y

00.FX² 01.F√X 02.П6 03.1 04.+05.F√X 06.ИП6 07.×08.П7 09.ИП6 10.FX² 11.1 12.=13. — 14.П8 15.ИП7 16.×17.2 18.×19.П7 20.ИП6 21.1 22.+23.ИП5 24.FX² 25.×26.ИП7 27.+28.П7 29.ИП8 30.FX² 31.ИП7 32.+33.П7 34.ИП1 35.ИП2 36.÷37.П9 38.ИП3-39.ИП4 40.— 41.П8 42.ИП9 43.×44.ИП5 45.×46.2 47.×48.ИП8 49.FX² 50.+51.ИП9 52.FX² 53.+54.П0 55.F√X 56.ИП7 57.F√X 58.×59.ИП5 60.×61.2 62.×63.ИП7 64.+65.ИП5 66.FX² 67.÷68.С/П. Инструкция: ввести x в П1, s_x в П2; y в П3, s_y в П4, r_{xy} в П5. Ответ на индикаторе (округлить до целого с избытком).

Таблица 3.3.10

Коэффициент смещения \bar{xy} относительно $\bar{x}\bar{y}$ в зависимости от r_{xy} , V_x и V_y

$\sqrt{V_x V_y} = V_{cp}$	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
2	4.6	4.2	3.8	3.4	3.0	2.6	2.2	1.8	1.4
1.5	3.0	2.8	2.6	2.4	2.1	1.9	1.7	1.5	1.2
1.2	2.3	2.2	2.1	1.9	1.7	1.6	1.4	1.3	1.1
1.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1
0.8	1.6	1.5	1.4	1.4	1.3	1.3	1.2	1.1	1.06
0.7	1.44	1.39	1.34	1.29	1.25	1.20	1.15	1.10	1.05
0.6	1.32	1.29	1.25	1.22	1.18	1.14	1.11	1.07	1.04
0.5	1.23	1.20	1.18	1.15	1.13	1.10	1.07	1.05	1.03
0.4	1.14	1.13	1.11	1.10	1.08	1.06	1.05	1.03	1.02
0.3	1.08	1.07	1.06	1.05	1.05	1.04	1.03	1.02	1.01
0.2	1.04	1.03*	1.02	1.02	1.02	1.02	1.01	1.01	1.004
0.1	1.009	1.008	1.007	1.006	1.005	1.004	1.003	1.002	1.001
0.005	1.002	1.002	1.002	1.002	1.001	1.001	1.001	1.0005	0.0003

* Обсуждаемый пример.

(хотя отличие не очень существенно). При отрицательной зависимости (табл. 3.3.1) область применения средней арифметической ограничена очень малыми выборками при очень высокой вариации переменных. В нашем примере безоговорочное предпочтение должно быть отдано средней взвешенной, погрешность которой (0.179) более чем в два раза меньше арифметической (0.364).

В заключение раздела заметим, что расчет средневзвешенного можно проводить как непосредственно усредняя произведения $X\bar{Y}_i$, так и по тождеству $\bar{X\bar{Y}} = \bar{X}\bar{Y} + \text{cov}_{xy}$.

§ 4. Повышение точности оценки произведения нескольких параметров без увеличения объема выборки

Внимание, которое было уделено сравнению среднеарифметических и средневзвешенных оценок, и в частности тождествам типа $\bar{xy} = \bar{x}\bar{y}$, не было оправдано, если бы мы ограничились только рекомендациями, когда какую оценку применять.

Нередко при разведке месторождений одни из изучаемых признаков могут быть определены в большем числе пунктов, чем другие. Например, из-за недостаточного выхода ядра содержание полезного ископаемого в некоторой части скважин оказывается непредставительным, в то время как мощность может быть определена с помощью каротажа достаточно точно. В горных выработках, пройденных в плоскости рудных тел (по простирацию или падению) и вскрывающих его на полную мощность, число проб для определения содержания ограничивается экономическими соображениями, а мощность может быть замерена по геологической документации практически в неограниченном числе точек, так как характер изменения можно непосредственно наблюдать. Таким образом, при разведке могут быть получены как парные данные, так по неко-

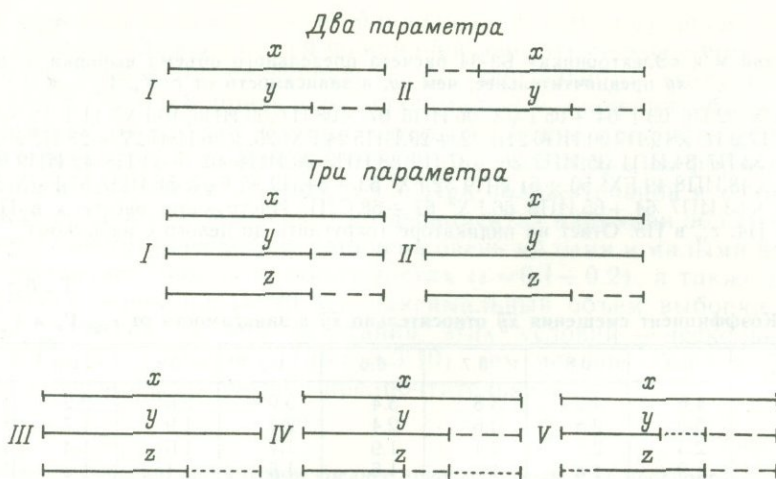


Рис. 3.4.1. Варианты некомплектности данных при двух и трех коррелирующих параметрах.

Сплошными линиями обозначены множества точек, в которых изучаемый параметр непосредственно измерен, штриховыми — доукомплектован с помощью парного уравнения регрессии на одну из измеренных переменных, пунктирными — доукомплектован с помощью множественного уравнения регрессии на две измеренные переменные.

тому числу точек и не парные, когда в точке наблюдения известна лишь одна из двух требующихся величин.

Если оцениваемый параметр есть произведение трех или большего числа величин, из которых в i -й точке измерены не все, то такие данные будем называть некомплектными. Вопрос методики уточнения запасов путем использования некомплектных данных в литературе не обсуждался. Он очень близок к вопросу об оценивании среднего значения какой-либо величины с помощью сфинкс-выборки. Отличие заключается в том, что вспомогательной величиной в данном случае является один из сомножителей, составляющих оцениваемый параметр.

Вопрос состоит в том, как получить более эффективную оценку $\overline{m\bar{c}}_N$, чем обычное средневзвешенное $\overline{m\bar{c}}_n$. Нельзя ли, воспользовавшись представлением величины $\overline{m\bar{c}}_n$ как суммы $\overline{m\bar{c}} + \text{cov}_{mc}$, улучшить оценку уточняя либо один из сомножителей первого слагаемого, либо второе слагаемое. Рассмотрим этот вопрос сначала для произведения двух величин. Здесь возможны два варианта: а) одна из величин (x) измерена во всех k элементах выборки, другая (y) — только в n элементах ($n < k$) (рис. 3.4.1, I); б) одна величина измерена в l элементах выборки объема k ($l < k$), другая — в m элементах ($m < k$), причем множества этих элементов пересекаются, объем пересекающейся части составляет n , т. е. обе они измерены в n элементах ($l > n, m > n$) (рис. 3.4.1, II).

Вариант (I). Запишем среднее значение произведения двух переменных в изучаемой выборке объема k в виде двух слагаемых. Первое — по n элементам выборки с известными x_i и y_i , второе — по $(k-n)$ элементам, в которых величина x измерена, а y определена по уравнению регрессии

$$(\bar{xy})_k = \frac{1}{k} \left(\sum_{i=1}^n x_i y_i + \sum_{i=n+1}^k \bar{x}_i \bar{y}_i \right). \quad (3.4.1)$$

Первое слагаемое заменим выражением по основному тождеству, во втором величину \bar{y}_i заменим ее выражением по уравнению регрессии, построенному по n элементам (в которых известны x и y): $\bar{y}_i = a_{yx} + b_{yx} x_i$, где

$$b_{yx} = \frac{\text{cov}(xy)_n}{s_{x_n}^2}, \quad a_{yx} = \bar{y}_n - b_{yx} \bar{x}_n = \bar{y}_k - \frac{\text{cov}_{xy_n}}{s_{x_k}^2} \bar{x}_k,$$

В результате получим

$$(\bar{xy})_n = \left(\frac{n}{k} \bar{x}_n \bar{y}_n + \frac{k-n}{k} \bar{x}_{k-n} \bar{y}_n \right) + \frac{\text{cov}_{xy_n}}{s_{x_n}^2} \cdot \left[\frac{n}{k} s_{x_n}^2 + \frac{(k-n)}{k} (\bar{x}_{k-n}^2 - \bar{x}_n \cdot \bar{x}_{k-n}) \right]. \quad (3.4.2)$$

Нижние индексы означают здесь число элементов, вошедших в расчет соответствующей статистики, и, кроме того, по смыслу — множество, которому эти элементы принадлежат. Вместо $s_{x_n}^2$ в (3.4.2) подставим $\bar{x}_n^2 - \bar{x}_n \cdot \bar{x}_n$:

$$\begin{aligned} (\bar{xy})_k = \bar{y}_n \left(\frac{n}{k} \bar{x}_n + \frac{k-n}{k} \bar{x}_{k-n} \right) + \frac{\text{cov}_{xy_n}}{s_{x_n}^2} \cdot \left[\left(\frac{n}{k} \bar{x}_n^2 + \frac{k-n}{n} \bar{x}_{k-n}^2 \right) + \right. \\ \left. + \bar{x}_n \left(\frac{n}{k} \bar{x}_n + \frac{k-n}{k} \bar{x}_{k-n} \right) \right] = \bar{x}_k \cdot \bar{y}_n + \frac{\text{cov}_{xy_n}}{s_{x_n}^2} (\bar{x}_n^2 - \bar{x}_n \cdot \bar{x}_k). \end{aligned}$$

Величину $\bar{x}_n^2 - \bar{x}_n \cdot \bar{x}_k$ назовем дисперсией по смешанному числу данных и обозначим $s_{x_n, k}^2$, тогда

$$(\bar{xy})_k = \bar{x}_k \cdot \bar{y}_n + \text{cov}_{xy_n} \frac{s_{x_n, k}^2}{s_{x_n}^2}. \quad (3.4.3)$$

Таким образом, устанавливается, что для уточнения среднего значения произведения можно использовать некомплектные данные, взяв произведения средних по одной переменной — по k данным, по другой — по n и введя поправку на зависимость, равную ковариации, скорректированной множителем $s_{x_n, k}^2 / s_{x_n}^2$. Проиллюстрируем это следующими примерами. В первом примере (табл. 3.4.1) мы наблюдаем, что при отсутствии корреляции между признаками среднее арифметическое ($\bar{m}\bar{c}$) является более эффективной оценкой величины $\bar{m}\bar{c}_N$, чем $\bar{m}\bar{c}$ (погрешности оценки равны 0.148 и 0.187 соответственно). Во втором примере между m и c существует тесная зависимость ($\rho = 0.841$). В табл. 3.4.2 приведены результаты исследования 9 выборок по 40 элементов. Несмотря на значительно меньшую вариацию $\bar{m}\bar{c}$ по сравнению $\bar{m}\bar{c}$ (0.374 против 0.491), погрешность оценки среднего метропроцента статистикой $\bar{m}\bar{c}$ почти в три раза выше, чем статистикой $\bar{m}\bar{c}$ (1.333 и 0.491 соответственно), за счет большого смещения.

Таблица 3.4.1

Среднее взвешенное, среднее арифметическое и погрешности их оценок при отсутствии корреляции между признаками в изучаемой совокупности

Номер выборки	Значения признаков из выборки $n=64$				cov _{mc}	$\bar{m}\bar{c} + \text{cov}_{mc}$
	\bar{m}	\bar{c}	$\bar{m}\bar{c}$	$\bar{m}\bar{c}$		
1	0.938	1.141	1.070	0.984	-0.137	0.933
2	1.031	1.078	1.111	0.891	-0.238	0.873
3	1.141	0.812	0.926	0.875	-0.052	0.874
4	0.906	0.828	0.750	0.813	0.062	0.812
5	1.062	0.875	0.929	0.938	0.023	0.952
6	1.125	0.969	1.090	0.969	-0.121	0.969
7	1.188	1.047	1.244	1.328	0.038	1.282
8	0.812	1.078	0.875	0.750	-0.126	0.749
9	0.984	0.984	0.968	1.031	0.062	1.030
10	0.844	0.922	0.778	0.672	-0.106	0.672
Среднее	1.003	0.979	0.974	0.925	-0.059	0.915
Эмпирические погрешности оценок	—	—	0.148	0.187	0.100	0.179

Таблица 3.4.2

Среднее взвешенное, среднее арифметическое и погрешности их оценок при высоком коэффициенте корреляции в совокупности ($\rho = +0.841$)

Номер выборки	Значение признаков из выборки $n = 40$				cov_{mc}	$\overline{m\bar{c}} + \text{cov}_{mc}$
	\bar{m}	\bar{c}	$\overline{m\bar{c}}$	$\overline{m\bar{c}}$		
1	2.206	0.874	1.928	2.927	1.050	2.978
2	1.956	0.814	1.592	2.563	1.000	2.592
3	1.949	0.723	1.409	2.497	1.122	2.531
4	2.134	0.722	1.541	3.537	2.052	3.593
5	2.546	0.946	2.408	3.678	1.309	3.717
6	2.256	0.864	1.949	2.670	0.740	2.690
7	2.377	1.037	2.465	3.729	1.295	3.760
8	2.223	0.886	1.970	3.192	1.255	3.225
9	2.441	0.895	2.185	3.484	1.329	3.514
Среднее	2.232	0.862	1.936	3.142	1.239	3.178
Эмпирические погрешности оценок	0.202	0.100	0.374	0.791	0.359	0.519
Эмпирические погрешности оценок величины $\overline{m\bar{c}}_M$			1.333	0.491		

В табл. 3.4.3 приведены результаты оценок среднего метропроцента с использованием некомплектных данных. В колонке (б) средний метропроцент оценивался как произведение средних плюс смещение. При этом среднее значение мощности принималось не по выборке из 40 элементов, а по неограниченному числу замеров, т. е. равным среднему значению по совокупности. Погрешность

Таблица 3.4.3

Сравнительная эффективность оценки среднего метропроцента ($\overline{m\bar{c}}$) по обычной методике и с использованием дополнительной информации по непарным данным ($\mu_m = 2.500$, $\rho = +0.841$)

Номер выборки	Значение метропроцента				
	обычное по выборке объемом $n = 40$ (а) ($\overline{m\bar{c}}_{40}$)	с использованием неограниченного числа замеров мощностей (б) ($\overline{m\bar{c}}_{1000} + \text{cov}_{m\bar{c}40}$)	с использованием известного коэффициента корреляции (в) ($\overline{m\bar{c}}_{40} + \text{cov}_{1000}$)	с использованием информации как в (б) и (в) (г) ($\overline{m\bar{c}}_{1000} + \text{cov}_{1000}$)	
1	2.927	3.235	3.095	3.352	
2	2.563	3.035	2.759	3.202	
3	2.497	2.929	2.576	2.974	
4	3.537	3.857	2.708	2.972	
5	3.678	3.674	3.575	3.532	
6	2.670	2.901	3.116	3.327	
7	3.729	3.887	3.632	3.759	
8	3.192	3.470	3.137	3.382	
9	3.484	2.567	3.352	3.405	
Среднее по выборкам	3.142	3.395	3.106	3.323	
Эмпирическая погрешность оценки	0.491	0.385	0.374	0.251	
Расчетная погрешность оценки	0.498	0.367	0.382	0.251	

такой оценки оказалась почти в полтора раза ниже традиционной. В колонке (в) в качестве оценки использовалось произведение средних по выборке $\bar{m}_{40}\bar{c}_{40}$, а в качестве поправки использовалась не выборочная, а истинная ковариация. Такой случай вполне возможен, например на нефтяных месторождениях. Средние значения пористости и остаточной водонасыщенности принимаются, как правило, по совокупности образцов или интервалов в контурах залежи в пределах ее эффективной мощности, а корреляция между этими характеристиками может быть установлена на основании значительно большего количества замеров. Погрешность оценки, как видно из табл. 3.4.3, уменьшается в 1.3 раза. Еще больший эффект повышения точности получается при одновременном использовании некомплектных данных по мощности и более точных данных по корреляции.

Описанный способ повышения точности оценки геологоразведочного параметра без увеличения объема выборки, насколько известно автору, в литературе не описан.

Вариант (II). Представим среднее значение произведения двух переменных в выборке объема k в виде трех слагаемых. Как и в варианте (а), первое слагаемое содержит данные по n элементам, в которых измерены обе величины — x и y , второе — по $l-n$ элементам, где измерен только x , третье — по $m-n$ элементам, где измерен только y :

$$(\bar{xy})_k = \frac{1}{k} \left(\sum_{i=1}^n x_i y_i + \sum_{i=n+1}^l x_i \bar{y}_i + \sum_{i=l+1}^k \bar{y}_i \bar{x}_i \right).$$

В отношении каждого слагаемого сделаем выкладки, аналогичные проведенным для варианта (а), в результате получим

$$\begin{aligned} (\bar{xy})_n &= \frac{1}{k} \left[\bar{x}_n \bar{y}_n + \text{cov}_{xy_n} + \sum_{i=n+1}^l x_i (B_{yx} x_i + \bar{y}_n - B_{yx} \bar{x}_n) + \right. \\ &\quad \left. + \sum_{i=l+1}^k y_i (B_{xy} y_i + \bar{x}_n - B_{xy} \bar{y}_n) \right] = \frac{(m-n) \text{cov}_{xy_n}}{k s_{y_n}^2} \cdot \\ &\quad \cdot (\bar{y}_{m-n}^2 - \bar{y}_n \cdot \bar{y}_{m-n}) + \frac{(l-n) \text{cov}_{xy_n}}{k s_{x_n}^2} (\bar{x}_{l-n}^2 - \bar{x}_n \cdot \bar{x}_{l-n}) + \\ &\quad + \frac{n \text{cov}_{xy_n}}{k} + \frac{n}{k} \bar{x}_n \bar{y}_n + \frac{m-n}{k} x_n y_{m-n} + \frac{l-n}{k} \bar{y}_n \bar{x}_{l-n}. \end{aligned}$$

Выражения в круглых скобках есть не что иное как дисперсия по смешанному числу данных. Проведя эту замену, получим

$$\begin{aligned} (\bar{xy})_n &= \left[\frac{m-k}{k} \bar{x}_n \bar{y}_{m-n} + \frac{n}{k} \bar{x}_n \bar{y}_n + \frac{(l-n)}{k} \bar{y}_n \bar{x}_{l-n} \right] + \\ &+ \text{cov}(xy)_k \left[\frac{m-k}{n} \cdot \frac{s^2(y)_{m-k,k}}{s^2(y)_k} + \frac{l-k}{n} \cdot \frac{s^2(x)_{l-k,k}}{s^2(x)_k} + \frac{k}{n} \right]. \end{aligned} \quad (3.4.4)$$

Введем несколько обозначений, упрощающих запись формулы (3.4.4) и проявляющих ее физический смысл. Величины $(m-n)/k$, $(l-n)/k$ и n/k являются долями изучаемой совокупности, в которых известны соответственно только значения y , только значения x и значения обеих величин. Тогда правую часть (3.4.4) можно интерпретировать как средневзвешенное из произведений $\bar{x}\bar{y}$ в трех подвыборках с поправкой в виде некоторой функции от ковариации cov_{xy_n} .

Погрешности оценок с использованием некомплектных данных рассчитываются по соответствующим формулам § 3, а входящие в них составляющие этой погрешности должны быть рассчитаны с учетом дополнительных данных.

Выражения среднего значения произведения трех и четырех входящих в формулы подсчета запасов величин очень громоздки, и мы их не приводим. К тому же число вариантов неукomплектованности данных велико (рис. 3.4.1). В этих случаях нет необходимости использовать обсуждаемые выражения, так как проще доукомплектовать выборку оцениванием недостающих данных по зависимости, т. е. использовать изложенные в предыдущей главе методы обработки сфинкс-выборок.

В самое последнее время интерес к доукомплектованию выборок, т. е. к созданию сфинкс-выборок, значительно возрос в связи с попытками унификации и стандартизации обработки экспериментальных данных вообще. Изучаются возможности обработки данных при наличии пропущенных значений, создается программа в СССР и США (Айвазян и др., 1983), поэтому разработанные и изложенные в этой главе методы могут иметь значение не только в геологии.

Глава 4

КОНТРОЛЬ ПОГРЕШНОСТЕЙ И ОПТИМИЗАЦИЯ ОБРАБОТКИ ПРОБ ПРИ РАЗВЕДКЕ МЕСТОРОЖДЕНИЙ И ПОДСЧЕТЕ ЗАПАСОВ

§ 1. Общие замечания

Изучение геологических объектов, в особенности месторождений полезных ископаемых, включает в себя опробование в широком смысле слова. Оно заключается во взятии образца или пробы, ее обработке, т. е. приведении к виду, пригодному для анализа, и в анализе, т. е. измерении требуемых свойств.

Проблемы опробования и обработки проб являются по существу экономическими, и все они в конечном счете сводятся к одному: дать оценку объекта или его части с заданной (контролируемой) точностью, затратив на это минимум времени, материальных и трудовых затрат. Достичь этого не просто, так как разработка методики оптимизации сама по себе требует немалых усилий. В общей схеме геологоразведочного процесса затраты на обработку проб не являются подавляющими, но в масштабах крупного геологического региона и всей страны они составляют сотни миллионов рублей. До 40 % затрат приходится на ручной труд, обычно тяжелый физически, даже если речь идет не о собственно ручном дроблении, а о подсобных операциях. Работа проводится в условиях запыленности, рабочие органы измельчителей сильно изнашиваются, контроль процесса обработки по точности сохранения содержания химических элементов или ингредиентов сложен и в большинстве случаев отсутствует.

Несмотря на значительный прогресс в метрологии испытаний неоднородных материалов, методы контроля обработки проб находятся в относительно неразвитом состоянии. В большинстве лабораторий и цехов пользуются (больше по традиции, чем по необходимости) формулой Чечотта. За последние 20 лет принципиально изменилась вычислительная оснащенность специалистов, в том числе и геологов. Прежде отчетливо ощущалось стремление к упрощению формул с целью сокращения расчетов. Сложным для оценки «вручную» было измерение фундаментальных характеристик анализируемых руд при дроблении. В связи с развитием и широким распространением ЭВМ вычислительные трудности отпали. Расчет необходимой массы проб с применением микрокалькуляторов или номограмм, полученных на ЭВМ, занимает несколько секунд. Учитывая, что такие расчеты проводят 1—2 раза для каждого типа руды, т. е. очень редко, говорить всерьез о «выразительной сложности» некоторых рекомендованных ранее формул, как написал Ф. Штаммбергер, например, о формуле Краснова, не приходится.

Существенно изменилась в настоящее время и формулировка задачи в области методики обработки проб. Если прежде она состояла в вычислении отдельной необходимой массы пробы при заданных условиях, то теперь — в построении оптимальной схемы обработки в целом. В этих условиях говорить

о большей или меньшей сложности формул вообще не имеет смысла, так как решение получают с применением ЭВМ и потребителю нет необходимости ни в каких ручных расчетах. В «худшем» случае он должен уметь снять с номограмм несколько параметров.

Бороздовую, штуфную, валовую или какую-либо иную пробу, полученную при опробовании массивов горных пород, месторождений, товарных рудных масс, угля, концентрата или продуктов технологических процессов, будем называть исходной. Масса исходной пробы изменяется в широких пределах — от нескольких граммов до 100 кг и более. Масса же конечной пробы, поступающей в лабораторию, не превышает 20—100 г, а непосредственно на анализ — и того меньше. Наибольший диаметр кусков в исходной пробе может достигать 100—300 мм, а диаметр частиц в конечной пробе не должен превышать 0.1—0.2 мм. Поэтому материал пробы, прежде чем поступить на анализ, подвергается обработке. С целью уменьшения затрат энергии и времени обработка проб проводится в несколько этапов.

Под этапом обработки проб понимается последовательное выполнение четырех операций: 1) дробление или измельчение материала пробы, 2) грохочение или просеивание, 3) перемешивание, 4) сокращение массы пробы. Первые три операции являются вспомогательными. Они обеспечивают проведение наиболее важной последней операции, уменьшая до контролируемых пределов погрешность изучаемых свойств пробы. Обработка проб производится по схеме, составляемой с учетом особенностей полезного ископаемого, задачи исследования, а также характеристики исходной пробы, в основном — ее массы. На схеме указываются масса проб после каждого сокращения и диаметры отверстий контрольных сит.

Разработка и внедрение новых технических устройств измельчения и сокращения, автоматизация управления процессом обработки проб влекут за собой необходимость решения ряда методических вопросов. Научно обоснованная система обработки проб включает две проблемы. Первая заключается в оценке погрешностей, образующихся при сокращении материала измельченной пробы. Сокращение производится после каждого этапа дробления (измельчения) и приводит к случайным погрешностям различной величины в зависимости от массы сокращаемой пробы и крупности ее частиц (что в итоге определяет число частиц в пробе), контрастности частиц по изучаемому компоненту и степени сокращения на данном этапе. Формулы, определяющие погрешности сокращения, непосредственно выводятся из законов теории вероятностей и в основе своей чрезвычайно просты, если иметь дело с бессточковыми смесями рудных и нерудных частиц правильной кубической или шарообразной формы.

Все операции процесса обработки, особенно дробление и измельчение, требуют значительных затрат энергии и времени. Поэтому вторая проблема обработки проб заключается в выборе числа этапов и в распределении степени измельчения и сокращения по этапам таким образом, чтобы суммарные затраты на обработку были минимальными.

§ 2. Обзор методов и формул контроля обработки проб полезных ископаемых

Почти все исследователи, занимавшиеся проблемой обработки проб, уделяли немалое внимание истории вопроса и анализу результатов предшественников. Краткие обзоры литературы имеются у К. Л. Пожарицкого, П. Л. Каллистова, Н. В. Барышева, Д. А. Краснова, Ф. Штаммбергера и других авторов. Однако исчерпывающего анализа сделано не было. Некоторые недостаточно обоснованные положения продолжают переходить из работы в работу.

Развернувшаяся в 30-х годах дискуссия между П. Л. Каллистовым и К. Л. Пожарицким по поводу экспериментального определения коэффициентов формулы $q = Kd^\alpha$ и так называемой физической интерпретации параметров K и α осталась незавершенной. Не подведен итог попыткам экспериментального определения надежной массы проб при сокращении. По существу еще не было проведено методически корректной проверки соответствия фактических погрешностей сокращения расчетным для различных формул, в том числе и для широко употребляющейся формулы Ричардса—Чечотта. Вместе с тем в достаточной степени не оценены и не использованы положительные элементы основоположника научного подхода к сокращению проб Д. Брунтона.

Все это заставляет нас обратиться к историческому обзору, интересному самому по себе и полезному с точки зрения многочисленных подходов к практическому решению важного и, казалось бы, простого в статистическом смысле вопроса.

Согласно сведениям, перешедшим из ранних работ в последующие, первая попытка установить принципы правильного сокращения проб принадлежит Везину (по некоторым источникам, — Везэну). Нам не удалось найти ни первоисточник этих сведений, ни оригинальную работу Везина, поэтому восстановить первоначальную формулировку принципа трудно. В дошедшем до нас виде она заключается в том, что если отбирать материал для исследования из исходной пробы, то масса этого материала, обеспечивающая надежные результаты, прямо пропорциональна кубу поперечника частиц. Другими словами, при прочих равных условиях погрешность определяется числом частиц в выборке. В этом понимании принцип Везина является правильным при условии, что исходная проба достаточно велика по числу зерен по сравнению с выборкой, т. е. степень сокращения очень велика.

Записанный в виде равенства

$$q_1/q_2 = d_1^3/d_2^3 \quad (4.2.1)$$

этот принцип гласит, что если исходная проба 2, аналогичная пробе 1, в n раз мельче ($d_1/d_2 = n$), то требуемая навеска будет в n^3 раз меньше ($q_1/q_2 = n^3$). В последующем интерпретация принципа изменилась, и формулу (4.2.1) стали понимать как соотношение отбираемых масс на этапах обработки в зависимости от диаметров измельчения на этих этапах. Принцип Везина стали понимать так, что для обеспечения требуемой надежности число частиц некоторой пробы в процессе ее обработки должно быть постоянным.

Принцип Везина в этом понимании записывается в виде формулы

$$q = Kd^3. \quad (4.2.2)$$

Обычно считается, что она дает завышенные (против необходимых) массы проб. Однако это не совсем верно. Если значение коэффициента K определено экспериментально для тонко измельченной пробы, то применение его для проб с большим размером частиц действительно будет давать завышенные массы. Наоборот, экстраполяция коэффициента, найденного при крупном дроблении, на тонкоизмельченные пробы приведет к ненадежной массе. Для правильной реализации принципа параметр K должен быть определен на каждом этапе обработки в соответствии со степенью измельчения d , а это полностью обесценивает его роль.

Н. А. Котляр (1926) провел ряд опытов с искусственными смесями рудных и нерудных частиц и показал, что между необходимой массой пробы, с одной стороны, и диаметром, плотностью и содержанием — с другой, существуют следующие зависимости: а) кубическая, б) прямая линейная, в) прямая линейная, если погрешность исчислять в абсолютных процентах, и обратная, если в относительных.

Фундаментальное исследование Д. Брунтонна развивает принцип Везина. Исследователи как-то упускают из вида, что формула Брунтонна является специализацией формулы Везина в двух отношениях. Во-первых, она раскрывает содержание параметра K , во-вторых, включает в явном виде параметр допустимой погрешности. Чтобы это стало очевиднее, запишем формулу Брунтонна в следующем виде:

$$q = 1/p \cdot (k-c)/cnf\gamma d^3, \quad (4.2.3)$$

где p является предельно допустимой погрешностью в процентах от массы металла в пробе, k — содержание металла в богатейшей частице пробы, c — среднее содержание металла, f — поправка за форму частицы, γ — плотность богатейшей частицы пробы. Таким образом, по Брунтонну, также $q = Kd^3$, где $K = n/p \cdot (k-c)/cf\gamma$. Определение f сводится к изучению массы и числа зерен узкого класса частиц. По первоначальным оценкам Брунтонна, при дроблении до 0.1—0.05 мм золотых и серебряных руд f колеблется от 1 до 6. Как указывают Н. В. Барышев и В. И. Красиков (1936), значение f изменяется от 1 до 6.6, причем при увеличении крупности частиц оно уменьшается. Коэффициентом, вызвавшим наибольшую критику, является n — число богатейших частиц пробы, находящихся в избытке или недостатке по сравнению со средним его значением в исходной пробе.

Главная практическая особенность формулы Брунтонна заключается в г а р а н т и р о в а н и и требуемой надежности. Величина k взята не как средняя по пробе с учетом сростков, а максимальная. Плотность частицы и ее размер также взяты максимальными для пробы. Все это неизбежно приводит к превышению необходимой массы, вызвавшему единодушную, но вряд ли справедливую критику специалистов. Главное же возражение состояло в невозможности или сложности определения величины n , без чего формула была неупотребимой.

Практического применения формула Брунтонна не получила, но послужила отправной точкой для многочисленных исследователей. И все-таки можно согласиться с Ф. Штаммбергером (ГДР), сожалеющим, что работы Брунтонна и его идеи нашли лишь незначительное эхо и дальнейшее развитие. Сейчас мы должны признать, что она содержит все параметры, от которых зависит требуемая масса пробы (кроме одного), а именно: куб диаметра частицы, поправка за ее форму, плотность, коэффициент контрастности и допустимая погрешность. Только исходная масса пробы в ней не присутствует. Другое дело, что конкретный вид указанных параметров в применении к пробе сложного гранулометрического и минерального (точнее «сросткового») состава в действительности оказался более сложным и изменяющимся по мере измельчения пробы. Так, для учета изменения контрастности Н. В. Барышев и В. И. Красиков предполагали ввести коэффициент k , равный единице для смеси полностью раскрытых рудных и нерудных частиц и равный нулю, когда вся руда состоит из частиц-сростков с одинаковым содержанием металла. Для «погашения» действия максимального диаметра они ввели коэффициент ϵ , равный единице для проб с узким классом крупности и меньший единицы — для переизмельченных проб. Другие параметры формулы также подвергались различным усовершенствованиям, зачастую противоречащим друг другу из-за их различной физической интерпретации. Покажем это на примере величины f . Некоторые авторы понимают ее как отношение объема средней реальной частицы к объему куба со стороной d или шара с диаметром d . Другие связывают ее с гранулометрическим составом измельченного материала и характером его прохождения через сито: как отношение объема частицы к объему шара с диаметром, равным диаметру отверстий сита. В первом понимании к величине f требуется еще поправка на гранулометрический состав материала, во втором эта поправка как бы уже учтена. В первом случае для игольчатых частиц f будет очень мало,

во втором — очень велико. Так, по данным К. Л. Пожарицкого (1935), для золотых руд при крупном дроблении (до 1.65 мм) $f=0.03$. Сравнить эти значения f с данными Д. Брунтон или Н. В. Барышева и В. И. Красикова трудно по указанной причине.

Единственное заключение Д. Брунтон оказалось безусловно ошибочным, что материал должен быть тем мельче раздроблен, чем меньше сокращаемая проба. Как показали современные исследования, вывод должен быть противоположным. Кроме того, ни Д. Брунтон, ни Везин не учитывали накопления погрешностей по мере сокращения пробы при многоэтапной ее обработке.

В 1904 г. Р. Ричардс в книге по обогащению руд приводит таблицу данных пяти инженеров по надежной обработке проб. Эти данные, по мнению Р. Ричардса, показывают, что в рассмотренных девяти случаях изменение необходимой массы пробы примерно пропорционально квадрату изменения диаметра частиц. В заключение Р. Ричардс говорит, что на основании данных этой таблицы можно получить массы, «которые, вероятно, встретят одобрение практиков». Они действительно встретили одобрение, и проф. Г. О. Чечотт (1932) выразил предложенную Р. Ричардсом зависимость в виде формулы

$$q=Kd^3, \quad (4.2.4)$$

где q — надежная масса пробы (в кг), d — диаметр отверстий контрольного сита (в мм). Им же впервые была составлена таблица для выбора коэффициента K в зависимости от содержания металла в руде и характера вкрапленности. В ней руды подразделены на шесть категорий, и для каждой рекомендованы предельные и средние значения K . С тех пор формула Ричардса—Чечотта получила универсальное применение среди геологов-практиков и более полувека оказывает решающее воздействие на направление дальнейших исследований.

Следует отметить, что формула для многих случаев давала надежные и практически приемлемые массы, а если этого не случалось, ошибку относили за счет неверного определения параметра K . Действительно, его правильный выбор представляет сложную задачу. Предел колебаний K (от 0.16 до 24) достаточно широк, а деление руд на категории условно. Шагом назад по сравнению с формулой Брунтон было отсутствие предельно допустимой погрешности.

Формулу Ричардса—Чечотта на практике применяют для определения надежной массы на каждом этапе сокращения, полагая, что обеспечение надежности на каждом этапе обработки автоматически обеспечивает ее для процесса в целом. На деле же погрешности этапов складываются по законам статистики, что ни в самой формуле, ни в рекомендациях по ее применению не отмечается.

В формуле Ричардса—Чечотта не учитывается зависимость допускаемой погрешности от сокращаемой массы, а эта зависимость существует, ибо если степень сокращения невелика, при прочих равных условиях будет меньше и погрешность сокращения. Таким образом, нельзя считать вполне обоснованным то значительное место, которое занимала и занимает эта формула в практике обработки проб.

Американские инженеры К. Демонд и Р. Хальфердаль (Demond, Halferdahl, 1922) обобщили принцип Ричардса—Чечотта, сделав его более гибким за счет изменения показателя степени диаметра частиц:

$$q=Kd^\alpha. \quad (4.2.5)$$

В формуле стало два параметра, которые требовалось определить для каждого вида руды. Не претендуя на теоретическое решение, авторы пошли путем эксперимента, решая систему из двух уравнений с двумя неизвестными K и α :

$$\begin{cases} q_1=Kd_1^\alpha \\ q_2=Kd_2^\alpha \end{cases}$$

Значения q_1 и q_2 являются надежными массами при заданных диаметрах d_1 и d_2 , определяемыми экспериментально, т. е. обработкой многих проб и изучением расхождений. Поскольку такие эксперименты не просты и дороги, К. Демонд и Р. Хальфердаль сделали допущение, что масса в 1 г всегда (независимо от свойств руды) надежна для пробы, измельченной до 0.1 мм, т. е. $q_2=0.001$ кг, $d_2=0.1$ мм (в раздробленной до 0.1 мм руде содержится не менее $2 \cdot 10^8$ частиц, и содержание может быть определено с точностью 0.03 %). Будем называть эту пару величин стандартом Демонда—Хальфердаля. Подставив стандарт Демонда—Хальфердаля во второе уравнение, надо провести эксперименты по определению надежного веса q_1 для какого-либо другого измельчения d_1 . Конкретно, предлагалось пробу весом 2 т дробить до размера $d_1=2$ дюйма, делить (сокращать) ее на 16 частей, причем исходную массу пробы выбирать так, чтобы после сокращения она стала равной q_1 . Все шестнадцать навесок истирались до 0.1 мм, сокращались до 1 г и анализировались. Если расхождения не превышали ошибки химического анализа, масса считалась надежной и подставлялась в указанную систему уравнений для определения K и α . Практически значения α , по Демонду и Хальфердалю, колеблются в пределах от 1.5 до 2.7.

В течение 30 лет советская наука об опробовании шла по этому направлению. Основным продолжателем идей Демонда—Хальфердаля был К. Д. Пожарицкий (1937, и др.). Логарифмируя основную формулу, он приводит ее к следующему виду: $\lg q = \lg K + \alpha \lg d$. Соответствующая прямая на двойном логарифмическом бланке-графике проходит через точку с координатами $q_2=0.001$, $d_2=0.1$ (стандарт Демонда—Хальфердаля). Другая точка определялась экспериментально, как у К. Демонда и Р. Хальфердаля, однако со следующим усовершенствованием. Каждая из восьми параллельных проб не истиралась сразу до 0.1 мм, а подвергалась нескольким последовательным этапам дробления и сокращения, причем на каждом этапе сокращение производилось в соответствии с задаваемыми априори, а затем проверяемыми значениями K и α , которые определяют вторую (пробную) точку на графике. Такой подход К. Л. Пожарицкий считал более приближенным к практике. В этом отношении он был прав, но возникает вопрос: что определяет формула по Пожарицкому — надежную массу для заданного акта сокращения или для всей схемы обработки, состоящей из переменного числа этапов, ошибки которых, как известно, складываются по закону аддитивности дисперсий?

К. Л. Пожарицкий на протяжении длительного периода дискуссии с П. Л. Каллистовым настаивает на полезности этого усовершенствования именно в связи с накапливанием погрешностей сокращения. Погрешности, накапливаемые в процессе обычной ступенчатой обработки, как указывает он, выпадают из результатов опыта при методе Каллистова.

С опорной 20-летней критикой описанного способа определения K и α выступал П. Л. Каллистов. Во-первых, он отрицал стандарт Демонда—Хальфердаля, считая его выхолащиванием формулы, и предполагал путем экспериментов определять не одну, а две точки на графике и, соединив их, находить K и α . П. Л. Каллистов замечает, что если прямые Пожарицкого продолжить левее опорной точки (стандарта Демонда—Хальфердаля), то прямые для руд разной контрастности поменяются местами, вследствие чего получается, что при дроблении менее 0.1 мм наиболее неравномерные руды становятся более равномерными. Это остроумный и справедливый довод, который, однако, нельзя принять всерьез, так как в схеме Демонда—Хальфердаля не предполагается пробу массой в 1 г и размером зерен в 0.1 мм измельчать и сокращать далее. Современные методы анализа, требующие микронавесок и более тонкого измельчения, привели бы к другому стандарту, скажем, 0.05 г и 0.01 мм. Во-вторых, П. Л. Каллистов предлагал однозначную физическую интерпретацию параметров K и α . По его мнению, K выражает число частиц в пробе, необходимых

для достижения желаемой точности. Но это число частиц зависит в свою очередь от среднего содержания изучаемого компонента, крупности его зерен, их плотности, неравномерности распределения компонента в пробе, а также от содержания его в несущем минерале. Можно легко доказать, как считает П. Л. Каллистов, не приводя, однако, доказательств, что K остается одним и тем же при всяком измельчении, т. е. является постоянной величиной.

Показатель α , по Каллистову, должен корректировать зависимость массы пробы от среднего диаметра частиц, «погашая» действие максимального диаметра. Он определяется механическими свойствами материала. Несмотря на очевидную полезность содержательной интерпретации параметров эмпирических формул, она редко, как и в данном случае, удается.

Попытку подвести итог дискуссии предпринял крупнейший специалист ГДР по опробованию и подсчету запасов Ф. Штаммбергер (1965). Полагая, что К. Демонд и Р. Хальфердаль внесли «скромный» результат в проблему, он удивляется тому большому месту, которое отводят их формуле советские специалисты. Ф. Штаммбергер считает интерпретацию параметров K и α , данную П. Л. Каллистовым, гораздо более важным вкладом в проблему, чем саму формулу.

На основании современных данных в отношении формулы Демонда—Хальфердаля можно сделать следующие выводы: 1) физическая интерпретация эмпирических коэффициентов K и α не удалась и в дальнейшем представляется бесперспективной; 2) аппроксимация модели сокращения экспериментальным подбором K и α оказалась недостаточно гибкой для практики в широком диапазоне измельчений; 3) формула Демонда—Хальфердаля по первоначальному замыслу должна давать приемлемую на практике верхнюю границу необходимой массы. В этом качестве она не потеряла своего значения и сейчас.

Новая волна интереса к проблеме сокращения проб поднялась в 1955—1965 гг. благодаря исследованиям французских (П. Жи) и советских (Д. А. Краснов, Ю. А. Ткачев) специалистов.

В обобщенном виде формула Жи имеет следующий вид:

$$\sigma^2 = cd^3(1/q - 1/Q), \quad (4.2.6)$$

где Q — опробуемая масса; q — масса пробы; c — параметр опробования; d — размер ячейки сита, задерживающего от 5 до 10 % просеиваемого материала. Многие параметры этой формулы порознь встречались в более ранних формулах, но вместе в правильной взаимной связи они появились впервые. К ним относятся заданная погрешность сокращения, масса сокращаемой пробы и группа параметров, характеризующих контрастность материала $c: c = fqlm$, где f — коэффициент формы зерен, q — гранулометрический показатель, l — степень раскрытия минералов, m — коэффициент минералогического состава. П. Жи не придавал большого значения массе сокращаемой пробы Q , полагая, что практически всегда $q \ll Q$, и анализировал в основном формулу

$$\sigma^2 = fqlmd^3/q. \quad (4.2.7)$$

Одновременно и независимо от П. Жи проблему сокращения проб подробно исследовал Д. А. Краснов (1956—1969 гг.) с экспериментальной и теоретической стороны. Опыты заключались в сокращении искусственной смеси зерен магнетита и пегматитовой породы желобчатым делителем на 32 или 16 параллельных навесок и в изучении отклонений числа зерен от среднего. Опыты показали, что относительная погрешность уменьшается пропорционально увеличению корня квадратного из числа зерен.

Деление смеси двух минералов трех классов крупности показало независимость поведения каждого класса, т. е. сокращение происходит как бы само-

стоятельно для каждой фракции пробы. Д. А. Краснов приходит к следующей формуле:

$$q = 10d_{\text{ср}}^3 t^2 / m_a^2 (\alpha - 1), \quad (4.2.8)$$

где q — необходимая масса пробы (в кг), $d_{\text{ср}}$ — средний диаметр частиц (в см), m_a — относительная допустимая погрешность (в %), t — коэффициент вероятности, α — весовая доля рудных частиц.

Для определения весового содержания рудных частиц α Д. А. Краснов делает следующие правдоподобные допущения: 1) весь рудный минерал распределен равномерно по классам крупности, в частности содержание его в классе $-D + d_{\text{вкр}}$, где $d_{\text{вкр}}$ — преобладающий размер вкрапленности, такое же, как в классе $-d_{\text{вкр}} + 0$; 2) в классе $-D + d_{\text{вкр}}$ весь минерал находится в сростках, а в классе $-d_{\text{вкр}} + 0$ — в свободном состоянии; 3) весовое содержание полезного компонента в сростках β принимается равным содержанию его в наиболее богатых кусках-сростках максимальной крупности D .

Весовую долю сростков α_c автор определяет как средневзвешенное по двум классам: $-D + d_{\text{вкр}}$ с выходом ω и $-d_{\text{вкр}} + 0$ с выходом $1 - \omega$:

$$\alpha_c = \frac{\alpha}{\beta} \omega + \alpha(1 - \omega), \quad \text{где } \omega = 1 - d_{\text{вкр}}^c / D^c = 1 - 1/r^c. \quad \text{Подставляя это в (4.2.8),}$$

$$\text{Д. А. Краснов получает} \quad \left(\frac{1}{\alpha_c} - 1 \right) = \left(\frac{r^c \beta}{(r^c - (1 - \beta))^\alpha} - 1 \right).$$

Окончательно формула Краснова представляется в следующем виде:

$$q = \left(\frac{c}{c+1} \right)^3 D^3 \frac{t^2}{m^2} \left(\frac{r^c}{r^c - 1 - \beta} \frac{\beta}{\alpha} - 1 \right). \quad (4.2.9)$$

Положительная сторона формулы заключается в возможности корректировки максимального диаметра D с помощью параметра c гранулометрической кривой. Весьма ценна идея выразить поправочный множитель к коэффициенту контрастности $r^c / [r^c - (1 - \beta)]$ через отношение максимального диаметра частиц к диаметру рудной вкрапленности $r = D/d_{\text{вкр}}$ с учетом гранулометрического параметра.

Однако вывод Д. А. Красновым коэффициента контрастности, в котором нет отношения плотностей рудных и нерудных зерен, неубедителен. Д. А. Краснов указывает, что при $\beta = \alpha$ (содержание в сростках и рудных частицах равно среднему по пробе) необходимая масса пробы должна быть равна нулю. Замечание верное, но из основной формулы Краснова оно не вытекает, что еще раз указывает на сомнительность рассмотренного выражения для коэффициента контрастности.

Для упрощения пользования формулой Д. А. Краснов подставляет вместо ряда параметров их наиболее употребительные значения и получает $q = 0.025D^3(\beta - \alpha)$ кг, где D — диаметр максимального зерна — отверстия сита (в мм), β — содержание полезного компонента в рудной частице, α — среднее содержание его в пробе. В таком варианте формула совпадает с формулой Везина $q = Kd^3$, где $K = 0.025(\beta - \alpha)$.

§ 3. Вывод формулы сокращения смеси «рудных» и «нерудных» зерен правильной формы

Совершаемую при сокращении случайную погрешность целесообразно измерять средним квадратическим отклонением или его квадратом — дисперсией. При этом будем различать четыре ее варианта: дисперсию числа частиц полезного компонента в сокращенной пробе σ_x^2 ; дисперсию массы металла (размерность — квадрат весовых единиц: г², кг² и т. д.):

$$\sigma_w^2 = \rho^2 \sigma_x^2, \quad (4.3.1)$$

дисперсию содержания металла:

$$\sigma_c^2 = \frac{1}{q^2} \sigma_w^2 = \frac{\rho^2}{q^2} \sigma_x^2 \quad (4.3.2)$$

(выражается в квадратах весовых долей или весовых процентов); дисперсию относительной ошибки в содержании металла:

$$\sigma_{отн}^2 = \frac{1}{c^2} \sigma_c^2 = \frac{\rho^2}{q^2 c^2} \sigma_x^2 \quad (4.3.3)$$

(выражается в квадратах относительных процентов или относительных долей). Обозначения к формулам (4.3.1) — (4.3.3) следующие: x — число рудных частиц в сокращенной пробе; ρ — масса рудных частиц, т. е. металла в пробе; q — масса сокращенной пробы; c — среднее содержание металла в пробе.

Основной математической моделью сокращения смеси рудных и нерудных минералов является выборка без возвращения из конечной генеральной совокупности. Распределение числа рудных частиц в такой выборке подчиняется гипергеометрическому закону, который хорошо аппроксимируется нормальным законом с дисперсией $\sigma_{np}^2 = p(1-p)n[(N-n)/(N-1)]$. Так как $N \gg 1$, то приближенно

$$\sigma_{np}^2 = p(1-p)n \left(1 - \frac{n}{N}\right), \quad (4.3.4)$$

где n и N — число частиц соответственно в выборке и в генеральной совокупности; p — доля рудных частиц в ней. Переходя к относительной дисперсии, получим $\sigma^2 = \sigma_{отн}^2 = 1/(np)^2 \sigma_{np}^2 = (1/p-1)(1/n-1/N)$. К этому же результату мы придем, если используем равенство (4.3.3) и заменим в нем среднее содержание \bar{c} отношением массы металла в пробе к массе пробы: $\bar{c} = n\rho\rho/q$, где ρ — масса рудной частицы. Тогда

$$\sigma^2 = \frac{\rho^2}{q^2 c^2} \sigma_{np}^2 = \frac{\rho^2}{q^2 (n\rho\rho/q)^2} \sigma_{np}^2 = \frac{1}{(np)^2} \sigma_{np}^2 = \left(\frac{1}{p}-1\right) \left(\frac{1}{n}-\frac{1}{N}\right). \quad (4.3.5)$$

Величину сокращаемой и сокращенной пробы целесообразно выразить в весовых единицах, поэтому, полагая для кубической модели $N = Q/d^3\gamma$ и $n = q/d^3\gamma$, получим

$$\sigma^2 = (1/p-1) d^3\gamma (1/q-1/Q) \quad (4.3.6)$$

(обозначения см. после формулы (4.3.7)).

Следует отметить одно допущение. Несмотря на то что дисперсия выражена как функция массы пробы после сокращения, контроль величины выборки в рассматриваемой модели сокращения осуществляется по числу частиц в ней. Теоретически выборка должна состоять из n частиц, и, отбирая от

пробы массу q , только приближенно можно утверждать, что там окажется n частиц: $M(n) = q/\rho$, так как колебания в доле рудных частиц приведут одновременно к колебаниям в средней плотности материала выборки γ . Однако этими колебаниями без заметных погрешностей можно пренебречь. Кроме того, при расчете дисперсии по сле анализа навески становится известным фактическое значение γ , которое используется в формулах для апостериорного расчета погрешности сокращения.

Заменим численную долю рудных частиц p их весовой долей: $p_w = p\gamma_p/\gamma$, откуда $p = p_w\gamma/\gamma_p$. Весовую долю рудных частиц заменим весовой долей металла c (т. е. содержанием металла), что практически более удобно: $c = p_w k = p k \gamma_p/\gamma$, $p = c\gamma/k/\gamma_p$. После проведенных замен получим основную формулу погрешности сокращения:

$$\sigma^2 = (k\gamma_p/c/\gamma - 1) \gamma d^3 (1/q - 1/Q), \quad (4.3.7)$$

где σ^2 — относительная ошибка весового содержания металла при сокращении доли единицы; Q и q — масса сокращаемой и сокращенной пробы (в г); γ_p — плотность рудных зерен (в г/см³); γ — средняя плотность зерен в пробе (в г/см³); k — содержание металла в рудной частице (в долях единицы); c — содержание металла в пробе (то же); d — диаметр частиц (в см). Обозначая $C = (k\gamma_p/c/\gamma - 1)$ и решая равенство (4.3.7) относительно q , получим

$$q = c\gamma d^3 Q / (\sigma^2 Q + C\gamma d^3), \quad (4.3.8)$$

откуда видно, что масса сокращенной пробы при заданной ошибке сокращения зависит от массы пробы до сокращения:

$$\lim_{Q \rightarrow 0} q = 0, \quad \lim_{Q \rightarrow \infty} q = \frac{C\gamma d^3}{\sigma^2}.$$

С уменьшением массы сокращаемой пробы при прочих равных условиях уменьшается и необходимая масса пробы после сокращения. С увеличением массы сокращаемой пробы необходимая масса пробы после сокращения стремится к пределу $q = C\gamma d^3/\sigma^2$.

Зависимость необходимой массы пробы от допустимой погрешности является обратно пропорциональной, а от диаметра частиц — кубической. Требуемая масса пробы прямо пропорциональна также средней плотности.

Особого рассмотрения требует группа величин, выражающих характер распределения полезного компонента в частицах пробы, т. е. коэффициент контрастности руды

$$C = k\gamma_p/c/\gamma - 1. \quad (4.3.9)$$

Из этого выражения видно, что коэффициент контрастности — это приближенное отношение объемной концентрации «металла» в рудной частице к средней концентрации его в пробе. Минимальное значение коэффициента контрастности равно нулю. Проба при этом является идеально гомогенной, и при сокращении можно обойтись какой угодно малой навеской, лишь бы она была достаточна для анализа. Наибольшее значение C практически не ограничено. При содержании золота в кварц-золотоносной пробе, равном 10 г/т, контрастность руды, раздробленной до стадии полного высвобождения золота, будет составлять $2 \cdot 10^6$. Зависимость необходимой массы пробы от коэффициента контрастности при больших его значениях является прямо пропорциональной, поэтому определение его значения в исходной пробе и установление закона изменения в процессе дробления пробы является одной из наиболее важных задач контроля обработки.

Реальная пропущенная через сито с диаметром отверстий d_c проба состоит из смеси частиц $d \leq d_c$. Формулу сокращения такой пробы можно получить,

рассматривая по отдельности сокращение отдельных фракций и определенным образом объединяя результат. Для этого нами получено выражение для погрешности в объединенной пробе, полученной из параллельных навесок:

$$\sigma^2 = \sum p_i^2 c_i^{*2} \sigma_i^2 = \sum \omega_i^2 \sigma_i^2, \quad (4.3.10)$$

где ω_i — весовая доля металла в i -м классе; c_i^* — коэффициент относительной концентрации полезного компонента в нем; σ_i — погрешность сокращения класса. Из (4.3.10) непосредственно следует, что $d_{\text{эКВ}}^3 = d_1^3 p_1 + d_2^3 p_2 + \dots + d_n^3 p_n$, где p_i — весовая доля класса i .

Применяя аналогичный прием по отношению к плотности γ и коэффициенту контрастности C , получим

$$\gamma_{\text{эКВ}} = \sum \gamma_i \omega_i, \quad C_{\text{эКВ}} = \sum C_i \omega_i.$$

Однако формулы эквивалентных параметров справедливы лишь каждая в отдельности, т. е. для случаев, когда по фракциям пробы изменяется только один ископаемый параметр, а остальные постоянны.

В противном случае усреднению подлежит произведение всех трех параметров: C , γ и d^3 , т. е. $(C\gamma d^3)_{\text{эКВ}} = \sum C_i \gamma_i d_i^3 \omega_i$, где C_i , γ_i , d_i — средние значения в пределах узких классов.

§ 4. Оценка погрешности сокращения при многоэтапной обработке проб

Погрешность содержания на каждом этапе обработки пробы зависит от погрешностей, которые вносятся сокращениями на всех предыдущих этапах. Поэтому необходимо установить связь между ними и выяснить влияние каждого этапа на накопление погрешности.

Пусть в начальной пробе имеется частиц N_0 , из них доля рудных равна p_0 . В сокращенной пробе общее число частиц равно N_1 . Тогда дисперсия доли рудных частиц p_1 в сокращенной пробе равна: $D [p_1] = p_0(1-p_0) (1/N_1 - 1/N_0)$. Материал сокращенной пробы подвергается дроблению, при этом число частиц в пробе возрастает и становится равным $N'_1 = a_1 N_1$, где коэффициент a_1 больше единицы. Затем проба сокращается до общего числа частиц N_2 , при этом значение p_1 влияет на долю рудных частиц p_2 в пробе после второго сокращения.

Математическое ожидание и дисперсия двух случайных величин p_1 и p_2 связаны между собой следующими соотношениями:

$$M [p_2] = M [M [p_2 | p_1]], \\ D [p_2] = D [M [p_2 | p_1]] + M [D [p_2 | p_1]],$$

где $M [p_2 | p_1]$ и $D [p_2 | p_1]$ — условные математическое ожидание и дисперсия соответственно. Очевидно, что $M [p_2 | p_1] = p_1$, а $M [p_2] = M [p_1] = p_0$. Применяя это, определим значение первого слагаемого в формуле для дисперсии:

$$D [M [p_2 | p_1]] = D [p_1] = p_0(1-p_0) (1/N_1 - 1/N_0).$$

Значение условной дисперсии $D [p_2 | p_1]$ находим, полагая, что p_1 фиксированная величина, поэтому $D [p_2 | p_1] = p_1(1-p_1) (1/N_2 - 1/N'_1)$. Отсюда второе слагаемое равно:

$$M [D [p_2 | p_1]] = (1/N_2 - 1/N'_1) \cdot M [p_1 - p_1^2] = \\ = (1/N_2 - 1/N'_1) (M [p_1] - M^2 [p_1] - D [p_1]).$$

Окончательно получим выражение для дисперсии случайной величины p_2 :

$$D [p_2] = p_0 (1 - p_0) \left[\left(\frac{1}{N_1} - \frac{1}{N_0} \right) + \left(\frac{1}{N_2} - \frac{1}{N'_1} \right) - \left(\frac{1}{N_2} - \frac{1}{N'_1} \right) \left(\frac{1}{N_1} - \frac{1}{N_0} \right) \right]$$

или, отбрасывая последнее слагаемое:

$$D [p_2] \leq p_0 (1 - p_0) \left[\left(\frac{1}{N_1} - \frac{1}{N_0} \right) + \left(\frac{1}{N_2} - \frac{1}{N'_1} \right) \right].$$

Величина пренебрегаемого слагаемого будет незначительна даже при небольшом числе частиц в пробе. Рассмотрим простой пример. Пусть $N_0 = 20$, $p_0 = 0.3$. Пробу сократили в четыре раза, т. е. $N_1 = 5$. Затем ее раздробили таким образом, что число частиц стало равным $N'_1 = 40$, и сократили еще в четыре раза, т. е. до $N_2 = 10$. Если воспользоваться точной формулой для дисперсии, то ее значение равно $D [p_2] = 0.045$, а приближенное значение $D' [p_2] = 0.047$, т. е. относительная ошибка составляет 4.4 %. Для значительного числа частиц в пробе, когда $N = 10^6 - 10^{10}$, как это и бывает в действительности, расхождение будет еще меньше, поэтому дисперсию можно принимать равной значению оценки по приближенной формуле.

Используя аналогичные выкладки, можно получить формулу дисперсии рудных частиц для обработки пробы за n этапов:

$$D [p_n] = p_0 (1 - p_0) [(1/N_1 - 1/N_0) + \dots + (1/N_n - 1/N'_{n-1})],$$

т. е. с достаточным приближением можно считать, что общая дисперсия обработки есть сумма дисперсий на отдельных этапах. Если рассматривать одноэтапный процесс сокращения пробы как развернутый многоэтапный процесс обработки без дробления, т. е. когда значения размеров частиц d_k не меняются и равны d , то получим равенство

$$\sigma^2 = C\gamma d^3 / Q_0 (m_1 - 1 + m_2 - m_1 + \dots + m - m_{n-1}) = C\gamma d^3 / Q_0 (m - 1).$$

Это выражение совпадает со значением дисперсии при единичном сокращении пробы от массы Q_0 до массы $Q_1 = Q_0/m$.

§ 5. Изменение контрастности в процессе обработки пробы

Для целей оптимизации процесса обработки проб коэффициент контрастности необходимо выразить как функцию диаметра измельчения, используя в качестве параметров этой функции такие известные по минералогическому описанию руд характеристики, как минеральный состав, структура и текстура руд, определяющие характер вкрапленности полезного компонента. Зная механические свойства рудных и вмещающих минералов и пород, можно с достаточной точностью предсказать изменение коэффициента контрастности в процессе дробления.

При дроблении пробы происходит раскрытие минеральных сростков, так что с уменьшением их размеров доля частиц, состоящих из чистых минеральных видов, увеличивается. Когда материал пробы измельчен до такой степени, что ни одна частица в пробе не является сростком рудного и нерудного ингредиента, коэффициент контрастности достигает максимальной величины и далее не изменяется. Это его значение назовем критическим (C^*). Расчет критического значения представляет определенную трудность, особенно если руда содержит несколько видов рудных минералов. Далее мы попытаемся выразить изменение коэффициента контрастности в процессе измельчения через его критическое значение и отношение размера частиц пробы к размеру рудной вкрапленности.

Коэффициент контрастности пробы C представляет собой относительную дисперсию количества металла в отдельных частицах пробы по отношению к среднему его количеству в частице:

$$C = \frac{1}{\bar{w}} M (\omega - \bar{\omega})^2 = \sum_{i=1}^n \left(\frac{\omega_i}{\bar{w}} - 1 \right)^2, \quad (4.5.1)$$

где ω_i — масса металла в i -й частице; \bar{w} — средняя масса металла в частице; n — число частиц в пробе. Средняя масса металла равна общей массе металла, деленной на число частиц, или среднему содержанию, умноженному на среднюю массу частицы: $\bar{w} = cd^3\gamma$.

В частном случае, когда проба состоит из рудных и нерудных частиц с частотами p_p и $p_n = 1 - p_p$ соответственно.

$$c = (\omega_p/\bar{w} - 1)^2 p_p + (0/\bar{w} - 1)^2 \cdot (1 - p_p) = (\omega_p/\bar{w} - 1)^2 + 1 - p_p.$$

Выразим p_p через \bar{w} и ω_p : $\omega = \omega_p \cdot p + 0(1 - p_p)$, откуда $p_p = \bar{w}/\omega_p$. Подставляя это выражение в (4.5.1), получим

$$C = (\omega_p/\bar{w} - 1)^2 \bar{w}/\omega_p + 1 - \bar{w}/\omega_p = \omega_p/\bar{w} - 1,$$

или, после выражения массы металла через содержание,

$$C = k\gamma_p/c\gamma - 1, \quad (4.5.2)$$

т. е. для пробы, состоящей из смеси рудного и нерудного минерала, коэффициент контрастности принимает уже знакомое нам (см. формулу (4.3.9)) выражение.

Аналогичные выкладки для случая, когда руда состоит из двух типов рудных зерен, дают для коэффициента контрастности выражение

$$C = R_1^2 R_2 / (R_2 - R_1) + R_2^2 R_1 / (R_1 - R_2), \quad (4.5.3)$$

где R — парциальный контраст частиц i -го минерала: $R_i = (\omega_i/\bar{w} - 1) = k_i\gamma_i/c\gamma - 1$. Для пробы с любым числом m минералов-носителей нами получено, что

$$C = \sum_{i=1}^m R_i^2 p_i, \quad (4.5.4)$$

где p_i — весовая доля i -го минерала-носителя, т. е. коэффициент контрастности пробы равен взвешенной сумме квадратов парциальных коэффициентов контрастностей всех минералов, слагающих пробу.

Очевидно, что контрастность по мере измельчения пробы увеличивается, и это явление в значительной степени определяет требуемую массу надежной пробы. Однако попыток описать количественно изменение контрастности в процессе измельчения было не много. П. Жи предложил следующий эмпирический способ. Максимально возможное для данной руды его значение, которое достигается при полном раскрытии минералов, он умножал на коэффициент раскрытия l , меняющийся от 0 до 1:

$$l = (c_6 - \bar{c}) / (k - \bar{c}), \quad (4.5.5)$$

где k — содержание металла в минерале-носителе, \bar{c} — среднее его содержание в пробе. Содержание металла в богатейшей частице c_6 определяется эмпирически с помощью анализа химического состава отдельных частиц или их микроскопического изучения.

Д. А. Краснов (1969) не выделяет в своей формуле специального параметра, определяющего коэффициент контрастности пробы. Выражением, ближе всего описывающим этот параметр, является $(1/\alpha_c - 1)$, или

$$C = c_6/\bar{c} - 1. \quad (4.5.6)$$

Влияние гранулометрической характеристики Д. А. Краснов учитывает с по-

мощью параметра h гранулометрической кривой $y = Ax^h$, где y — весовой выход класса мельче x . При этом

$$C = r^h \beta / (r^h - 1 + \beta) \cdot 1/\alpha - 1,$$

где $r = d/d_{\text{вкр}}$. Второй сомножитель сопоставляется с критическим значением коэффициента контрастности и является постоянной величиной. Первый сомножитель — изменяющаяся, как и у П. Жи, от 0 до 1 при дроблении часть. Действительно, при полном высвобождении минералов $\beta = 1$ и $r^n \beta / (r^n - 1 + \beta) = 1$. Коэффициент контрастности при этом достигает критического значения. К сожалению, для гомогенной пробы ($\beta = \alpha$) нулевого значения C , как должно быть, не получается.

Указанное не позволяет воспользоваться коэффициентом контрастности Д. А. Краснова, но его идея использовать в выражении для C отношение размера частиц к размеру вкрапленности и показатель гранулометрической кривой весьма интересна.

Мы рассмотрели изменение коэффициента контрастности в предположении, что: а) дробление частиц происходит независимо от расположения поверхностей минеральных сростаний, но так, что рудные вкрапления в частицах дроблению не подвергаются; б) частица пробы содержит не более одного рудного вкрапления (Ткачев, Шейн, 1987). Получили следующее выражение для C :

$$C = k\gamma_p/c/\gamma (d_{\text{вкр}}/d)^3 = k\gamma_p/(c\gamma r^3) - 1. \quad (4.5.7)$$

Этот результат любопытен тем, что в довольно широком диапазоне размеров измельчения он предсказывает парадоксальное явление — независимость требуемой массы от диаметра частиц. Действительно, коэффициент контрастности по (4.5.7) изменяется практически обратно пропорционально кубу диаметра. Требуемая масса, будучи прямо пропорциональной кубу диаметра, при этом не изменяется! Это становится понятным, если учесть, что число рудных частиц при дроблении в некотором диапазоне $d_0 > d \geq d_{\text{вкр}}$ остается постоянным. Бедные золотые руды с мелкой вкрапленностью золота хорошо иллюстрируют это явление.

Заменив предположение (б) тем, что распределение числа рудных вкраплений в частицах пробы удовлетворяет закону Пуассона, мы получили достаточно громоздкое выражение

$$C = \sum_{x=0}^m \left\{ \left(\frac{x\gamma_p k}{r^3 \gamma c} - 1 \right)^2 \left(\frac{r^3 c \gamma}{k \gamma_p} \right) \exp \left(- \frac{r^3 c \gamma}{k \gamma_p} \right) \right\} \frac{1}{x!}, \quad (4.5.8)$$

где m — общее число рудных вкраплений. Нанеся на график в логарифмическом масштабе зависимость C от d , обнаруживаем удивительный результат: все расчетные значения C в точности лежат на прямых с угловым коэффициентом $-1/3$. Это указывает на обратную пропорциональность коэффициента контрастности от куба диаметра измельчения. Таким образом, результат, полученный в предположении, что в частице пробы может иметься не более одного рудного включения, совпадает с результатом, полученным без этого ограничивающего предположения. Общим в рассуждениях, приводящих к этому результату, остается предположение, что рудные вкрапления не подвергаются разделению на части. Число рудных вкраплений остается постоянным, что и приводит к независимости требуемой массы пробы от диаметра измельчения. Такие руды мы не рекомендуем сокращать до тех пор, пока не начнут интенсивно разрабатываться рудные вкрапления. Другими словами, условием, определяющим возможность сокращения, является не вообще уменьшение размера частиц и увеличение их общего числа, а увеличение числа рудных частиц.

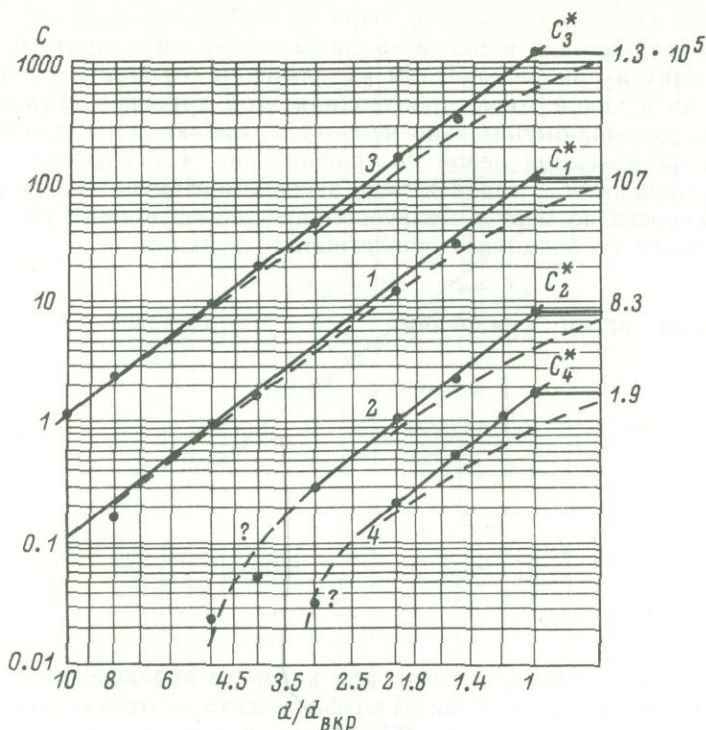


Рис. 4.5.1. Изменение нижней границы коэффициента контрастности руды (C) от относительного размера кусков ($d/d_{\text{вкр}}$).

Фундаментальная точка C^* характеризует предельно раскрытую руду. Сплошные линии — для случая, когда рудные включения не дробятся, штриховые — когда дробление происходит независимо от расположения включений. 1—4 — различные руды по номерам примеров, охарактеризованных в тексте. Для примера № 3 шкалу C умножить на 10^2 .

Рассмотрим более общую модель дробления, при которой расколы кусков происходят по поверхностям, не зависящим от расположения рудных включений. Полная модель такого процесса сложна для анализа, поэтому воспользуемся моделью со следующими ограничениями: 1) все рудные включения перед дроблением имеют поперечник $d_{\text{вкр}}$, все частицы материала — d ; 2) частицы материала либо не имеют включений, либо имеют по одному включению; 3) включение расположено в пределах частицы случайным образом, причем все ее положения равновероятны; 4) частицы материала имеют форму куба со стороной d , а рудные включения — форму куба со стороной $d_{\text{вкр}}$. В результате для коэффициента контрастности получается очень громоздкое выражение (Ткачев, Шейн, 1987).

Значения C нанесены на графики. Анализ графиков (рис. 4.5.1) показывает, что при больших $d/d_{\text{вкр}}$ они практически совпадают с графиками для моделей измельчения без разрушения рудных включений. Действительно, при этом вероятность, что включения не разрушаются, близка к единице. При малых $d/d_{\text{вкр}}$, когда начинает интенсивно уменьшаться размер включений, дробление как средство гомогенизации пробы достигает своей основной цели и требуемая масса уменьшается с уменьшением диаметра частиц. Реальные горные породы и руды раскалываются преимущественно по границам минералов, поэтому анализируемые графики (рис. 4.5.1) указывают на нижнюю границу коэффициента контрастности. На практике более важной задачей является оценка верхней границы значений C , применение которой с гарантией

обеспечило бы заданную точность. При этом мы исходим из следующих предположений: а) кривая верхней границы должна проходить через фундаментальную точку и при дальнейшем измельчении совпадать с горизонталью $C = C_{\max}$; б) ни в какой точке производная этой кривой в логарифмических координатах не должна быть меньше чем -3 ; в) кривая должна быть выпуклой вверх; г) исходной (самой левой на графике рис. 4.5.1) точкой для кривой должно быть значение C_0 , определенное экспериментально; д) скорость изменения контрастности по мере измельчения материала должна убывать.

Перечисленным требованиям удовлетворяет функция

$$\lg C = a (\lg r)^\alpha + \lg C_{\max} \quad (4.5.9)$$

при $a > 1$. Решая систему уравнений

$$\begin{cases} \lg C_0 = a (\lg r_0)^\alpha + \lg C_{\max}, \\ \alpha a (\lg r)^\alpha = -3, \end{cases}$$

получаем:

$$\alpha = \frac{\lg r_0^3}{\lg \frac{C_{\max}}{C_0}}, \quad a = -\frac{3}{\alpha (\lg r_0)^{\alpha-1}}, \quad (4.5.10)$$

где $r_0 = d_0/d_{\text{вкр}}$, d_0 — размер частиц, для которого определено C_0 .

Проведем расчет верхней границы коэффициента контрастности для галенитовой руды с содержанием свинца 2%. Пусть размер галенитовых вкраплений равен 3 мм. В пробе, раздробленной до 30 мм ($r = \frac{30}{3} = 10$), богатейшие куски содержат до 12% галенита, т. е. $12 \cdot 0.86 \approx 10\%$ свинца. Плотность богатейшей частицы $\gamma_p = 7.5 \cdot 0.12 + 2.7 \cdot 0.82 \approx 3.2$; $C_{\max} = k\gamma_p/c\gamma - 1 = 0.86 \cdot 7.5/0.02 \cdot 2.7 - 1 \approx 107$. Отсюда $C_0 = c_0\gamma_p/c\gamma - 1 = 10 \cdot 3.2/2 \cdot 2.7 - 1 = 4.9$; $C_{\max}/C_0 \approx 107/4.9 = 21.8$. Определим показатель кривой $\alpha = 3 \lg r / \lg C_{\max}/C_0 = 3 \lg 10 / \lg 21.8 = 2.24$. Угловый коэффициент кривой равен $a = -3/\alpha (\ln r)^{\alpha-1} = -3/2.24 (\ln 10)^{1.24} = -0.48$. Таким образом, $\ln C = \ln C_{\max} - 0.48 (\ln r)^{2.17}$. Определим коэффициент контрастности для различной степени измельчения:

d , мм	30	18	12	9	6	4.5	3
C	4.9	18.4	40.5	60.8	90	103.2	118.4

На рис. 4.5.2. изображены результаты расчетов для примера 1. Заштрихована область возможных значений коэффициентов контрастности данной руды по мере ее измельчения. Если требуемая относительная среднеквадратическая точность сокращения по свинцу составляет 10%, то требуемая масса пробы после сокращения, определяемая по формуле (без учета влияния начальной массы) $q = Cd^3\gamma/\sigma^2$, будет равна:

d , мм	3.0	1.8	1.2	0.9	0.6	0.45	0.3
$q_{10\%}$, г	$36 \cdot 10^3$	403	260	164	72	35	10

Если соединить прямой фундаментальную точку графика, имеющую координаты ($C = C_{\max}$, $d/d_{\text{вкр}} = 1$), с начальной экспериментальной точкой ($C = C_0$, $\frac{d}{d_{\text{вкр}}} = \frac{d_0}{d_{\text{вкр}}}$), то эта прямая будет соответствовать уравнению

$$\lg C = \lg C_{\max} - a \left(\lg \frac{d}{d_{\text{вкр}}} \right)^1 = \lg C_{\max} + a \lg d_{\text{вкр}} - a \lg d, \quad (a < 3)$$

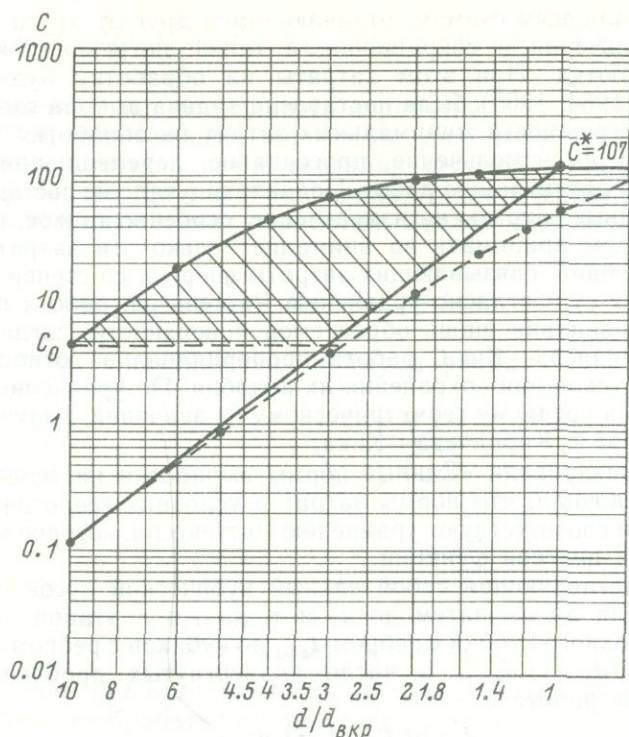


Рис. 4.5.2. Нижняя и верхняя границы контрастности руды, охарактеризованной в примере № 1. C_0 — контрастность исходной руды, определенная эмпирически или расчетом. Заштрихована область возможных значений при дальнейшем измельчении. Остальные обозначения см. рис. 4.5.1.

или, после потенцирования, уравнению

$$C = C_{\max} \text{const} / d^a.$$

Величина $C_{\max} \text{const}$ является константой. Подставив ее в нашу формулу сокращения (без учета исходного веса), получим

$$q = \text{const} \gamma / \sigma^2 \cdot d^3 / d^a = K d^{3-a}.$$

Этот результат является ни чем иным, как формулой Демонда—Хальфердаля, т. е. ее можно рассматривать как частный случай нашей формулы в предположении, что коэффициент контрастности руды изменяется по (4.5.9) с параметром $\alpha=1$. Теперь мы можем уверенно поставить точку в дискуссии об определении значений и физической интерпретации параметров K и a в формуле Демонда—Хальфердаля. Коэффициент K определяется критическим значением коэффициента контрастности размеров вкрапленников ($d_{\text{вкр}}$) и скоростью их высвобождения a . Показатель степени a есть по существу $3-a$, где величина a может быть получена на основании одной экспериментальной точки, для чего достаточно визуальным или химическим анализом определить наибольшее содержание компонента в частице (куске).

§ 6. Методы оптимизации обработки проб

Было показано (см. § 4), что квадрат случайной погрешности обработки пробы равен сумме квадратов случайных погрешностей, совершаемых на отдельных этапах. Поэтому, задавшись допустимой погрешностью, пробу можно об-

работать по различным схемам, отличающимся друг от друга промежуточной степенью измельчения и сокращения, а также погрешностью, допускаемой на этапе обработки. При этом затраты на обработку будут различными. Нами (Ткачев, 1964, 1965) была поставлена задача выбора схемы, требующей допускаемой погрешности минимальных затрат на обработку. Они складываются из затрат на измельчение, просеивание, перемешивание, сокращение, упаковку проб и организацию работ. Подавляющую долю составляют энергетические и временные затраты на измельчение, особенно тонкое, поэтому в дальнейшем мы будем принимать во внимание только эти затраты. Существует несколько уравнений, связывающих затраты энергии со степенью измельчения (см. БСЭ, изд. 2-е). Согласно уравнению Риттенгера, работа по измельчению прямо пропорциональна вновь образуемой поверхности; согласно объемному уравнению Кирпичева—Кика, работа пропорциональна отношению объемов измельченных кусков при сохранении их подобия. По уравнению Бонда работа пропорциональна среднему геометрическому из значений, получаемых по уравнению Риттенгера и Кирпичева—Кика.

Мы проанализировали «Единые нормы выработки на геологоразведочные работы» и установили, что нормы затрат в условиях геологической разведки почти в точности соответствуют уравнению Риттенгера, которое мы и принимаем при составлении целевой функции.

Проба, представляющая собой единый кубический кусок со стороной d_0 , дробится сначала до d_1 , затем до d_2 и т. д. — в n этапов до размеров d_n . При дроблении одного кубика с ребром d_{k-1} до кубиков с ребром d_k работа будет равна $6K(d_{k-1}/d_k - 1)d_{k-1}^2$, а число подвергнутых дроблению кубиков — d_0^3/d_{k-1}^3 , поэтому работа

$$A_k = 6Kd_0^3(1/d_k - 1/d_{k-1}), \quad (4.6.1)$$

где K — коэффициент, зависящий от механических свойств породы. Перед дроблением на k -м этапе число частиц в пробе $l_k = Q_0/\gamma d_{k-1}^3 m_{k-1}$, $k = 2, \dots, n-1$, а суммарная работа при n -этапной обработке

$$\begin{aligned} A &= 6KQ_0 \sum_{k=1}^n \left(\frac{1}{d_k} - \frac{1}{d_{k-1}} \right) \frac{1}{m_{k-1}} = \\ &= 6KQ_0 \left(\frac{1}{d_1} - \frac{1}{d_0} \right) + \sum_{k=2}^n \left(\frac{1}{d_k} - \frac{1}{d_{k-1}} \right) \frac{1}{m_{k-1}}. \end{aligned} \quad (4.6.2)$$

Постоянный множитель $6KQ_0$ можно опустить и в качестве целевой функции оптимизации обработки проб рассматривать такую:

$$A = \frac{1}{d_1} - \frac{1}{d_0} + \sum_{k=2}^n \left(\frac{1}{d_k} - \frac{1}{d_{k-1}} \right) \frac{1}{m_k} \quad (4.6.3)$$

при следующих линейных ограничениях:

$$\begin{aligned} d_k - d_{k-1} &\leq 0, \quad k = 1, \dots, n, \\ d_n &\geq \delta, \\ 1 - m_1 &\leq 0, \\ m_{k-1} - m_k &\leq 0, \\ m_{n-1} - m &\leq 0, \end{aligned} \quad (4.6.4)$$

где δ — технически достижимый (предельно малый) диаметр дробления, m — общая степень сокращения пробы. Из (4.6.2) вытекает следующее нелинейное ограничение:

$$d_1^3(m_1-1) + \sum_{k=2}^{n-1} d_k^3(m_k - m_{k-1}) + d_n^3(m - m_{n-1}) - \sigma^2 \leq 0, \quad (4.6.5)$$

где σ — заданная итоговая погрешность обработки.

Таким образом, оптимизация обработки проб есть задача нелинейного программирования, требующая минимизации целевой функции от $(2n - 1)$ переменных при наличии $(2n + 2)$ ограничений. Аналитическое решение удастся получить только при двух этапах обработки. Введем следующие обозначения:

$$\bar{D} = \frac{\sigma^2}{C}, \quad x_i = \frac{d_i}{d_0}, \quad \bar{D}_i = \frac{\sigma_i^2}{C}, \quad i = 1, 2, \dots, n. \quad (4.6.6)$$

Тогда задача состоит в нахождении минимума функции

$$A_2 = \frac{1}{x_1} \cdot 1 + \left(\frac{1}{x_2} - \frac{1}{x_1} \right) \cdot \frac{1}{m_1} \quad (4.6.7)$$

при ограничениях

$$\begin{aligned} 1 - x_1 &\geq 0, & m_1 - 1 &\geq 0, \\ x_1 - x_2 &\geq 0, & m_2 - m_1 &\geq 0, \\ x_2 &> 0, \\ x_1^3(m_1 - 1) + x_2^3(m_2 - m_1) - \bar{D} &= 0. \end{aligned} \quad (4.6.8)$$

Используя равенство (4.6.8), а также то обстоятельство, что средние квадратические погрешности обработки на отдельных этапах — положительные величины, можно сузить область поиска минимума:

$$\sqrt[3]{\bar{D}/(m^2 - 1)} \leq x_1 \leq 1, \quad 0 < x_2 \leq \sqrt[3]{\bar{D}/(m^2 - 1)}. \quad (4.6.9)$$

Из сущности задачи вытекает, что x_1 больше, чем x_2 . Поэтому можно находить минимум функции

$$A_2 = 1/x_1 - 1 + 1/x_2 m_1 \quad (4.6.10)$$

с ограничением в виде равенства

$$x_1^3 m_1 + x_2^3 m_2 - \bar{D} = 0. \quad (4.6.11)$$

В этом случае точка минимума находится методом неопределенных множителей Лагранжа:

$$x_1^* = \sqrt[4]{\bar{m}}, \quad \sqrt[3]{\bar{D}/4m_2}, \quad x_2^* = \sqrt[3]{3\bar{D}/(4m_2)}, \quad m_1^* = 3\sqrt[4]{m_2}. \quad (4.6.12)$$

Точки минимума, вычисленные по формулам (4.6.12), были проверены методом сетки на ЭВМ. Результаты практически совпали. При $D = 500$, $d_0 = 1$, $m = 10^7$ аналитически были получены следующие значения: сокращение на первом этапе $m_1^* = 5.08 \cdot 10^2$, диаметр на первом этапе $d_1^* = 0.9$, на втором — 0.02 мм.

В наиболее распространенных случаях, когда число этапов обработки больше двух, оптимизация обработки требует применения численных методов выпуклого анализа. При большом (как у нас) числе переменных это приводит к огромным затратам машинного времени. Поэтому мы применили принцип декомпозиции, заключающийся в данном случае в следующем: а) фиксируем степени сокращения на этапах m_i и решаем задачу нахождения оптимальных размеров дробления d_i , обеспечивающих минимум работы при заданной точности; б) фиксируем размеры дробления d_i и решаем задачу нахождения

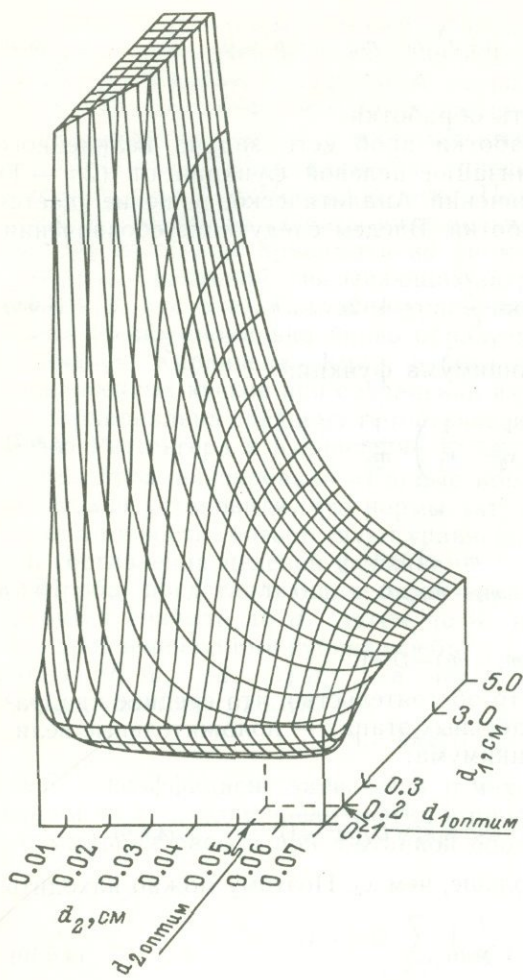


Рис. 4.6.1. Вид поверхности целевой функции (затрат на измельчение) для двухэтапной обработки в координатах «степень измельчения на первом этапе—степень измельчения на втором этапе». Масштаб логарифмический.

оптимальных степеней сокращений. Затем переходим к пункту (а) и продолжаем процесс, пока решения не перестанут существенно (более чем на 3—5 %) отличаться. У нас нет математических доказательств, что этот процесс будет сходиться к глобальному минимуму. Однако проверка сеточным методом показала, что процесс решения быстро завершается. По рис. 4.6.1, выполненному на ЭВМ, устанавливается, что целевая функция образует лишь один глобальный минимум и не образует никаких локальных минимумов или других особенностей. Глобальный минимум находится в узком «желобе», вытянутом вдоль оси d_2 . В широких окрестностях минимума затраты на обработку изменяются незначительно, поэтому с практической точки зрения вполне удовлетворительна приближенная оптимизация. В областях, далеких от оптимальной точки, затраты возрастают в десятки раз, что свидетельствует о важности выбора оптимальных схем.

§ 7. Составление надежных и экономически оптимальных схем обработки проб

С помощью методов, изложенных в § 6 и более подробно — в другой нашей работе (Ткачев, Шеин, 1987), показано, что оптимальная схема обработки пробы всегда существует. Если ввести понятие критического размера (диаметра) измельчения

$$d_{кр} = \sqrt[3]{Q_0 \sigma_{доп}^2 / C_{max} / (m-1)}, \quad (4.7.1)$$

где Q_0 — начальная масса пробы, $\sigma_{доп}$ — допустимая погрешность обработки, C_{max} — максимальный для данной руды коэффициент контрастности, m — общая степень сокращения пробы, то решение задачи и представление результатов значительно упрощаются.

Для расчета оптимальной схемы обработки требуются теперь только два параметра — требуемая степень общего сокращения и критический диаметр. Первый из этих параметров всегда известен по исходной массе и конечной навеске. По своему характеру это конструктивный параметр, ибо именно он

определяет «траекторию» обработки: с его изменением изменяется соотношение степеней сокращения на этапах и соотношение диаметров измельчения. Вторым параметром — критический диаметр — выступает здесь в качестве масштабного множителя. Он определяет абсолютные размеры измельчений. Критический размер обладает следующим свойством. Если пробу нельзя (по техническим причинам) измельчить до $d_{кр}$, выдержать допустимую погрешность невозможно. Если ее можно измельчить только до $d_{кр}$, возможна обработка только в один этап. Выражение для критического диаметра (4.7.1) мы преобразуем в следующий вид:

$$d_{кр} = \sqrt[3]{1/T}, \quad \text{где } T = \frac{C}{\sigma^2 q_k} \left(\frac{m_k - 1}{m_k} \right). \quad (4.7.2)$$

Величину T назовем тяжестью обработки пробы. Из приведенного выражения видно, что тяжесть обработки увеличивается с увеличением начальной массы пробы и с уменьшением конечной навески q_k . Она увеличивается пропорционально коэффициенту контрастности и уменьшается с увеличением допустимой погрешности.

Результаты расчетов на ЭВМ оптимальных схем обработки проб сведены в номограммы, по одной на фиксированное число этапов: от 2 до 5 (рис. 4.7.1, а—г, см. схему 2). Они позволяют сделать следующие выводы и предложить практическую методику составления оптимальных схем.

Если принимать в расчет только затраты на дробление, обработка в два этапа всегда выгоднее, чем в один. Однако обычно степень измельчения загружаемого в аппарат материала при эффективном его использовании должна быть не менее трех-пяти. Поэтому практически включение второго этапа становится эффективным, если требуемая общая степень сокращения не менее 100. При увеличении m_k конечный размер измельчения пробы должен быть меньше, но никогда не опускается ниже $0.63d_{кр}$. Этот размер можно принять как контрольный при составлении оптимальных схем в два этапа. Рассмотрим подробно процесс составления оптимальной схемы обработки с использованием построенных номограмм.

Первый шаг состоит в выборе числа этапов обработки. Он определяется единственным параметром — общей требуемой степенью сокращения пробы $m_k Q_0 / q_k$, где Q_0 — начальная масса, q_k — навеска, идущая в анализ. В табл. 4.7.1 во второй колонке приведено число этапов, полученное в соответствии с принципом эффективного использования измельчителей. В каждом из них степень измельчения должна достигать в среднем 3—5, но не опускаться ниже двух. Так, при общей степени сокращения $m_k = 200$ обработку необходимо проводить в два этапа, при $m_k = 5 \text{ кг/1 г} = 5 \cdot 10^3$ требуется три этапа и т. д. Рекомендуем следовать простому эмпирическому правилу: число этапов обработки равно целой части десятичного логарифма степени сокращения.

Второй шаг в составлении оптимальной схемы обработки заключается в выборе степени сокращений на этапах. Для этого выбирается номограмма для соответствующего числа этапов. В нашем случае ($m_k = 200, n = 2$) необходимо использовать номограмму, данную на рис. 4.7.1, а. По горизонтальной оси находим значение 200 и для него снимаем степени сокращения на этапах: $m_1 = 8, m_2 = m_k = 200$ (для руд с переменной контрастностью аналогичным способом используются рис. 4.7.2, а—г; см. вкладку 3). Для случая $m_k = 5 \cdot 10^3$ применяется номограмма на рис. 4.7.1, б и для значения 5000 снимаем степень сокращения на этапах: $m_1 = 6, m_2 = 80, m_3 = 5000$. Для приближенного определения степеней сокращения можно использовать табл. 4.7.1 или для варианта с изменяющейся контрастностью — табл. 4.7.2. Для $m_k = 200$ ($n = 2$) из второй строки табл. 4.7.1 устанавливаем, что на первом этапе необходимая степень сокращения должна составить 39 % общей величины, на втором — 61 %. Для

a

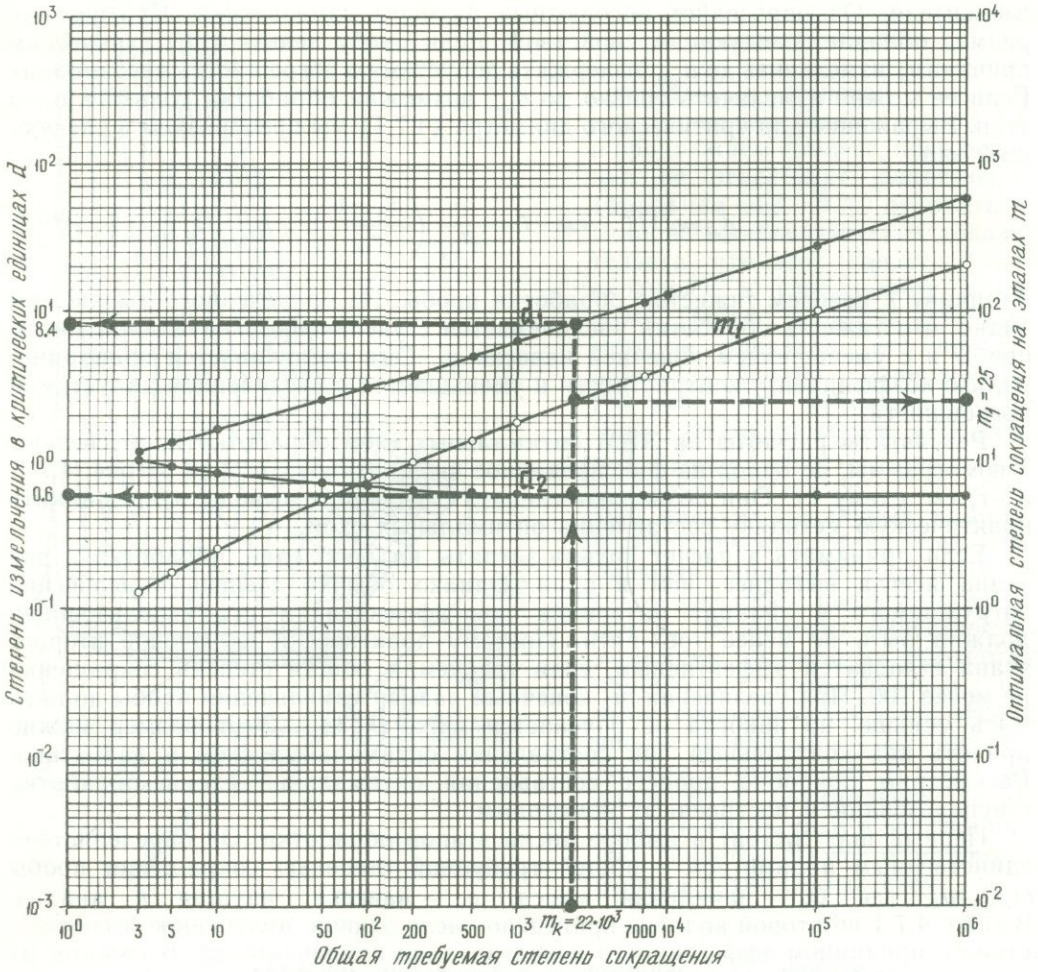


Рис. 4.7.1. Номограмма оптимальных схем обработки проб с постоянной контрастностью.

На горизонтальной оси откладывают общую степень сокращения $2.2 \cdot 10^3$ и восстанавливают из точки m_k перпендикуляр. Оптимальные сокращения находят, проектируя точку пересечения перпендикуляра с линией m_i на ось справа, оптимальные дробления d_1 и d_2 , проектируя точки пересечения перпендикуляра с линиями d_1 и d_2 на ось слева. *a* — обработка проб в два этапа, *b* — в три этапа, *в* — в четыре этапа, *г* — в пять этапов. Штриховыми линиями со стрелками показан порядок использования номограммы для одного из примеров.

б

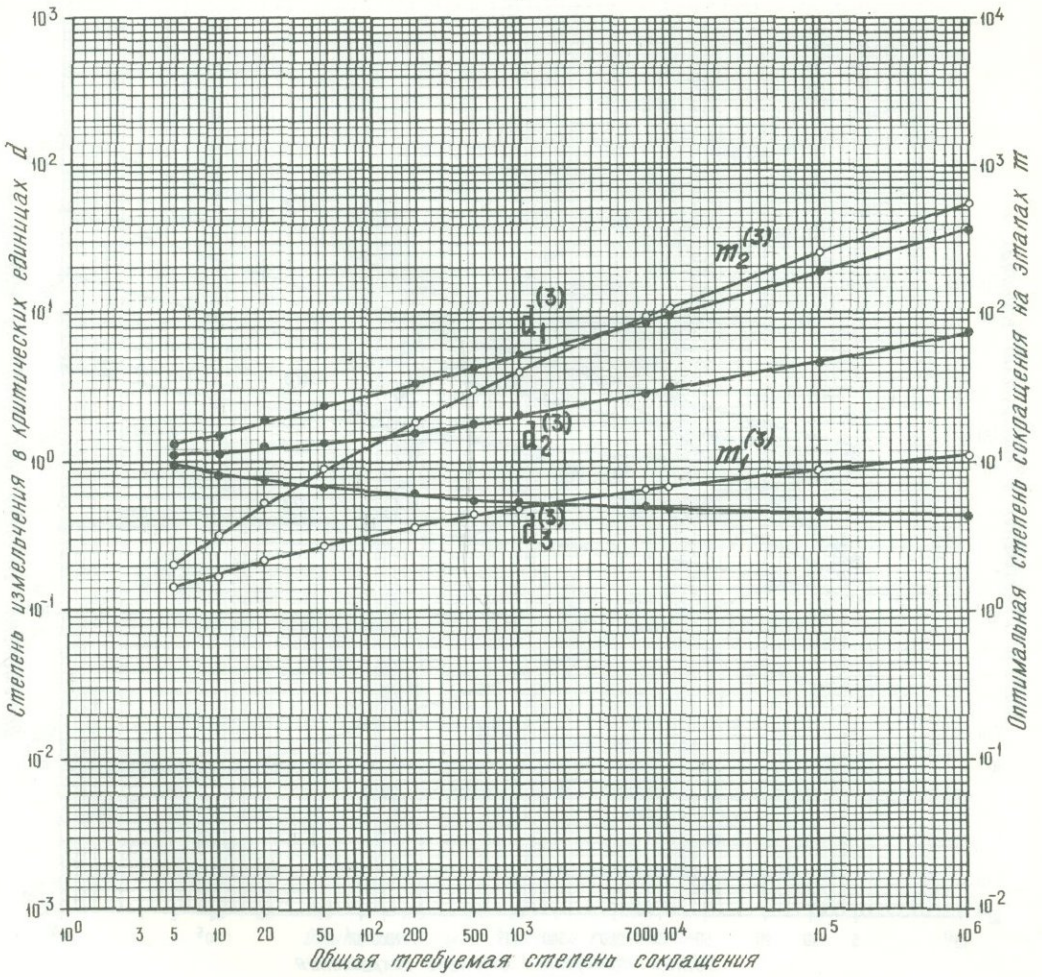


Рис. 4.7.1 (продолжение).

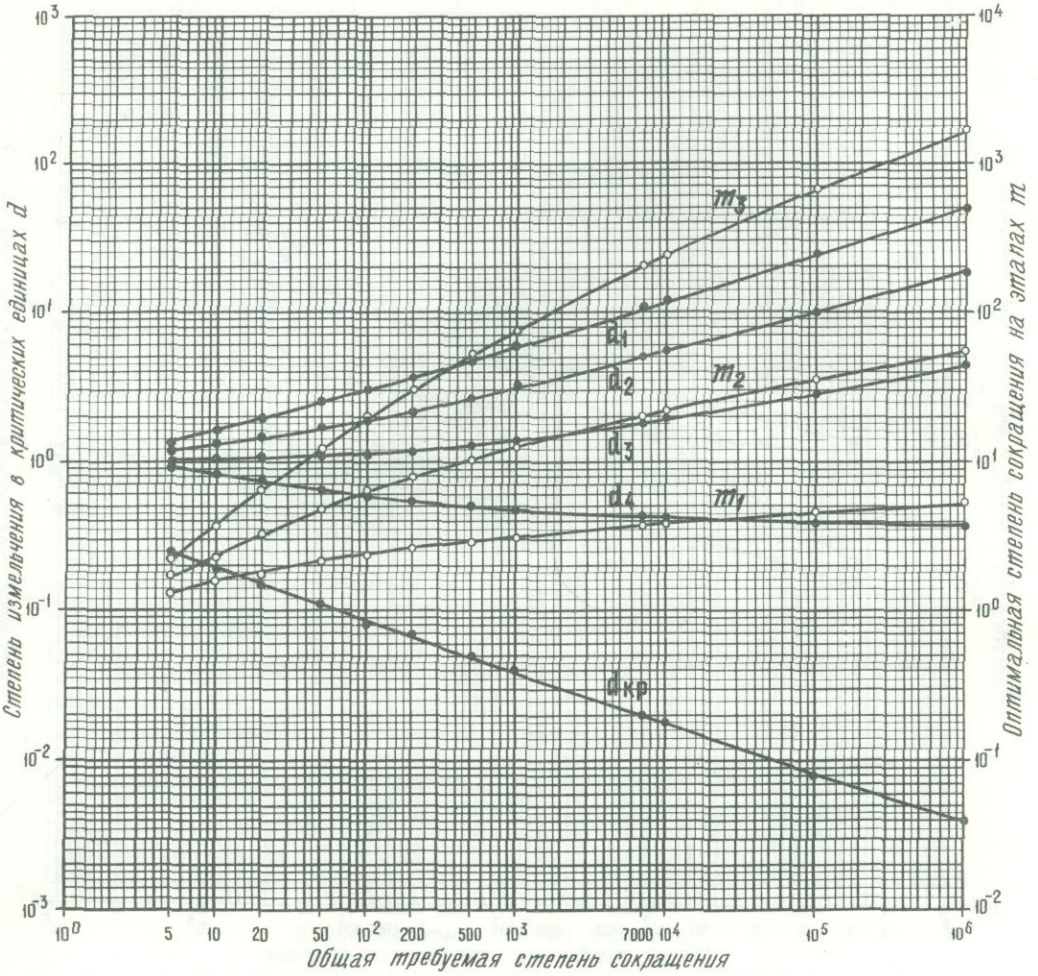


Рис. 4.7.1 (продолжение).

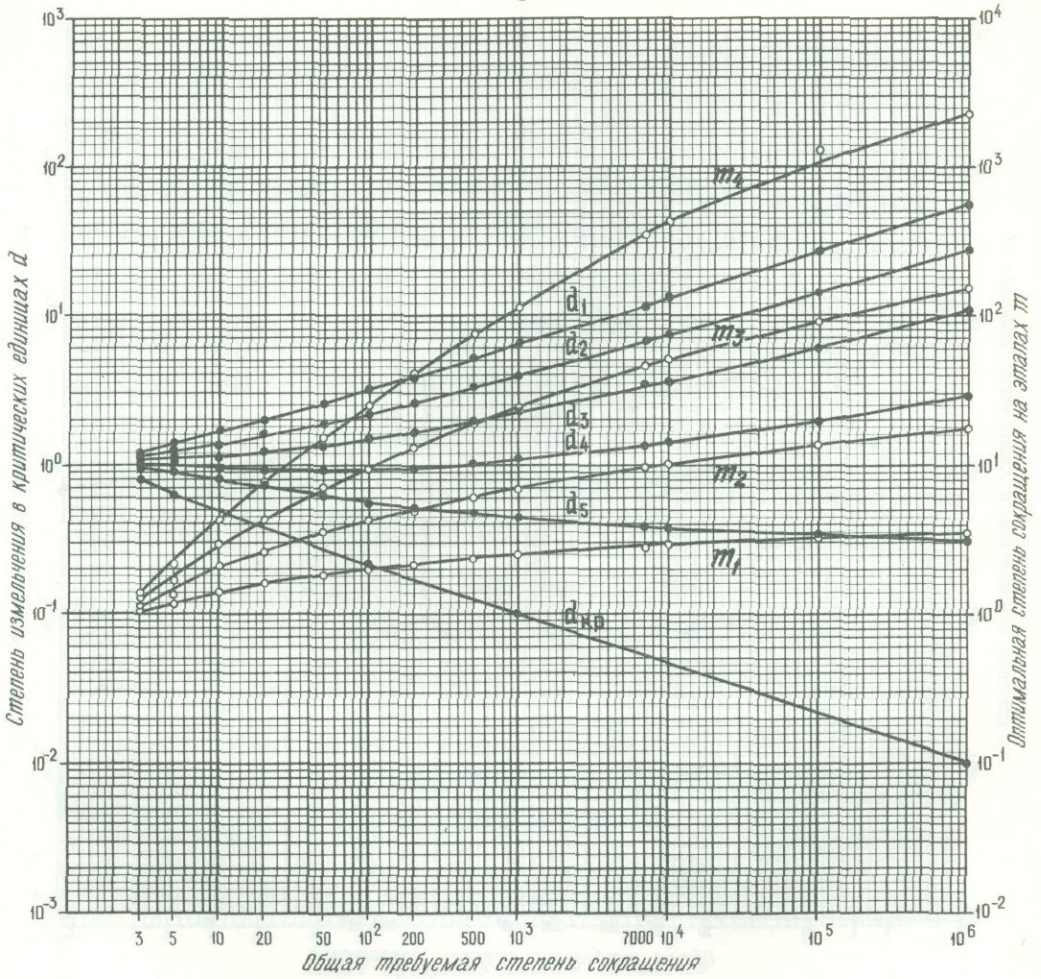


Рис. 4.7.1 (продолжение).

а

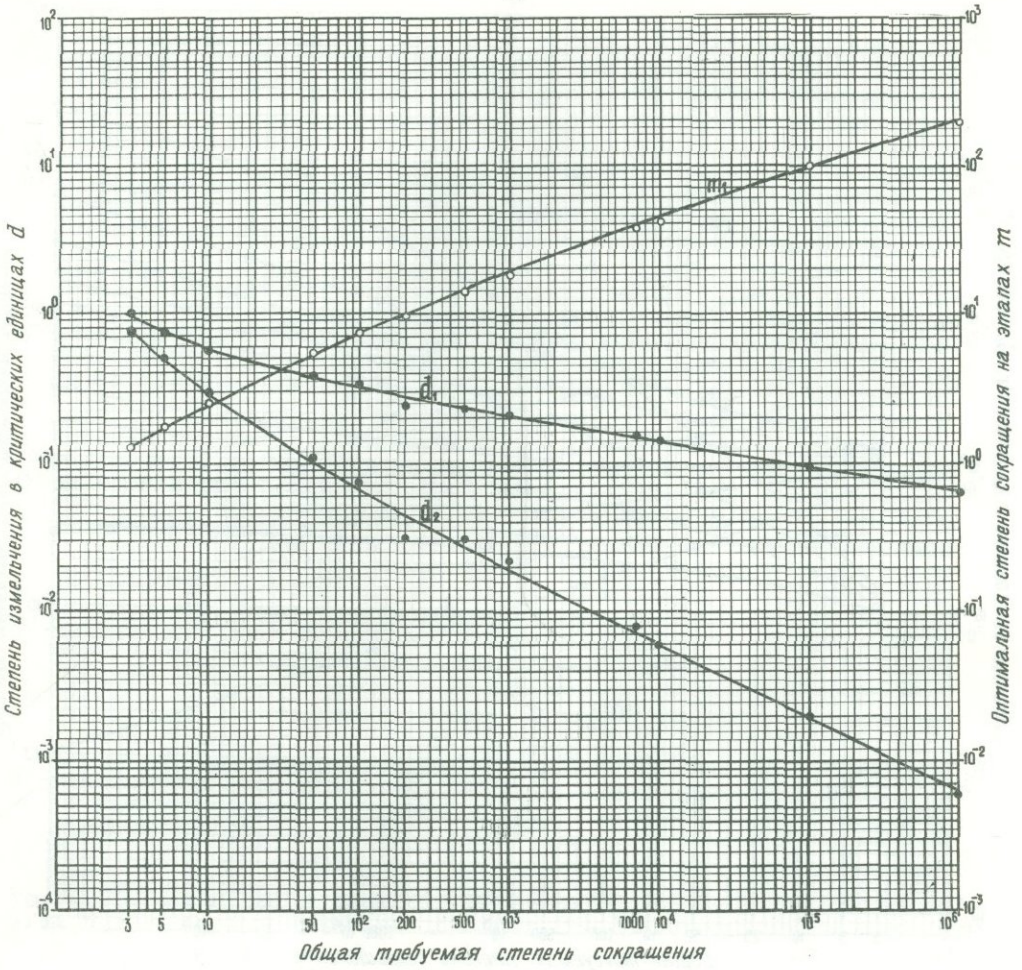


Рис. 4.7.2. Номограмма оптимальных схем обработки проб с переменной контрастностью.
 Обозначения — те же, что и под рис. 4.7.1. Порядок использования номограмм тот же.

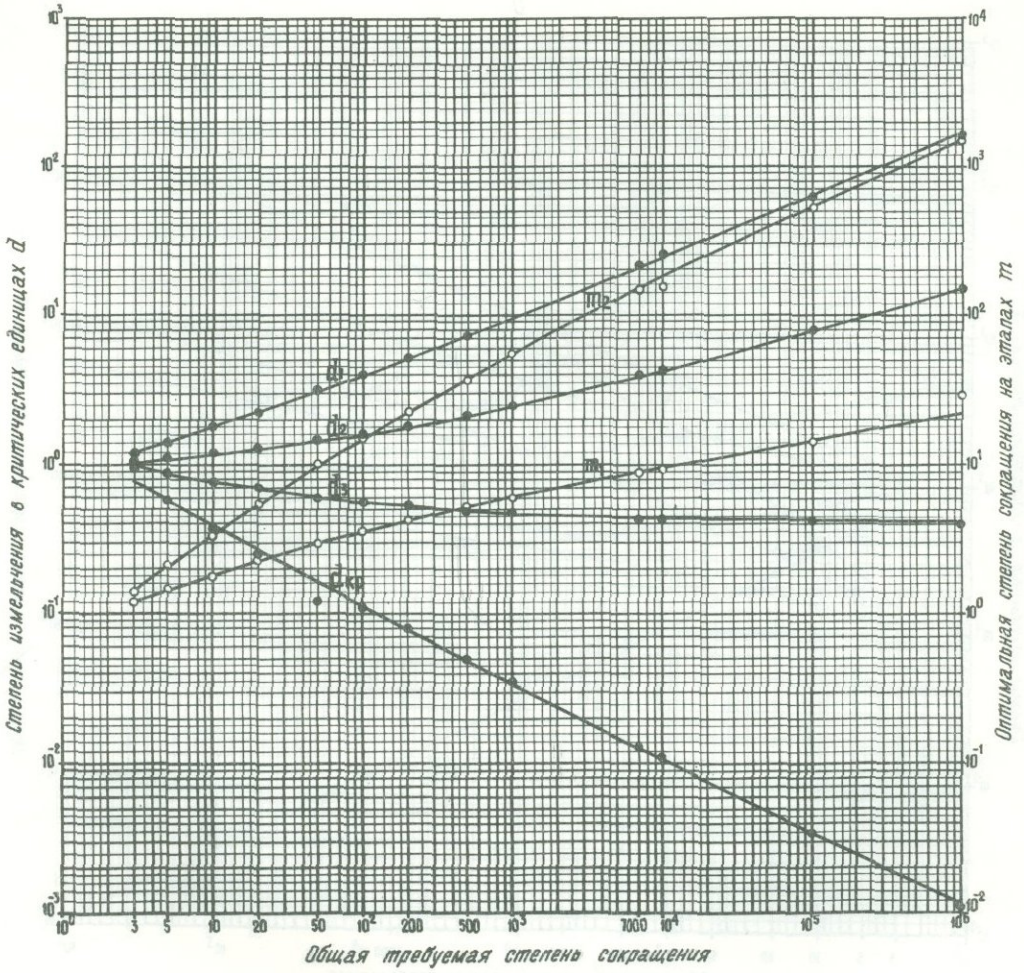


Рис. 4.7.2 (продолжение).

в

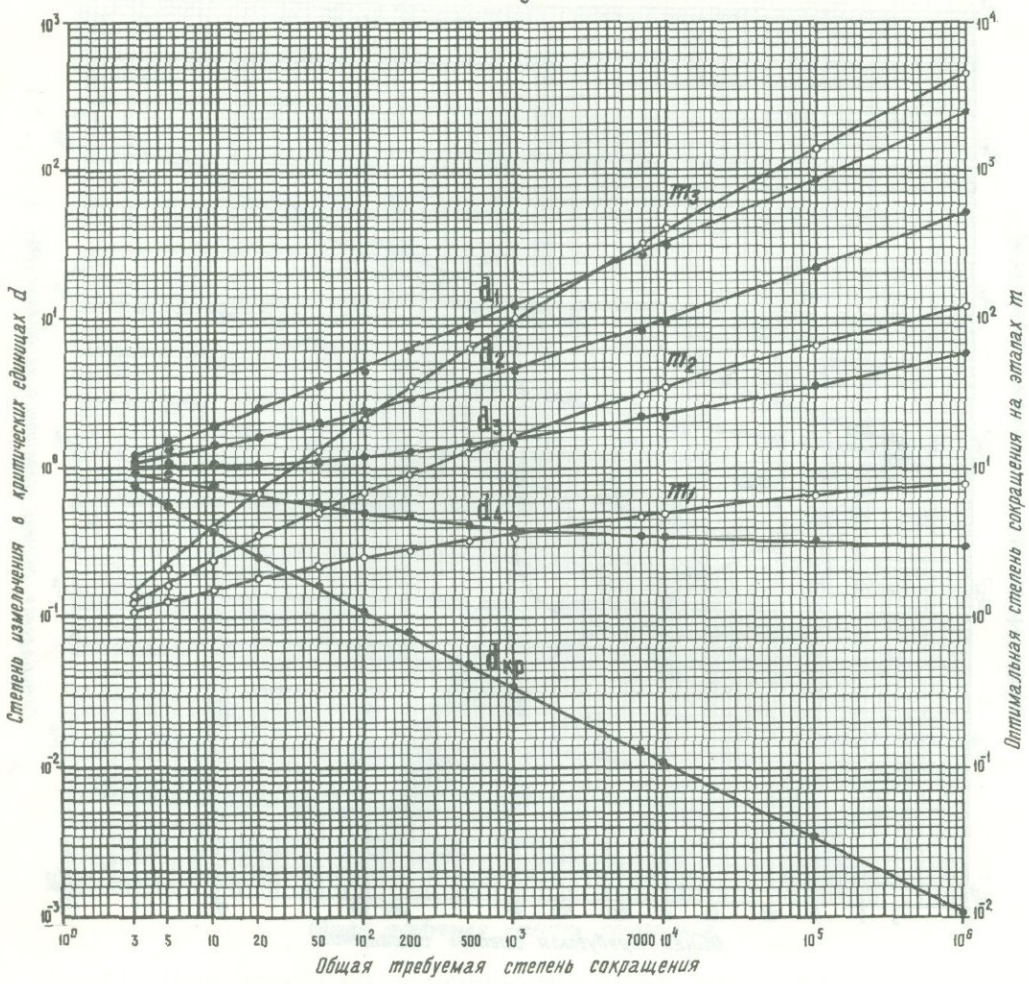


Рис. 4.7.2 (продолжение).

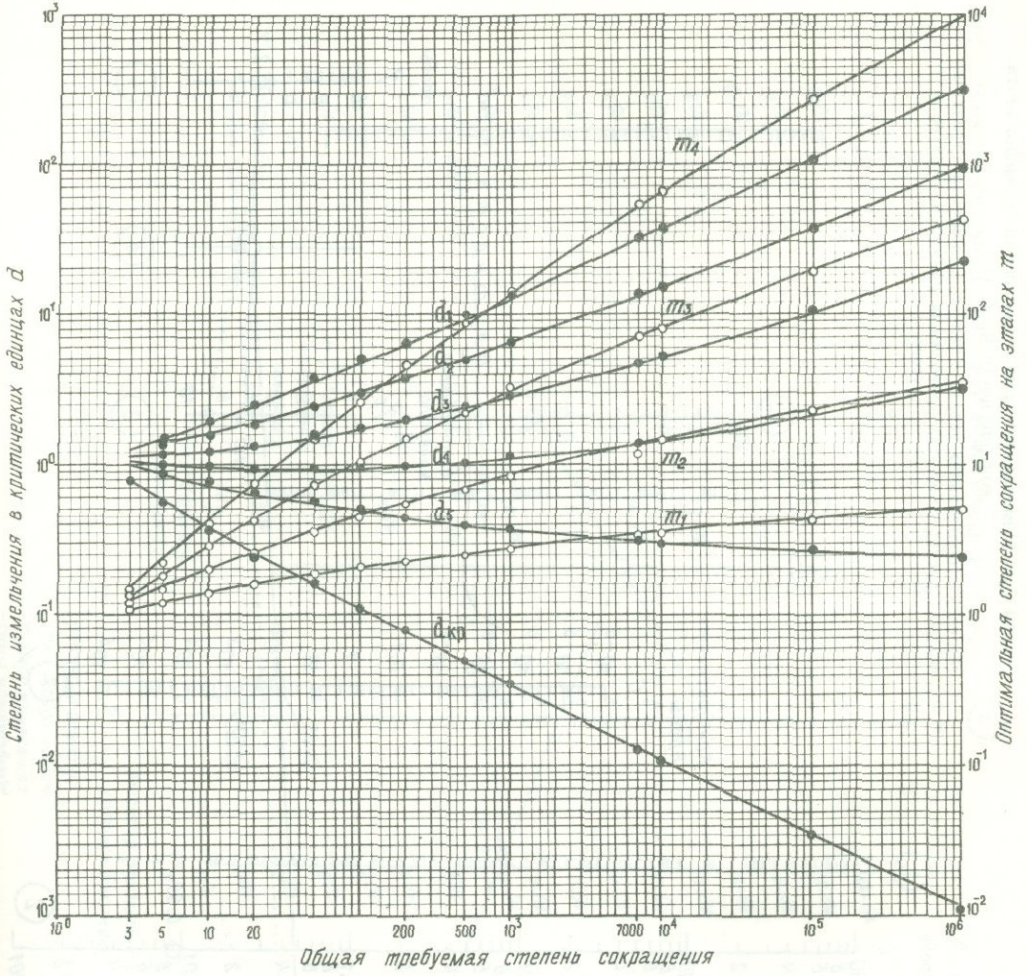


Рис. 4.7.2 (продолжение).

Содержание полезного компонента, %
в богатейшем минерале

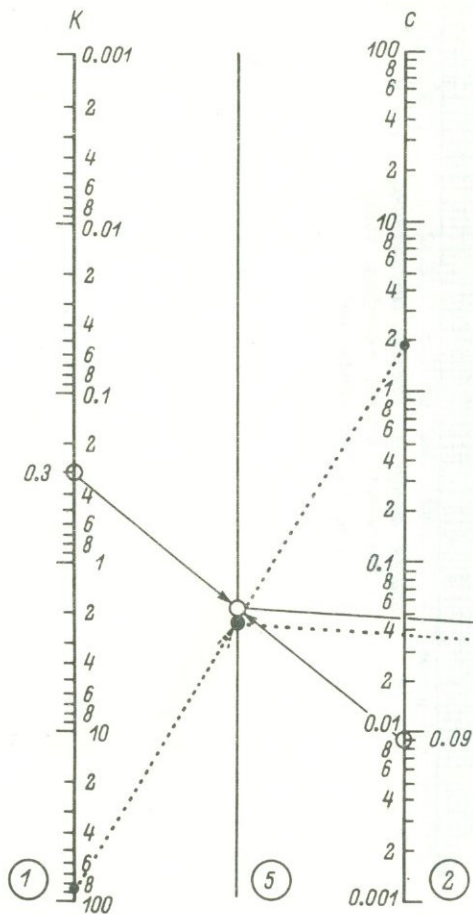
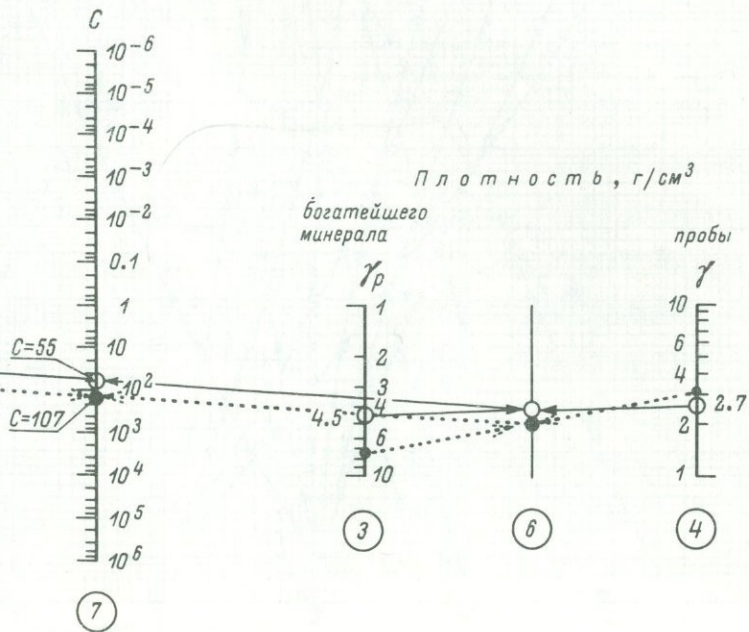


Рис. 4.7.3. Номограмма для определения коэффициента контрастности проб (C).

Порядок использования указан в тексте. Сплошными стрелками показан расчет одного из приводимых примеров, пунктиром — другого.

Контрастность
пробы



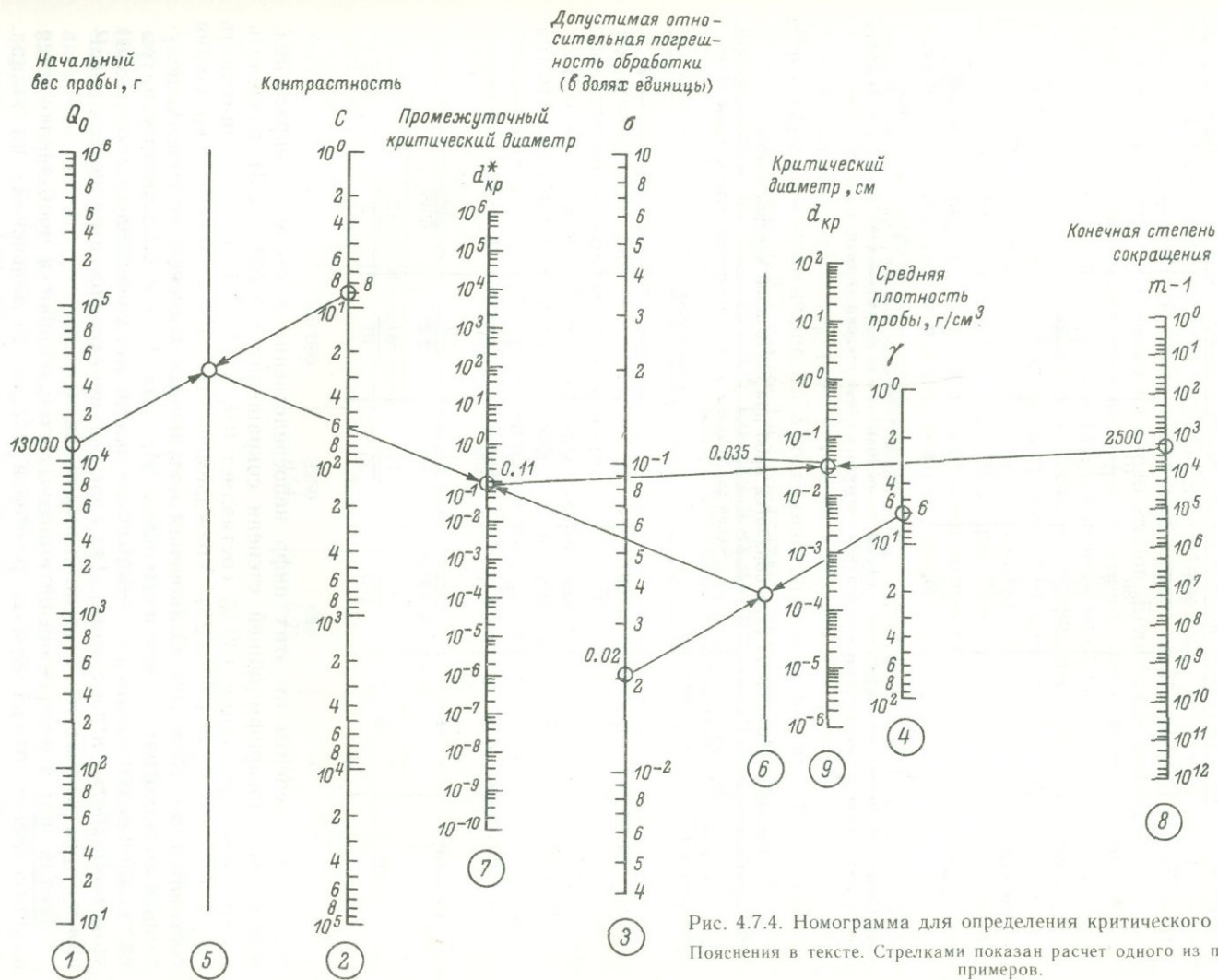


Рис. 4.7.4. Номограмма для определения критического диаметра. Пояснения в тексте. Стрелками показан расчет одного из приводимых примеров.

Таблица оптимальных схем обработки проб при изменяющейся контрастности

Общая степень сокращения m_k	Рекомендуемое число этапов n	Доля степени сокращения (в %), реализуемая на каждом этапе (над чертой), и размер измельчения в единицах критического диаметра (под чертой)				
		1	2	3	4	5
10^2	1	100	—	—	—	—
		1.0	—	—	—	—
10^2-10^3	2	39	61	—	—	—
		2.5	0.7	—	—	—
10^3-10^4	3	23	30	47	—	—
		5	2	0.5	—	—
10^4-10^5	4	15	17	27	41	—
		12	5.5	1.9	0.4	—
$>10^5$	5	10	13	16	22	39
		25	14	6	2	0.3

Примечание. Здесь и в табл. 4.7.2 логарифм общей степени сокращения разделяется на слагаемые в указанных пропорциях, потенцируется и получается частная степень сокращения m_i ; для получения общей степени сокращения m_k на i -м этапе следует перемножить частные степени от первой до i -й.

Таблица 4.7.2

Таблица оптимальных схем обработки проб при постоянной контрастности

Общая степень сокращения m_k	Рекомендуемое число этапов n	Доля степени сокращения (в %), реализуемая на каждом этапе (над чертой), и размер измельчения в единицах критического диаметра (под чертой)				
		1	2	3	4	5
10^2	1	100	—	—	—	—
		1.0	—	—	—	—
10^2-10^3	2	42	58	—	—	—
		2.5	0.6	—	—	—
10^3-10^4	3	26	32	42	—	—
		9.5	2.5	0.45	—	—
10^4-10^5	4	15	21	25	35	—
		12	10	2.2	0.35	—
$>10^5$	5	12	14	20	12	32
		100	35	10	2	0.25

того чтобы перейти от этих цифр непосредственно к степени сокращения, необходимо логарифм общей степени сокращения ($\lg 200 = 2.3$) разделить в указанной пропорции (39 % составляет 0.9, 61 % — 1.4) и потенцировать ($m_1 = 10^{0.9} \approx 8$, $m_2 = 10^{1.4} \approx 25$). Убеждаемся, что общая степень сокращения составляет $8 \cdot 25 = 200$. Отклонения полученных значений от точных, полученных по номограмме, незначительные. Из табл. 4.7.1 и 4.7.2 легко усматривается общая тенденция в степенях сокращения: они увеличиваются от первых этапов обработки к последним. Для проб с изменяющейся в процессе дробления контрастностью эта тенденция выражена меньше.

Третий шаг в построении оптимальных схем обработки проб заключается в установлении относительных размеров частиц (измельчений) на этапах. Оно аналогично описанному для степеней сокращения; значения снимаются с графиков, обозначающих размер, а не степень сокращения (кривые d на номограммах и левые шкалы для считывания значений). В нашем случае ($m_k = 200$)

на рис. 4.7.1,а находим $d_1^* = 2.6$, $d_2^* = 0.69$. Для приближенного решения используем табл. 4.7.1, где значения размеров помещены под чертой: $d_1^* = 2.5$, $d_2^* = 0.7$. Общая качественная картина распределения измельчений такова. При многоэтапной обработке степень измельчения (отношения размеров частиц до и после дробления на этапе) на первых этапах невелика и увеличивается к последним этапам. При увеличении числа этапов конечный размер уменьшается от 1 (для одного этапа) до 0.25 (для пяти этапов) критического. Конечная степень измельчения (в критических диаметрах) приближенно равна критическому диаметру, деленному на число этапов. Если такое измельчение невозможно по техническим причинам, число этапов необходимо уменьшить по сравнению с тем, которое рекомендуется исходя из общей степени сокращения.

Четвертый шаг заключается в расчете коэффициента контрастности. Для быстрого приближенного его расчета приводим номограмму (рис. 4.7.3).

Пятый шаг заключается в расчете тяжести обработки T и критического диаметра $d_{кр}$ по формуле (4.7.2). Мы приводим номограмму, реализующую эту формулу (рис. 4.7.4). Для иллюстрации четвертого и пятого шагов, а также всего процесса в целом рассмотрим следующий пример.

Требуется обработать кварц-галенитовую пробу с содержанием свинца около 2.0 %. Начальная масса пробы — 13 кг, навеска, идущая в анализ, — 1 г, допустимая погрешность обработки — 10 %. Коэффициент контрастности такой пробы $C = k\gamma_p / (c\gamma) - 1 = 86 \cdot 7.5 / 2 \cdot 3 - 1 = 107.5 - 1 = 106.5$. Расчет C по номограмме рис. 4.7.3 показан штриховыми линиями. Исходные значения на шкалах 1 и 2, а также 3 и 4 соединяют прямыми, получая промежуточные значения на шкалах 5 и 6. Соединяя их прямой, на шкале 7 читают ответ.

Тяжесть обработки этой пробы $T = C / \sigma^2 q_k \cdot (m_k - 1) / m_k = 107 / 0.1^2 \cdot 1 = 10700 \approx 11 \cdot 10^3$. Критический диаметр $d_{кр} = \sqrt[3]{1/T} = \sqrt[3]{1/11 \cdot 10^3} = 1/10 \times \sqrt[3]{11} = 1/22.2 = 0.045$ (см.).

На номограмме рис. 4.7.4 расчет $d_{кр}$ указан штриховыми линиями. Шкалы задействуют в порядке их нумерации.

В соответствии с общей степенью сокращения ($m_k = 13000 : 1 = 1.3 \cdot 10^4$) оптимальная обработка такой пробы должна проводиться в 4 этапа. Степени сокращения и измельчения считываем с номограммы рис. 4.7.1, в или приближенно выбираем из табл. 4.7.1. Полученную схему обработки изобразим в виде

Показатель	Этапы обработки				
	Исходная проба	1	2	3	4
Частная степень сокращения на этапе	—	4	5	13	50
Масса, г	13000	3250	650	50	1
Размер, см	—	0.54	0.25	0.09	0.02

Еще раз проиллюстрируем, как изменяется число этапов, степени сокращения и измельчения по этапам, модифицируя пример, описанный на с. 80. При массе аналогичной по свойствам пробы 2 кг степень сокращения равна 200. Число этапов для такой степени сокращения равно двум. По табл. 4.7.1 определяем, что степень сокращения на первом этапе должна составить 8, на втором — 25. На первом этапе необходимо измельчить пробу до $d_1 = 2 \cdot d_{кр} = 0.0677 \times 2 = 0.14$ см, на втором — до $d_2 = 0.7 \cdot d_{кр} = 0.05$ см. Если бы исходная масса такой же пробы была 20 кг, то общая степень сокращения составила бы 2000, оптимальное число этапов — 3 со следующим распределением сокращений и измельчения (округленно):

Показатель	Этап обработки			
	Исходная проба	1	2	3
Сокращение	—	6	10	33
Масса, г	$2 \cdot 10^4$	3333	33	10
Измельчение, см	—	0.3	0.14	0.03

При исходной массе 200 кг необходимо 4 этапа обработки по следующей схеме:

Показатель	Этап обработки				
	Исходная проба	1	2	3	4
Частная степень сокращения	—	4	5	15	67
Масса, г	$2 \cdot 10^5$	50 000	10 000	670	10
Измельчение, см	—	0.8	0.4	0.13	0.03

Поскольку степень измельчения от этапа к этапу здесь не очень велика, почти равноценной была бы схема из трех этапов:

Показатель	Этап обработки		
	1	2	3
Сокращение	10	20	100
Масса, г	$2 \cdot 10^4$	1000	10
Измельчение, см	0.3	0.14	0.03

Глава 5

ПОГРЕШНОСТИ ИЗМЕРЕНИЙ И АНАЛИЗОВ В РАЗВЕДКЕ МЕСТОРОЖДЕНИЙ, ИХ КОНТРОЛЬ И УЧЕТ ПРИ ПОДСЧЕТЕ ЗАПАСОВ

§ 1. Общие замечания

Геологические объекты по сравнению с другими природными объектами оценки отличаются большой изменчивостью изучаемых свойств. Поэтому погрешностями анализов или измерений по причине их малости по сравнению с естественной изменчивостью обычно пренебрегают, однако во многих случаях это приводит к существенному искажению результатов. Знание метрологических характеристик применяемых при разведке методов измерений позволяет корректно выбрать дальнейшую обработку данных.

Понятие «измерение» в данной работе не нуждается в определении. Заметим лишь, что в отличие от анализа под измерением в узком смысле слова мы понимаем установление численного значения интересующей нас величины непосредственным сравнением со стандартной шкалой, например измерение мощности пласта рулеткой. Под анализом понимается измерение, требующее сложной подготовки, например химического разложения образца (пробы). Непосредственному измерению при анализе подвергается так называемый аналитический сигнал. По одному или нескольким аналитическим сигналам как функция от них вычисляется аналитический параметр, а измеряемая (анализируемая) величина снимается с градуировочного графика как функция аналитического параметра. Под такую общую схему подпадают практически все геологические и геофизические измерения и анализы. В одном из разделов этой главы будут рассмотрены способы повышения точности анализов, связанные с построением градуировочных графиков.

§ 2. Характеристики качества анализа

Каждый метод измерения характеризуется сходимостью, воспроизводимостью, правильностью и точностью. ГОСТом 10 263-70 официально закреплены только два термина: сходимость — качество измерений, отражающее близость друг к другу результатов измерений, выполняемых в одинаковых условиях; воспроизводимость — то же, для измерений, выполняемых одним и тем же методом, но в различных условиях (в различное время, разных местах, различными средствами). На практике этих крайних видов характеристики качества измерений недостаточно, так как уровень воспроизводимости сильно зависит от того, какая часть условий анализа стабилизирована. В литературе не раз отмечалась необходимость различать несколько видов воспроизводимости, соответствующих определенной степени стабилизации условий анализа (Налимов, 1960; Дубов, 1972; Ткачев, Юдович, 1975), однако мы не встречали рекомендаций, какие из этих видов действительно полезны и в каких случаях их следует использовать при расчетах.

Рассмотрим виды воспроизводимости на примере количественного спектрального анализа, проводимого по следующей методике: на пластинку фотографируются спектры партии проб, нескольких эталонов и марки интенсивностей. По последним строится характеристическая кривая (ХК) проявленной фотоэмульсии данной пластинки. С помощью ХК почернения спектральных линий пересчитываются в интенсивности. Аналитический параметр рассчитывается как функция интенсивностей линий определяемого элемента, элемента сравнения и фона около этих линий. Градуировочный график строится для каждой пластинки в координатах: аналитический параметр — логарифм содержания в эталонах. Содержания в пробах оцениваются по этому графику. Для такой методики анализа практически полезно выделять следующие виды воспроизводимости.

1. *Сходимость*, т. е. воспроизводимость для пары проб, снятых на одной пластинке рядом. Так иногда снимают две реплики одной пробы для получения наилучшей сходимости как предельной возможности метода. Соответствующая этой сходимости погрешность — наименьшая из возможных для данной методики анализа, так как максимально реализует постоянство условий анализа: условия горения электрической дуги максимально близки, интенсивности рассчитываются по одной характеристической кривой, содержания — по одному градуировочному графику. К сожалению, область применения такой характеристики как сходимость при интерпретации результатов ограничена сравнением содержаний в парах рядом снятых проб. Например, если необходимо сравнить основное и контрольное опробование, то для уменьшения влияния на выводы погрешностей анализа пробы целесообразно анализировать не в случайном порядке и, тем более, не двумя сериями — основные, а затем контрольные, а разбив их на пары — основная и соответствующая ей контрольная. Порядок съемки в паре желательно рандомизовать. Такая простая и не требующая дополнительных затрат процедура позволяет «выжать» из анализа его предельную точность, не прибегая к более точным дорогим методам.

2. *Воспроизводимость* на пластинке, т. е. воспроизводимость результатов анализов проб при их съемке на одну пластинку в случайном порядке (на одну пластинку должна быть снята вся совокупность реплик проб и эталонов). Этот вид воспроизводимости можно распространить и на случай совместной фотообработки нескольких экспонированных «за один прием» пластинок при использовании одной для всех характеристической кривой и одного градуировочного графика. Необходимое условие использования этого варианта воспроизводимости — непременная рандомизация порядка съемки. Нестабильность горения дуги за время съемки на пластинку существенно снижает величину этого вида воспроизводимости по сравнению со сходимостью. Тем не менее сравнительно малая величина этой погрешности воспроизводимости создает благоприятные условия для получения статистически значимых выводов. Естественно, что эти выводы могут касаться только совокупности проб, помещенных на эту пластинку. Емкость пластинки ограничена, в связи с чем редко удается широко использовать при разведке месторождений этот уровень точности.

3. *Воспроизводимость* на пачке, т. е. в пределах совокупности проб, снятых на пластинки одной пачки или партии одного номера эмульсии при неизменности условий съемки и настройки прибора. Величина этого вида погрешности воспроизводимости выше предыдущей в основном за счет вклада погрешностей построений градуировочного графика и характеристической кривой. При кажущейся искусственности выделения этого вида воспроизводимости в практической ее полезности не приходится сомневаться, так как грамотный аналитик старается использовать пластинки из одной пачки для анализа проб данной геохимической серии, одного заказчика, одного месторождения или его определенной части.

4. Лабораторная воспроизводимость в своей максимальной величине проявляется в длинном или растянутом во времени ряду анализов с дополнительной подстройкой приборов, при использовании разных партий пластинок и других средств анализа. В ее состав входит также составляющая, возникающая при отборе аналитической пробы (обычно несколько десятков миллиграммов) из «лабораторной» пробы — раздробленного образца весом 50—200 г.

Таким образом, главное при выделении предлагаемых видов воспроизводимости — не продолжительность охватываемого рядом анализов периода времени, как иногда принимается в методической литературе (например, Налимов, 1960; Ткачев, Юдович, 1975), а степени изменения условий анализа.

В большинстве случаев геолог-разведчик или геохимик продолжает исследование объекта в течение ряда лет, и наиболее употребительной для него является лабораторная воспроизводимость. Однако не следует пренебрегать возможностью обработки отдельных частей своего материала, проанализированного в такой последовательности, которая позволяет использовать при интерпретации другой, более выигрышный вид воспроизводимости, т. е. меньшее значение погрешности анализа. Вытекающая отсюда рекомендация — сохранять подробный «аналитический адрес» каждого полученного результата для последующего «выжимания» из анализов максимально значимых статистических выводов. Используя этот же принцип, легко наметить виды воспроизводимости и для других способов анализа вещества.

В связи с изложенным можно сделать интересное замечание по поводу полезности уменьшения случайной погрешности анализа: с точки зрения потребителя важна не сама по себе ее малость, а то, насколько длинные серии проб можно анализировать с определенным малым уровнем погрешности. Учитывая, что среднее из n измерений имеет случайную погрешность σ_B/\sqrt{n} , «потребительскую» ценность методики можно охарактеризовать величиной \sqrt{n}/σ_B , где n — предельная длина серии проб, которую можно проанализировать с погрешностью σ_B . Например, пусть имеется два метода анализа — первый, позволяющий анализировать серии по 1000 проб со средней квадратической относительной погрешностью 10 %, и второй, позволяющий анализировать серии по 100 проб с погрешностью 5 %. Тогда ценность первого ($\sqrt{1000}/10 \approx 3.16$) выше ценности второго ($\sqrt{100}/5 = 2$) в 1.58 раза, и грамотный геолог, оперирующий большими массивами данных, должен выбрать первый, используя второй для решения более узких задач, требующих незначительного числа анализов. Изложенное достаточно убедительно свидетельствует о том, что геолог для аналитика должен быть не пассивным заказчиком, а обязательным участником в организации анализов.

Численная оценка воспроизводимости. Воспроизводимость численно характеризуется средним квадратическим расхождением между результатами анализа одной и той же пробы — σ_B . Несмотря на кажущуюся простоту, ее практический расчет в ряде ситуаций встречает методические трудности. Рассмотрим его для случаев многократного анализа одной пробы и двукратного анализа каждой пробы в двух вариантах — с рандомизацией порядка съемки реплик одной и той же пробы и в варианте, когда повторные анализы составляют самостоятельную серию.

Расчет s_B^2 как оценки σ_B^2 по данным многократного анализа одной пробы наиболее прост, и формулы для него хорошо известны (Налимов, 1960; Дубов и др., 1970; Дубов, 1972; Ткачев, Юдович, 1975):

$$s_B = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (5.2.1)$$

где n — число повторностей, x_i — результаты анализов. Недостатком этого метода расчета s_B является сомнительная представительность полученной величины, поскольку выбранная для многократного анализа проба может не вполне соответствовать остальным пробам анализируемой совокупности по минеральному составу, содержанию изучаемого компонента и тщательности подготовки. Кроме того, этот метод непригоден, если воспроизводимость зависит от концентрации.

Более рационально пользоваться двукратными анализами x_{i1} и x_{i2} всех или большей части проб изучаемой совокупности, так как s_B зависит в общем случае не только от метода, но и от анализируемых содержаний. Расчет производится по формуле

$$s_B = \sqrt{\frac{1}{2n} \sum_{i=1}^n (x_{i1} - x_{i2})^2}. \quad (5.2.2)$$

Приводя эту известную формулу, мы хотим обратить внимание на то, что она применима лишь в случае рандомизации порядка основных и повторных анализов. Для распространенного случая, когда повторные анализы образуют самостоятельную серию, Р. И. Дубов (Дубов, Фурсов, 1966; Дубов и др., 1970; Дубов, 1972) рекомендует формулу, учитывающую систематическую ошибку Δ одной серии повторных анализов относительно другой:¹

$$s_B = \sqrt{\frac{1}{2(n-1)} \sum (x_{i1} - x_{i2} - \Delta)^2}, \quad \Delta = \left| \frac{1}{n} \sum_{i=1}^n (x_{i1} - x_{i2}) \right|. \quad (5.2.3)$$

Рекомендованный Р. И. Дубовым учет разности в сериях, уменьшающий расчетное значение воспроизводимости по сравнению с формулой (5.1.2), может быть и правильным, и ошибочным в зависимости от обстоятельств. Например, применительно к упомянутой методике спектрального анализа, если первые реплики проб сняты на одной пластинке, вторые — на другой, то формула (5.2.3) будет верной для оценки воспроизводимости «на пластинке», а формула (5.2.2) — для оценки воспроизводимости для пачки. В большинстве случаев геологу важнее вторая, так что безоговорочно применять формулу (5.2.3) не следует.

Р. И. Дубовым (Дубов и др., 1970; Дубов, 1972) обсуждается весьма интересный, но не получивший широкого распространения метод определения s_B , когда первая и вторая серии повторных анализов не равноточны. Он основан на сравнении дисперсии содержаний в одной и той же совокупности проб по результатам анализов первых s_1^2 и вторых s_2^2 реплик проб. Их разность — не что иное, как разность дисперсий погрешностей анализов $s_1^2 - s_2^2 = s_{a1}^2 - s_{a2}^2$, поскольку «природная» дисперсия содержаний для них одинакова. Когда серии повторных анализов неравноточны, формулы (5.2.2) и (5.2.3) фактически дают полусумму аналитических дисперсий основной (первой) и контрольной (второй) серий. Поэтому можно записать два уравнения с двумя неизвестным s_{a1}^2 и s_{a2}^2 :

$$\begin{cases} 1/2 (s_{a1}^2 + s_{a2}^2) = s_B^2 \\ s_1^2 - s_2^2 = s_{a1}^2 - s_{a2}^2 \end{cases} \quad (5.2.4)$$

где s_1^2 и s_2^2 — дисперсии содержаний по результатам анализов. Однако определить s_{a1}^2 и s_{a2}^2 из этой системы уравнений возможно лишь с большими погрешно-

¹ Выявление систематической ошибки метода анализа или метода опробования будет рассмотрено в специальном разделе.

стями ввиду малости дисперсий погрешностей анализов по сравнению с природной дисперсией содержаний. Метод можно рекомендовать лишь для случая, когда одна из повторностей является контрольной и выполнена более точным методом.

В практике нередко встречается расчет воспроизводимости по многократным (m) анализам нескольких (n) проб:

$$s_B^2 = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{m-1} \sum_{j=1}^m (x_{j,i} - \bar{x}_i)^2 \right]. \quad (5.2.5)$$

При этом важно, чтобы выбранные для анализа пробы в нужных пропорциях представляли изучаемую совокупность. Если этого сделать нельзя, мы рекомендуем за оценку воспроизводимости брать средневзвешенное из $s_{B_i}^2$ для отдельных содержаний с весами, пропорциональными встречаемости проб с соответствующими содержаниями (Ткачев, Юдович, 1975).

Расчет доверительных интервалов. Часто в литературе (например, Усиков, 1983) поднимают вопрос, применимы ли приведенные выше методы и формулы для расчета воспроизводимости анализов, если распределение погрешностей отличается от нормального. В частности, Ю. Т. Усиков (1983, с. 23) утверждает, что условием применимости этих формул является нормальность распределения погрешностей и независимость их дисперсии от уровня измеряемой величины.

Ставить такой вопрос неправомерно, так как независимо от распределения дисперсия по определению есть средний квадрат отклонения от среднего. Другое дело — вопрос о том, как и для чего дальше использовать эту дисперсию. Известно, что исчерпывающей характеристикой случайной величины, в том числе и погрешности, является вид функции распределения (закон распределения) и ее параметры. Но такая характеристика в общем громоздка и не наглядна. На практике очень удобна и наглядна такая статистическая характеристика, как доверительный интервал для заданной доверительной вероятности. Для одного метода анализа можно указать сколько угодно доверительных интервалов с соответствующими им доверительными вероятностями. Наиболее распространены одно-, двух- и трехсигмовые интервалы, соответствующие для нормального распределения доверительным вероятностям 68.26, 95.46 и 99.74 %. Употребляются также, особенно в практике аналитических лабораторий, 50-процентные доверительные интервалы. Расчет доверительных интервалов для нормального распределения погрешностей достаточно подробно описан (например, Налимов, 1960) и не встречает затруднений. Для него необходимо знание лишь двух величин — воспроизводимости s_B и числа n данных, по которым она рассчитана.

Большой интерес представляет расчет доверительных интервалов без использования предположения о нормальности распределения погрешности анализа. Если нет никаких предположений о характере распределения погрешностей, то доверительные интервалы a для содержаний с доверительной вероятностью P можно получить из неравенства Чебышева $P\{|x - c| < a\} \geq 1 - s_B^2/a^2$, откуда $a = \pm s_B/\sqrt{1-P}$. Эти доверительные интервалы вряд ли представляют интерес, так как дают слишком большой «запас прочности». Н. А. Макулов и др. (1974) указывают, что если распределение погрешности симметрично и имеет один максимум, то имеет место более сильное неравенство, дающее в 1.5 раза более узкий доверительный интервал, чем чебышевское. Для распределений, медиана которых совпадает с математическим ожиданием $M(x)$, т. е. удовлетворяющих более слабому условию, чем условие симметричности, Н. А. Макулов рекомендует следующий простой непараметрический способ характеристики доверительных интервалов. Так как $P\{x_{\min} > M(x)\} = 1/2^n = P\{x_{\max} < M(x)\} = 1/2^n$, то $M(x)$ попадает в интервал $[x_{\min}, x_{\max}]$ с доверительной вероятностью $P^* = 1 - 2P = 1 - 1/2^{n-1}$.

Сравнение длины этих интервалов с интервалами для нормального распределения показало, что первые шире всего лишь в 1.18 раза при $n=2$ и в 1.67 раза при $n=15$ (n — число повторностей анализа). Следовательно, можно рекомендовать этот непараметрический способ даже для нормальных распределений, так как его применение не требует расчета воспроизводимости. Правда, при таком способе не получают доверительный интервал для заданной вероятности, как принято, а рассчитывают доверительную вероятность для некоторого интервала, равного размаху R результатов n повторных анализов. Большим неудобством является также то, что в совокупности, например дважды проанализированных проб, доверительные интервалы и доверительные вероятности для истинных содержаний будут различны для каждой пробы и привести их к «общему знаменателю» заданной доверительной вероятности невозможно.

Вторая часть обсуждаемого вопроса — справедливы ли приведенные формулы оценки воспроизводимости и методы построения по ним доверительных интервалов, если воспроизводимость сама зависит от анализируемых содержаний. В этом случае речь может идти: 1) либо о воспроизводимости для данного узкого класса содержаний, 2) либо о средней воспроизводимости метода для совокупности анализируемых проб с заданными частотами содержаний. В первом случае изложенное справедливо для заданного класса содержаний, во втором — доверительные интервалы будут верными в среднем для партии проб, но для отдельных классов содержаний в эти средние доверительные интервалы будет попадать число результатов, не соответствующее доверительной вероятности.

§ 3. Характеристика воспроизводимости анализа, когда она зависит от содержания

Вековая практика измерений послужила эмпирической основой для классической теории измерений. Она утверждает, что при правильно организованном измерении, когда результат формируется под воздействием большого числа случайных факторов и ни один из них не выделяется решающим образом среди других, распределение погрешностей будет подчиняться нормальному закону. Однако для некоторых видов анализа распределение погрешностей существенно отличается от нормального. Из литературы известно (Лончих, 1965; Дубов, 1972; Ткачев, Юдович, 1975; Иванова, 1976; Ткачев, 1981, и др.), что распределение погрешностей количественного спектрального анализа является логарифмически нормальным. Абсолютная случайная средняя погрешность анализа при этом увеличивается прямо пропорционально содержанию, а относительная остается постоянной. Ю. Т. Усиков (1983) утверждает, что «в практике геолого-разведочных работ мы имеем дело с совокупностью промежуточных случаев, когда и абсолютные, и относительные погрешности могут зависеть от уровня измеряемых свойств» (с. 18). Но автор не приводит ни теоретических соображений, что так должно быть, ни конкретных примеров, где это наблюдается в действительности (в его табл. 1 и 2 не указан документальный источник данных, а сами данные, как выясняется из рисунков, характеризуются логнормальной (Усиков, 1983, рис. 5) и нормальной (там же, рис. 6) моделями.

Попытаемся выяснить, чем вызвано отклонение распределения погрешностей анализа от нормального и не является ли это отклонение причиной изменения средней погрешности с измеряемой концентрацией. Модель измерения или анализа и порождаемую ею модель формирования погрешностей мы предлагаем обобщенно изобразить в виде следующей схемы (табл. 5.3.1). Содержание этой схемы заключается в том, что в процессе анализа измеряемая величина возбуждает некоторый физический сигнал, который преобразуется в некоторый другой сигнал и т. д., пока не будет получен сигнал (или сигналы), который подвергается непосредственному измерению. Длина цепочки преобра-

Таблица 5.3.1

	Модель анализа	Модель формирования погрешностей анализа		Пример модели количественного спектрального анализа	Пример модели формирования погрешностей количественного спектрального анализа		
x	Измеряемая величина	Искажение измеряемой величины в измеряемом объекте	$x + \Delta x$	Содержание в пробе	Погрешность подготовки навески	$c + \Delta c$	
$f(x)$	Возбуждение физического сигнала	Искажение физического сигнала	$f + \Delta f$	$I = Kc$	Оптическое излучение химического элемента в плазме дуги (искры)	$I + \Delta I$	
$A(f(x))$	Преобразование физического сигнала в аналитический	Искажение аналитического сигнала	$A + \Delta A$		$S = K \lg I$	Экспонирование фотоматериала	Отражение, рассеивание, излучения в фотоэмульсии, неоднородность и другие дефекты фотоматериала
$A(f(x))$	Фиксация и (или) измерение аналитического сигнала	Погрешность непосредственного измерения аналитического сигнала	$A + \Delta A + \Delta' A$	Проявление фотоматериала, фотометрирование спектральных линий		Нестандартность и неоднородность условий проявления, ошибки фотометрирования	$S + \Delta S$
$x = F(A)$	Расчет численного значения измеряемой величины	Погрешности расчета измеряемой величины	$x + \Delta x$	Расчет градуировочного графика, расчет содержаний		Погрешности расчета	

зований сигнала может быть различной. Каждое преобразование происходит по определенному физическому закону, которому соответствует определенная математическая формула. Поэтому в дальнейшем мы будем говорить о преобразовании математических величин, соответствующих физическим сигналам. Погрешность возникает от искажения сигналов на всех стадиях анализа. Пусть измеряемая величина x сначала преобразуется в величину $f(x)$, затем в величину $\varphi[f(x)]$, которая непосредственно измеряется и содержит погрешность всех стадий анализа $\varphi[f(x+\Delta x) + \Delta f] + \Delta\varphi$. Наш постулат заключается в том, что при правильно организованном анализе все случайные величины (Δx , Δf , \dots , $\Delta\varphi$) распределены по нормальному закону с дисперсиями $\sigma_{\Delta x}^2$, $\sigma_{\Delta f}^2$, \dots , $\sigma_{\Delta\varphi}^2$ и нулевыми средними. В соответствии с правилом переноса ошибок (см., например, Худсон, 1967) погрешность оценки измеряемой величины $\Delta = \Delta x + \Delta f/f'(x) + \Delta\varphi/\varphi'f'f'x$. В простейшем случае никаких преобразований с измеряемой величиной не происходит, измеряется непосредственно она, поэтому $\Delta = \Delta x$. В случае линейных преобразований сигнала в процессе анализа (а большинство методов анализа организовано именно так) имеем $\Delta = \Delta x + k_1\Delta f + \dots + k_2\Delta\varphi$, т. е. погрешность измерения будет распределена нормально с некоторой дисперсией $\sigma_{\Delta}^2 = \sigma_{\Delta x}^2 + k_1^2\sigma_{\Delta f}^2 + \dots + k_2^2\sigma_{\Delta\varphi}^2$, где k — коэффициенты линейных преобразований сигнала. Этот случай также соответствует классической теории ошибок измерений.

В уже упоминавшемся количественном спектральном анализе первая стадия анализа заключается в подготовке пробы. Изучаемая величина c — концентрация химического элемента — преобразованию не подвергается, но может искажаться на величину Δc (погрешность сокращения пробы, неточность взятия навески для анализа, разбавления ее буферной смесью и т. д.). На второй стадии анализа возбуждается оптическое излучение атомов элемента в дуге, интенсивность которого прямо пропорциональна концентрации $I = kc$. Интенсивность искажается на величину ΔI , а содержание — на величину $c = I/k\Delta I$. На третьей стадии анализа излучение фиксируется на фотоэмульсии и измеряется ее почернение, пропорциональное логарифму интенсивности $S = \log I$. Почернение в процессе фиксации и измерения искажается на величину ΔS . В итоге будем иметь

$$\Delta = \Delta c + \frac{1}{k} \Delta I + \frac{\Delta S}{\lg'(I) \cdot I'(c)} = \Delta c + \frac{1}{k} \Delta I + \frac{1}{k} \Delta S \cdot c.$$

Опуская постоянные коэффициенты, зависящие от выбора единиц измерения, можно записать

$$\Delta = \Delta c + \Delta(\log c) \cdot c. \quad (5.3.1)$$

Таким образом, общая погрешность измерения концентрации в данном анализе содержит две компоненты. Одна из них распределена нормально, другая — логнормально. Вторая компонента прямо пропорциональна концентрации. Распределение общей случайной погрешности будет композицией нормального и логнормального распределения. Степень его близости к нормальному или, напротив, к логнормальному зависит от соотношения искажений, вносимых на «нормальной» (до логарифмического преобразования сигнала) и «логнормальной» стадиях анализа. Так как первое нормально распределенное слагаемое в (5.2.6) не зависит от c , а второе — прямо пропорционально ему, то с увеличением содержания при прочих равных условиях распределение погрешностей будет приобретать более «логнормальный» облик.

Практические данные по успешной аппроксимации погрешностей количественного спектрального анализа логнормальным распределением указывают на то, что основные погрешности вносятся на логнормальной стадии анализа. В дальнейшем мы будем иметь в виду именно этот случай.

Таким образом, теоретическое рассмотрение показало, что если в процессе анализа или измерения имеют место только линейные преобразования сигнала,

распределение погрешностей будет нормальным с параметрами, не зависящими от уровня (значения) измеряемой величины. Такова большая часть «классических» анализов (например, химический весовой). Если имеют место и линейные, и нелинейные преобразования, то распределение погрешностей будет композицией нормального и отличного от нормального, но всегда можно подобрать функцию, нормализующую последнее распределение. Эта функция, как правило, известна из физической сущности анализа. Эти результаты позволяют подбирать функции для аппроксимации распределения погрешностей исходя из содержательного анализа, что особенно важно в косвенных геофизических методах измерения подсчетных параметров нефтегазоносных пород в скважинах.

Полученные выводы не имеют ничего общего с предположением Ю. Т. Усикова (1983) о всеобщей применимости для аппроксимации распределения погрешностей измерений трехпараметрического логнормального распределения, хотя нельзя отрицать возможность хорошей аппроксимации им во многих практических случаях.

§ 4. Способы выражения и расчета воспроизводимости при логнормальном распределении погрешностей

Расчет и использование воспроизводимости, описанные выше, полезны лишь тогда, когда погрешность существенно не зависит от содержания. В противном случае при ее расчете и дальнейшей статистической обработке данных возникают дополнительные трудности. На практике распространены следующие соотношения между анализируемой концентрацией и воспроизводимостью метода анализа (табл. 5.4.1). Прежде всего следует отметить, что определение вида распределения погрешностей анализа по многократным анализам одной пробы в рассматриваемом случае ненадежно и предпочтение должно быть отдано двукратным анализам проб в широком диапазоне концентраций (Дубов, 1972; Иванова, 1976).

Таблица 5.4.1

Соотношения между анализируемой концентрацией и воспроизводимостью метода анализа

Зависимость среднеквадратической случайной погрешности анализа от концентрации	Распределение погрешностей анализа	
	нормальное	логнормальное
Не зависит от концентрации	Весовой химанализ (Налимов, 1960), рентгеноспектральный, квантометрический (Ткачев, Юдович, 1975)	В § 3 показано, что таких случаев не может быть
Прямо пропорциональна концентрации	В § 3 показано, что таких случаев не может быть	Спектральный анализ (Гавришин, 1980; Налимов, 1960, и др.; Дубов и др., 1970; Дубов, 1972), колориметрический, полярографический (Дубов, Фурсов, 1966)

Заменив содержания их логарифмами, произведем расчет воспроизводимости одним из изложенных выше способов. Полученную величину называют логарифмической погрешностью воспроизводимости S_{lg} . Как записать теперь доверительные интервалы для истинного содержания? Построим доверительные интервалы для логарифмов содержаний обычным известным в математической статистике способом:

$$\lg c = \lg x \pm t S_{lg} / \sqrt{n}, \quad (5.4.1)$$

где $\lg x$ — среднее арифметическое логарифма содержания из n параллельных определений; S_{\lg} — оценка логарифмической погрешности воспроизводимости σ_{\lg} ; t — коэффициент, определяемый для заданной доверительной вероятности P и числа степеней свободы f по таблицам распределения Стьюдента. Если S_{\lg} подсчитывалась по формуле (5.2.2), то $f=k$, где k — число дважды проанализированных образцов, если же S_{\lg} подсчитывалась по формуле (5.2.1), то $f=m-1$, где m — кратность анализа пробы.

Величина S_{\lg} и построенные по ней доверительные интервалы позволяют аналитику проводить метрологический контроль работы аналитической лаборатории, а геологу — решать по результатам анализов свои задачи. Строго говоря, нет необходимости для характеристики воспроизводимости переходить от логарифмов обратно к содержаниям. Однако для геологов-разведчиков, геохимиков такое представление результатов непривычно, ненаглядно и не вписывается в укоренившиеся традиции. Возникает необходимость уметь быстро переходить от логарифмической воспроизводимости и логарифмических доверительных интервалов к соответствующим характеристикам процентных концентраций.

Поэтому результат с учетом доверительных интервалов в некоторых работах предлагается записывать $\bar{x} \cdot 10^{\pm S_{\lg}}$ (Дубов и др., 1970; Ткачев, 1979), где $\bar{x} = 10^{\lg x}$. Среди геологов большой популярностью как мера изменчивости пользуется коэффициент вариации. В аналитической практике аналогом коэффициента вариации служит относительная средняя квадратическая погрешность, выраженная в процентах, — коэффициент вариации воспроизводимости V_a . При нормальном распределении численное его значение равно половине 68 %-го доверительного интервала, деленной на содержание. Мы предлагаем использовать это правило для вычисления коэффициента вариации воспроизводимости при любых других распределениях, т. е. коэффициентом вариации воспроизводимости считать половину 68 %-го доверительного интервала. При логнормальном распределении погрешностей длину этого интервала вычислим с помощью S_{\lg} . В большую сторону она равна $x \cdot 10^{S_{\lg}} - x = x(10^{S_{\lg}} - 1)$, в меньшую $x - x \cdot 10^{-S_{\lg}} = x(1 - 1/10^{S_{\lg}}) = x(10^{S_{\lg}} - 1/10^{S_{\lg}})$. Половина их суммы равна

$$\frac{1}{2} \left[x(10^{S_{\lg}} - 1) + x \left(\frac{10^{S_{\lg}} - 1}{10^{S_{\lg}}} \right) \right] = \frac{1}{2} x(10^{S_{\lg}} - 1) \left(1 + \frac{1}{10^{S_{\lg}}} \right).$$

Отсюда коэффициент вариации воспроизводимости V_a выражается через логарифмическую погрешность воспроизводимости следующим образом:

$$V_a = 100/2 (10^{S_{\lg}} - 1) (1 + 1/10^{S_{\lg}}). \quad (5.4.2)$$

Для малых погрешностей $10^{S_{\lg}}$ есть число положительное, не очень отличающееся от единицы, поэтому с некоторым «запасом» можно записать (Ткачев, 1979):

$$V_a = (10^{S_{\lg}} - 1) \cdot 100 \%, \text{ или } V_a = (e^{S_{\ln}} - 1) \cdot 100 \%. \quad (5.4.3)$$

При логнормальном распределении доверительный интервал не симметричен относительно измеренного значения и, делая упрощение, приводящее к (5.4.3), мы за требующуюся нам половину доверительного интервала берем половину доверительного интервала в большую от среднего значения сторону. Для среднего из n -кратных анализов коэффициент вариации

$$V_{a,n} = (10^{S_{\lg}/\sqrt{n}} - 1) \cdot 100 \%. \quad (5.4.4)$$

Коэффициенты вариации воспроизводимости V_a при логарифмически-нормальном распределении погрешностей анализа

Формулы для расчета	Значения V_a (в %) для различных значений логарифмической воспроизводимости S_{lg}							
	$S_{lg}=0.01$	$S_{lg}=0.02$	$S_{lg}=0.03$	$S_{lg}=0.05$	$S_{lg}=0.1$	$S_{lg}=0.2$	$S_{lg}=0.3$	$S_{lg}=0.4$
$V_a=100/2(10^{S_{lg}}-1)(1+1/10^{S_{lg}})$	2.3	4.6	6.9	11.5	23.2	47.7	74.7	105.7
Коэффициент вариации среднего из n анализа по формуле $V_{a,n}=100/2(10^{S_{lg}/\sqrt{n}}-1)(1+1/10^{S_{lg}/\sqrt{n}})$	1.63	3.26	4.9	8.15	16.35	33.1	50.8	69.8
То же, по формулам: $V_{a,n}=V_a/\sqrt{n}$ $V_a=(10^{S_{lg}}-1) \cdot 100$ $V_a=\sqrt{10^{2.3S_{lg}}-1} \cdot 100$	1.63 2.3 2.3	3.25 4.7 4.6	4.9 7.1 6.9	8.13 12.2 11.5	16.4 25.8 23.3	33.7 58.4 48.6	52.8 99.5 78.1	74.7 151.2 115.5
Приближенная формула Б. Я. Юфы (1979): $V_a=2.5S_{lg}$	2.5	5.0	7.5	12.5	25.0	50.0	75.0	100.0

Применяя правило « \sqrt{n} », т. е. рассчитывая V_a по формуле $V_{a,n}=V_a/\sqrt{n}$ (табл. 5.4.2), мы получаем практически те же самые результаты. Некоторые авторы (Юфа, 1979) полагают, что поскольку погрешность анализа распределена логарифмически-нормально, это автоматически обеспечивает применение формул, связывающих параметры распределения случайной величины с параметрами распределения ее логарифма: $S_x^2=\bar{x}^2(\exp S_{lg}^2-1)$, откуда

$$V_x=\sqrt{e^{S_{ln}^2}-1} \cdot 100=\sqrt{e^{(5.3S_{lg}^2)}-1} \cdot 100=\sqrt{10^{(2.3S_{lg}^2)}-1} \cdot 100 \% \quad (5.4.5)$$

Отождествление V_x с V_a , т. е. расчет коэффициента воспроизводимости по формуле (5.4.5), строго говоря, неверен. Однако, как показывает табл. 5.4.2, отличия в результатах практически незначительны и появляются только для очень больших значений V_a , превышающих 100 %.

Рассмотрим теперь не столько сложный, сколько коварный вопрос о соотношении параметров распределения (среднего значения и дисперсии) случайной величины (на примере содержания) и параметров распределения ее логарифма при логарифмически-нормальном распределении погрешности. Формулы, определяющие эти соотношения, хорошо известны из математической статистики: если величина x распределена логнормально, то справедливы следующие соотношения:

$$\mu_x=\exp(\mu_{ln}+1/2\sigma_{ln}^2), \quad (5.4.6)$$

$$\sigma_x^2=\mu_x^2(\exp\sigma_{ln}^2-1), \quad (5.4.7)$$

$$V_x=\sqrt{e^{\sigma_{ln}^2}-1} \cdot 100 \% ,$$

где μ_{ln} и σ_{ln}^2 — параметры (среднее и дисперсия) нормального распределения натуральных логарифмов содержаний. При использовании десятичных логарифмов эти равенства следующие:

$$\mu_x=\exp\left(\frac{\mu_{lg}}{0.4343}+\frac{1}{2}\sigma_{lg}^2\frac{1}{0.4343}\right), \quad (5.4.8)$$

$$\sigma_x^2=\mu_x^2[\exp(\sigma_{lg}^2/0.4343^2)-1], \quad (5.4.9)$$

$$V_x=\sqrt{\exp(\sigma_{lg}^2/0.4343^2)-1} \cdot 100=\sqrt{e^{5.3\sigma_{lg}^2}-1} \cdot 100=\sqrt{10^{2.3\sigma_{lg}^2}-1} \cdot 100 \% \quad (5.4.10)$$

Корректное использование этих формул при интерпретации данных требует содержательного анализа ситуации. Интерес для анализа представляют три основных случая:

1) изучаемой совокупностью является множество проб, в которых содержание подчиняется логнормальному распределению; 2) изучаемой совокупностью являются результаты анализа одной и той же пробы с погрешностью, распределенной по логарифмически нормальному закону; 3) изучается совокупность проб с некоторым распределением с помощью анализов, погрешность которых распределена логнормально или имеет логнормально распределенную составляющую.

В первом случае оценка среднего и дисперсии содержания может быть произведена либо по «обычным» формулам оценки математического ожидания и дисперсии $\bar{x} = 1/n \sum x_i$, $S_x^2 = 1/(n-1) \sum (x_i - \bar{x})^2$ либо по формулам (5.4.6). Подробный анализ того, какой способ предпочесть, дан в нашей более ранней работе (Ткачев, Юдович, 1975).

Во втором случае для расчета среднего формулы первого случая неприменимы (хотя наблюдаемые результаты имеют такое же распределение!). Правильной оценкой среднего здесь является медиана, которую при логнормальном распределении оценивают по формуле

$$Me = \bar{x} = \exp(\overline{\ln x}) = 10^{\overline{\lg x}}, \quad (5.4.11)$$

т. е. среднее значение результатов анализов следует рассчитывать как среднее геометрическое. Из сопоставления (5.4.11) и (5.4.6) видно, что применение ко второму рассматриваемому случаю формул, верных для первого случая, привело бы к завышению результата в $1/2 S_{\ln}^2$ раза. Расчет дисперсии для второго случая был рассмотрен выше.

Третий случай требует более подробного рассмотрения. Начнем его с упоминания одного геохимического феномена, заключающегося в том, что чем выше погрешность применяемого анализа, тем большими получаются кларки элементов, хотя сами анализы, как показывает внешний контроль, систематического смещения не дают. Эффект, следовательно, появляется на стадии статистической обработки.

Выше было показано, что среднее значение подсчитывается по-разному в зависимости от того, что из себя представляет обрабатываемая совокупность результатов: если это истинные (без погрешностей) содержания в пробах — то как среднее арифметическое, если это результаты анализов одной и той же пробы с логнормальными погрешностями, — то как среднее геометрическое. Практическая ситуация всегда является средним из этих крайних случаев: истинные содержания в пробах отягощены погрешностями. Здесь ни среднее арифметическое, ни среднее геометрическое не являются верными оценками. Первую известную в геолого-геохимической литературе попытку решить этот вопрос для частного случая — логнормального распределения содержаний в совокупности изучаемых проб, отягощенных логнормальными же погрешностями, сделал Р. И. Дубов (Дубов, Фурсов, 1966). Он предлагает использовать для этого формулу (5.4.8), в которую на место σ_{\lg}^2 необходимо подставить природную дисперсию логарифмов содержаний, которая получается как разность между общей наблюдаемой дисперсией логарифмов содержаний $\sigma_{\lg_0}^2$ и логарифмической дисперсией воспроизводимости $\sigma_{\lg_s}^2$:

$$\sigma_{\lg}^2 = \sigma_{\lg_0}^2 - \sigma_{\lg_s}^2. \quad (5.4.12)$$

В результате будет получена несмещенная оценка содержания в совокупности проб, которая, как правило, ниже, чем употребляемое обычно среднее арифметическое. Не ясно, подходит ли этот способ для произвольного распределения признака в совокупности анализируемых проб. В работе Т. И. Ивановой (1976)

сделана попытка решить этот вопрос по-другому. Развивая дальше этот метод, нам удалось получить полное решение вопроса.

Пусть c_i — истинные содержания в пробах, тогда истинное среднее арифметическое по совокупности проб равно $c = 1/n \sum c_i$. Пусть далее x_i — результат анализа i -й пробы с погрешностями, распределенными логнормально. Введем величины $K_1 = c_i/x_i$, если $c_i > x_i$, и $K_2 = x_i/c_i$, если $c_i < x_i$; $\lg K_{1i} = \lg c_i - \lg x_i$, $\lg K_{2i} = \lg x_i - \lg c_i$. При логнормальном распределении погрешностей, в силу того что логарифмическая погрешность не зависит от содержания, величины $\lg K_{1i}$ и $\lg K_{2i}$ не зависят от c_i и образуют две половины нормального распределения с нулевым средним и дисперсией σ_{\lg}^2 , причем распределение величины K_1 усечено при $c = 0$ справа, а величины K_2 — слева. Тогда

$$M(x) = M \left\{ \left(\frac{1}{n} \sum_{i=1}^n K_{1i} c_i I_1 + \frac{1}{n} \sum_{i=1}^n K_{2i} c_i I_2 \right) \right\}, \quad (5.4.13)$$

где I_1 и I_2 — индикаторы множества

$$I_1 = I(c_i > x_i) = \begin{cases} 1, & \text{если } c_i > x_i \\ 0, & \text{» } c_i < x_i, \end{cases}$$

$$I_2 = I(c_i < x_i) = \begin{cases} 0, & \text{если } c_i > x_i \\ 1, & \text{» } c_i < x_i. \end{cases}$$

Далее $M(x) = M\{1/n \sum K_{1i} I_{1i} c_i\} + M\{1/n \sum K_{2i} I_{2i} c_i\} = 1/n M\{\sum K_{1i} I_{1i} c_i\} + 1/n M\{\sum K_{2i} I_{2i} c_i\} = M\{(K_1 I_1) c\} + M\{(K_2 I_2) c\} = M(K_1 I_1) \cdot M(c) + \text{cov}_{(K_1 I_1), c} + M(K_2 I_2) \cdot M(c) + \text{cov}_{(K_2 I_2), c}$. В силу попарной независимости I от K и от c в рамках соответствующих подмножеств имеем: $\text{cov}_{(K_1 I_1), c} = 0$, $\text{cov}_{(K_2 I_2), c} = 0$. Тогда $M(x) = M(K_1 I_1) \cdot M(c) + M(K_2 I_2) \cdot M(c) = M(K_1 I_1) \cdot c + M(K_2 I_2) \cdot c = c \cdot [M(K_1 I_1) + M(K_2 I_2)]$. В силу независимости K от I в рамках своих подмножеств (т. е. там, где определены K_1 и K_2) имеем $M(x) = c [M(K_1) \geq M(K_2) \cdot M(I_2)]$, но $M(I_1) = M(I_2) = P(c_i > x_i) = P(c_i < x_i) = 0.5$ (в силу симметричности распределения логарифмов погрешностей). Таким образом,

$$M(x) = c \cdot 1/2 [M(K_1) + M(K_2)] = c \delta_3, \quad (5.4.14)$$

где $\delta_3 = 1/2 [M(K_1) + M(K_2)]$. Величины $M(K_1)$ и $M(K_2)$ легко табулировать как математическое ожидание усеченных логнормально распределенных случайных величин:

$$M(K_1) = \frac{1}{\sqrt{2\pi} \sigma_{\lg}^2} \int_{-\infty}^1 e^{-\frac{1}{2} \frac{(\lg K)^2}{\sigma_{\lg}^2}} dK / \frac{1}{\sqrt{2\pi} \sigma_{\lg}^2} \int_{-\infty}^1 \frac{1}{K} e^{-\frac{1}{2} \frac{(\lg K)^2}{\sigma_{\lg}^2}} dK;$$

$$M(K_2) = \frac{1}{\sqrt{2\pi} \sigma_{\lg}^2} \int_1^{\infty} e^{-\frac{1}{2} \frac{(\lg K)^2}{\sigma_{\lg}^2}} dK / \frac{1}{\sqrt{2\pi} \sigma_{\lg}^2} \int_1^{\infty} \frac{1}{K} e^{-\frac{1}{2} \frac{(\lg K)^2}{\sigma_{\lg}^2}} dK. \quad (5.4.15)$$

Практически можно ограничиться пределом интегрирования в первом случае от $-3\sigma_{\lg}$, во втором — до $+3\sigma_{\lg}$. Результаты расчетов приведены в табл. 5.4.3. Проще, величина $1/2 [M(K_1) + M(K_2)]$ является математическим ожиданием логнормально распределенной случайной величины, имеющей среднее, равное единице, и дисперсию σ_{\ln}^2 :

$$M(\delta_3) = \exp(0 + 1/2 \sigma_{\ln}^2), \quad (5.4.16)$$

Величина коэффициента завышения δ среднего арифметического в зависимости от погрешностей анализа в условиях логнормального их распределения

σ_{lg}	$V_a, \%$	δ	σ_{lg}	$V_a, \%$	δ
0.01	2.3	1.00025	0.1	23	1.026
0.02	4.6	170010	0.2	48	1711
0.03	6.9	1.0023	0.3	75	1.26
0.04	10	1.0041	0.4	106	1.52
0.05	12	1.0065	0.5	142	1.87
0.06	14	1.0093	0.6	186	2.45
0.07	16	17013	0.7	241	3.23
0.08	19	1.017	0.8	308	4.11
0.09	21	1.021	0.9	391	4.71

Примечание. V_a рассчитывался по формуле (5.4.2).

или в десятичных логарифмах

$$M(\delta_3) = \exp\left(\frac{1}{2} \frac{\sigma_{lg}^2}{(0.4343)^2}\right) = \exp\left(\frac{\sigma_{lg}^2}{0.377}\right), \quad (5.4.17)$$

что дает результаты, аналогичные приведенным в таблице.

Из таблицы видно, что для лучших методик количественного спектрального анализа ($v_a \leq 10\%$) завышением среднего арифметического можно пренебречь: оно составляет не более 0.5 отн $\%$. В области реальных погрешностей рядовых лабораторий ($10\% < v_a \leq 25\%$) завышение, доходящее до 3 $\%$, может учитываться в особо важных случаях. Для полуколичественных анализов завышение резко возрастает. Так, при хорошем полуколичественном спектральном анализе, дающем 5—6 баллов на порядок концентрации ($v_a \approx 150\%$), завышение среднего арифметического составляет 1.9 раза, при 4 баллах на порядок ($v_a \approx 200\%$) — 3.0 раза.

Изучение сильно изменчивых объектов, каковым являются геологические образования, экономически более выгодно проводить экспрессными, не отличающимися высокой точностью методами, добиваясь необходимой достоверности увеличением числа проб. В этих условиях завышение, достигающее полутора-трехкратной величины, автоматически переходит на вычисляемые средние значения и следующие из них геологические выводы, поэтому учет предлагаемой поправки здесь необходим. Эти выводы применимы также к случаю, когда погрешность анализа содержит нормально и логнормально распределенные компоненты. При расчете завышения результатов на место σ_{lg}^2 надо поставить логнормальное слагаемое дисперсии анализа. В самом общем случае необходимо следующее.

1. Из физического содержания метода анализа выяснить, имеет ли место нелинейное преобразование измеряемого сигнала. Если да, то определить, какой функции $f(x)$ оно соответствует.

2. Рассмотреть нормальное распределение функции $f(x)$ измеряемой величины и установить зависимость между медианой и математическим ожиданием M_0 этого распределения: $M_e = f(M_0, \sigma_f)$, где σ_f — средняя квадратическая погрешность, вносимая на «нелинейной стадии» анализа. Ее оценка — задача не простая, но важная для метрологической характеристики метода анализа.

3. Установить и вводить поправку в средние из результатов анализов совокупностей проб, равную разности $M_0 - M_e$ (или M_0/M_e) как функции от σ_f .

4. Доверительные интервалы рассчитывать по функции, являющейся композицией нормального и f -нормального распределений.

Изложенная процедура не проста и доступна лишь совместным усилиям геолога или аналитика, знающего методику измерений, и профессионального статистика.

§ 5. Внешний контроль анализов. Выявление систематических погрешностей

Под внешним контролем понимается повторный анализ (обычно более точным методом) некоторой части проб, проанализированных основным методом. Как правило, он проводится другими исполнителями в другой лаборатории. Задача внешнего контроля — выявление систематических ошибок с целью внесения поправок в результаты основного метода. В ГОСТ 16 263-70 установлены следующие определения. Систематическая погрешность измерения — составляющая погрешности измерений, остающаяся постоянной или закономерно изменяющаяся при повторных измерениях одной и той же величины. Поправка — значение величины, одноименной с измеряемой, прибавляемой к полученному при измерении значению с целью исключения систематической погрешности. Поправочный множитель — число, на которое умножается результат измерения с целью исключения систематической погрешности. В геологии и геохимии широко распространены также синонимы — постоянная и пропорциональная поправка.

Таким образом, уже по характеру определений ясно, что изучение зависимости между основными и контрольными анализами ограничивается линейным случаем, и этого вполне достаточно для геологоразведочной практики. Проблема внешнего контроля в геологии возникла при подсчете запасов и оценке месторождений. Диапазон интересующих геолога содержаний обычно был невелик, поэтому ранние работы были направлены на выявление постоянной составляющей. В последующем положение существенно изменилось.

Обзор проблемы был дан в нашей более ранней работе (Ткачев, Юдович, 1975). В настоящем изложении мы дополняем его появившимися новыми результатами. В 1950—1960-х годах среди геологоразведчиков оживленно обсуждалась проблема математической обработки данных внешнего контроля (Юфа, 1951, 1954, 1958; Прокофьев, 1955, 1962; Смирнов и др., 1960; Раевский, Шурубор, 1958; Шашкин, 1956; Шараров, 1954; Королев, 1969; Кельин, Михайлович, 1969; Родионов, 1964, и др.).

В свое время Н. В. Барышев (1948) предложил использовать известную в математической статистике формулу критерия t (Стьюдента) для оценки существенности разности средних двух выборок с попарно сопряженными нормально распределенными случайными величинами:

$$t = \frac{\bar{x} - \bar{y}}{s\Delta} = \frac{\bar{x} - \bar{y}}{s_x^2 + s_y^2 - 2s_x s_y \cdot r_{xy}}, \quad (5.5.1)$$

где \bar{x} — среднее из основных анализов n проб; \bar{y} — то же, из контрольных; s_x^2 , s_y^2 — дисперсии содержаний в контролируемой совокупности проб соответственно по основным и контрольным анализам; r_{xy} — коэффициент корреляции между рядами основных и контрольных анализов. Н. В. Барышев предлагал систематическую ошибку считать незначимой, если $t < 2$. Правильно было бы сравнивать значение t с табличным значением $t_{\alpha, f}$ для задаваемого α -процентного уровня значимости и $f = n - 1$ степеней свободы и делать такое заключение при $t < t_{\alpha, f}$. Этот метод корректен, если основные и контрольные анализы могут отличаться только на постоянную величину, а это выполняется обычно для узких интервалов концентраций.

Значительная часть более поздних методов либо аналогична указанному, либо хуже. Так, Д. А. Родионов (1964) предложил определять наличие систематической ошибки по критерию $t = \bar{\Delta} / S_{\bar{\Delta}}$, где $\bar{\Delta} = \frac{1}{n} \sum \Delta_i$, $\Delta_i = x_i - y_i$. Тождество с (5.5.1) очевидное.

И. П. Шараров (1954) предложил вычислять «более простую» статистику

$$k = \frac{\sum (y_i - x_i)}{\sum |y_i - x_i|}, \quad (5.5.2)$$

в принятых обозначениях приобретающую вид $k = \bar{\Delta} / |\bar{\Delta}|$, где $|\bar{\Delta}|$ — среднее арифметическое абсолютных отклонений. Систематическую ошибку предлагалось считать существенной, если $|k| > 0.6$. Позже В. П. Королев (1969) обратил внимание на то, что при нормальном распределении разностей выполняется известное статистическое равенство $s_{\Delta} = 1.25 |\bar{\Delta}|$, и отождествил тем самым критерий Шаропова с критерием t путем замены $k = 1.25$. Следует заметить, что эффективность оценки s_{Δ} посредством $|\bar{\Delta}|$ ниже, чем по формуле $s_{\Delta} = \sqrt{\frac{1}{n} \sum (x_i - y_i)^2}$, и применение критерия k ведет к потере надежности статистического вывода.

Некоторые другие авторы, в основном сотрудники Государственной комиссии по запасам (ГКЗ), высказывались против статистических расчетов при выявлении систематической ошибки: «В подавляющем большинстве случаев при наличии систематической погрешности. . . видно преобладание одних знаков разностей, . . . и подтверждать ее какими-либо расчетами нет необходимости» (Прокофьев, 1955, с. 28). Мерой систематической ошибки эти авторы (Прокофьев, 1955, 1962, и др.) считают среднее арифметическое отклонение $\bar{\Delta}$, мерой случайной ошибки — среднее арифметическое из абсолютных отклонений $|\bar{\Delta}|$, т. е. стоят на позициях И. П. Шаропова, который оперирует отношением этих величин.

В настоящее время все измерения и оценки параметров при разведке месторождений и подсчете запасов в соответствии с требованиями Министерства геологии и ГКЗ должны сопровождаться метрологическими характеристиками точности, в том числе, как мы полагаем, пределами неучтенных систематических ошибок (ГОСТ 8.207-76).

Если желательно максимальное упрощение расчетов, вместо критерия t можно рекомендовать критерий знаков, для которого требуется найти лишь число положительных k и отрицательных $n-k$ отклонений. Нулевые разности, т. е. совпадения основного и контрольного анализов, отбрасываются и в общее число наблюдений не входят. Критические значения k находят по таблицам (например, Мюллер и др., 1982; табл. 16). Этот критерий не требует никаких (кроме непрерывности) предположений о распределении погрешностей. Однако при нормальном распределении эффективность его составляет 0.64 критерия t , т. е. для заключения с одинаковой значимостью требуется почти вдвое больше наблюдений. Известный из геологической литературы (Крейтер, 1960, с. 271) способ Б. И. Галкина является ни чем иным, как критерием знаков, представленным в виде специальной номограммы. Входами в номограмму служат объем выборки проанализированных проб (n) и разность между числом отклонений разного знака. По этим данным находят вероятность (p) того, что расхождения не случайны.

При отсутствии систематической ошибки вероятности положительных и отрицательных отклонений одинаковы, поэтому можно воспользоваться критерием хи-квадрат:

$$\chi^2 = \sum \frac{(np_i - np)^2}{np} = \sum_{i=1}^2 \frac{(k_i - 0.5n)^2}{0.5n}, \quad (5.5.3)$$

где n — число контрольных проб; k_1, k_2 — числа положительных и отрицательных отклонений соответственно. В отличие от критерия t здесь не требуется нормальности распределения разностей, лишь бы вероятности положительных и отрицательных погрешностей как основных, так и контрольных методов анализа были одинаковы. Эта особенность критерия важна при сравнении спектральных анализов с химическими, ибо в каком бы масштабе мы ни выражали погрешности — логарифмическом или обычном, одно из распределений погрешностей не будет симметричным и разности не будут распределены нормально.

Что касается способа введения поправки в результаты основных анализов, то Н. В. Барышев, а затем А. П. Прокофьев (1955), В. И. Смирнов и А. П. Прокофьев (Смирнов и др., 1960) предлагали умножать результаты основных анализов на коэффициент $K_a = \bar{y}/\bar{x}$. И. П. Шарапов (1954, и др.) указывал, что поправки для разных интервалов содержаний могут иметь различные значения и предлагал вычислять K_a отдельно для разных классов содержаний. В. Л. Шашкин (1956) в развитие этой идеи предложил выборку для контрольных анализов составлять не только по разным классам содержаний, но и по разным типам руд, а каждую контрольную пробу составить из вещества нескольких основных проб. В этом последнем есть дополнительное преимущество (не отмеченное автором), что случайная погрешность основного анализа в сборной пробе будет в \sqrt{m} раз меньше (m — число частных проб, вышедшее в сборную); распределение ее будет более близко к нормальному, чем распределение погрешности анализа частных проб. Это существенно повышает эффективность критериев выявления систематической ошибки и уточняет коэффициент K_a .

Д. А. Родионов (1964) выступил против применения коэффициента K_a , так как введение такой поправки якобы искажает дисперсию исследуемых концентраций ($s_x^2 = K_a^2 s_x^2$), что «весьма нежелательно». Однако если основные анализы действительно содержат систематическую ошибку, то параметры распределения (среднее и дисперсия) также неверны и именно нуждаются в исправлении. Просто вопрос об исправлении дисперсии, которое должно сопровождать исправление среднего, не столь актуален и обычно выпадает из поля зрения.

Разногласия между поборниками умножения на переводной коэффициент $K_a = \bar{y}/\bar{x}$ и сторонниками прибавления постоянного смещения $\Delta = \bar{y} - \bar{x}$ не могут быть решены перечисленными выше средствами, ибо они упираются в вопрос, постоянна ли систематическая ошибка (в этом случае надо прибавлять) или пропорциональна (надо умножать). Эти вопросы могут быть решены с помощью корреляционно-регрессионного анализа. Впервые такие предложения выдвинул Б. Я. Юфа (1951). Он предложил рассчитать уравнение регрессии результатов основных анализов на контрольные (x на y). По результатам основных анализов непосредственно с графика снимались «исправленные» значения. В терминах регрессионного анализа процедура заключалась в построении уравнения регрессии

$$x = a + by, \quad (5.5.4)$$

откуда

$$y = 1/b(x - a) = 1/bx - 1/ba = A + Bx$$

(a и b — параметры линейного уравнения регрессии), и оценке «исправленных» значений $x_{\text{испр}} = A + Bx$.

Это предложение подвергалось неквалифицированной критике (Шарапов, 1954; Шашкин,¹ 1956), сводившейся в основном к тому, что нельзя исправлять (корректировать по уравнению регрессии) индивидуальные значения x . Однако именно в этом и заключается достоинство метода (усреднить исправленные значения всегда можно), избавляющее от необходимости следить за репрезентативностью контрольной выборки при ее составлении или корректировать K_a по мере изменения частот содержаний в исследуемом объекте. Главное же преимущество методики Б. Я. Юфы (отметим заранее — если бы она была корректна в основной своей части) — в отдельном выявлении постоянной (a) и пропорциональной (b — в формуле (5.5.4)) составляющих систематической ошибки. В. И. Раевский и Ю. В. Шурубор (1958), совершенствуя метод Б. Я. Юфы, предложили оценивать статистическую значимость параметров a и b путем проверки гипотезы $H_0: b=1$ и $H_0: a=0$.

¹ В глубокой и содержательной статье В. Л. Шашкина, в значительной своей части посвященной критике работ Б. Я. Юфы, этот пункт является единственным, не выдерживающим обратной критики.

В 1970-х годах упомянутый «корреляционный» метод выявления систематических ошибок получил широкое распространение и одновременно был подвергнут уничтожающей критике. А. М. Марголин (1972) отметил, что «регрессионный анализ предназначен для случая, когда значения признака-функции приближены, а значения признака-аргумента (результаты контрольных анализов, — Ю. Т.) заданы точно» (с. 282). Следовательно, его применение должно ограничиваться случаем, когда контрольные анализы не содержат случайных погрешностей. Действительно, если и основные, и контрольные анализы не имели бы случайных погрешностей, а основной анализ не имел бы никаких систематических ошибок, то на графике все результаты ложились бы на прямую с параметрами $a=0$, $b=1$. Как только в контрольных анализах появятся случайные погрешности, точки графика образуют эллипс рассеивания и прямая регрессии x на y пройдет положе, т. е. в «норме» будет $a>0$, $b<1$. А. М. Марголин тем самым дал правильное объяснение известному аналитикам из практики факту, что регрессионный анализ приводит к ложному представлению о систематической завышенности основных анализов в области высоких и заниженности — в области низких содержаний.

Исчерпывающую критику «корреляционного» метода выявления систематической ошибки дал Ю. В. Рошин (1968а, 1968б). Он показал, что указанная задача решается не построением уравнения регрессии результатов основного анализа на контрольный, а построением прямой функциональной зависимости между математическими ожиданиями основного и контрольного анализов, что является задачей не регрессионного, а конъюэнтного анализа. В методе Ю. В. Рошина предполагается выполнение следующих условий.

1. Случайные погрешности измерений и основного, и контрольного метода распределены нормально с нулевыми средними.

2. Дисперсии случайных погрешностей одинаково зависят от измеряемой величины

$$\begin{aligned}\sigma^2[\varepsilon_i(x)] &= \delta_x^2 h^2(\varphi_i), \\ \sigma^2[\varepsilon_i(y)] &= \delta_y^2 h^2(\varphi_i),\end{aligned}$$

где $\varepsilon_i(x)$ — случайная погрешность результата основного анализа x ; $\varepsilon_i(y)$ — то же, контрольного; h — известная функция, φ_i — истинное содержание в i -й пробе.

3. Известно одно из значений δ или отношение дисперсий сравниваемых методов $\lambda = \delta_x^2/\delta_y^2$.

Предварительно вычисляются следующие статистики:

$$\begin{aligned}\bar{x} &= \frac{1}{\omega_i} \sum_{i=1}^n \omega_i x_i, & \bar{y} &= \frac{1}{\omega_i} \sum \omega_i y_i, & \Sigma_{xx} &= \Sigma \omega_i (x_i - \bar{x})^2, \\ \Sigma_{yy} &= \Sigma \omega_i (y_i - \bar{y})^2, & \Sigma_{xy} &= \Sigma \omega_i (x_i - \bar{x})(y_i - \bar{y}),\end{aligned}$$

где $\omega_i = 1/h_2(\varphi_i)$.

Искомые оценки коэффициентов α и β функционального отношения результатов основного и контрольного анализов

$$Y_i = \alpha + \beta X_i \quad (5.5.5)^2$$

вычисляются по формулам $\beta = \theta + \sqrt{\theta^2 + \lambda}$, где $\theta = (\Sigma_{yy} - \lambda \Sigma_{xx})/2\Sigma_{xy}$. В том случае, когда известно не отношение λ , а величина s_y^2 , $\beta = (\Sigma_{yy} - n\delta_y^2)/\Sigma_{xy}$. Затем $\alpha = \bar{y} - \hat{\alpha}\bar{x}$. Доказательства изложенного даны Д. Линдли (Lindley, 1947). Систематическая ошибка основного анализа относительно контрольного устанавливается проверкой значимости отклонений α от нуля (постоянная составляющая)

² У Ю. В. Рошина обозначения коэффициентов другие.

Схема дисперсионного анализа для проверки совместности значения β с экспериментальными данными

Тип изменчивости	Число степеней свободы	Сумма квадратов	Дисперсия
Регрессия	1	$\frac{(\sum_{xy} - (\sum_{yy} - \lambda \sum_{xx}) - \lambda \sum_{xy})^2}{\sum_{yy} + 2\lambda \sum_{xy} + \lambda^2 \sum_{xx}} = A$	$\hat{\sigma}_1^2 = A$
Остаточная	$n-2$	$\frac{\lambda^2 (\sum_{xx} \cdot \sum_{xy} - \sum_{xy}^2)}{\sum_{yy} + 2\lambda \sum_{xy} + \lambda^2 \sum_{xx}} = B$	$\hat{\sigma}_2^2 = \frac{B}{n-2}$
Полная	$n-1$	$\sum_{yy} - 2\sum_{xy} + \sum_{xx} = A+B$	

и β от единицы (пропорциональная составляющая) с помощью дисперсионного анализа по схеме, представленной в виде табл. 5.5.1.

Отношение $\hat{\sigma}_1^2/\hat{\sigma}_2^2$ подчиняется распределению F_p с $f_1 = 1, f_2 = n - 2$ степенями свободы. Таким образом, если $\hat{\sigma}_1^2/\hat{\sigma}_2^2 < F_{\alpha, (1, n-2)}$, то β несущественно отличается от единицы и пропорциональную составляющую систематической ошибки нельзя считать установленной. Ю. В. Рошиным предложен приближенный способ проверки существенности отклонения α от нуля в предположении, что $\beta=0$, не отличающийся, правда, простотой. Если результаты одного из методов, неважно какого — основного или контрольного, не содержат случайных погрешностей, описанный метод совпадает с «корреляционным». При этом уравнение регрессии следует строить «на результаты» метода анализа, не содержащего случайных погрешностей.

К сожалению, приходится согласиться с Ю. В. Рошиным, что описанная методика не может быть применена для большого числа практических случаев. В настоящее время не имеется метода сравнения химического анализа со спектральным, так как функции h для них различны. Если для химического $h(c)=1$ (погрешность не зависит от концентрации), то для спектрального $h(c)=c$ (погрешность прямо пропорциональна концентрации). Нормальность распределения случайных погрешностей (одно из условий применения описанного метода), как было показано в предыдущем разделе, несовместима с тем, чтобы $h(c)$ не была константой. Это положение практически сводит область применения этого изящного и корректного метода к единичным случаям.

Попытку решить проблему выявления систематических ошибок для всех возможных в геологии случаев предпринял Ю. Т. Усиков (1983), процитировав (без ссылки) начало нашего (Ткачев, Юдович, 1975) обзора проблемы, но уйдя от содержательного анализа современного ее состояния.

Автор использовал свою идею об аппроксимации случайных погрешностей анализа трехпараметрическим логарифмически-нормальным распределением (см. § 4). Его способ заключается в следующем.

1. Вместо содержаний x_j изучается величина $u = \log(x_j + c)$, где константа c подбирается так, чтобы погрешность измерения была распределена нормально.

2. Рассчитывается среднее отклонение между преобразованными результатами основных анализов и истинными «эталонными» преобразованными содержаниями:

$$\bar{\Delta}_c = \frac{1}{n} \sum_{j=1}^n [\log(x_j + c) - \log(y_j + c)]. \quad (5.5.6)$$

3. Значимость смещения преобразованной переменной Δ_c устанавливается по критерию Стьюдента.

4. Если смещение значимо, оценка исправленных значений производится по формуле

$$x_{\text{испр}j} = \beta + \gamma x_j, \quad (5.5.7)$$

где

$$\gamma = 1/\text{ant}(\bar{\Delta}_c),^3 \quad \beta = c(\gamma - 1). \quad (5.5.8)$$

Во-первых, отметим, что предложение Ю. Т. Усикова касается лишь самого простого частного случая, когда контрольные анализы не отягощены случайными погрешностями. Во-вторых, предлагаемое преобразование переменной ($x \rightarrow \log(x+c)$) не гарантирует, что знаки отклонений $\Delta_i = \log(x_i+c) - \log(y_i+c)$ будут одинаковы для различных содержаний. Следовательно, незначимость отклонения $\bar{\Delta}_c$ от нуля вовсе не означает отсутствия систематического смещения. Более того, формулы (5.5.7) и (5.5.8) верны только в том единственном и маловероятном случае, когда точки $(\log(x_i+c), \log(y_i+c))$ ложатся на прямую, расположенную под углом 45° к оси $\log(y+c)$. Таким образом, предложение Ю. Т. Усикова не содержит ни способа выявления систематических погрешностей, ни способа оценки ее существенности (ее постоянной и пропорциональной составляющей) даже для того частного случая, для которого оно по замыслу предназначено. Само преобразование измеряемой величины полезно, если оно нормализует распределение погрешностей, но дальнейшая обработка данных по выявлению систематической ошибки должна вестись по схеме регрессионного анализа и лишь в том случае, если контрольные анализы не содержат случайных погрешностей. Применить методику Ю. В. Рощина к преобразованной переменной в противном случае представляется затруднительным или невозможным, так как в этом более общем случае результаты основных и контрольных анализов должны подвергаться преобразованию с различными значениями c (которые надо эффективно оценить) и функция h для них будет различной. Создание универсальной теории выявления систематических погрешностей нам представляется бесперспективным, а практически решить проблему можно с помощью численного моделирования на ЭВМ.

§ 6. Сопоставление измерений пористости по керну и скважинно-геофизическими методами. Комплексная оценка пористости

Сопоставление различных методов опробования и выявление систематических ошибок одних методов относительно других, принятых за контрольные, включает в себя все сложности, присущие внешнему контролю анализов. Кроме того, возникают дополнительные сложности, связанные с большими различиями в функциях распределения измеряемой величины по данным различных видов опробования, в объеме и «геометрии» проб сравниваемых методов, в расположении их друг относительно друга. При контрольных анализах имеют дело с одной и той же пробой и различие в результатах целиком и полностью вызвано различием в анализах. При контрольном опробовании анализируются различные пробы, отличающиеся по содержанию, и тем сильнее, чем изменчивее изучаемое свойство.

Особый интерес представляет сопоставление двух методов, один из которых не содержит систематических ошибок, но характеризуется большими случайными погрешностями, а для другого характерно обратное соотношение погрешностей. Важный из нефтяной геологии пример этому — сопоставление результатов кернавого опробования и геофизических исследований (измерений) скважин (ГИС). Оценка коллекторских свойств по керну систематических

³ ant в обозначениях Ю. Т. Усикова — антилогарифм.

погрешностей обычно не имеет, но из-за немногочисленности образцов и их редкого расположения страдает большими случайными погрешностями. Геофизические измерения, напротив, непрерывны по стволу скважины, хорошо воспроизводимы, но, будучи косвенными, могут содержать систематическую ошибку, сильно зависящую от эталонировки и выбранных для расчета констант и зависимостей.

К сожалению, в геологоразведочной нефтяной литературе имеются лишь отрывочные и зачастую противоречивые сведения о погрешностях анализов и измерений. С. И. Шишигин (1975) приводит средние относительные погрешности определения в образцах керна терригенных пород пористости (1—1.5%), проницаемости (3—7%), нефтенасыщенности (2—7%) и нефтеотдачи (10—20%). Для погрешности извлекаемого нефтесодержания в породе он приводит значения от 8 до 23%. Г. Д. Сараева (1976) приводит для погрешности анализа пористости значения 0.6—0.9%, а В. М. Бортницкая (1977) — плохо согласующиеся с предыдущими, на порядок большие погрешности — 11—17%. Такой разницей в значениях тем более удивителен, что расчет лабораторной погрешности анализа керна не вызывает особенных затруднений. Дело, по-видимому, заключается в неосведомленности геологов, использующих эти анализы.

Еще сложнее выяснить по литературным данным изученность погрешностей геофизических измерений в скважинах. По данным Т. Г. Мамедова и др. (1979), относительная погрешность определения мощности коллектора составляет 10%, пористости и нефтегазонасыщенности — 10—15%. При этом неясно, к какому объекту относятся указанные погрешности. Цирюльников В. П. (1980) делает очень важный вывод, что «при сравнительно хороших показателях аппаратурных погрешностей низкая точность измерений объясняется неудовлетворительным состоянием эталонировки и стандартизации» (с. 147). Анализ распределения погрешностей и их влияния на точность последующих оценок ни по керну, ни по геофизике в цитированных работах не дается.

Пусть имеется совокупность k скважин на месторождении, в которых пористость x измерена с помощью ГИС и представлена в виде средних значений x_i^r со средней квадратической погрешностью S_{i_i} по определенным интервалам глубин,¹ которые мы будем для краткости называть интервалами коллектора. Из некоторых интервалов коллектора отобраны образцы керна; пористость в них измерена лабораторным методом x_{ij}^k со средней квадратической погрешностью s_{ij}^k , где i — номер интервала коллектора из общей их совокупности, j — номер образца в i -м интервале. Точные места (глубины) отбора образцов неизвестны, а известна лишь принадлежность образцов к определенному интервалу. Задача заключается в выявлении и учете систематического смещения геофизических данных относительно керновых путем сопоставления керновых и геофизических данных и в получении комплексной оценки пористости интервалов, обладающей минимально возможной погрешностью.

Аналогичная задача — соединить высокую воспроизводимость одного метода с отсутствием систематической ошибки другого — встречается в геологии часто. Решение усложняется тем, что «керновое» среднее по интервалу относится к столбу породы с поперечником, равным диаметру керна, «геофизическое» — к некоторому (значительно большему) объему породы, расположенному за пределами стенок скважины (рис. 5.6.1).

Комплексирование данных при оценке коллекторских свойств — задача известная: Л. Ф. Дементьев и В. И. Азаматов (1961) производили «визуальное» сравнение керновых и геофизических средних по месторождению, но комплексной оценки из них не вычисляли. В. Л. Комаров (1961а, 1961б) считает, что сравнивать керновые и геофизические данные надо по всему разрезу месторож-

¹ Мы не касаемся специфических вопросов интерпретации каротажных диаграмм, а принимаем указанные значения за исходные данные.

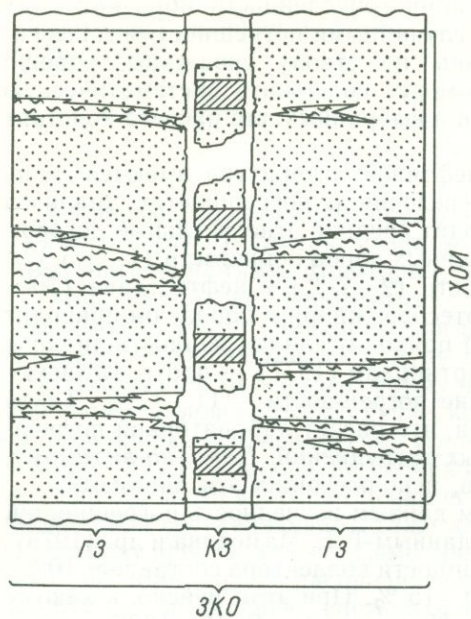


Рис. 5.6.1. Взаимное расположение зон пласта породы-коллектора вокруг скважины, в которых пористость оценивается различными способами.

ХОИ — хорошо опробованный интервал; КЗ — зона оценки керном, ГЗ — зона оценки по ГИС, ЗКО — зона комплексной оценки.

дения в целом и по всей площади. Им сравниваются функции распределения частот пористости по критерию Колмогорова и делается вывод, что «геофизические» распределения имеют меньшие дисперсии, а средние по ним существенно выше. Ю. Н. Батурин (1965) предлагает керновые и геофизические оценки по залежи усреднять с весовыми коэффициентами, равными либо числу измерений каждым из методов, если средние разнятся несущественно, либо произведению числа измерений на точность соответствующего метода, если средние отличаются существенно. Из работы остается неясным, что понимать под числом измерений, например для ГИС. А. Я. Фурсов (1968) общее среднее рекомендует находить как среднее арифметическое из керновой и геофизической оценок по залежи. Позже (Калабин и др., 1975) стали применять дисперсионный анализ и при отсутствии существенного различия средних «объединять оценки», но что понимать под «объединением», — неясно. В большинстве случаев подсчета запасов керн используется преимущественно для контроля данных промысловой геофизики.

Получение комплексной оценки включает решение следующих вопросов: 1) выбор элементарных единиц (объектов) сопоставления; 2) расчет среднего значения свойства и его погрешности каждым методом в пределах выбранных единиц сопоставления; 3) установление статистической зависимости между керновой и геофизической пористостью по совокупности выбранных объектов сопоставления, т. е. расчет параметров уравнения регрессии пористости по керну на пористость по геофизике, выявление систематического смещения результатов геофизических измерений относительно керновых данных и установление поправок к геофизическим данным; 5) расчет комплексной оценки.

В качестве единицы сопоставления нами принят объект, названный «хорошо опробованным интервалом» (ХОИ). Под ним подразумевается интервал коллектора (или его однородная часть), хорошо охарактеризованный керном (не менее 5 образцов на 1 м эффективной мощности), с высоким качеством геофизических материалов. Интервалы очень большой мощности, даже если они однородны, рекомендуется разделить на несколько меньших частей.

Расчет среднего значения пористости в пределах ХОИ по данным кернового опробования производится либо как среднее арифметическое, если точные места отбора образцов внутри ХОИ неизвестны, либо как среднее взвешенное с весовыми коэффициентами, пропорциональными длине «зон влияния» образцов. При большом количестве образцов на интервал возможны аппроксимация пористости какой-либо кривой и вычисление среднего как интеграла аппроксимирующей функции. Вычисление погрешности оценки пористости интервала по керновым данным будет рассмотрено отдельно.

Расчет средней в ХОИ пористости по данным геофизических измерений производится непосредственно интегрированием кривой, так как измерения

непрерывны по разрезу. Случайная погрешность этой средней определяется погрешностью измерений примененного геофизического метода. Особенность ситуации заключается здесь в том, что она зависит от длины интервала измерения: чем он больше, тем случайная погрешность меньше. Здесь мы делаем допущение, что случайная погрешность измерения в некотором интервале глубин не зависит от погрешности измерения в соседнем, не пересекающемся с ним, или более удаленном интервале. Тогда $s_{x_{r3}}^2 = s_{x_{(0)}}^2 / l$, где $s_{x_{r3}}^2$ — случайная погрешность измерения пористости в ХОИ мощностью l ; $s_{x_{(0)}}^2$ — то же, единичной мощности (если за единичную мощность принимается мощность в 1 м, то l выражается в метрах). Известно, что участки керна с самой высокой пористостью избирательно выкрашиваются при бурении, поэтому средняя пористость интервала по результатам анализа оставшихся образцов занижена. Чтобы исключить или уменьшить занижение, расчеты статистик зависимости проводятся только для тех интервалов из числа ХОИ, пористость которых, по геофизическим данным, не превышает максимальной пористости по керну.

По парам значений пористости сформированной таким образом совокупности ХОИ (x_i^k — по керну, x_i^r — по геофизике) рассчитываются коэффициент корреляции r_{rk} и параметры уравнения регрессии пористости по керну на пористость по геофизике:

$$x^k = a + bx^r. \quad (5.6.1)$$

В большинстве публикаций, затрагивающих сравнение керновых и геофизических данных, поправки в последние вводятся с помощью уравнения регрессии, аналогичного (5.6.1). В предыдущем разделе была показана ошибочность такой методики при сопоставлении основных и контрольных анализов. Измерения пористости по геофизике можно сопоставить с основными анализами, по керну — с контрольными, поэтому применение параметров уравнения регрессии в качестве поправок и здесь было бы ошибочным, если можно так выразиться, но даже более ошибочно по следующим причинам.

1. Керновое опробование, выступая в качестве контрольного ввиду отсутствия или малости систематических ошибок, характеризуется высоким уровнем случайных погрешностей, более высоким, чем геофизические измерения.

2. Результаты кернового опробования характеризуют другой объем породы, чем данные геофизических измерений, поэтому даже в том случае, когда ни керновые, ни геофизические данные не были бы отягощены случайными погрешностями, коэффициент корреляции между ними был бы существенно меньше единицы ввиду наличия пространственной изменчивости пористости. Заметим, что коэффициент корреляции между результатами основного и контрольного анализов, откорректированный по формуле (2.2.5) с учетом погрешностей измерений, был бы равен единице.

Для того чтобы применить к данному случаю конфлюэнтный метод Ю. В. Роцина выявления систематической ошибки и расчета поправок, необходимо следующее. Во-первых, — выбрать единый для геофизических измерений и кернового опробования объект оценки пористости. По разрезу скважины он, естественно, совпадает с i -м интервалом коллектора. В латеральном направлении это будет зона комплексной оценки (ЗКО, см. рис. 5.6.1). Во-вторых, — рассчитать погрешности оценки пористости в объеме ЗКО: а) геофизическими измерениями (которые охватывают некоторую зону вокруг ствола скважины, обозначенную нами как ГЗ); б) керновым опробованием, охватывающим керновую зону (КЗ). В-третьих, — рассчитать общие погрешности оценок, названных в (а) и (б), с учетом погрешностей измерений (для (а)) и лабораторных анализов (для (б)). Принимая эти общие погрешности оценок за аналоги погрешностей измерений основных и контрольных анализов, можно применить какой-либо метод (в частности, метод Ю. В. Роцина) выявления систематических погрешностей.

Таким образом, сопоставление основного и контрольного анализов отличается от сопоставления основного и контрольного опробования тем, что в последнем случае необходимо принимать во внимание не только погрешности измерений, но и погрешности оценок изучаемого свойства в некотором объеме породы, включающем как основную, так и контрольную пробы.

Расчет последней упомянутой погрешности является типичной задачей геостатистики. В настоящее время геостатистические характеристики (полу-вариограмма) нефтяных месторождений в отношении таких свойств, как пористость, для расстояний порядка диаметра ЗКО не изучены, поэтому здесь предлагается более простой приближенный метод, основанный на сравнении эмпирических дисперсий пористости по результатам кернового опробования и геофизических измерений и коэффициента корреляции между этими результатами. Наблюдаемая (эмпирическая) дисперсия пористости в совокупности ХОИ по результатам кернового опробования s_{κ}^2 может быть представлена в виде суммы

$$s_{\kappa}^2 = s_{x_{(ЗКО)}}^2 + s_{x_{(ЗКО, КЗ)}}^2 + s_{x_{(КЗ)}}^2, \quad (5.6.2)$$

где $s_{x_{ЗКО}}^2$ — дисперсия истинных значений пористости ЗКО совокупности ХОИ; $s_{x_{ЗКО}}^2$ — дисперсия распространения² значения пористости в керновой зоне на зону комплексной оценки; $s_{x_{КЗ}}^2$ — дисперсия распространения значения пористости в образцах керна на керновую зону (с учетом погрешности анализа образцов). Равенство (5.6.2) справедливо, так как упомянутые ошибки распространения не зависят друг от друга и от пористости ЗКО.

Аналогичное равенство можно написать и для дисперсии пористости по результатам геофизических измерений $s_{г}^2$:

$$\frac{1}{\beta^2} s_{г}^2 = s_{x_{(ЗКО)}}^2 + s_{x_{(ЗКО, ГЗ)}}^2 + \frac{1}{\beta^2} s_{x_{(ГЗ)}}^2, \quad (5.6.3)$$

где $s_{x_{(ЗКО, ГЗ)}}^2$ — дисперсия распространения значения пористости в геофизической зоне на зону комплексной оценки; $s_{x_{(ГЗ)}}^2$ — дисперсия, характеризующая погрешность геофизических измерений пористости в геофизической зоне. Величина β в (5.6.3) представляет собой искомую пропорциональную поправку к результатам геофизических измерений пористости (постоянная поправка не меняет дисперсии результатов измерения).

В методе Ю. В. Рошина (адаптированном к данному случаю) необходимо знать либо одну из сумм $s_{x_{(ЗКО, КЗ)}}^2 + s_{x_{(КЗ)}}^2$, $s_{x_{(ЗКО, ГЗ)}}^2 + s_{x_{(ГЗ)}}^2$, либо их отношение λ . Величины $s_{x_{(КЗ)}}^2$ и $s_{x_{(ГЗ)}}^2$ нам известны, а непосредственное вычисление величин $s_{x_{(ЗКО, КЗ)}}^2$ или $s_{x_{(ЗКО, ГЗ)}}^2$ сопряжено с упоминавшимися трудностями. Мы можем сделать весьма правдоподобное предположение об отношении последних величин: дисперсия распространения обратно пропорциональна объему пробы v и прямо пропорциональна выражению $1-f$, где f — коэффициент отбора ($f=v/V$). Таким образом

$$\frac{s_{x_{(ЗКО, КЗ)}}^2}{s_{x_{(ЗКО, ГЗ)}}^2} = \frac{(1 - V_{КЗ}/V_{ЗКО}) v_{ГЗ}}{V_{КЗ} (1 - V_{ГЗ}/V_{ЗКО})}.$$

Так как $V_{ЗКО} = v_{КЗ} + v_{ГЗ}$, то $s_{x_{(ЗКО, КЗ)}}^2 / s_{x_{(ЗКО, ГЗ)}}^2 = v_{ГЗ}^2 / v_{КЗ}^2$. Объемы керновой, геофизической зон и зоны комплексной оценки пропорциональны величинам $d_{КЗ}^2$, $d_{ЗКО}^2$ — $d_{КЗ}^2$, $d_{ЗКО}^2$, где $d_{КЗ}$, $d_{ЗКО}$ соответственно диаметры керновой зоны и зоны комплексной оценки. Отсюда

² Погрешность, вносимая при распространении содержания в пробе на зону ее влияния, в геостатистике характеризуется дисперсией распространения (Карлье, 1966, с. 323).

$$v_{гз}^2/v_{кз}^2 = \left(\frac{d_{зко}^2 - d_{кз}^2}{d_{кз}^2} \right)^2 = \left(\frac{d_{зко}^2}{d_{кз}^2} - 1 \right)^2 = (\eta^2 - 1)^2,$$

где $\eta = d_{зко}/d_{кз}$ — отношение диаметра столба породы, «прослушиваемого» геофизическим методом к диаметру керна, который мы принимаем приближенно равным диаметру скважины.

Тогда можно записать следующее равенство:

$$s_{\bar{x}_{(кз, гз)}}^2 = s_{\bar{x}_{(зко, гз)}}^2 (\eta^2 - 1)^2. \quad (5.6.4)$$

На основании равенства (2.2.5) можно также записать

$$r_{x_r x_k} = r_{x_{зко} x_{зко}} s_{x_{зко}} / s_{x_k} \cdot s_{x_{зко}} / 1 / \beta s_{x_r}, \quad (5.6.5)$$

где $r_{x_{зко} x_{зко}}$ — коэффициент корреляции между рядами пористости ЗКО по керну и геофизике в случае, если бы ни один из этих методов оценки пористости не содержал случайных погрешностей, а систематические расхождения между ними содержали только постоянную и пропорциональную (значению пористости) составляющие, т. е. $r_{x_{зко} x_{зко}} = 1$.

Равенства (5.6.2—5.6.5) образуют систему уравнений

$$\left. \begin{aligned} s_{x_k}^2 - s_{x_{(зко)}}^2 - s_{x_{(зко, кз)}}^2 - s_{x_{(кз)}}^2 &= 0, \\ s_{x_r}^2 - \beta^2 s_{x_{(зко)}}^2 - \beta^2 s_{x_{(зко, гз)}}^2 - s_{x_{(гз)}}^2 &= 0, \\ s_{x_{(зко, кз)}}^2 - s_{x_{(зко, гз)}}^2 (\eta^2 - 1)^2 &= 0, \\ r_{x_r x_k} s_{x_k} s_{x_r} - \beta s_{x_{(зко)}}^2 &= 0, \end{aligned} \right\} \quad (5.6.6)$$

которую легко решить, так как она является: а) линейной относительно трех неизвестных: $s_{x_{(зко)}}^2$ — природной дисперсии пористости в совокупности ХОИ, $s_{x_{(зко, гз)}}^2$ и $s_{x_{(зко, кз)}}^2$ — дисперсий распространения пористости в геофизической и керновой (соответственно) зонах на зону комплексной оценки и б) квадратной относительно β — искомой пропорциональной поправки к данным геофизических измерений. Можно остановиться на этом значении или по теперь уже известным величинам $s_{x_{(зко, кз)}}^2$, $s_{x_{(зко, гз)}}^2$, $s_{x_{(кз)}}^2$ и $s_{x_{(гз)}}^2$ определить λ и воспользоваться методом Ю. В. Рошина для определения α и β .

Дисперсии $s_{x_{(зко, кз)}}^2$ и $s_{x_{(зко, гз)}}^2$ используются не только в качестве промежуточных величин при расчете поправок к данным геофизических измерений. Они служат весовыми коэффициентами при объединении керновых и геофизических данных, поэтому мы рассмотрим еще один метод их расчета, опирающийся на предположения, отличные от тех, которые используются при составлении системы уравнений (5.6.6).

Оценивание пористости в ЗКО «распространением» на эту зону значения пористости в КЗ основана на корреляции между этими величинами, т. е. на том, что $r_{кз, зко} > 0$, где $r_{кз, зко}$ — коэффициент корреляции по совокупности ХОИ между пористостью зоны комплексной оценки и пористостью керновой зоны. Распространяя на этот случай известное равенство, можно записать

$$s_{\bar{x}_{(зко, кз)}}^2 = (1 - r_{кз, зко}^2) \cdot s_{x_{(зко)}}^2. \quad (5.6.7)$$

Задача сводится теперь к выражению величины $r_{кз, зко}$ через коэффициент корреляции между результатами измерения пористости ХОИ по керну и геофизике — $r_{x_k x_r}$. Прежде всего исключим с помощью равенства (2.2.5) влияние на этот коэффициент погрешностей измерения, т. е. вычислим такой коэффициент, который характеризовал бы зависимость между x_k и x_r , если бы они были получены без случайных погрешностей. Обозначим этот коэффициент через

$$r_{кз, гз} = r_{x_k x_r} \frac{s_{x_k} s_{x_r}}{\sqrt{s_{x_k}^2 - s_{x_{(кз)}}^2} \sqrt{s_{x_r}^2 - s_{x_{(гз)}}^2}} \quad (5.6.8)$$

Если окажется, что $r_{кз, гз} \geq 1$, то это будет свидетельствовать, что декларируемые лабораторией и (или) геофизиками значения погрешностей измерений³ не соответствуют фактическим, а значительно превышают их. Может оказаться, что $r_{кз, гз} \geq 1$, даже если одну из погрешностей принять за нулевую. Тогда декларируемое значение другой из них определенно превышает фактическую. Во всех этих случаях (рис. 5.6.2) следует тщательнее рассмотреть метрологические характеристики применяемых методов измерений. В дальнейшем мы будем предполагать, что $r_{кз, гз} < 1$.

Разобьем геофизическую зону на участки, каждый из которых по объему и форме идентичен керновой зоне, т. е. в данном случае — на столбцы породы длиной, равной длине ХОИ, и диаметром $d_{кз}$. Пусть таких участков будет n . Тогда $r_{кз, гз}$ можно рассматривать как коэффициент корреляции между пористостью в фактической керновой зоне $x_{кз}$ и средней пористостью из n условных керновых зон, намечаемых в пределах ГЗ — $x'_{кз_i}$ ($i=1, n$), т. е. $x_{кз} = x_{кз}$, $x_{гз} = 1/n (x'_{кз_1} + x'_{кз_2} + \dots + x'_{кз_n})$. Коэффициент корреляции между величинами, расположенными в левых частях этих равенств, необходимо выразить через коэффициенты корреляции между величинами правых частей. Для этого сделаем предположение, что $r_{кз, кз_i} = r_{кз_i, кз_j} = \text{const} = r_{кз, кз}$ ($i=j$). Тогда можно написать

$$r_{кз, гз} = \sum_{i=1}^{n-1} \frac{s_{кз}}{s_{гз}} r_{кз, кз_i} \quad (5.6.9)$$

где $s_{кз}^2$ — дисперсия пористости керновых зон в совокупности ХОИ, $s_{гз}^2$ — то же геофизических зон. Дисперсию $s_{гз}^2$ находим по эмпирической дисперсии пористости ХОИ по керну ($s_{x_k}^2$), вычитая из нее дисперсию случайной погрешности измерения пористости с помощью керна в керновой зоне ($s_{x_{(кз)}}^2$):

$$s_{кз}^2 = s_{x_k}^2 - s_{x_{(кз)}}^2 \quad (5.6.10)$$

Величина $s_{гз}^2$ есть дисперсия суммы $k = n - 1$ коррелированных с коэффициентом корреляции $r_{кз, кз}$ случайных величин, каждое из слагаемых которой имеет дисперсию $s_{кз}^2$. Применяя к этому случаю известное равенство из теории вероятностей $s_{x+y}^2 = s_x^2 + s_y^2 + 2r_{xy} s_x s_y$ получим

$$s_{гз}^2 = k s_{кз}^2 [1 + (k-1) r_{кз, кз}] \quad (5.6.11)$$

Подставляя это выражение для $s_{гз}^2$ в (5.6.9), получим:

$$r_{кз, гз} = \frac{\sqrt{n-1} r_{кз, кз}}{\sqrt{1 + (n-2) r_{кз, кз}}} \quad (5.6.12)$$

Аналогичным путем находим коэффициент корреляции пористости в керновой зоне с пористостью в зоне комплексной оценки:

$$r_{кз, зко} = \frac{\sqrt{n} r_{кз, кз}}{\sqrt{1 + (n-1) r_{кз, кз}}} \quad (5.6.13)$$

Теперь необходимо решить (5.6.12) относительно величины $r_{кз, кз}$ и использовать ее в (5.6.13) для расчета величины $r_{кз, зко}$, которая используется в (5.6.7) для расчета искомой дисперсии:

³ Имеются в виду средние квадратические погрешности.

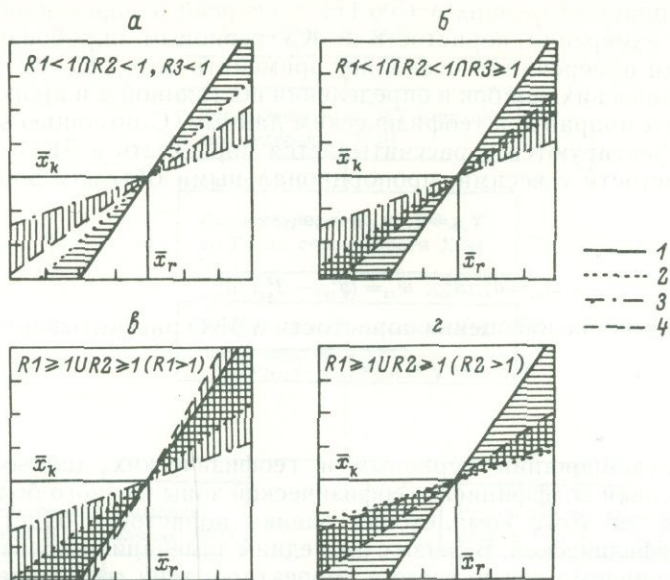


Рис. 5.6.2. К анализу соответствия декларированных погрешностей измерения пористости по керну (x_k) и геофизическим методом (x_r) фактическим геологоразведочным данным.

а — декларированные погрешности не противоречат фактическим данным; *б* — погрешность одного из методов завышена против фактической; *в* — декларированная погрешность измерений по ГИС сильно завышена против фактической; *г* — то же, измерений по керну. 1 — эмпирическое уравнение регрессии x_r на x_k ; 2 — то же, после элиминирования погрешности оценки по керну; 3 — эмпирическое уравнение регрессии x_k на x_r ; 4 — то же, после элиминирования погрешности оценки по ГИС. R — условные коэффициенты корреляции пористости по керну с пористостью по ГИС: $R1$ — при условии отсутствия случайных погрешностей измерений по керну; $R2$ — то же, по ГИС; $R3$ — то же, и по керну, и по ГИС.

$$r_{k3, k3} = \frac{1}{2} \left[\left(1 - \frac{1}{n-1} \right) r_{k3, r3}^2 \pm r_{k3, r3} \sqrt{r_{k3, r3}^2 \left(1 - \frac{1}{n-1} \right) + \frac{4}{n-1}} \right]. \quad (5.6.14)$$

Проверкой при $n \rightarrow \infty$ устанавливаем, что в (5.6.14) необходимо взять знак плюс, иначе получится абсурдный в данном случае результат $r_{k3, k3} = 0$. В (5.6.7) нам остается неизвестной дисперсия пористости зоны комплексной оценки в совокупности ХОИ $s_{x(3KO)}^2$. Ее определяем из равенства, отражающего аддитивность дисперсий:

$$s_{k3}^2 = s_{x(3KO)}^2 + s_{x(3KO, k3)}^2. \quad (5.6.15)$$

Действительно, дисперсия свойства в совокупности проб (по одной из каждого из оцениваемых объектов) больше дисперсии средних значений свойства в совокупности этих объектов на величину средней дисперсии оценки среднего в объеме одной пробой.

Подставляя выражение для $s_{x(3KO)}^2$ из (5.6.15) в (5.6.7), получим $s_{x(3KO, k3)}^2 = (1 - r_{k3, 3KO}^2) (s_{k3}^2 - s_{x(3KO, k3)}^2)$, откуда

$$s_{x(3KO, k3)}^2 = \frac{(1 - r_{k3, 3KO}^2) s_{k3}^2}{(2 - r_{k3, 3KO}^2)}. \quad (5.6.16)$$

Выражение для s_{k3}^2 подставляем из (5.6.10). Погрешность оценивания пористости ЗКО распространением на эту зону значения пористости в ГЗ ($s_{x(3KO, r3)}^2$) определим из тех же предположений, что и в предыдущем способе, т. е. будем считать, что

$$s_{x(3KO, k3)}^2 = s_{x(3KO, r3)}^2 \cdot (\eta^2 - 1)^2. \quad (5.6.17)$$

Знания обеих присутствующих в (5.6.17) дисперсий распространения, а также погрешностей измерения пористости в КЗ керновым опробованием и в ГЗ геофизическими измерениями позволяет применить метод Ю. В. Рощина выявления систематических ошибок и определения постоянной α и пропорциональной β составляющих поправок к геофизическим данным. С помощью этих поправок последние корректируются и рассчитывается пористость в ЗКО объединением значений пористости с весами, пропорциональными объемам зон:

$$x_{\text{ЗКО}} = x_{\text{ГЗ}} \omega_{\text{ГЗ}} + x_{\text{КЗ}} \omega_{\text{КЗ}}, \quad (5.6.18)$$

$$\omega_{\text{КЗ}} = d_{\text{КЗ}}^2 / d_{\text{ЗКО}}^2, \quad \omega_{\text{ГЗ}} = (d_{\text{ЗКО}}^2 - d_{\text{КЗ}}^2) / d_{\text{ЗКО}}^2. \quad (5.6.19)$$

Погрешность комплексной оценки пористости в ЗКО рассчитывается по формуле

$$s_{\hat{x}}(\text{ЗКО}) = \sqrt{s_{\hat{x}}^2(\text{КЗ}) \omega_{\text{КЗ}}^2 + s_{\hat{x}}^2(\text{ГЗ}) \omega_{\text{ГЗ}}^2}. \quad (5.6.20)$$

При комплексировании керновых и геофизических данных, как видно из (5.6.19), весовой коэффициент геофизической зоны намного больше такового керновой зоны, поэтому комплексная оценка пористости будет практически совпадать с геофизической. В связи с последним замечанием пользу от изложенного мы усматриваем прежде всего в корректировании геофизических данных.

В настоящее время благодаря развитию геофизических методов исследования скважин и широкому использованию ЭВМ появилась возможность практически без дополнительных затрат получить оценку пористости (равным образом и других интересующих геолога характеристик) несколькими методами (ядерными, электрическими, акустическими и др.). Изложенная методика с соответствующими коррективами применима и к этому случаю. Отличия заключаются главным образом в том, что геофизические зоны практически перекрывают друг друга и комплексную оценку после корректировки по керну результатов всех методов следует рассчитывать по формулам гл. 6.

Блок-схема алгоритма изложенного метода дана на рис. 5.6.3. Программы для ЭВМ, разработанные в Институте геологии Коми филиала АН СССР, включены в комплекс автоматизированного подсчета запасов и позволяют повысить точность оценки коллекторских свойств разведываемых нефтяных месторождений Тимано-Печорской провинции на 10—15 %.

В заключение рассмотрим характерный численный пример из практики автоматизированного подсчета запасов. По залежи D_{2st} одного из месторождений нефти для сопоставления пористости по керну по данным акустического каротажа в автоматическом режиме была сформирована следующая (табл. 5.6.1) совокупность ХОИ. Результаты первичной статистической обработки этих данных на ЭВМ приведены в табл. 5.6.2 (фрагмент машинной распечатки). Результаты сопоставления описанным методом приведены на рис. 5.6.4, также являющемся машинной распечаткой. Условные коэффициенты корреляции, рассчитанные по формуле, подобной (5.6.8), показывают, что декларированные погрешности измерений, которые определяют погрешности оценки пористости геофизикой в ГЗ и керном в КЗ, не противоречат разведочным данным. Высокий коэффициент корреляции между пористостью в КЗ и ГЗ (0.978) свидетельствует о незначительной латеральной изменчивости содержания в пределах ЗКО. Пропорциональная ($\beta=1.493$) и постоянная ($\alpha=-6.807$) поправки к геофизическим данным существенно отличаются от параметров уравнения регрессии «керновой» пористости на «геофизическую» (1.27 и -4.012 соответственно).

Сравним эти результаты с результатами сопоставления основного и контрольного анализов методом Ю. В. Рощина в обеих его модификациях. Для их реализации на микрокалькуляторах составлены программы (табл. 5.6.3

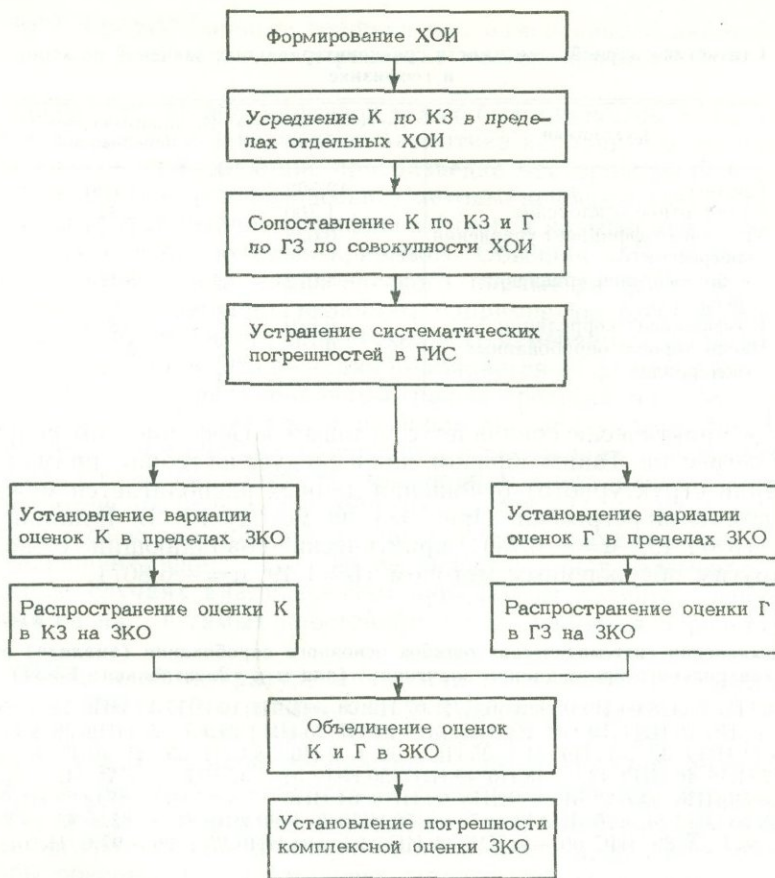


Рис. 5.6.3. Блок-схема алгоритма расчета комплексной оценки пористости по керну и с помощью геофизических исследований скважин.

и 5.6.4). Обсчет данных из табл. 5.6.1 показал следующее: если оценку пористости ЗКО с помощью ГИС считать не содержащей случайных погрешностей, а случайную компоненту расхождений между x_r и x_k относить на счет случайных погрешностей ядерной оценки ($\sigma_{\text{керна}} = \Gamma.109$), то поправки к «геофизике» совпадут с параметрами уравнения регрессии «керна на геофизику» ($\beta = 1.27$, $\alpha = -3.985$). Напротив, если $\sigma_{\text{керна}} = 0$ (или, что то же самое, $\lambda = 0$), то

Таблица 5.6.1

Средние значения пористости в хорошо опробованных интервалах, используемых для сопоставления ядерных данных с геофизическими

Номер интервала по порядку	Среднее по геофизике	Среднее по керну	Номер интервала по порядку	Среднее по геофизике	Среднее по керну
1	14.7	15.2	6	15.5	15.2
2	12.1	12.8	7	12.6	12.5
3	12.6	13.5	8	9.3	8.1
4	11.4	9.9	9	12.7	10.6
5	11.8	9.2			

Статистики парной зависимости среднеинтервальных значений по керну и геофизике

Статистики	По геофизическим данным	По данным кернового опробования
Среднее	12.522	11.890
Стандартное отклонение	1.700	2.427
Угловой коэффициент уравнения регрессии	0.623	1.270
Свободный член уравнения регрессии	-4.012	5.118
Коэффициент корреляции	0.889	0.889
Число хорошо опробованных интервалов	9	9

$\beta = 1.6 \approx \frac{1}{0.623}$ практически совпадает с угловым коэффициентом сопряженного уравнения регрессии. Таким образом, как и следует из теории, прямая функционального (или структурного) отношения данных располагается между сопряженными прямыми регрессии. При $\lambda=1$ по методу Ю. В. Рошина получаем результат ($\beta=1.49$, $\alpha=-6.74$), практически совпадающий с полученным на ЭВМ другим обсужденным методом ($\beta=1.49$, $\alpha=-6.807$).

Таблица 5.6.3

Программа выявления систематических ошибок основного опробования (анализа) относительно контрольного при заданном значении λ (для м/к «Электроника» БЗ-34)

00.П5 01.П6 02.П7 03.П8 04.П9 05.П0 06.С/П 07.П2 08.↔09.П1 10.ИП1 11.ИП5 12.+13.П5 14.ИП2 15.ИП6 16.+17.П6 18.ИП1 19. FX^2 20.ИП7 21.+22.П7 23.ИП2 24. FX^2 25.ИП8 26.+27.П8 28.ИП1 29.ИП2 30.×31.ИП9 32.+33.П9 34.1 35.ИП0 36.+37.П0 38.С/П 39.БП 40.07 41.ИП5 42.ИП0 43.÷44.П5 45.ИП6 46.ИП0 47.— 48.П6 49.ИП7 50.ИП0 51.÷52.ИП5 53. FX^2 54.— 55.П7 56.ИП8 57.ИП0 58.— 59.ИП6 60. FX^2 61.— 62.П8 63.ИП5 64.ИП6 65.×66.ИП0 67.×68.ИП9 69.↔70.— 71.П9 72.ИПА 73.ИП7 74.×75.ИП8 76.↔77.— 78.ИП0 79.×80.ИП9 81.— 82.2. 83.— 84.ПС 85. FX^2 86.ИПА 87.+88. $F\sqrt{X}$ 89.ИПС 90.+91.ПВ 92.ИП5 93.×94.ИП6 95.↔96.—97.С/П.

Таблица 5.6.4

Программа выявления систематических ошибок основного опробования (анализа) относительно контрольного при известном значении погрешности контрольного s_x^2 (для м/к «Электроника» БЗ-34)

00.П5 01.П6 02.П7 03.П8 04.П9 05.П0 06.С/П 07.П2 08.↔09.П1 10.ИП1 11.ИП5 12.+13.П5 14.ИП2 15.ИП6 16.+17.П6 18.ИП1 19. FX^2 20.ИП7 21.+22.П7 23.ИП2 24. FX^2 25.ИП8 26.+27.П8 28.ИП1 29.ИП2 30.×31.ИП9 32.+33.П9 34.1 35.ИП0 36.+37.П0 38.С/П 39.БП 40.07 41.ИП5 42.ИП0 43.— 44.П5 45.ИП6 46.ИП0 47.— 48.П6 49.ИП7 50.ИП0 51.— 52.ИП5 53. FX^2 54.— 55.П7 56.ИП8 57.ИП0 58.— 59.ИП6 60. FX^2 61.— 62.П8 63.ИП5 64.ИП6 65.×66.ИП0 67.×68.ИП9 69.↔70.— 71.П9 72.ИПА 73. FX^2 74.ИП8 75.↔76.— 77.ИП9 78.— 79.ИП0 80.×81.ПВ 82.ИП5 83.×84.ИП6 85.↔86.— 87.С/П.

Порядок счета по программам табл. 5.6.3 и 5.6.4. 1) Ввести λ (или $\sigma_{\text{ан. контр}}$ для варианта 5.6.4) в ПА. Набрать О, нажать В/О, С/П. На индикаторе появится 0. 2) После останова набрать x (основной анализ), |. Набрать y (контрольный), С/П. Останов, на индикаторе — число набранных пар. 3) Исчерпались ли пары значений? Если да, то перейти к п. 2, если нет, то к п. 4. 4) Набрать БП 41, С/П. На индикаторе прочесть значение постоянной составляющей поправки α , в ПВ — пропорциональной составляющей β , в П5 — \bar{x} , в П6 — \bar{y} , в П7 — s_x^2 , в П8 — s_y^2 , в П9 — $\text{cov}_{xy} \cdot n$, $r_{xy} = \frac{\text{cov}_{xy}}{s_x \cdot s_y}$. Пример: $x=1$, $y=2.1$; $x=2$, $y=3.9$; $x=3$, $y=6.1$. Ответ: $\alpha=0.0226$, $\beta=2.005$, $x=2$, $y=4.0(3)$, $s_x^2=0.(6)$, $s_y^2=2.68$, $\text{cov}_{xy} \cdot n=4$, $r_{xy}=0.998$. При необходимости произвести пересчет с тем же набором данных, но с другими λ или $\sigma_{\text{ан. контр}}$ заново вводить x и y не требуется. Необходимо ввести новое значение λ (или $\sigma_{\text{ан. контр}}$) в ПА и нажать счет с БП 72, С/П.

§ 7. Градуировочные графики при измерениях и анализах в геологии и разведке месторождений

Правильные результаты анализов могут быть получены только при тесном сотрудничестве геолога-заказчика и аналитика-исполнителя. Кроме того, геологи и геофизики, применяющие при разведке месторождений и опробовании различные физические (геофизические) методы, непосредственно сталкиваются с построением градуировочных графиков (иногда называемых аналитическими графиками или калибровочными кривыми). Обычная общая схема измерений заключается в получении аналитического параметра (АП) и установлении зависимости АП от измеряемой величины, например содержания. В качестве АП выступает функция от одного или нескольких измеряемых в процессе анализа аналитических сигналов. Эта функция подбирается из физических соображений таким образом, чтобы градуировочный график представлял собой прямую, т. е. подбор АП является в то же время подбором линеаризующего преобразования. Например, при геофизических измерениях в скважинах широко используется двойной разностный параметр. В эмиссионном спектральном анализе почти точной аналогией его является параметр типа $АП = (I_A - I_\Phi) / (I_C - I_\Phi)$, где I_A — интенсивность спектральной линии определяемого элемента (вместе с фоном), I_Φ — интенсивность фона, I_C — интенсивность линии элемента сравнения. В простейших случаях аналитическим параметром является непосредственно измеряемый сигнал, в самых простейших он же является определяемой величиной.

В опубликованной литературе, прямо или косвенно относящейся к геологии, имеется много сведений о градуировочных графиках (Гусинский и др., 1972; Смирнова и др., 1973; Макулов, 1974; Орлов, 1977; Каган, 1978а, б), однако сколько-нибудь законченной теории не существует. Она должна ответить на следующие практические вопросы: как выбрать АП; как аппроксимировать зависимость «АП—измеряемая величина», чтобы обеспечить минимальный вклад ошибки градуировки в общую ошибку анализа; как поставить эксперимент, чтобы обеспечить максимальную надежность градуировочного графика при минимуме анализов стандартных образцов или эталонов. Последний вопрос особенно актуален в геологии, когда условия проведения измерений «в поле» или в недрах трудно стабилизировать и когда на одно измерение (или небольшую группу измерений объекта из исследуемой совокупности) приходится делать несколько измерений стандартных образцов. В количественном спектральном анализе серия стандартных образцов снимается на каждую пластинку. В ГИС хорошо опробованные интервалы или опорные пласты подбираются для каждого каротажа «от одного числа».

При построении градуировочных графиков в геологии встречаются две сильно различающиеся ситуации. В первой для калибровки используются паспортизованные стандартные образцы (СО) с практически точно установленным значением измеряемой величины. Здесь аналитик ограничен в выборе значений измеряемой величины в СО, используемых для построения градуировочного графика. Если измеряемая величина — содержание, то это ограничение снимается возможностью разбавить СО до необходимых концентраций. К этой ситуации применима теория оптимального планирования экспериментов.

Во второй ситуации, весьма распространенной при измерениях в геологии, для построения градуировочного графика используется выборка из элементов (например, проб) изучаемой совокупности, проанализированная другим, эталонным методом. Это влечет два следствия. Во-первых, содержание становится случайной величиной с некоторым распределением, являющимся выборочным из распределения в изучаемой совокупности (в первой ситуации содержание в СО было «математической переменной», численные значения которой устанавливались исследователем). Во-вторых, содержания в этой выборке проб не точ-

ные, а имеющие некоторую погрешность. Ситуация напоминает характерную при сравнении результатов основного анализа (в его роли выступает АП) и контрольного (в его роли выступает содержание, определенное эталонным методом) и по существу идентична ей.

Обратим внимание читателя также на связь между построением градуировочного графика во второй ситуации и построением уравнения регрессии и, тем самым, — на связь между анализом и оценкой по регрессии. Параметры обеих «конструкций» оцениваются методом наименьших квадратов. Вспомогательная переменная в оценке по регрессии может быть сопоставлена с АП. Отличия здесь лишь количественные: коэффициент корреляции между АП и измеряемой величиной всегда близок к единице, а между вспомогательной и измеряемой величинами — обычно меньше. Чем же в таком случае отличается градуировочный график от уравнения регрессии? Почему для оценки по регрессии используется уравнение регрессии оцениваемой величины на вспомогательную, а для оценки результатов анализа градуировочный график строится по уравнению регрессии аналитического параметра на содержание, т. е. вспомогательной величины на оцениваемую?

Попытаемся ответить на эти вопросы. В одной из работ автора (Ткачев, Юдович, 1975) и в предыдущих разделах данной работы была обоснована идея о том, что уравнение регрессии является такой конструкцией, которая учитывает при оценке оцениваемой переменной две стороны объекта или явления: во-первых, зависимость между величинами, во-вторых, среднее значение оцениваемой величины в изучаемой совокупности. Информацию и о первом, и о втором несет выборка, на основании которой строится уравнение регрессии. Действительно, не принимая во внимание вспомогательную величину, а только на основании распределения оцениваемой величины в выборке мы можем судить о значении оцениваемой величины (y) в некотором элементе совокупности: мы можем утверждать, что $y = \bar{y}$ (\bar{y} — выборочное среднее) и средняя квадратическая погрешность этой оценки будет численно равна s_y . Чем ниже коэффициент корреляции между оцениваемой и вспомогательной величинами, тем выше статистический вес указанной информации. При $|r_{xy}|$ близких к единице, напротив, сильно увеличивается вес информации о зависимости. Таким образом, уравнение регрессии включает в себе сведения не только о зависимости, но и о распределении оцениваемой величины в изучаемой совокупности. Один и тот же физический закон, связывающий две величины x и y линейной функциональной зависимостью, будет выражаться различными уравнениями регрессии y на x в зависимости от двух обстоятельств: от степени влияния различных факторов, отклоняющих измеренные значения этих величин от прямой, характеризующей эту зависимость, и от распределения величины y в изучаемой совокупности. При его изменении изменяются и параметры уравнения регрессии. Чем больше будет дисперсия распределения, тем меньшим окажется угловой коэффициент прямой регрессии и больше свободный член. Чем больше будет \bar{y} , тем больше — свободный член (подробнее об этом см. Ткачев, Юдович, 1975, с. 193).

Вернемся к градуировочному графику. В условиях, когда для его построения в качестве СО используется репрезентативная выборка из изучаемой совокупности проанализированных эталонным методом проб и когда этот график будет использоваться при анализе проб из этой же совокупности, градуировочный график совпадает с уравнением регрессии изучаемой величины на АП. В этой ситуации именно он обеспечивает минимум средней квадратической случайной погрешности анализа. Поэтому мы (Иванова, Ткачев, 1982) называем его ситуационным. Такое название оправдано тем, что для того же метода анализа, с применением тех же материалов и аппаратуры, но для другой совокупности («в другой ситуации») градуировочный график будет иным.

Ситуационному графику мы противопоставляем фундаментальный, параметры которого не зависят от анализируемой совокупности. К фундаментальному графику весьма близок обычно применяемый в практике аналитических лабораторий график, минимизирующий отклонения АП, т. е. уравнение регрессии АП на измеряемую величину. Если последняя измерена в СО или в группе проб, их заменяющих, практически без погрешностей, то это уравнение регрессии в точности соответствует фундаментальному градуировочному графику. В общем случае он должен строиться как график структурного уравнения между АП и измеряемой величиной, например по методике сравнения основных и контрольных анализов, изложенной в предыдущем разделе. При этом аналогом результата основного анализа является АП, контрольного — соответствующее ему значение измеряемой величины в СО. Обычно средняя квадратическая погрешность измеряемой величины в СО бывает известна, поэтому применима более простая, вторая из изложенных, модификация. Мы рекомендуем еще более простой ее эквивалент, удобный тем, что требует лишь обычных стандартных программ статистической обработки. Вычисляются параметры уравнения регрессии АП на измеряемую величину x :

$$\text{АП} = a + bx, \quad (5.7.1)$$

затем по формуле, аналогичной (2.2.6), вычисляется угловой коэффициент зависимости АП от содержания b_{Φ} :

$$b_{\Phi} = b \frac{s_x^2}{s_x^2 - s_{\text{ан } x}^2}, \quad (5.7.2)$$

где s_x^2 — эмпирическая дисперсия измеряемой величины в совокупности СО, $s_{\text{ан } x}$ — средняя квадратическая погрешность оценки x в СО. При производстве анализов измеряемая величина вычисляется по значению АП:

$$x_i = a + b_{\Phi} \cdot \text{АП}_i, \quad (5.7.3)$$

где

$$b_{\Phi} = 1/b_{\Phi}, \quad a = (\overline{\text{АП}} - b_{\Phi} \bar{x}) / b_{\Phi} = \beta \overline{\text{АП}} - \bar{x}.$$

В аналитической практике, как правило, расчеты ограничиваются оценкой a и b в уравнении (5.7.1), решением его относительно x и оценкой x_i по АП_i , что неверно. Например, А. Г. Орлов (1977) в обобщающей работе, специально посвященной методам расчета в спектральном анализе, приводит методику и следующий практический пример (с. 133, формула VI-28): $\Delta s = -1.28 + 0.91 \lg C$, в котором параметры этого уравнения рассчитаны методом наименьших квадратов, как в (5.7.1), и которое непосредственно служит для оценки $\lg C$ по Δs , хотя не указано, что содержания в эталонах определены без существенных погрешностей. О пересчете углового коэффициента уравнения регрессии в угловой коэффициент структурного уравнения (5.7.2) необходимо заботиться особенно тогда, когда в качестве СО используются «свои» пробы, не охарактеризованные паспортными содержаниями.

Если аналитик имеет возможность произвольно выбрать для построения градуировочного графика серию стандартных образцов (или контрольных проб) с любыми требуемыми содержаниями, то теория оптимального планирования эксперимента позволяет значительно повысить точность градуировочного графика специальным подбором содержаний в СО. Из физических соображений, в результате предшествующего этапа или, наконец, после предварительных поисковых экспериментов должно быть установлено, полиномом какой степени аппроксимируется градуировочный график. Н. А. Макулов (1974) приводит следующие оптимальные планы:

для линейного градуировочного графика

p_i	1/2	1/2		
x_i	-1	+1		

для градуировочного графика второго порядка

p_i	1/3	1/3		1/3
x_i	-1	0		+1

для градуировочного графика третьего порядка

p_i	1/4	1/4	1/4	1/4
x_i	-1	-0.477	+0.477	+1

Здесь p_i — доля СО с $x=x_i$ (x_i — «узел» плана), $p_i=n_i/N$, x — известная функция содержания, N — общее число СО. При условии, что во всех узлах плана средние квадратические погрешности измерения АП одинаковы и равны $\sigma_{\text{АП}}$ и что содержание в СО определены без погрешностей, в теории оптимального планирования эксперимента получены следующие формулы оценки параметров α и β структурного уравнения АП $=\alpha+\beta x$:

$$\beta = \frac{\overline{\text{АП}}_2 - \overline{\text{АП}}_1}{x_2 - x_1} = \frac{\overline{\text{АП}}_2 - \overline{\text{АП}}_1}{2}, \quad (5.7.4)$$

$$\alpha = \frac{\overline{\text{АП}}_1 + \overline{\text{АП}}_2}{2}. \quad (5.7.5)$$

Погрешности их оценок вычисляются по формулам

$$\sigma_\alpha = \sigma_\beta = \sigma_{\text{АП}} / \sqrt{N}. \quad (5.7.6)$$

Дисперсия оценки содержания в пробе определяется по формуле

$$s_c^2 = (s_1^2 + s_2^2) / \left(\frac{dx}{dc} \Big|_{x=x^{\text{нр}}} \right), \quad (5.7.7)$$

здесь $s_1^2 = \frac{\sigma_{\text{АП, нр}}^2}{\beta^2}$, $s_2^2 = \frac{\sigma_{\text{АП, СО}}^2 + \beta^2 \sigma_{\text{СО}}^2}{\beta^2} \left(\frac{x^{\text{нр}} - \bar{x}_{\text{СО}}}{\sum (x_{\text{СО}} - \bar{x}_{\text{СО}})^2} + \frac{1}{N} \right)$,

где s_y^2 — дисперсия отсчета аналитического параметра в данной пробе,¹ $\sigma_{\text{АП}}^2$ — дисперсия отсчета аналитического параметра при анализе СО (обычно $s_y^2 = \sigma_{\text{АП}}^2$), β — параметр структурного уравнения, $(\sigma_x^{\text{СО}})^2$ — дисперсия оценки содержания (паспортного) в СО; x — содержание (или функция от содержания, например, $\lg C$), для которого ведется расчет s_c^2 . В частном случае количественного спектрального анализа $x = M \lg C + K$, где M — масштабный множитель, после некоторых упрощений Н. А. Макулов получает из (5.7.7) выражение для коэффициента вариации воспроизводимости выражение $v \approx 4\sigma_{\text{АП}}/\beta M$ 100%. В цитированной работе он дает также результаты для случая, когда градуировочный график аппроксимируется полиномом второй и третьей степеней. Л. М. Каган (1978а, 1978б), взяв за основу (5.7.7), устанавливает вклад градуировки s_2 в общую погрешность анализа s_c . Обозначая $L = (x_{\text{нр}} - \bar{x}_{\text{СО}})^2 / \sum (\bar{x}_i - \bar{x}_{\text{СО}})^2$, автор при равномерном распределении СО на всем интервале концентрации получает

$$L = 3(N-1)/(N(N+1)) \quad \text{и} \quad s_2^2/s_1^2 = 2(2N-1)/(N(N+1)),$$

т. е. при $N=3 \div 8$ $s_2^2/s_1^2 = 83-42\%$.

Полагая, что все анализируемые содержания (точнее 99.7%) укладываются

¹ Если АП — функция от аналитического сигнала, то s_y^2 и $\sigma_{\text{АП}}^2$ вычисляются с учетом переноса ошибок.

в интервал $\bar{x} \pm 3s_x$, длину интервала концентраций примем равной $6s_x$. Вычислим среднее значение дисперсии градуировки s_r^2 для партии проб с данным средним квадратическим отклонением содержаний s_x :

а) для набора СО, являющегося случайной выборкой из анализируемой совокупности проб

$$s_r^2 = \sigma_{\text{АП}}^2 / \beta^2 (s_x^2 / (Ns_x^2) + 1/N) = 2\sigma_{\text{АП}}^2 / (N\beta^2);$$

б) для равномерного на интервале $\bar{x} \pm 3s_x$ распределения СО (дисперсия равномерного распределения равна $h^2/12 = (6s_x)^2/12 = 6s_x^2$):

$$s_r^2 = \sigma_{\text{АП}}^2 / \beta^2 (s_x^2 / (3Ns_x^2) + 1) = 4/3 \sigma_{\text{АП}}^2 / (N\beta^2);$$

в) для случая, когда все СО — на границах изучаемого интервала концентраций (N — четное, $\sum_N (x_i - \bar{x}_{\text{СО}})^2 = N(3s_x)^2$):

$$s_r^2 = \frac{\sigma_{\text{АП}}^2}{\beta^2} \left(\frac{s_x^2}{9Ns_x^2} + \frac{1}{N} \right) = \frac{10}{9} \frac{\sigma_{\text{АП}}^2}{\beta^2 N}.$$

При равном числе СО и прочих равных условиях отношение погрешностей градуировки для рассмотренных случаев будет как 1.8 : 1.2 : 1, т. е. градуировочный график по случайной выборке проб в качестве СО почти в два раза проигрывает графику по оптимальному плану.

Каким количеством стандартных образцов можно ограничиться при построении градуировочного графика? Сравнительно легко на этот вопрос можно ответить, решая задачу минимизации погрешности анализа при заданных затратах на анализ и варьируя при этом числом эталонов и числом повторностей анализов. Практически в таком скрупулезном подходе нет необходимости, и можно ограничиться уменьшением погрешности градуировки до пренебрежимого уровня, когда отношение $s_2^2 / (s_1^2 + s_2^2)$ станет близко к единице. Для оптимального плана $s_2^2 / (s_1^2 + s_2^2) = \frac{10}{9N} / (\frac{10}{9N} + 1) \approx \frac{1}{N+1}$, что дает следующие значения вклада градуировки:

Число стандартных образцов (N)	2	3	4	5	9
Доля градуировки в общей дисперсии анализа, %	30	25	20	17	10

Совершенно не оправдана широко распространенная в геологии практика, когда выборка проб, используемых для построения градуировочных графиков, формируется ни по случайному принципу, ни по оптимальному плану. Например, Л. И. Вознесенский и др. (1971 г.) для калибровки рентгенофлуоресцентного анализа используют пробы, проанализированные химическим способом и приблизительно равномерно представляющие весь диапазон встречающихся содержаний. Судя по их градуировочному графику, он может хорошо аппроксимироваться параболой 2-го порядка, и по оптимальному плану необходимо взять пробы с тремя уровнями содержаний: 5, 20 и 35 %.

При анализе партии проб, как правило, имеется некоторая информация о концентрациях в ней анализируемых компонентов. Например, могут быть известны параметры распределения содержания в совокупности проб или пределы, за которыми концентрации химических элементов теоретически невозможны. Заметим, что полное отсутствие информации соответствует, по Байесу, гипотезе о равномерном распределении концентраций на всей числовой прямой.

Попытаемся ответить на вопрос, как использовать эту априорную информацию для повышения точности оценки результатов анализа. Аналитик, вообще говоря, может всегда придерживаться гипотезы о равномерном распределении, переложив заботу о возможной переоценке содержаний с учетом априорной

информации на заказчика, однако проблема при этом не исчезает. Кроме того, не всякий аналитик решится выдавать анализы, среди которых, например, будут встречаться значения больше 100 %.

Нередко анализу подвергаются материалы, имеющие заранее известный предел концентраций определяемых элементов: кремнистые породы или бокситы, в которых содержания SiO_2 и Al_2O_3 соответственно не могут превышать 100 %; минералы, в которых содержания минералообразующих элементов ограничены элементами, допустимыми стехиометрическими отношениями, а содержания элементов-примесей — пределом «изоморфной емкости»; сплавы, для которых известны пределы концентраций вводимых компонентов и т. д. Однако вследствие погрешностей анализа значения содержаний, снимаемые с градуировочного графика, могут оказаться выше предельных. Очевидно, что такая оценка содержаний заведомо неверна. В литературе неизвестны методы, позволяющие дать более правильную оценку с учетом информации о предельных концентрациях.

Самый простой способ решить поставленную задачу — заменить все оценки содержания выше предельных предельными. Нетрудно доказать, что такое «решение» неизбежно ведет к завышению результатов как для каждой отдельной пробы, так и в среднем для группы проб с содержаниями, близкими к предельному. Это завышение будет тем существеннее, чем больше погрешность анализа и чем ближе содержание в анализируемых пробах к предельному.

В общем виде задача заключается в построении ситуационного графика, т. е. уравнения регрессии $x = \varphi(y)$ (содержания на аналитический сигнал) по известным: 1) фундаментальному градуировочному графику; 2) предельным содержаниям c_{\max} , c_{\min} ; 3) аналитическим погрешностям, измеряемым дисперсией σ_x^2 . Искомое уравнение регрессии $x = \varphi(y)$ является линией математических ожиданий содержаний, соответствующих заданному значению аналитического сигнала. Из этого вытекает следующий эмпирический способ его построения: объединяются анализы эталонов с заданными аналитическими сигналами y_i ; подсчитываются для них средние значения содержаний, т. е. величины $\bar{x}_{y=y_i}$; через точки с координатами $(\bar{x}_{y=y_i}, y_i)$ проводится искомый график. Такой способ построения неудобен, так как требует анализа большого числа эталонов с содержаниями, близкими к предельному. Обычно в спектральном анализе стремятся ограничиться минимумом эталонов (2—3 на один порядок концентраций), поскольку фотографирование большого числа эталонов на одну пластинку соответственно уменьшает количество анализируемых проб на этой пластинке и, следовательно, снижает производительность анализа. Кроме того, полученный график будет всего лишь приближенной оценкой ситуационного графика. Представляется более целесообразным аналитический (в математическом смысле) способ, позволяющий получить более надежные результаты.

Пусть зависимость между аналитическим параметром (почернением спектральной линии) и содержанием описывается уравнением

$$y = a + bx + \varepsilon_y, \quad (5.7.8)$$

где ε_y — случайная погрешность, характеризуемая дисперсией σ_ε^2 . Тогда дисперсия разброса точек по оси абсцисс (т. е. по оси содержаний) относительно прямой, описываемой уравнением (5.7.8), будет иметь вид

$$\sigma_x = \sigma_\varepsilon / b. \quad (5.7.9)$$

Если величина ε_y распределена нормально, а уравнение зависимости линейно, то погрешность оценки содержаний по аналитическому сигналу также распределена нормально с дисперсией σ_x^2 , определяемой по формуле (5.7.9). Поэтому математические ожидания содержания x_i при $y = y_i$ в условиях отсутствия пре-

дела лежат на линии регрессии (5.7.8). Распределение погрешностей содержания ε_x при заданном аналитическом сигнале в этом случае симметрично относительно точки x_i на прямой, описываемой уравнением (5.7.8). Отсюда математическое ожидание величины ε_x равно

$$M(\varepsilon_x) = \int_{-\infty}^{+\infty} \varepsilon_x f(\varepsilon_x) d\varepsilon_x = 0, \quad (5.7.10)$$

где $f(\varepsilon_x)$ — плотность вероятности нормального распределения. При наличии пределов содержания c_{\min} и c_{\max} математическое ожидание погрешности для каждого значения y (и, следовательно, x) в общем случае не равно нулю и определяется из уравнения

$$M^*\{\varepsilon_x\} = \int_{c_{\min}}^{c_{\max}} \varepsilon_x f(\varepsilon_x) d\varepsilon_x. \quad (5.7.11)$$

Преобразовать уравнение регрессии y на x в сопряженное уравнение регрессии x на y в этих условиях можно решением уравнения $y = ax + b$ относительно x :

$$x = 1/b y - a/b$$

и прибавлением к результату величины $M^*\{\varepsilon_x\}$:

$$x^* = x + M^*\{\varepsilon_x\}. \quad (5.7.12)$$

Величину $M^*\{\varepsilon_x\}$ необходимо рассматривать как функцию трех аргументов: x , σ_{ε_x} и $c_{\text{пред}}$. Это достигается следующим преобразованием равенства (5.7.11):

$$M^*\{\varepsilon_x\} = \int_{t_1}^{t_2} \varepsilon_x f(\varepsilon_x) d\varepsilon_x, \quad (5.7.13)$$

где $t_1 = (c_{\min} - x)/\sigma_x$, $t_2 = (c_{\max} - x)/\sigma_x$.

Таким образом, $M^*\{\varepsilon_x\}$ является математическим ожиданием величины, усеченной в точках t_1 и t_2 , распределенной по нормальному закону. Учитывая, что ширина интервала содержаний $[c_{\min}, c_{\max}]$ велика по сравнению с σ_{ε_x} , достаточно ограничиться односторонним усечением, считая при этом, что, когда x приблизится к c_{\min} , распределение усечено только слева, когда к c_{\max} , — справа. Значения $M^*\{\varepsilon_x\}$ для такого одностороннего усечения приведены в табл. 5.7.1. Из формул (5.7.11) и (5.7.13), а также из таблицы видно, что $M^*\{\varepsilon_x\}$ равна нулю только в точке x_0 , равноудаленной от пределов содержания, т. е. построенный ситуационный график теоретически совпадает с фундаментальным только в этой точке. С приближением точки x_0 к одному из пределов модуль величины $M^*\{\varepsilon_x\}$, а вместе с ней и $|x - x^*|$ увеличивается. При беспредельном увеличении аналитического сигнала значение x^* асимптотически стремится к c_{\max} , и ситуационный график все более и более отклоняется от фундаментального. Практически заметные расхождения наблюдаются для участков, на которых $|c_{\text{пред}} - x| < 2\sigma_x$.

Методика преобразования фундаментального градуировочного графика в ситуационный, учитывающий предельные концентрации, показана на рис. 5.7.1 и заключается в следующем. На фундаментальный график наносится вертикальная прямая $x = c_{\max}$. Вблизи точки пересечения графика с этой прямой проводится несколько вспомогательных прямых, параллельных оси концентраций. На каждой вспомогательной прямой измеряется длина отрезка, заключенного между графиком и прямой $x = c_{\max}$, и делится на σ_x , т. е. определяется величина t . Если $x > c_{\max}$, то t отрицательна, иначе — положительна. По табл. 5.7.1 для каждого t определяется значение $M^*\{\varepsilon_x\}$. Умноженные на σ_x , они

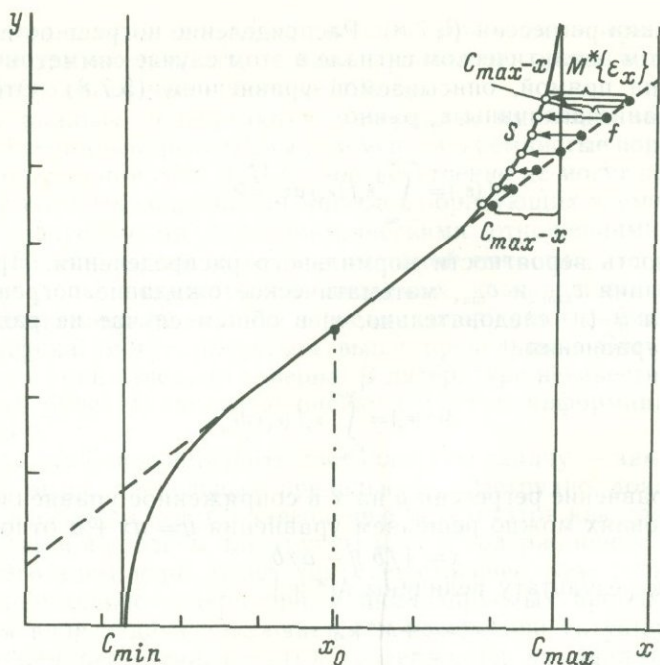


Рис. 5.7.1. Преобразование фундаментального градуировочного графика f в ситуационный S при наличии пределов содержаний c_{min} , c_{max} .

Стрелками показан перенос точек фундаментального графика (черные кружки) на ситуационный (светлые кружки). Величина переноса вычисляется по табл. 5.7.1.

определяют величину смещения влево соответствующих точек фундаментального графика. Для c_{min} преобразование графика производится аналогичным образом: смещение точек вправо.

Практически при достаточно малых σ_x полученный ситуационный график на большей части своей длины совпадает с фундаментальным, и лишь при содержаниях, приближающихся к предельным, резко отклоняется от него, асимптотически приближаясь к вертикалям $x=c_{пред}$.

Предложенный метод построения градуировочного графика был применен нами для количественного спектрального определения SiO_2 в нерастворимых остатках карбонатных пород (рис. 5.7.2). Они часто имели преимущественно

Таблица 5.7.1

Абсолютное значение математического ожидания усеченной нормально распределенной величины $M^*\{\epsilon_x\}$

Целая часть t	Десятые доли t									
	0	1	2	3	4	5	6	7	8	9
3	0.004	0.003	0.002	0.002	0.001	0.001	—	—	—	—
2	0.056	0.045	0.036	0.029	0.023	0.018	0.014	0.010	0.008	0.006
1	0.289	0.253	0.221	0.191	0.164	0.140	0.118	0.099	0.083	0.068
0	0.798	0.736	0.676	0.618	0.563	0.510	0.460	0.413	0.369	0.328
-0	0.798	0.863	0.930	1.000	1.071	1.144	1.219	1.295	1.373	1.452
-1	1.532	1.614	1.697	1.781	1.867	1.955	2.039	2.127	2.214	2.302
-2	2.392	2.482	2.571	2.661	2.749	2.838	2.926	3.011	3.096	3.171
-3	3.242	—	—	—	—	—	—	—	—	—

Примечание. Если $t=t_2$, то значение $P^*\{\epsilon_x\}$ отрицательно, если $t=-1$ — положительно.

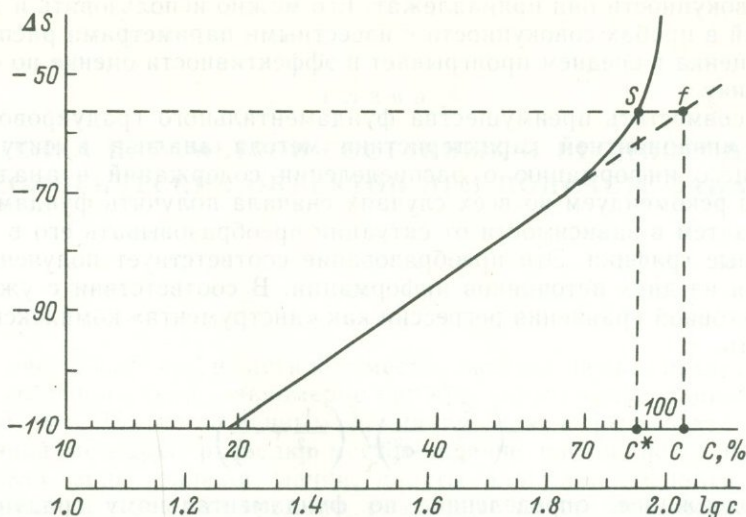


Рис. 5.7.2. Практический пример ситуационного графика при анализе кремнеземистых нерастворимых остатков карбонатных пород.

C — содержание, считываемое с фундаментального графика f (возможны значения больше 100 %); C^* — содержание для такого же значения аналитического сигнала, считываемое с ситуационного графика S .

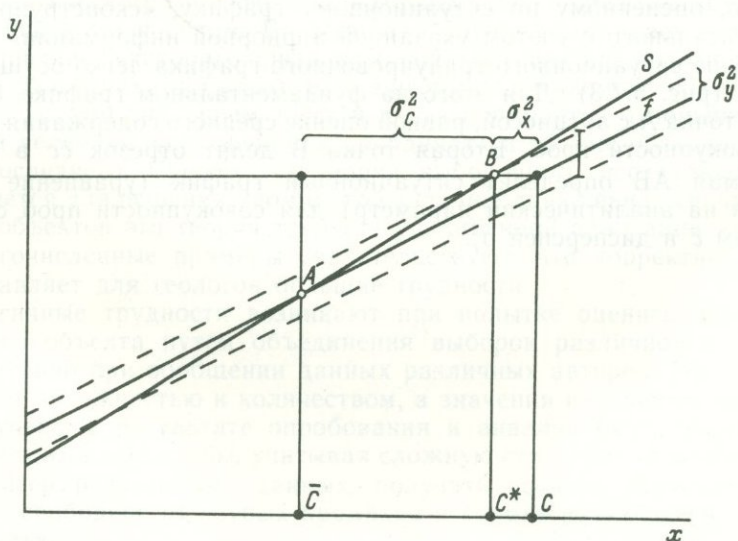


Рис. 5.7.3. Преобразование фундаментального градуировочного графика f в ситуационный S при нормальном распределении содержаний в совокупности анализируемых проб.

Точка A соответствует среднему содержанию в совокупности представляемых для анализа проб, точка B получена делением отрезка длины $C - \bar{C}$ в пропорции σ_c^2 / σ_x^2 .

кварцевый состав, и содержание кремнезема в них было близко к 100 %. По обычным графикам в этих случаях мы получали до 110 % SiO_2 , что заведомо неверно. Изложенный метод позволил получить более корректные результаты, никогда не выходящие за допустимые пределы.

Фундаментальный график, вообще говоря, предназначен для оценки содержаний в пробах, проанализированных данным методом, независимо от того,

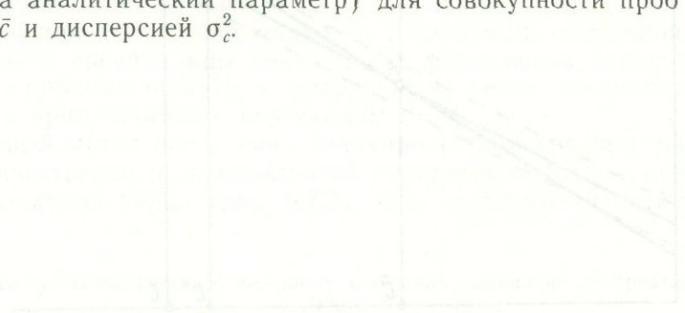
к какой совокупности они принадлежат. Его можно использовать и для оценки содержаний в пробах совокупности с известными параметрами распределения, но такая оценка в среднем проигрывает в эффективности оценке по ситуационному графику.

Чтобы совместить преимущества фундаментального градуировочного графика как инвариантной характеристики метода анализа и ситуационного, учитывающего информацию о распределении содержаний в анализируемых пробах, мы рекомендуем во всех случаях сначала получить фундаментальный график, а затем в зависимости от ситуации преобразовывать его в различные ситуационные графики. Это преобразование соответствует получению оценки содержания из двух источников информации. В соответствии с уже излагавшейся трактовкой уравнения регрессии как «инструмента» комплексной оценки будем иметь

$$c^* = \left(\frac{\bar{c}}{\sigma_c^2} + \frac{c}{\sigma_x^2} \right) / \left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_x^2} \right), \quad (5.7.14)$$

где c — содержание, определенное по фундаментальному градуировочному графику; σ_x^2 — дисперсия оценки этого содержания ($\sigma_x^2 = \sigma_{a.c}^2 / b^2$, где $\sigma_{a.c}^2$ — дисперсия аналитического сигнала при $c = \text{const}$, b — угловой коэффициент градуировочного графика); \bar{c} — среднее содержание в совокупности анализируемых проб; σ_c^2 — дисперсия содержаний в ней (две последние величины и составляют априорную информацию). Величина c^* будет соответствовать содержанию, оцененному по ситуационному графику, «сконструированному» из фундаментального с учетом указанной априорной информации.

Построение ситуационного градуировочного графика легко осуществляется графически (рис. 5.7.3). Для этого на фундаментальном графике (прямая f) отмечается точка A с абсциссой, равной оценке среднего содержания в анализируемой совокупности проб. Вторая точка B делит отрезок $\bar{c}c$ в пропорции σ_c^2 / σ_x^2 . Прямая AB определит ситуационный график (уравнение регрессии содержания на аналитический параметр) для совокупности проб со средним содержанием \bar{c} и дисперсией σ_c^2 .



ОЦЕНКА ПАРАМЕТРОВ СОСТАВНЫХ И НЕОДНОРОДНЫХ
ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ ПРИ ПОДСЧЕТЕ ЗАПАСОВ

Геологический объект, в частности месторождение, нельзя представить в виде одного однородного тела, равномерно опробованного (разведанного) по всему своему объему. По естественным причинам или в результате применения определенной методики разведки месторождение расчленяется на отдельные части — тела, зоны, залежи, пачки, пласты или блоки, фланги, горизонты, этажи, панели и т. п., как правило, различные по статистическим свойствам изучаемых параметров и разведанные с различной точностью. Например, месторождения нефти состоят из отдельных залежей, залежи — из пластов, а в пределах пласта такой параметр, как пористость, может оцениваться отдельно для разных типов коллекторов. Нередки ситуации, когда объектом изучения является ряд месторождений региона, класс месторождений определенного генетического типа, формации горных пород в пределах крупных геоблоков и земной коры в целом. Во всех этих случаях возникает задача оценить среднее значение некоторого признака (свойства), не совсем точно называемого геологами-разведчиками параметром, в целом по изучаемому объекту. Как правило, под средним значением здесь и далее понимается отношение интеграла изучаемой величины (как функции координат) в объеме объекта к массе или объему последнего. Существует теория выборочного исследования, которая и применяется для решения упомянутой задачи, однако ввиду специфики геологических объектов эта теория требует существенной адаптации и специализации. Многочисленные примеры свидетельствуют, что корректное применение ее представляет для геологов большие трудности.

Аналогичные трудности возникают при попытке оценить параметры геологического объекта путем объединения выборок различной представительности, особенно при обобщении данных различных авторов. При этом данные отличаются надежностью и количеством, а значения изучаемых величин могут быть получены в результате опробования и анализа различными методами. Задача состоит в том, чтобы, учитывая сложную структуру объекта и разнородный характер разведочных данных, получить среднее значение изучаемого признака,¹ свободное от грубых промахов и характеризующееся минимальной погрешностью.

Среднее, полученное в результате объединения различных данных или характеризующее составной объект, будем называть сложным. Хотя необходимость сложных средних в геологии существовала всегда, особенно в региональной геохимии, металлогении и разведке месторождений, их получение превратилось в специальную задачу начиная от работ Д. А. Родионова и В. В. Иванова (1967).

Способам получения сложных средних в геологии посвящена обширная литература. В литературе по нефти и газу среди этой литературы значительная часть посвящена оцениванию коллекторских свойств залежей по ядерным и гео-

¹ Вместо этого словосочетания в дальнейшем будем часто писать коротко: среднее.

физическим данным. В этой главе мы делаем попытку систематически изложить методы оценивания сложных средних в геологии на примере рудных и нефтяных месторождений при подсчете запасов, а также в геохимии. В литературе по математической статистике оценкой (параметра) называют величину, а среди геологов, особенно геолого-экономического направления, оценкой называют процедуру. Поэтому нам придется отказаться от строгости в терминологии и писать «геолого-экономическая оценка (вместо — оценивание) месторождения», но «оценивание содержания».

§ 1. Иерархия объектов изучения в геологии и разведке месторождений. Структура геологоразведочной выборки

При изучении геологических объектов выделяют многочисленные уровни их строения от минерального до формационного и глобального: так называемые уровни неоднородности. В связи с опробованием горных пород при инженерно-геологических изысканиях уровни неоднородности детально рассматривались в монографии М. В. Раца (1968). Применительно к разведке и опробованию месторождений полезных ископаемых уровни неоднородности подвергались исследованию Л. И. Четвериковым (1968). В нефтяной геологии они неоднократно обсуждались В. Я. Воробьевым и М. М. Элланским (1979).

Месторождения полезных ископаемых рассматриваются перечисленными и другими исследователями как сложные системы, состоящие из отдельных элементов различного масштаба. Ортодоксальные представители такого направления (см., например, Давтян, 1978) считают, что на разных стадиях разведки в процессе ее проведения объектами изучения становятся все более мелкие структурные единицы геологических тел. На стадии промышленной детальной разведки действительно имеет место изучение мелких структурных составных единиц месторождения, в связи с чем неизбежно возникает необходимость оценивания параметров сложной системы тел, которые сами состоят из более мелких единиц.

Уровни строения тел выделяются, конечно, не только в геологии. Часто говорят об уровнях строения биологических объектов: уровне клеток, тканей, органов или систем органов (система кровообращения). Особенностью геологических объектов является то, что их уровни строения и выделяемые структурные единицы не только определяются природными факторами, но зависят также от способа их изучения. Методика разведки месторождений в значительной степени определяет выделение составных частей разведкуемого тела для их раздельного оценивания. Так, при разведке вертикальными скважинами выделяются залежи, пласты, пропластки. Выделить отдельные тела в латеральном направлении редкой сетью скважин практически невозможно. При разведке горизонтами горных выработок, напротив, хорошо выделяются отдельные зоны концентраций и т. д., а в вертикальном направлении по необходимости приходится ограничиваться средними характеристиками. Многозалежное нефтяное месторождение состоит из отдельных залежей, залежи — из пластов, пласты при определенной методике разведки расчленяются на отдельные пористые и проницаемые интервалы коллектора. Методикой разведки определяется подразделение последних на интервалы долбления, в которых группируются образцы керна. Самостоятельными объектами оценивания могут быть отдельные блоки или фланги месторождения. При необходимости на рудных месторождениях оценка может проводиться раздельно для различных типов руд и вмещающих пород, а на месторождениях нефти — для различных типов коллекторов (Митрофанов, Тульбович, 1974). Показательный пример раздельной оценки коллекторов месторождения с последующим усреднением содержится в статье А. Е. Нелепченко и Т. Л. Гореловой (1975).

Рассмотрим понятие структуры выборки. Мы не ставим себе целью

дать формальное определение этого понятия (хотя на каком-то этапе развития теории опробования месторождений оно станет необходимым), а ограничимся лишь содержательным описанием примеров структуры выборок. Случайная выборка из такой, например, обследуемой совокупности, как урна с шарами, конечно является простой. Систематическая выборка тоже может иметь простую структуру (в выборку отбирается каждый n -ый элемент пронумерованной совокупности).

Если изучаемую совокупность разбить на подсовкупности и из каждой подсовкупности получить случайную или систематическую выборку, то общая выборка получится сложной. При этом неважно, по какому принципу произведена разбивка на подсовкупности, быть может даже случайным образом (хотя при этом теряется смысл разбиения на подсовкупности). Выборки могут быть получены не из всех подсовкупностей. Предварительно может быть получена случайная или систематическая выборка из этих подсовкупностей.

Рассмотрим важнейшие особенности геологических совокупностей и выборок из них. Во-первых, значение изучаемого признака является функцией координат. Такие величины получили название пространственных переменных. В общем случае значения этих переменных коррелируют друг с другом, причем зависимость усиливается с уменьшением расстояния между точками (пробами). Во-вторых, элементами совокупностей и выборок при разведке месторождений могут быть как естественные объекты (тела, жилы, минеральные зерна и т. д.), так и, чаще, искусственно выделяемые участки (пробы, горные выработки) заданного объема и формы и ориентировки, как называют это в геостатистике, — заданной геометрии. Без описания физической сущности элементов и их геометрии любые статистические характеристики и основные на них выводы теряют всякий смысл. В-третьих, выборочная процедура, которая заключается в разведке и опробовании месторождений, накладывает экономические ограничения на взаимное расположение проб. Доступ в недрах к точке отбора пробы пока невозможен без неизбежного при этом доступа к множеству других точек, соединяющих ее с дневной поверхностью. Это очевидное обстоятельство определяет сложность и характер структуры реальной разведочной выборки. Пробы тесно располагаются по горным выработкам и скважинам, которые в свою очередь нанизаны на другие горные выработки и т. д. — в несколько ступеней. Обычно при разведке месторождений образуется двух-, трех- или четырехступенчатая выборка. Например, при разведке крутопадающего пластового месторождения большой мощности образуется три ступени: пробы — элементы первой ступени, группируются в пластопересечения — элементы второй ступени, которые физически представляют собой орты, или рассечки. Последние проходят из штреков или штолен, образующих элементы третьей ступени.

При разведке нефтяных месторождений естественно образуется следующая двух- или трехступенчатая структура: керновые пробы (или интервалы геофизических измерений) — элементы первой ступени, скважины — элементы второй ступени. Если скважины расположены по профилям, последние образуют элементы третьей ступени. Если каждая проба в целях получения более достоверных результатов подвергается анализу несколько раз, то повторности анализов мы вправе считать разведочными элементами и выделять для них самостоятельную ступень. Аналогично, если проба состоит из стандартного числа штуфов или секций, последние также можно выделить в особую ступень. Характерная структура выборки изображена на рис. 6.1.1.

На практике структура геологоразведочной выборки индивидуальна на каждом месторождении. На нефтяных месторождениях усложняется расчленением пласта на интервалы коллектора. На нее накладывает отпечаток технология бурения: ввиду подъема керна отдельными рейсами координаты образцов керна определяются интервалом долбления. Так как существенная часть керна

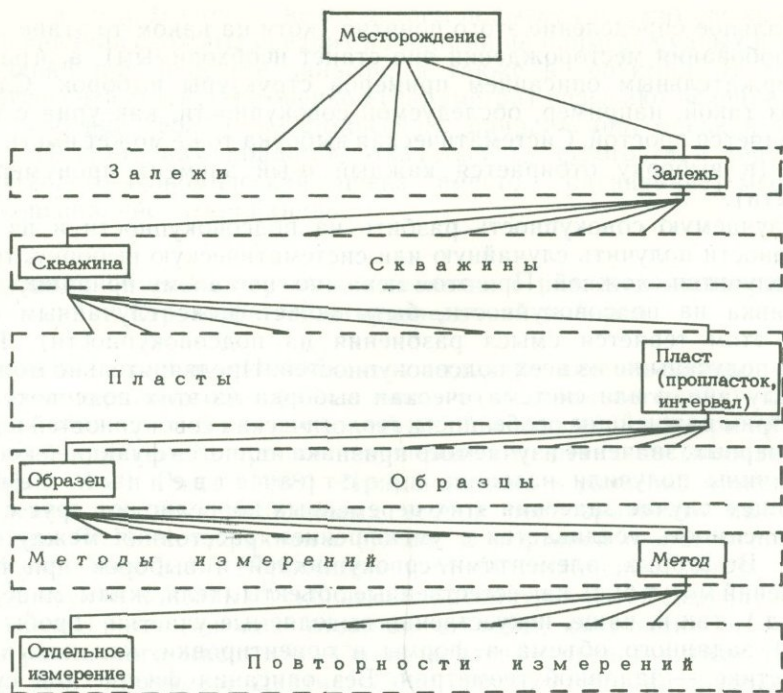


Рис. 6.1.1. Упрощенная структура геологоразведочной выборки на месторождениях нефти.

не выносятся, в пределах этого интервала может быть установлена только последовательность образцов и почти никогда не бывают известны глубины отбора.

Выделение ступеней в значительной степени условно и зависит от цели. С одной стороны, структура выборки, отражая строение изучаемого объекта и методику его разведки, определяет методику расчета среднего значения свойства по всему объекту. С другой стороны, структура является объектом оптимизации. Для этой последней цели, т. е. для определения наилучшего при фиксированных затратах соотношения числа элементов различных ступеней, их выделение, естественно, не должно носить условного характера. Каждая ступень должна быть представлена реальными объектами, т. е. выработками, при проходке которых совершаются затраты. Например, группирование скважин вдоль профилей не дает оснований для выделения особой ступени.

Другое дело, если структура выборки рассматривается в связи с выбором способа усреднения. При оценке параметра по выборке с простой структурой расчет среднего ведется в один этап. При сложной структуре выборки лучшая оценка параметра может быть получена усреднением по ступеням. Например, вначале определяется оценка среднего по пропласткам, затем — по каждой скважине, наконец по среднескважинным получают оценку общего среднего. Статистические веса каждого элемента каждой ступени определяются в зависимости от геометрии сети опробования и других факторов. Выделенные ступени в этом случае не обязательно должны представлять собой физические реальности.

Особенностью геологоразведочной выборки является также фиксированный или почти фиксированный порядок получения элементов выборки. Скважины на месторождении бурятся не все сразу, а в определенной последовательности по утвержденному проекту, допускающему лишь незначительные отклонения.

При бурении каждой скважины последовательность извлечения образцов определяется глубиной их отбора.

Наконец, важной особенностью геологоразведочной выборки является невозможность постепенного увеличения плотности опробования (или плотности разведочной сети). Если при разведке используется равномерная сеть, наиболее рациональная почти во всех отношениях, то на каждом последующем шаге разведки ее можно уплотнить лишь в целое число раз. При этом получается слишком большой скачок точности разведки и слишком большое ее удорожание — удвоение стоимости разведки после каждого шага уплотнения, что обесценивает тонкие методы расчета оптимальной плотности сети. Если же на деле осуществлять сформулированный В. М. Крейтером принцип последовательных приближений,¹ т. е. проводить анализ плотности сети после каждой или нескольких пробуренных скважин и увеличивать детальность разведки добавлением единичных скважин в сеть, нарушается крейтеровский принцип равномерной разведанности месторождения. Несмотря на очевидность перечисленных особенностей структуры и других свойств геологоразведочной выборки, во многих работах даются рекомендации по «необходимому числу» проб, основанные на представлении об одноступенчатой простой выборке проб из месторождения, не имеющие ничего общего с реальным положением вещей.

Как частный случай особой структуры выборки необходимо рассматривать «параллельные» наборы данных различных авторов или разных методов исследования, относящихся к одному и тому же объекту, и наборы, имеющие одинаковую или различную структуру. Сравнение и объединение данных различных авторов выходят за рамки задач геологической разведки. В частности, на актуальность сравнения экспериментальных данных двух литературных источников указывает А. Г. Орлов (1978).

Ниже мы предлагаем классификацию и анализ ситуаций оценки сложных средних, наиболее часто встречающихся в геологии и при разведке месторождений. В основе даваемых при этом рекомендаций лежит известная в математической статистике теория слоевого опробования, и свою роль мы усматриваем лишь в соединении известного с известным, дополняя необходимые звенья.

§ 2. Оценка среднего по данным нескольких серий измерений

Речь будет идти о расчете среднего значения некоторой изучаемой изменчивой величины в месторождении по данным измерений, выполненных либо различными авторами, либо различными методами, либо и то, и другое. Чтобы точнее идентифицировать случай, который будет описан в данном разделе главы, напомним, что в математической статистике различают два рода средних. Пусть несколько измерений подвергается одна и та же величина. Каждый элемент образцовой при этом совокупности случайных величин есть сумма истинного значения и погрешности измерения. Среднее значение первого рода есть оценка этого истинного значения. Не будь погрешностей измерений, истинное значение можно было бы получить с помощью единичного измерения. Примеры: плотность данного минерального зерна, содержание компонента в данной пробе, мощность пласта в данном пластопересечении, максимальная протяженность рудного тела и т. д.

Среднее второго рода характеризует совокупность измеренных объектов или частей изменчивого объекта. Примером может служить средняя мощность протяженного рудного тела, средняя высота некоторого участка гористой

¹ Не следует путать с математическим методом последовательных приближений.

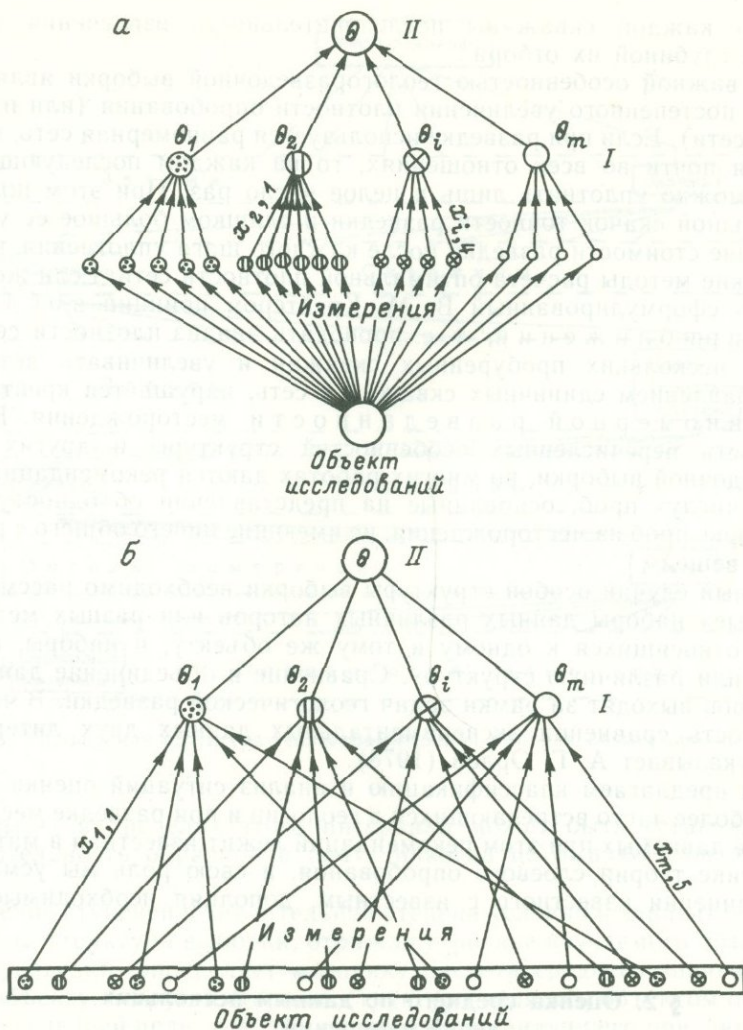


Рис. 6.2.1. Схема, иллюстрирующая объединение оценок.

а — единично-измеряемого объекта несколькими сериями измерений; *б* — совокупно-измеряемого объекта несколькими системами опробования; θ — общая оценка, $\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_m$ — оценки по сериям, $x_{i,j}$ — отдельные измерения; I, II — уровни обобщения данных.

местности, среднее содержание в совокупности проб. Такое подразделение условно и зависит от принятых представлений об объекте изучения. Но практически оно однозначно. Никто не будет утверждать, что n взвешиваний одного и того же валуна (для увеличения точности) то же самое, что и взвешивание n систематически отобранных валунов для оценки среднего веса их в данной аллювиальной свите. Этот раздел посвящен средним первого рода.

С целью повышения точности измерения проводят несколько раз или выполняют несколько серий измерений различными методами (каждую серию — своим методом). Пусть имеется m серий измерений одной и той же величины различными методами (рис. 6.2.1). Это могут быть радиометрические и химические анализы одной и той же пробы, измерения мощности пласта в скважине с помощью различных видов каротажа, определение глубины залегания интересующего нас горизонта различными геофизическими методами и т. д.

Каждая серия может состоять из одного или n_i измерений; по данным всех серий измерений необходимо получить среднее значение, обладающее минимальной дисперсией. Известно (например, Урбах, 1963), что это достигается усреднением оценок средних θ_x , вычисленных для отдельных серий измерений, с весовыми коэффициентами, обратно пропорциональными дисперсиям этих оценок. Значение θ_i по каждой серии определяется из n_i измерений очевидным образом:

$$\theta_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (6.2.1)$$

Дисперсия оценки среднего по i -й серии $\sigma_{\theta_i}^2$ определяется дисперсией i -го метода измерения σ_i^2 и числом измерений n_i . Таким образом

$$\hat{\theta} = \left(\sum_{i=1}^m \frac{\theta_i}{\sigma_{\theta_i}^2} \right) / \left(\sum_{i=1}^m \frac{1}{\sigma_{\theta_i}^2} \right) = \left(\sum_{i=1}^m \frac{n_i \theta_i}{\sigma_i^2} \right) / \left(\sum_{i=1}^m \frac{n_i}{\sigma_i^2} \right) \quad (6.2.2)$$

Если все применяемые методы равноточны, то весовые коэффициенты равны числу измерений в каждой серии, что соответствует вычислению среднего арифметического по объединенной выборке. При неравноточности применяемых методов среднеарифметическая оценка может сильно исказить истинное значение.

Для дисперсии оценки параметра, вычисленного по формуле (6.2.2), Д. А. Родионов и В. В. Иванов (1967) получили выражение

$$\sigma_{\hat{\theta}}^2 = 1 / \sum_{i=1}^m \frac{n_i}{\sigma_i^2}, \quad (6.2.3)$$

являющееся обратной величиной суммы весовых коэффициентов. Рассматривая эффективность оценки среднего как величину, обратную квадрату погрешности, можно сказать, что эффективность оценки среднего (при условии взаимной независимости серий измерений) равна сумме эффективностей оценок по данным каждой из этих серий. Легко убедиться, что дисперсия среднего арифметического из m серий

$$\sigma_x^2 = \frac{1}{m^2} \sum_{i=1}^m \sigma_{\theta_i}^2 \quad (6.2.4)$$

Таблица 6.2.1

Пример оценивания среднего по данным двух разноточных видов анализа

Характеристики	Метод 1	Метод 2
Число повторений	8	10
Средняя квадратическая погрешность метода	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$
Дисперсия воспроизводимости метода	$1 \cdot 10^{-6}$	$4 \cdot 10^{-6}$
Среднее в серии	0.012	0.009
Расчет общего среднего	$\hat{\theta} = \frac{0.012 \cdot 8 / 1 \cdot 10^{-6} + 0.009 \cdot 10 / 4 \cdot 10^{-6}}{8 / 1 \cdot 10^{-6} + 10 / 4 \cdot 10^{-6}} = 0.0173$	
Расчет дисперсии оценки общего среднего	$\sigma_{\hat{\theta}}^2 = 1 / (8 / 1 \cdot 10^{-6} + 10 / 4 \cdot 10^{-6}) = 1 \cdot 10^{-7}$	
Средняя квадратическая погрешность общего среднего	$\sigma_{\hat{\theta}} = \sqrt{1 \cdot 10^{-7}} \approx 3.2 \cdot 10^{-4}$	
Запись результата, %	$\hat{\theta} = 0.0113 \pm 0.0032$	

не меньше, чем σ_0^2 , и равна ей только при равноточных сериях измерений. Этим и оправдывается применение формулы (6.2.2) для оценки средних значений.

Пример. Определить содержание урана в эталонной пробе по данным двух различных методов радиометрического анализа. Исходные данные, схема расчета и его результаты приведены в табл. 6.2.1.

§ 3. Оценка среднего по объединенным данным независимых систем опробования

В этом разделе мы рассматриваем оценку среднего значения какого-либо свойства в изменчивом объекте, например среднего содержания химического элемента в массиве горной породы, средней пористости пласта или запаса нефти в месторождении, т. е. оценку среднего второго рода. Ее нельзя определить одним измерением (по крайней мере, при современном уровне техники), а получают путем усреднения большого числа измерений или анализов проб, отобранных из различных участков объекта. Задача заключается в расчете общей оценки среднего при объединении результатов, полученных по данному объекту различными исследователями, — каждый своей системой опробования, расположенной независимо от системы опробования других авторов. Данные могут отличаться методикой их получения и достоверностью, причем в литературе приводятся не исходные материалы, а только средние значения и погрешности их оценок. Задача заключается в расчете среднего с использованием всех данных.

Среднее значение и погрешность его оценки по объединенной выборке определяются по формулам (6.2.2) и (6.2.3) предыдущего раздела. Отличительная особенность: погрешности оценок частных средних $^1 \sigma_{\theta_i}$ здесь вовсе не равны погрешностям оценки отдельных значений, деленным на число наблюдений. Величины $\sigma_{\theta_i}^2$ по каждой системе опробования должны определяться в соответствии с основными положениями геостатистики. Их расчет представляет самостоятельную и часто очень сложную задачу, которая здесь не рассматривается. Лишь в частном случае, когда между значениями признака в соседних пробах не наблюдается существенной автокорреляции, будет выполняться равенство $\sigma_{\theta_i} = \sigma_i / \sqrt{n_i}$, где σ_i^2 — дисперсия признака в совокупности проб i -й системы опробования.

Уточним требование независимости серий измерений или систем опробования. Независимость друг от друга двух серий измерений равносильна независимости случайной погрешности каждого измерения первым методом от случайной погрешности каждого измерения вторым методом, т. е. все φ_{ij} случайных погрешностей должны быть попарно независимы ($i=1, m, j=1, n_i$). Из этого условия вытекает необходимая для применения формул (6.2.2) и (6.2.3) независимость случайной погрешности оценки среднего одним методом от таковой — другим методом. Для проверки этого условия потребовалось бы несколько раз повторить серии измерений обоими методами. В геологической разведке эту проверку сделать невозможно по техническим и экономическим причинам, и вряд ли она необходима — обычно из физических соображений независимость очевидна.

Независимость систем опробования заключается в том, что положение точек опробования одной системы не зависит от расположения точек отбора проб (измерений) другой системы. Если на месторождении, разбуренном скважинами, два исследователя определяли пористость в образцах, «независимо» друг

¹ Под частным средним (в отличие от общего среднего) в этой главе мы будем подразумевать среднее, полученное или одним из методов опробования, или одним из авторов, или по одной из частей опробуемого объекта.

от друга отобранных из керна скважин, то требуемой независимости для оценки среднего по месторождению нет: положение образцов в месторождении для обоих исследователей определяется скважинами и наличием столбиков неразрушенного керна. Здесь независимость выполняется только в пределах отдельных скважин, в связи с чем *находить совокупную оценку пористости по данным двух авторов надо по каждой скважине отдельно, а затем получить оценку по залежи.* Объединять две оценки по залежи без учета зависимости между этими оценками было бы некорректно. Сильно коррелируют, например, оценки среднескважинных значений пористости по керну и геофизическим измерениям, поэтому здесь также нельзя проводить объединение керновой и геофизической оценок пористости на уровне средней по залежи. Повторное «независимое» опробование тех же горных выработок незначительно уменьшает случайную погрешность общей оценки среднего по месторождению именно вследствие высокой коррелированности объединяемых оценок. Дисперсию оценки среднего по месторождению можно подразделить на два слагаемых: дисперсию оценки среднего в объеме горных выработок и дисперсию распространения этого среднего на весь объем месторождения. При повторном опробовании горных выработок первое слагаемое уменьшается, второе — нет. Вообще следует придерживаться общего принципа — объединять оценки, если их несколько, на возможно более низких ступенях многоступенчатой выборки.

Пример. Предварительная разведка месторождения, представленного меденосным пластом, осуществлялась сетью скважин, а детальная разведка — с помощью горных выработок. Их сеть расположена случайно по отношению к сети скважин.

Первоначально вычисляются содержание и его дисперсия отдельно по сети скважин и горным выработкам. В данном примере эти величины будем считать заданными. Требуется объединить эти данные и получить среднее содержание меди и оценку его точности (табл. 6.3.1).

Таблица 6.3.1

Пример оценивания среднего по данным двух видов опробования

Характеристики	По скважинам	По горным выработкам
Среднее содержание, %	2.16	1.83
Дисперсия оценки содержания	$4.0 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$
Расчет общего среднего	$\hat{\mu} = \frac{2.16/4 \cdot 10^{-4} + 1.83/1.4 \cdot 10^{-2}}{1/4 \cdot 10^{-2} + 1/1.4 \cdot 10^{-2}} = 1.92$	
Расчет дисперсии его оценки	$\sigma_{\hat{\mu}}^2 = 1/(1/1.4 \cdot 10^{-2} + 1/4 \cdot 10^{-2}) = 1.03 \cdot 10^{-2}$	
Среднеквадратическая погрешность оценки общего среднего	$\sigma_{\hat{\mu}} = \sqrt{1.03 \cdot 10^{-2}} = 1.0 \cdot 10^{-1}$	
Рекомендуемая форма записи результата, %	$\hat{\mu} = 1.92 \pm 0.10$	

Из приведенного примера видно, что среднее содержание, вычисленное по объединенным результатам, незначительно отличается от среднего значения по горным выработкам. Это становится понятным при сравнении дисперсий оценок средних: разведка горными выработками дала почти в три раза более точные результаты, которые «перевесили» при объединении данных.

На практике могут встретиться случаи, когда зависимость оценок существенна и «обойти» ее невозможно. Характерным примером может служить оценка нефтенасыщенности породы в ее естественном залегании в недрах с помощью нескольких косвенных методов, т. е. по регрессии. По каждой из вспомогательных величин нефтенасыщенность оценивается с погрешностью s_i , но между са-

мими вспомогательными величинами установлена зависимость, поэтому применение формул (6.2.2) и (6.2.3) неверно. Затруднение можно обойти следующим образом. Для определенности упорядочить вспомогательные величины x_i так, чтобы $s_1 < s_2 < s_3, \dots, s_n$. Оценку y по регрессии с первой переменной (с погрешностью s_1) взять за основу, вычислить «остатки» оцениваемой величины (т. е. $\Delta y_i = a_1 + b x_{1i} - y_i$), некоррелированные с первой переменной, и найти уравнение регрессии этих остатков на вторую переменную. Аналогичным образом будем поступать, пока не исчерпаются вспомогательные величины. Дисперсия комплексной оценки s^2 по такой схеме будет равна

$$s^2 = \frac{1}{\frac{1}{s_1^2} + \frac{1}{s_2^2(1-r_{1|2})} + \frac{1}{s_3^2(1-r_{12|3})} + \dots + \frac{1}{s_n^2(1-r_{123\dots|i})}}, \quad (6.3.1)$$

где s_i — погрешность оценки по i -й вспомогательной переменной, $r_{123\dots|i}$ — множественный коэффициент корреляции между i -й переменной и множеством переменных с номерами до $i-1$ включительно. Мы предполагаем, что эта формула может применяться не только в этом специальном случае, но и в более общих.

§ 4. Оценка среднего в объекте по данным, относящимся к составным частям этого объекта

Часто исследуемые объекты состоят из разных частей: пластов, если это осадочная формация, участков, если это массив, или месторождение минеральных видов, если это горная порода, и т. д. Объект может быть разделен на части условно в соответствии с последовательностью изучения или применяемыми методами. Иногда части являются самостоятельными объектами, но наступает момент, когда необходимо объединить данные по ним с получением результатов по объекту в целом. Здесь мы рассмотрим случай, когда исследованные части при их объединении составляют весь исследуемый объект. Ситуацию можно рассматривать либо как объединение разных опробованных совокупностей в одну и ее оценивание, либо как разделение изучаемой совокупности на части и раздельное опробование частей. В математической статистике это известно как опробование расслоенных конечных совокупностей, или с л о е в о е о п р о б о в а н и е.

Изучаемые признаки можно подразделить на экстенсивные и интенсивные. Смысл их близок к смыслу соответствующих термодинамических терминов. К экстенсивным признакам в геологической разведке отнесем линейные размеры рудных залежей и тел, их площадь, объем, количество компонента (запасы), к интенсивным — содержание, пористость, нефтенасыщенность, проницаемость, упругость, плотность и т. д. Смысл экстенсивных или интенсивных признаков практически аналогичен таковому в термодинамике.

Оценки экстенсивных признаков складываются при объединении частей объекта

$$\hat{\theta} = \sum_{i=1}^m \hat{\theta}_i. \quad (6.4.1)$$

В соответствии с законами статистики дисперсия оценки суммы равна сумме дисперсий оценок слагаемых:

$$\sigma_{\hat{\theta}}^2 = \sum_{i=1}^m \sigma_{\hat{\theta}_i}^2, \quad (6.4.2)$$

Эта формула по своей сути аналогична уже известной геологам формуле для расчета дисперсии относительной погрешности количества металла в объединя-

емых пробах, которая приводится у Э. Карлье (1966). Применение формулы требует независимости случайных погрешностей объединяемых оценок Δ_i ($\Delta_i = \theta_i - \hat{\theta}_i$). Эту независимость следует понимать примерно так, что положительная ошибка в оценке одной части тела не изменяет условной вероятности положительной или отрицательной ошибки в оценках других частей. На практике это условие, как правило, выполняется и не требует специальной проверки, разве что в особом случае, на который стоит указать. Например, зона обогащения протягивается через все отдельно оцениваемые части рудного тела. Горная выработка случайно прошла вдоль зоны обогащения. В результате погрешности оценок частей тела окажутся все положительными и сильно коррелированными. Применение правильной методики разведки позволяет исключить такие явления.

Интенсивный признак представляет собой отношение двух экстенсивных признаков.¹ Пусть объем порового пространства (экстенсивный параметр) — W , объем залежи (экстенсивный параметр) — V , тогда пористость (интенсивный параметр) $K = W/V$. По (6.4.1) $W = \sum W_i$, откуда, деля обе части равенства на V , получим $K = W/V = 1/V \sum W_i = 1/V \sum K_i V_i = \sum K_i V_i / V = \sum K_i p_i$, где p_i — доли частей объекта по объему, т. е. для интенсивного параметра имеем

$$\hat{\theta} = \sum_{i=1}^m p_i \theta_i, \quad (6.4.3)$$

а для дисперсии оценки соответственно

$$\sigma_{\hat{\theta}}^2 = \sum_{i=1}^m p_i^2 \sigma_{\theta_i}^2, \quad (6.4.4)$$

где $\sigma_{\theta_i}^2$ является функцией плотности сети и изменчивости признака в i -й части геологического объекта.

Если мы выступаем не в пассивной роли объединяющего данные, а сами планируем разведку, то в соответствии с теорией слоевого опробования необходимо распределять пробы между «слоями» так, чтобы их число n_i было пропорционально объему слоя и внутрислойной изменчивости признака. Большие и сильно изменчивые части объекта должны быть опробованы большим числом проб:

$$n_i = n \frac{V_i \sigma_i}{V} / \sum_{i=1}^m \frac{V_i \sigma_i}{V}. \quad (6.4.5)$$

Пример 1. Месторождение разведано на трех горизонтах. По каждому горизонту рассчитаны запасы, среднее содержание и его дисперсия. Требуется определить среднее содержание и его дисперсию по всему месторождению (табл. 6.4.1).

Пример 2. Проведены анализы монофракций перидотита на окись титана. Требуется оценить его содержание в породе и дать оценку его точности (табл. 6.4.2).

Пример 3. Изучена мощность ореолов аномальных содержаний фтора вокруг пегматитовых тел. Требуется определить среднюю мощность ореолов данной группы образований. Отличительной особенностью данного примера является то, что веса p_i , которые следует использовать при расчетах, равны между собой и не зависят от величины пегматитовых тел (табл. 6.4.3).

Частным случаем изложенного является оценка среднего в группе объектов по значениям частных средних в каждом объекте группы. Нас может интере-

¹ Это заявление можно подтвердить хотя бы тем, что размерность известных нам интенсивных признаков есть отношение размерностей экстенсивных: кг/см², м/с, Ом/м и т. д.

Пример оценивания среднего по результатам опробования составных частей объекта

Характеристики	Значения характеристик по горизонтам		
	1	2	3
Среднее содержание, %	3.42	2.50	1.76
Средняя квадратическая погрешность оценки среднего	0.35	0.41	0.50
Дисперсия оценки	0.12	0.17	0.25
Весовая доля горизонта, p_i	0.62	0.23	0.15
Расчет среднего содержания по месторождению	$\hat{\theta} = 0.62 \cdot 3.42 + 0.23 \cdot 2.50 + 0.15 \cdot 1.76 = 2.96$		
Расчет дисперсии его оценки	$\sigma_{\hat{\theta}}^2 = 0.38 \cdot 0.12 + 0.053 \cdot 0.17 + 0.022 \cdot 0.25 = 0.06$		
Средняя квадратическая погрешность	$\sigma_{\hat{\theta}} = \sqrt{0.06} = 0.24$		
Рекомендуемая форма записи результата, %	$\hat{\theta} = 2.96 \pm 0.24$		

Таблица 6.4.2

Пример оценивания среднего в породе по результатам анализов ее минералов

Характеристики	Значения характеристик в минералах		
	оливин	пироксен	магнетит
Содержание в минерале, %	0.01	0.025	3.7
Средняя квадратическая погрешность оценки содержания	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	0.2
Дисперсия оценки	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$4 \cdot 10^{-2}$
Содержание минералов в породе, %	65.3	35.3	0.4
Расчет среднего содержания в породе	$\hat{\theta} = 0.653 \cdot 0.01 + 0.353 \cdot 0.025 + 0.004 \cdot 3.7 = 0.030$		
Расчет дисперсии его оценки	$\sigma_{\hat{\theta}}^2 = 0.425 \cdot 1 \cdot 10^{-6} + 0.124 \cdot 1 \cdot 10^{-6} + 16 \cdot 10^{-6} \cdot 4 \times 10^{-2} \cong 1.2 \cdot 10^{-6}$		
Средняя квадратическая погрешность среднего	$\sigma_{\hat{\theta}} = \sqrt{1.2 \cdot 10^{-6}} = 1.1 \cdot 10^{-3}$		
Рекомендуемая форма записи результата, %	$\hat{\theta} = 0.030 \pm 0.001$		

Таблица 6.4.3

Пример оценивания среднего по результатам измерений в различных объектах

Характеристики	Значение характеристик в телах				
	1	2	3	4	5
Мощность ореолов, м	2.15	0.83	1.74	2.45	1.29
Средняя квадратическая погрешность оценки мощности	0.41	0.09	0.16	0.38	0.22
Дисперсия оценки	0.170	0.008	0.026	0.145	0.047
Весовые коэффициенты	0.2	0.2	0.2	0.2	0.2
Расчет средней мощности, м	$\hat{\theta} = 0.2(2.15 + 0.83 + 1.74 + 2.45 + 1.29) = 1.69$				
Расчет дисперсии оценки средней мощности	$\sigma_{\hat{\theta}}^2 = 0.04(0.170 + 0.008 + 0.026 + 0.145 + 0.047) = 0.016$				
Средняя квадратическая погрешность средней мощности	$\sigma_{\hat{\theta}} = \sqrt{0.016} = 0.13$				
Рекомендуемая форма записи результата, %	$\hat{\theta} = 1.69 \pm 0.13$				

Пример оценивания среднего в толще по результатам опробования всех ее пластов (1-й этап, терригенные пласты)

Характеристики	Значения характеристик по терригенным пластам			
	1	2	3	4
Содержание по пласту, %	0.035	0.041	0.029	0.932
Дисперсия его оценки	$4 \cdot 10^{-6}$	$5.7 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$	$4.4 \cdot 10^{-6}$
Доля пласта в мощности терригенной части	0.14	0.25	0.28	0.33
Расчет среднего содержания, %	$\hat{\theta}_1 = 0.14 \cdot 0.035 + 0.25 \cdot 0.041 + 0.28 \cdot 0.029 + 0.33 \cdot 0.032 = 0.0341$			
Расчет дисперсии его оценки	$\sigma_{\hat{\theta}_1}^2 = 2 \cdot 10^{-2} \cdot 4 \cdot 10^{-6} + 6.2 \cdot 10^{-2} \cdot 5.7 \cdot 10^{-6} + 7.8 \cdot 10^{-2} \cdot 2.9 \cdot 10^{-6} + 10.9 \cdot 10^{-2} \cdot 4.4 \cdot 10^{-6} = 1.1 \cdot 10^{-6}$			
Средняя квадратическая погрешность	$\sigma_{\hat{\theta}_1} = \sqrt{1.1 \cdot 10^{-6}} = 1 \cdot 10^{-3}$			
Рекомендуемая форма записи результата, %	$\hat{\theta}_1 = 0.034 \pm 0.001$			

Таблица 6.4.5

Пример оценивания среднего в толще по результатам опробования всех ее пластов (1-й этап, карбонатные пласты)

Характеристики	Значение характеристик по карбонатным пластам		
	1	2	3
Содержание по пласту, %	0.056	0.074	0.048
Дисперсия его оценки	$12 \cdot 10^{-6}$	$9 \cdot 10^{-6}$	$6.8 \cdot 10^{-6}$
Доля пласта в мощности карбонатной части	0.23	0.16	0.61
Расчет среднего содержания, %	$\hat{\theta}_2 = 0.23 \cdot 0.056 + 0.16 \cdot 0.074 + 0.61 \cdot 0.048 = 0.054$		
Расчет дисперсии его оценки	$\sigma_{\hat{\theta}_2}^2 = 5.29 \cdot 10^{-2} \cdot 12 \cdot 10^{-6} + 2.56 \cdot 10^{-2} \cdot 9 \cdot 10^{-6} + 37.21 \cdot 10^{-2} \cdot 6.8 \cdot 10^{-6} = 3.4 \cdot 10^{-6}$		
Средняя квадратическая погрешность	$\sigma_{\hat{\theta}_2} = \sqrt{3.4 \cdot 10^{-6}} = 1.8 \cdot 10^{-3}$		
Рекомендуемая форма записи результата, %	$\hat{\theta}_2 = 0.054 \pm 0.002$		

Таблица 6.4.6

Пример оценивания среднего в толще по результатам опробования всех ее пластов (2-й этап)

Характеристики	Значения характеристик по литотипам	
	терригенный	карбонатный
Содержание в литотипе, %	0.034	0.054
Дисперсия его оценки	$1.1 \cdot 10^{-6}$	$3.4 \cdot 10^{-6}$
Доли литотипа в мощности толщи	0.71	0.29
Расчет среднего содержания, %	$\hat{\theta} = 0.71 \cdot 0.03 + 0.29 \cdot 0.054 = 0.0369$	
Расчет дисперсии его оценки	$\sigma_{\hat{\theta}}^2 = 50.4 \cdot 10^{-2} \cdot 1.1 \cdot 10^{-6} + 8.4 \cdot 10^{-2} \cdot 3.4 \cdot 10^{-6} = 0.84 \cdot 10^{-6}$	
Средняя квадратическая погрешность	$\sigma_{\hat{\theta}} = \sqrt{0.84 \cdot 10^{-6}} \cong 1 \cdot 10^{-3}$	
Рекомендуемая форма записи результата, %	$\hat{\theta} = 0.040 \pm 0.001$	

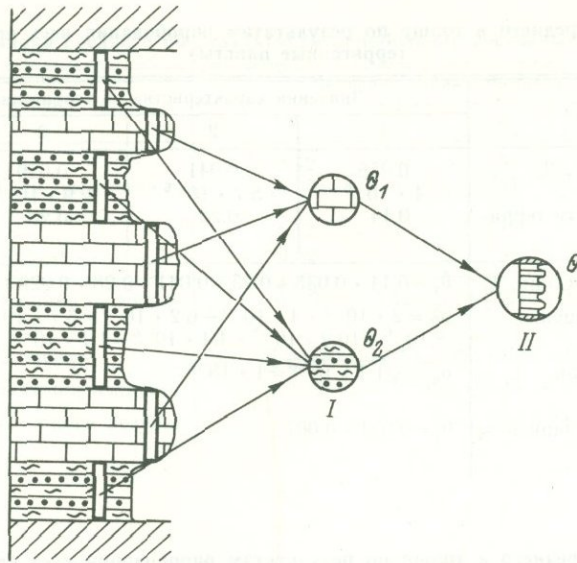


Рис. 6.4.1. Иллюстрация ступеней обобщения данных к примеру 4.

I, II — ступени обобщения данных; θ_1, θ_2 — средние по литотипам; θ — общее среднее по толще.

совать средняя пористость залежей нефти нескольких однотипных месторождений. Единственная, но важная особенность этого случая состоит в выборе весовых коэффициентов p . Они могут быть величины $1/N$, где N — число объектов в группе, или q_i/Q , где q_i — масса (или объем) i -го объекта, Q — масса (или объем) объектов всей группы. Первый вариант более подходит для изучения характеристик геологических процессов, когда более важное значение имеет число реализаций. Второй рекомендуется при определении средних содержаний компонентов, особенно при подведении геохимического баланса элементов, при расчете кларков и т. д. Весовые коэффициенты могут быть выбраны и из других соображений.

Принципиально ничем не отличается случай, когда отдельные объекты, составляющие интересующую нас группу, объединены в несколько подгрупп. Если имеются исходные данные по отдельным объектам, объединение в подгруппы можно игнорировать, т. е. свести ситуацию к уже описанной. Однако лучше оценку средних и их погрешностей провести в два этапа. На первом этапе слоями будут объекты, оцениваемыми совокупностями — подгруппы. На втором этапе применяются те же формулы, слоями будут подгруппы, оцениваемой совокупностью — группа.

Например, при изучении терригенно-карбонатной толщи все пласты известняков могут быть объединены в одну группу, песчаников — в другую, глин — в третью. Пласты одного и того же литотипа меньше отличаются по значениям изучаемых признаков, чем пласты различных литотипов, поэтому такой путь оценки в общем случае наиболее предпочтителен. Но выигрыша в точности, если мы сначала проведем расчет средних содержаний по отдельным типам пород (т. е. по подгруппам), а затем объединим подгрупповые средние в среднее по группе в целом, в данном случае не получится. Преимущество этого более правильного методического приема выявится тогда, когда опробованы не все объекты слоя (в данном случае — не все пласты данного литотипа), а только выборки из него.

Пример 4. На содержание стронция опробованы два литотипа карбонатно-терригенной толщи. Карбонатная часть толщи представлена тремя пластами,

терригенная — четырьмя. Объемный вес разновидностей принят одинаковым. Требуется определить среднее содержание стронция в толще на опробованной территории и дать оценку точности получаемого содержания. На первом этапе «слоями» будут пласты, оцениваемой совокупностью — литотипы. Первый этап расчетов по терригенным пластам дан в табл. 6.4.4, по карбонатным — в табл. 6.4.5. На втором этапе оценки слоями будут литотипы, генеральной совокупностью — вся толща. Исходные данные, расчеты и их результаты приведены в табл. 6.4.6. Пример иллюстрируется рис. 6.4.1.

Следует отметить, что расчет среднего содержания в толще среднеарифметическим способом по всему числу отобранных проб может привести в общем случае к грубым ошибкам, а расчет средневзвешенным способом, где весами для каждой пробы являются зоны их влияния без учета литологических различий, — к еще большим промахам и значительно менее достоверной оценке, чем по рекомендуемой здесь методике.

§ 5. Оценка среднего в объекте по данным, относящимся к опробованным частям этого объекта

От описанного в предыдущем разделе этот случай отличается тем, что части, по которым имеются оценки средних, не исчерпывают объекта целиком. Например, объектом оценки может служить пористость пласта коллектора по скважине. Пласт в этой скважине состоит из пяти прослоев, а керн отобран только из трех. Объектом оценки может быть совокупность сходных геологических тел, например залежей, из которых разведана только часть.

Получение общего среднего распадается на два этапа. Сначала оцениваются параметры в опробованной части объекта, или соответственно в опробованной подсовкупности оцениваемой совокупности объектов. Для этого используются формулы (6.4.1) и (6.4.2) для экстенсивного параметра или (6.4.3) и (6.4.4) — для интенсивного.

На втором этапе среднее значение интенсивного параметра в опробованной части объекта распространяется без изменения на весь объект. Среднее значение экстенсивного параметра необходимо увеличить в пропорции $v/\sum v_i$, где v — объем всего объекта, $\sum v_i$ — сумма объемов его опробованных частей.

Основная сложность на втором этапе возникает при расчете погрешности оценки параметра. По закону аддитивности дисперсий она состоит из двух слагаемых. Первое представляет собой погрешность оценки опробованной части объекта и берется из формулы (6.4.4), второе — погрешность оценки всего объекта с помощью изученной его части. В геостатистике ошибку, вносимую распространением содержания в «пробе» на зону ее «влияния», характеризуют дисперсией распространения. Проведя аналогию, с нашим случаем можно сопоставить «пробу» с опробованной частью объекта, а зону ее «влияния» — с целым объектом. Тогда второе слагаемое сопоставляется с погрешностью распространения. Заметим, что первое слагаемое в таком случае поставляется с погрешностью измерения содержания в «пробе». Рассчитать погрешность распространения без применения геостатистических методов невозможно, а изложение их не является темой этой книги. Поэтому мы изложим приближенный способ, основанный на предположении, что средние значения параметра не коррелируют в рядом расположенных частях объекта. Пусть m_s — число примерно одинаковых опробованных частей объекта, m — общее их число. Тогда дисперсия распространения

$$\sigma_p^2 = (1-f) \frac{1}{m_s} \sigma_m^2 \quad (6.5.1)$$

где f — коэффициент отбора ($f = m_s/m$), σ_m^2 — дисперсия частных средних (средних в опробованных частях объекта θ_i по отношению к общему среднему $\bar{\theta}$);

$$\sigma_m^2 = \frac{1}{m_s} \sum_{i=1}^{m_s} (\bar{\theta}_i - \bar{\theta})^2.$$

Таким образом

$$\sigma_{\bar{\theta}}^2 = \sum_{i=1}^{m_s} p_i^2 \sigma_{\bar{\theta}_i}^2 + \sigma_p^2. \quad (6.5.2)$$

При $f=1$ рассматриваемый случай редуцируется в предыдущий ($\sigma_p^2=0$), если же опробованные части составляют незначительную долю целого объекта, то, как показывают многочисленные примеры, второе слагаемое в (6.5.2)

$$\sigma_p^2 = \sigma_m^2/m_s \quad (6.5.3)$$

становится определяющим. При оценке совокупности объектов по опробованной их подсовокупности σ_m^2 является дисперсией средних между объектами, m и m_s — их общее и опробованное число.

Встает вопрос: что рациональнее при фиксированных затратах — охватить опробованием большее число частей объекта (и уменьшить тем самым второе слагаемое дисперсии оценки) или полнее опробовать каждую часть при меньшем их числе и уменьшить тем самым первое слагаемое? Решение зависит от соотношения изменчивости свойства внутри частей объекта и изменчивости средних по отдельным частям в пределах изучаемого объекта (совокупности) и будет обсуждаться позже как составная часть проблемы оптимизации структуры выборки.

В том случае, когда части объекта не сопоставимы по размеру, приближенную оценку дисперсии можно получить, искусственно подразделяя опробованные части объекта на еще более мелкие приблизительно одинаковые части.

Пример 1. В качестве основы для данного примера взят предыдущий, видоизмененный таким образом, чтобы опробованные пласти составляли лишь часть разреза. Пусть изучаемая толща состоит из 18 пластов терригенных отложений. Из них опробованы известные нам из табл. 6.4.4 четыре и из десяти пластов известняков, из которых опробованы указанные в табл. 6.4.5, — три (рис. 6.5.1). Мощности терригенной и карбонатной частей разреза относятся по-прежнему как 0.71:0.29. Расчеты параметров по опробованным подсовокупностям пластов каждого литотипа совпадают с проведенными в указанных таблицах. Каждое полученное среднее распространяется на «свой» литотип, а расчет среднего по толще совпадает с приведенным в табл. 6.4.6. Однако расчет дисперсии оценки содержания по каждой литологической разновидности производится уже по другой формуле (6.5.2). Первое входящее в нее слагаемое по каждому литотипу уже известно из табл. 6.4.4 и 6.4.5. Второе слагаемое возникает из-за того, что была опробована только часть пластов каждой литологической разновидности. Расчет этого слагаемого проведен в табл. 6.5.1 для карбонатной части разреза и 6.5.2 — для терригенной. Расчет дисперсии оценки среднего содержания стронция в толще в целом с использованием данных табл. 6.5.1 и 6.5.2 производится по формуле (6.4.4): $\sigma_{\bar{\theta}}^2 = 0.71^2 \cdot 5.4 \cdot 10^{-6} + 0.29^2 \cdot 21 \cdot 10^{-6} = 30 \cdot 10^{-6}$; $\bar{\theta} = 0.040 \pm 5.5 \cdot 10^{-3}$.

Дисперсия оценки среднего содержания в этом примере ($30 \cdot 10^{-6}$) возросла по сравнению с предыдущим примером ($0.86 \cdot 10^{-6}$) более чем в тридцать раз! Такое значительное уменьшение точности оценки (более чем в пять раз) по тем же данным опробования обусловлено тем, что теперь оценка относится не только к опробованным пластам, но ко всем (двадцати восьми) пластам данной толщи. На наш взгляд, показательный пример того, как

Пример оценивания среднего в толще по выборке пластов из нее (карбонатная часть)

Номер пласта i	Среднее содержание по пласту (из табл. 6.4.5)	Доля пласта в общей мощности опробованной карбонатной части	$\theta_i - \theta$	$(\theta_i - \theta)^2$	$p_i (\theta_i - \theta)^2$	Расчет дисперсии оценки содержания в карбонатной части. Первое слагаемое берется из табл. 6.4.5
1	0.056	0.23	$2 \cdot 10^{-3}$	$4 \cdot 10^{-6}$	$0.2 \cdot 10^{-6}$	$\sigma_{\theta_2}^2 = 3.4 \cdot 10^{-6} + (1 - 3/10) \cdot 25.5 \times 10^{-6} = 0.21 \cdot 10^{-4}$
2	0.074	0.16	$20 \cdot 10^{-3}$	$400 \cdot 10^{-6}$	$12 \cdot 10^{-6}$	
3	0.048	0.61	$6 \cdot 10^{-3}$	$36 \cdot 10^{-6}$	$13.3 \cdot 10^{-6}$	
Среднее	0.054		$\hat{\theta}_2 = 0.064 \pm \pm 0.005$		$\sum_{i=1}^3 = 25.5 \cdot 10^{-6}$	$\sigma_{\theta_2} = \sqrt{0.21 \cdot 10^{-4}} = = 0.47 \cdot 10^{-2}$

изменяется точность оценки в зависимости от того, что подразумевалось под объектом оценки.

В различных случаях две составляющие дисперсии оценки находятся в различном соотношении. В данном примере основной вклад в общую дисперсию вносят большие колебания содержания от пласта к пласту. Другой составной частью — дисперсией оценки содержания в отдельных пластах, обусловленной изменчивостью содержания внутри пласта и плотностью сети опробования, практически можно пренебречь.

Рассмотрим вопрос, что произойдет с оценкой среднего и ее погрешностью, если бы мы игнорировали подразделение толщи на литотипы. Во-первых, изменилась бы оценка общего среднего ввиду изменения весовых коэффициентов, с которыми учитывались частные средние по пластам. Во-вторых, сильно увеличилась бы погрешность оценки, так как колебания средних пластовых содержаний между пластами внутри одного литотипа значительно меньше, чем между пластами всей толщи. На этом примере мы видим, что информация, полученная из геологической документации или картирования типов руд, разновидностей горных пород, т. е. более дешевым способом, чем опробование, увеличивает в конечном счете точность подсчета запасов.

§ 6. Оценка среднего в совокупности объектов, по средним в отдельных группах объектов, выделенных случайным образом

В ряде случаев группа объектов, исследованная определенным автором, не отличается каким-нибудь характерным признаком от групп объектов, исследованных другими авторами. Объединение объектов в группы (слои) вызвано только необходимостью использовать уже обобщенные данные. Объемы слоев, т. е. объемы тех частей общей совокупности объектов, к которым относятся данные каждого автора, в этом случае не поддаются оценке. Тогда мы рекомендуем выделить из обследуемой совокупности подсовокупность, состоящую только из опробованных всеми авторами объектов. При этом будем иметь подсовокупность, разделенную на «слои» (данные различных авторов), причем в каждом слое опробованы все объекты, составляющие данный слой. Оценка среднего и его дисперсия в подсовокупности сведется к случаю § 4. Выбор объектов из совокупности в слои, как это следует из существа рассматриваемого варианта, случаен, поэтому дисперсию оценки среднего можно представить себе как сумму дисперсии оценки среднего в подсовокупности и дисперсии распространения этого среднего на всю совокупность. Первая рассчитывается по формуле (6.4.2) или (6.4.4). Вторая равна:

$$\sigma_p^2 = (1-f) \frac{1}{m_s} \sigma_m^2 = (1-f) \frac{1}{m_s^2} \sum_{i=1}^{m_s} (\hat{\theta}_i - \bar{\theta})^2, \quad (6.6.1)$$

где m_s — число слоев в подсовокупности, f — доля опробованной подсовокупности по отношению ко всей совокупности. Если всем объектам по смыслу придаются одинаковые веса, то $f = m_s/M$, где m_s и M — число объектов: опробованных и всех соответственно.

В пределе, когда $f=1$, этот случай вырождается в случай § 4. Когда f мало по сравнению с единицей, второе слагаемое дисперсии становится обратно пропорциональным объему подсовокупности. Следует еще раз подчеркнуть, что оценка среднего будет менее точна, чем по схеме § 5, которую мы не можем применять из-за отсутствия информации о весовых коэффициентах отдельных «слоев».

Пример. В основу примера опять положим предыдущий. Представим себе, что тремя исследователями опробованы три различные выборки пластов (в общем случае в каждой выборке имеются и карбонатные, и терригенные пласты). При публикации результатов каждый автор указал лишь среднее содержание в своей группе пластов, дисперсию оценки этого содержания и мощность опробованной части разреза. Для получения сопоставимых с предыдущими примерами результатов средние содержания в пластах и дисперсии их оценок оставим, как в предыдущем примере (табл. 6.6.1). Предположим теперь, что авторами опубликованы лишь итоговые данные по своим группам пластов, как в верхней части табл. 6.6.2.

Рассмотренный пример показывает, что в данном случае из двух составляющих дисперсии — дисперсии оценки среднего в подсовокупности и дисперсии распространения этого среднего на всю совокупность — последняя имеет подавляющее значение, а первой практически можно пренебречь. Кроме того, мы видим, что точность оценки по сравнению с предыдущим случаем значительно понизилась, что хорошо видно из следующего сопоставления:

	Слоевое исчерпы- вающее опробова- ние	Слоевое неисчер- пывающее опробова- ние	Опробова- ние слу- чайными группами
Среднее	0.040	0.040	0.048
Квадратическая погрешность оценки среднего	$1.0 \cdot 10^{-3}$	$5.5 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$

В случае, когда опробованы все объекты (пласты) каждого «слоя», а «слои» полностью составляют изучаемую совокупность (пример из § 4), дисперсия оценки равна $0.84 \cdot 10^{-6}$. Если «слои» состоят не только из опробованных пластов, а включают большее их число, как в первом случае § 5, соответствующая дисперсия равна $30 \cdot 10^{-6}$. Наконец, если разными авторами опробовано по несколько терригенных и карбонатных пластов, но результаты приведены в виде средних по изученным группам, то дисперсия оценки в толще равна $36.9 \cdot 10^{-6}$. Картина была бы еще более показательной, если бы в составе опробованных каждым автором пластов преобладала какая-нибудь одна литологическая разновидность. Очевидно, что отступление от принципа «слоевого» опробования и формирование выборки без учета однородности (в данном случае — литологической) слоев приводят при том же количестве проб и анализов к значительной потере информации и уменьшению точности оценок.

Остается рассмотреть довольно экзотический случай, когда по среднему в выборке объема n мы делаем попытку предсказать среднее в другой

Пример оценивания среднего по данным различных авторов (1-й этап)

Характеристики	Значения характеристик по данным различных авторов						
	1-й автор			2-й автор		3-й автор	
Индекс пласта	1 ^К	2 ^Т	3 ^Т	3 ^К	4 ^Т	2 ^К	1 ^Т
Содержание в пласте	0.056	0.041	0.029	0.048	0.032	0.074	0.035
Дисперсия оценки содержания	$12 \cdot 10^{-6}$	$5.7 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$	$6.8 \cdot 10^{-6}$	$4.4 \cdot 10^{-6}$	$9 \cdot 10^{-6}$	$4 \cdot 10^{-6}$
Доля пласта в общей опробованной данным автором мощности	0.12	0.42	0.46	0.44	0.56	0.41	0.59
Расчет среднего содержания по опробованной каждым автором группе пластов	$\hat{\theta}_1 = 0.12 \cdot 0.056 + 0.42 \times 0.041 + 0.46 \cdot 0.029 = 0.037$			$\hat{\theta}_2 = 0.44 \cdot 0.048 + 0.56 \cdot 0.032 = 0.039$		$\hat{\theta}_3 = 0.41 \cdot 0.074 + 0.59 \cdot 0.035 = 0.051$	
Расчет дисперсии его оценки	$\sigma_{\theta_1}^2 = 10^{-8} (12 \cdot 1.2^2 + 4.22 \times 5.7 + 4.6^2 \cdot 2.9) = 1.8 \cdot 10^{-6}$			$\sigma_{\theta_2}^2 = 10^{-8} (4.4^2 \times 6.8 + 5.6^2 \cdot 4.4) = 2.7 \cdot 10^{-6}$		$\sigma_{\theta_3}^2 = 10^{-8} (4.1^2 \times 9 + 5.9^2 \cdot 4) = 2.9 \cdot 10^{-6}$	

Примечание. Буквенные индексы: К — карбонатный пласт, Т — терригенный; цифра — номер пласта по табл. 6.4.4 (для терригенного) или 6.4.5 (для карбонатного).

Таблица 6.6.2

Пример оценивания среднего по данным различных авторов (2-й этап)

Характеристики	Свободные значения характеристик		
	1-й автор	2-й автор	3-й автор
Среднее по группе	0.037	0.039	0.051
Дисперсия оценки среднего по группе	$1.8 \cdot 10^{-6}$	$2.7 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$
Доля мощности толщи опробованной данным автором	0.13	0.11	0.76
Расчет среднего содержания в толще	$\hat{\theta} = 0.13 \cdot 0.037 + 0.11 \cdot 0.039 + 0.76 \cdot 0.051 = 0.048$		
Расчет дисперсии его оценки	$\sigma_{\theta}^2 = 10^{-8} (1.3^2 \cdot 1.8 + 1.1^2 \cdot 2.7 + 7.6^2 \cdot 2.9) + (1-f) \cdot (11^2 + 9^2 \times 3^2) \cdot \frac{1}{6} \cdot 10^{-6}$ Если $f \ll 1$, то $\sigma_{\theta}^2 = 36.9 \cdot 10^{-6}$, $\sigma_{\theta} = 6.1 \cdot 10^{-3}$		

предполагаемой выборке объема k из той же совокупности. Определив, например, пористость пяти эксплуатируемых залежей нефти, мы задаемся вопросом, с какой средней квадратической погрешностью (обозначим ее $s_{n,k,N}$ и назовем погрешностью переноса) мы можем предсказать среднее значение пористости семи новых планируемых к разведке залежей такого же типа. Можно показать, что для неповторных выборок из конечных совокупностей объемом N

$$s_{n,k,N}^2 = s^2 (1/n + 1/k - 2/N), \quad (6.6.2)$$

где s^2 — дисперсия пористости в выборке залежей. При $k = N$ величина $s_{n,k,N}^2$ редуцируется в уже рассмотренную выше дисперсию распространения. Формула (6.6.2) применима, если полученная и проектируемая выборки независимы, т. е. нет никакой информации о том, какая часть проектируемой выборки (объема k) повторит имеющуюся выборку.

Практически интересен еще случай, когда указанные выборки содержат общую часть объема l , причем в остальной части (кроме заведомо общей части l) выборки независимы:

$$s_{n, k, N, l}^2 = 1/k^2 s^2 [(1/l - 1/n) l^2 - (1/n + 1/(k-l) - 2/N) (k-l)^2]. \quad (6.6.3)$$

При $l=0$ получаем формулу (6.6.2), при $l=k$, $k < n$ получим $s^2 = s^2 (1/k - 1/n)$, т. е. как для бесповторной выборки объема k из совокупности объема n .

§ 7. Анализ применяющихся способов оценки средних значений коллекторских свойств нефтяных месторождений

Рассмотрим один из актуальных вопросов методики подсчета запасов нефтяных месторождений: как усреднять данные измерений коллекторских свойств (на примере пористости), чтобы получить наиболее надежное среднее значение по залежи. Остановимся здесь на двух аспектах этого вопроса: 1) полезны ли подразделение коллекторов на типы и отдельный подсчет запасов для отдельных типов, 2) каким способом усреднять данные.

Одним из методов повышения достоверности оценки служит выделение типов (классов) коллекторов и отдельная оценка запасов нефти в них (Фурсов, 1968; Гудошников, Югин, 1977; Родыгин, Сливков, 1977, и др.). Такая методика неявно содержит предположение о том, что доли типов коллекторов известны достаточно надежно, хотя точной формулировки требований к надежности оценки долей ни в одной из перечисленных работ не дано. Самое важное здесь заключается в том, что оценивание относительных объемов типов пород должно опираться на существенно иную информацию, чем частоты их встречаемости, установленные по той же самой выборке, которая является основой при оценке коллекторских свойств. В противном случае будет получена оценка, совпадающая с прежней. Например, А. Я. Фурсов (1968) предлагает устанавливать ее применением комплекса геофизических методов. Многие авторы (например, Иоффе, 1968) полагают, что «представительная» (т. е. равномерная по площади и пропорциональная эффективной мощности по разрезу) выборка кернa обеспечивает правильную оценку параметра и без подразделения на типы коллекторов.

При разведке месторождений его разделение на залежи, участки или блоки имеет характер назначения, и весовые коэффициенты как бы задаются геологом-разведчиком. При четком и однозначном подразделении осадочного разреза на пласты или пачки определить соотношение их мощностей с высокой точностью также не представляет затруднений. В большинстве других случаев погрешность оценки соотношения типов пород становится одного порядка с погрешностью оценки самого параметра. Поскольку оценка среднего рассчитывается как функция произведения двух случайных величин (параметра и весового коэффициента), которые есть основания считать независимыми, то дисперсия оценки среднего, рассчитанного как взвешенное, равна

$$s_{\bar{s}}^2 = \sum p_i^2 s_{\theta_i}^2 + \sum \theta_i^2 s_{p_i}^2 \quad (6.7.1)$$

Таким образом, выделение типов пород (коллекторов) при оценке среднего целесообразно лишь в том случае, если геологическая информация обеспечивает такую малость погрешностей весовых коэффициентов s_{p_i} , что $s_{\bar{s}}^2$ по (6.7.1) становится существенно меньше дисперсии оценки параметра без деления объекта на части.

Показательно, что выделение частей объекта и уменьшение погрешностей весовых коэффициентов выступает как форма учета геологических данных для повышения надежности средних. К сожалению, определение погрешностей весо-

вых коэффициентов при картировании, составлении разрезов, геохимическом опробовании и петрофизических исследованиях не практикуется и методика его не разработана.

Влияние структуры выборки на способ усреднения обсуждался в литературе по нефти неоднократно. Например, Ю. И. Батурин (1965, с. 114) отмечает, что «выборка имеет двойной характер» (образцы из скважин, скважины из залежи, — Ю. Г.), и приходит к выводу, что «погрешность определения выборочной средней пористости для площади зависит от ошибки средних по каждой скважине и погрешности выборки за счет наличия существенных изменений пористости между скважинами» (с. 119). Вместе с тем до сих пор появляются работы с «обоснованием» применимости среднеарифметического, рассчитанного по всей совокупности керновых образцов месторождения (Ящук, 1974; Мончак, 1972; Каневский, Кочев, 1976, и др.). В практике подсчета запасов нефтяных месторождений этот способ все еще широко применяется (и не встречает сопротивления экспертов ГКЗ). Погрешность среднего при этом рассчитывается по формуле

$$\sigma_{\bar{\sigma}} = \sigma / \sqrt{n}, \quad (6.7.2)$$

где σ — стандартное отклонение пористости в выборке керновых проб.

Задавшись допустимой погрешностью оценки средней пористости $\sigma_{\bar{\sigma}}$, авторы упомянутых работ решают равенство (6.7.2) относительно n , определяя «необходимое» для оценки пористости «с заданной надежностью» число образцов. Стоит ли доказывать, что реальная структура выборки керна не имеет ничего общего с той простой моделью, которая положена в основу (6.7.2). Чем более скученно будут отобраны образцы (все из одной скважины и одного интервала), тем меньше будет σ , и это приведет к абсурдной рекомендации о достаточном значении n .

Рассмотрим вопрос, следует ли взвешивать при расчете средних по керновым образцам на длины зон влияния этих образцов. Теоретически это необходимо, но практически опыт показывает, что при современном состоянии техники бурения и отбора керна на этот вопрос необходимо ответить отрицательно. Неполный вынос керна (40—80 %), раздробленность его, несовпадение так называемых интервалов долбления с принятыми к подсчету запасов интервалами коллектора, выделенными по данным геофизики, приводит к тому, что положение большинства образцов керна в пределах подсчетных интервалов и расстояния между ними остаются в значительной степени неопределенными. Таким образом, на первом этапе усреднения среднее значение, например пористости, в интервале (пласте, пропластке коллектора) мы вынуждены подсчитывать как среднее арифметическое из образцов. В. Л. Комаров (1961а, 1961б) для оценки погрешности определения средней пористости в интервале рекомендует применять формулу, основанную на (6.7.2), но с поправкой на конечность изучаемой совокупности (месторождения):

$$\sigma_{\bar{\sigma}}^2 = \sigma^2 (1/n - 1/N),^1 \quad (6.7.3)$$

где n — число образцов керна в интервале (у автора — в скважине); N — общее число образцов, которые могли бы быть отобранными из интервала, при сплошном опробовании; σ^2 — дисперсия пористости по образцам данного интервала. Н. Н. Емельянов и М. Н. Кочетов (1966) рекомендуют такую же формулу, и тоже не для среднеинтервального значения, а для среднескважинного, но с поправкой на возможную закономерность изменения параметра

¹ В оригинале формула записана решенной относительно n . Кроме того, она рекомендована не для среднеинтервальных, а для среднескважинных, что можно считать корректным лишь в частном случае, когда скважиной вскрыт лишь один интервал.

по разрезу, которая выразится в появлении существенного значения коэффициента корреляции r между значениями свойства в соседних образцах. При таком условии авторы дополняют формулу (6.7.3) множителем $(1-r^2)$. Впервые в формулы для расчета погрешности среднего множитель $(1-r^2)$ ввел А. П. Грудев (1961). Характеризуя изменчивость параметра месторождения традиционным для геологов коэффициентом вариации, автор записывает его с поправкой $v(1-r^2)$. Для корректировки коэффициента корреляции в зависимости от расстояния между пробами l А. П. Грудев применяет экспоненциальную функцию $r(l) = \exp(-l/a)$. Теперь мы можем лишь сожалеть, что эти предложения, являющиеся по существу одним из истоков геостатистики, не получили тогда дальнейшего развития.²

При решении вопроса об оценке погрешности среднего к изложенному необходимо добавить следующее: 1) эмпирическая — расчетная по выборке (керновых образцов) дисперсия искажена (увеличена) погрешностями анализа керна; 2) по этой же причине рассчитанный коэффициент корреляции меньше истинного; 3) составной частью погрешности оценки является погрешность анализа, которая не исчезает даже при сплошном опробовании интервалов. С учетом этого предлагается следующая формула:

$$\sigma_{\bar{g}}^2 = (\sigma^2 - \sigma_{\text{ан}}^2) \left(1 - r^2 \frac{\sigma^2}{(\sigma^2 - \sigma_{\text{ан}}^2)} \frac{l}{n} - \frac{l}{N}\right) + \frac{\sigma_{\text{ан}}^2}{n}, \quad (6.7.4)$$

где σ^2 — наблюдаемая дисперсия пористости в керновой выборке, $\sigma_{\text{ан}}$ — погрешность единичного анализа керна.

Задавшись предельно допустимым значением $\sigma_{\bar{g}}^2$, можно, казалось бы, решая (6.7.4) относительно n , определять необходимое число образцов из интервала коллектора в данном его пересечении. Но такая задача не представляет интереса, так как интервалов много, все они отличаются по мощности, числу, расположению их в скважине и залежи и соответственно вносят различные вклады в точность оценки среднего по всей залежи.

Вопрос, как усреднять керновые данные — «валом», сразу по залежи без учета структуры выборки или через среднеинтервальные и среднескважинные — некоторые авторы (Батурин, 1965а, 1965б; Белов, 1972, и др.) предлагают решать с помощью однофакторного дисперсионного анализа разведочных данных. Если нулевая гипотеза о равенстве среднескважинных³ не отвергается, то, утверждают они, усреднение можно (в некоторых работах утверждается, что нужно) проводить среднеарифметическим способом по всем образцам залежи. Однако такой результат дисперсионного анализа, как показали наши расчеты по многим десяткам месторождений, чрезвычайно редок. К тому же, если гипотеза о равенстве скважинных средних не отвергается, оба способа дадут статистически идентичные результаты (Ткачев, 1978).

Нахождение среднескважинной пористости по средним в отдельных интервалах коллектора не содержит проблем: все авторы (Дементьев, Азаматов, 1961; Жданов и др., 1967, и др.) единодушны в том, что здесь надо применять средневзвешенное по мощности интервалов из среднеинтервальных значений, полученных либо по керновым, либо по геофизическим данным.

Оценка погрешности среднескважинного значения должна производиться по одной из следующих, охарактеризованных выше схем, как при оценке среднего в объекте по средним в его частях, исчерпывающих (когда опробованы все

² Экспоненциальный тип полувариограмм является одним из самых распространенных, экспоненциальная зависимость r от расстояния широко используется при аппроксимации геологических поверхностей.

³ В скважинах предполагается по одному интервалу коллектора в залежи, поэтому среднеинтервальные являются в то же самое время и среднескважинными.

интервалы коллектора в скважине) или неисчерпывающих (когда есть непробо- ванные интервалы) весь объект. Для более надежной оценки среднескважин- ного и его погрешности в случае длинных, неравномерно опробованных керном интервалов коллектора, последние лучше разделить на опробованные и непро- бованные части и расчет вести по второй из указанных схем.

Наибольшие трудности встречает усреднение среднескважинных по залежи. Взвешивать или не взвешивать среднескважинные, взвешивать их на мощность залежи или на зоны «влияния» скважин, и если взвешивать, то как понимать зоны «влияния» — как площади или как объемы ближайших районов? Ана- логичные вопросы обсуждаются и в литературе по твердым полезным иско- паемым. Они тесно связаны с вопросами геометризации параметра по площади залежи, которые будут обсуждаться позже. Здесь мы только отметим, что под- счет интегрированием по площади функции, аппроксимирующей параметр, дает по существу взвешенное среднее, причем весовые коэффициенты к сред- нескважинным определяются выбранной для аппроксимации математической моделью.

По Л. Ф. Дементьеву и В. И. Азаматову (1961), выигрыш в точности при расчете среднего по картам изолиний получается только при большом числе (30—50) скважин и плавном изменении параметра (т. е. если изолинии вполне надежны, — Ю. Т.), иначе взвешивание среднескважинных теряет смысл. Для обоснования необходимости взвешивания по площади применялся дисперсион- ный анализ (Батурин, 1965; Белов, 1972; Саввин и др., 1975, и др.). По их мне- нию, отсутствие существенных (статистически значимых) отличий среднесква- жинных яковы свидетельствует в пользу применения среднего арифметиче- ского. В более простом варианте строить карты параметра в изолиниях и подсчитывать средневзвешенное со взвешиванием предлагается при условии $\sigma_{\text{плоч}} \geq 2\bar{\sigma}_{\text{скв}}$ (Карелина, 1965). Наконец, В. А. Саввин с соавторами (1975) предлагают критерий выбора способа усреднения по принципу минимизации дисперсии оценки параметра по залежи. В правильном понимании этот крите- рий является единственно правильным, соответствующим конечной цели оценки. В геостатистике, например, процедура крайгинга построена именно таким обра- зом, что коэффициенты крайгинга рассчитываются из условия минимизации дисперсии оценки параметра. Однако авторы понимают минимизацию диспер- сии оценки как выбор такого способа усреднения, при котором рассчитанная авторами дисперсия оценки (рассчитанная способом, соответствующим способу расчета среднего) наименьшая. При этом наименьшее р а с ч е т н о е (не истин- ное!) значение дисперсии может соответствовать способу расчета, по смыслу не подходящему к данному случаю. Пусть, например, имеется три измерения пористости тремя методами — 15.2, 14.8, 14.7 со среднеквадратическими по- грешностями измерений $\sigma_1=0.22$, $\sigma_2=0.32$, $\sigma_3=0.32$. Среднеарифметический способ дает $\bar{m}=14.9$, $S_m^2=0.037$, $S_m^2=0.037/3=0.012$. Правильный, соответ- ствующий указанной в условии ситуации способ дает

$$\bar{m} = (15.2/0.22^2 + 14.8/0.32^2 + 14.7/0.32^2) / (1/0.22^2 + 1/0.32^2 + 1/0.32^2) = 15.0, \\ S_m^2 = 1 / (1/0.22^2 + 1/0.32^2 + 1/0.32^2) = 0.025,$$

т. е. в два раза ббльшую дисперсию оценки среднего. Авторы, следуя своему принципу, выбрали бы первую оценку, что неверно, и приписали бы ей погреш- ность $\sqrt{0.012}=0.11$, что еще более неверно. Более подробный анализ дискуссии о способах оценки среднего по нефтяной залежи опубликован в работах автора (Ткачев, 1976, 1978).

Применим схемы расчета средних и их погрешностей, изложенные в этой главе, к нефтяным залежам. Прежде необходимо выбрать одну из следующих альтернативных предпосылок: а) среднескважинные являются независимыми оценками параметра во всей залежи; б) они являются оценками параметра

в своих зонах «влияния» (очерченных, например, по методу ближайших районов).

В случае (а) используем схему § 3. Оценкой среднего значения параметра является здесь средневзвешенное из среднескважинных по мощности, дисперсия этой оценки равна

$$s_{\theta}^2 = s_{\text{пл}}^2 / N, \quad (6.7.5)$$

где N — число скважин; $\tilde{s}_{\text{пл}}^2$ — наблюдаемая дисперсия среднескважинной пористости по площади залежи, которая состоит из природной части (части, связанной с естественной изменчивостью параметра) $s_{\text{пл}}^2$ и средней дисперсии оценки среднескважинного значения $\tilde{s}_{\theta_{\text{скв}}}^2$. Последняя в частном случае равенства числа образцов по скважинам n и дисперсии признака в образцах из скважины ($s_{\text{обр}}^2$), а также когда по разрезу скважины не наблюдается автокорреляции пористости, а также когда погрешностью анализа образцов можно пренебречь, равна $s_{\theta_{\text{скв}}}^2 = s_{\text{обр}}^2 / n$, откуда

$$s_{\theta}^2 = \frac{s_{\text{пл}}^2 + s_{\theta_{\text{скв}}}^2}{N} = \frac{s_{\text{пл}}^2}{N} + \frac{s_{\text{обр}}^2}{Nn}. \quad (6.7.6)$$

В формуле (6.7.6) погрешность оценки среднего по залежи на первый взгляд как бы не зависит от точности оценки среднескважинных. Однако это не так, поскольку $\tilde{s}_{\text{пл}}^2$ уже включает в себя величину $s_{\theta_{\text{скв}}}^2$, увеличиваясь при уменьшении точности оценки. Формула (6.7.6) внешне совпадает с формулой, приведенной у А. А. Карелиной (1965, с. 121), Ю. Н. Батурина (1965, с. 119), А. В. Белова (1972) и некоторых других авторов, но у всех у них под $s_{\text{пл}}^2$ понимается *н а б л ю д а е м а я* (эмпирическая) дисперсия среднескважинных по площади следовательно их формулы *д в а ж д ы* включают величину $s_{\theta_{\text{скв}}}^2$, что неверно.

В случае (б) используется схема § 4, для которой необходимо знать весовые коэффициенты (зоны «влияния» скважин) и дисперсии оценок зон «влияния» соответствующими скважинами. Пути определения последних неясны. Геостатистические методы здесь трудно применимы ввиду очень малого числа скважин и крайней ненадежности полувариограмм пористости по залежи. Приемлемой, хотя и завышенной оценкой, является дисперсия среднескважинных в пределах залежи ($s_{\text{пл}}^2$), поэтому с перестраховкой можно записать

$$\sigma_{\theta}^2 = \sum p_i^2 s_{\text{пл}}^2 = s_{\text{пл}}^2 \sum p_i^2. \quad (6.7.7)$$

При равных весах для скважин $p_i = 1/N$ эта формула совпадает с (6.7.5):

$$\sum p_i^2 s_{\text{пл}}^2 = \sum_{i=1}^N \frac{1}{N^2} s_{\text{пл}}^2 = \frac{s_{\text{пл}}^2}{N^2} \sum_{i=1}^N 1 = \frac{s_{\text{пл}}^2}{N}.$$

Легко убедиться, что (6.7.7) при неравных весах всегда дает бóльшую погрешность, чем при равных. Например, при скученном расположении скважин, когда все веса малы по сравнению с каким-нибудь одним из них, получим оценку, близкую к оценке залежи одной скважиной. Этот вывод находится в полном согласии с рекомендуемым Г. А. Габриэлянцем с соавт. (1978) принципом размещения скважин: равные запасы на каждую скважину.

При наличии значительных участков залежи, на которых нет скважин, не рекомендуется «разрезать» этот участок на зоны «влияния» близлежащих скважин и присоединять их к соответствующим зонам влияния. Лучше воспользоваться схемой § 5, т. е. считать, что изученные части не исчерпывают залежь. Дисперсия оценки при этом будет содержать дополнительное слагаемое (см.

формулу (6.7.1)), расчет которого упирается в определение степени изменчивости средних значений параметра от части к части. Приблизительно, с заведомой перестраховкой мы можем взять за оценку этой изменчивости дисперсию среднескважинных. Тогда

$$s_{\bar{\theta}}^2 = \sum_{i=1}^N p_i^2 s_{\theta_i}^2 + (1/v^* - 1/v) s_{\text{пл}}^2 / N, \quad (6.7.8)$$

где $s_{\theta_i}^2$ — дисперсия оценки скважиной зоны ее влияния (принимается равной $\bar{s}_{\text{пл}}^2$); v^* — суммарный объем частей, опробованных скважинами; v — общий объем залежи. Дисперсия, рассчитанная по (6.7.8), в общем случае больше, чем определенная по (6.7.7) из-за скупивания скважин и «появления белых пятен» на площади залежи.

В заключение главы затронем интересный вопрос о влиянии априорной информации на способы и результаты усреднения данных. Сопоставим, в частности, схему §§ 3 и 4 этой главы: стоит только принять гипотезу, что весь объект статистически однороден относительно частей, как мы обязаны применять вариант § 3. На практике не всегда удается установить, точнее — решить, относить ли отдельные измерения к части объекта или ко всему объекту. А может быть, нельзя ограничиваться только этой альтернативой, и имеется множество промежуточных вариантов, соответствующих гипотезам о частичном соответствии частей объекта целому?

Представим себе традиционную в теории вероятностей урну с долей ценных шаров в ней p . Урну перегородили на две равные части так, что $N_1 = N_2 = 1/2N$, где общее N — число шаров. Доля ценных шаров в одной половине оказалась равной p_1 , в другой — p_2 , т. е. $[p = 1/2(p_1 + p_2)]$. Шары в каждой половине урны тщательно перемешиваются, но через перегородку не проникают. Отбирается n шаров (для определенности пусть n — четное число). Схема отбора включает условие, чтобы из каждой половины было отобрано одинаковое число шаров $n_1 = n_2 = 1/2 \cdot n$.

Поскольку распределение ценных шаров между отсеками урны было случайным, эта схема отбора эквивалентна случайному отбору n из N элементов. Погрешность определения доли ценных шаров

$$s = \sqrt{pq/n} = \sqrt{p(1-p)/n}, \quad s^2 = p(1-p)/n.$$

Раздельно оценивая погрешности p_1 и p_2 в отсеках урны

$$s_1 = \sqrt{\frac{p_1(1-p_1)}{n/2}}, \quad s_2 = \sqrt{\frac{p_2(1-p_2)}{n/2}},$$

а затем оценивая среднее значение p по схеме § 4, получим

$$\bar{s}^2 = \frac{1}{2} \cdot s_1^2 + \frac{1}{2} \cdot s_2^2 = \frac{1}{2n} \{p_1(1-p_1) + p_2(1-p_2)\}.$$

Заменяя $p_2 = 2p - p_1$, получим $\bar{s}^2 = [p(1-p) - (p_1 - p)^2]/n$, т. е. $\bar{s}^2 \leq s^2$, так как $(p_1 - p)^2 \geq 0$, причем равенство достигается при $p_1 = p_2 = p$. Например, если $p = 0.6$, $p_1 = 0.4$, $p_2 = 0.8$, то $s^2 = 0.24/n$, $\bar{s}^2 = 0.20/n$. Почему по второй схеме расчета погрешность оценки доли шаров оказалась меньше при том же n ? Этот парадокс, на наш взгляд, разрешается тем, что во второй схеме используется дополнительная информация — о перегородке и отборе n_1 и n_2 шаров из каждого отсека. Фактически был проведен систематический отбор, а при расчете по первой схеме это обстоятельство игнорировалось.

СПОСОБЫ ВЫРАЖЕНИЯ СТЕПЕНИ РАЗВЕДАННОСТИ И ТОЧНОСТИ ОЦЕНКИ ПАРАМЕТРОВ МЕСТОРОЖДЕНИЙ

§ 1. Анализ категорий запасов как показателя степени изученности и точности оценки параметров

Категории запасов месторождений полезных ископаемых являются в настоящее время единственным официально установленным мерилем, характеризующим целый ряд таких признаков, как степень изученности, вероятность их действительного нахождения в недрах, погрешность в определении количества и качества, степень «вскрытости» и подготовленности для промышленного освоения, их производственное назначение. В геологоразведочной литературе интенсивно и постоянно обсуждаются вопросы: какими основными количественными показателями можно охарактеризовать категории запасов, чтобы исключить неопределенность толкования инструкций и неоднозначность практической их квалификации при подсчете и утверждении в ГКЗ; как использовать категории запасов в качестве экономического рычага, стимулирующего совершенствование разведочного процесса, проектирования разработки, добычи полезного ископаемого и охраны недр?

В этом разделе мы проанализируем принципы классификации запасов в недрах, тенденции развития классификаций, связь категорий запасов со стадиями освоения месторождений, соотношение изученности и точности запасов, связь этих характеристик с производственным их назначением — все это с точки зрения оптимального управления геологической разведкой.

Как из самих классификаций запасов (Инструкция. . ., 1960, 1972, 1984), так и из литературы, посвященной их анализу, известно, что официальным основным принципом, положенным в их основу, является степень изученности. Строгого определения этого понятия не дается, но из последующего многократного его применения можно установить, что оно включает а) характер объекта, к которому относятся запасы (территория, структура, месторождение, залежь, часть залежи); б) перечень параметров, подлежащих установлению; в) надежность количественной или качественной оценки этих параметров; г) степень пространственной локализации (детализации) запасов и их отдельных параметров. Рассмотрим, как реализуются отдельные пункты этого перечня на примере классификации 1983 г. для нефти.

Характер объекта. Запасы категорий D_1 и D_2 (именуемые прогнозными ресурсами) относятся в целом к литолого-стратиграфическим комплексам в пределах крупных региональных структур. Перспективные ресурсы (категория C_3) относятся к отдельным площадям (структурам). Однако тут же в классификации указывается, что категория C_3 — это также запасы в не вскрытых бурением пластах разведанных месторождений (структур), т. е. по существу — запасы залежей, а залежь по отношению к структуре — объект низкой ступени классификации. Промышленные запасы категории C_2 относятся к залежи (ее части). Меньшего объекта, чем залежь (ее часть), в классификации не преду-

сматривается, и все последующие категории (C_1 , В, А) относятся к объекту — залежь. В этом положении, на наш взгляд, заключено противоречие. На деле запасы категорий C_1 , В, А относятся к частям залежи, все более мелким при переходе от C_1 к В и далее к А. Хотя формально при подсчете запасов нефти высших категорий залежь может и не подразделяться на мелкие части (блоки) с отдельным подсчетом запасов в них, фактически более детальная геометризация параметров позволяет надежно оценивать запасы и их свойства во все более мелких частях залежи.

Как бы точно ни было известно среднее значение любого подсчетного параметра в целом по залежи, это не дает оснований для перевода их в категорию В или А без изучения пространственной локализации запасов. Следовательно, при дальнейшем совершенствовании классификации есть необходимость прямо указать, что категории C_1 , В и А относятся к частям залежи, и установить абсолютные или относительные размеры частей залежи, к которым относятся эти запасы. В целом по залежи без реальной возможности соотнесения запасов с отдельными ее частями могут пониматься запасы только категории C_2 . Нет необходимости комментировать изменение следующего аспекта изученности — расширение перечня параметров при переходе от категорий Д к А.

Более подробного анализа требует надежность оценки параметров. В классификации она охарактеризована качественными показателями без применения каких-либо количественных критериев. Более того, особенностью классификации является то, что надежность характеризуется посредством рекомендуемой для заданной категории запасов плотности сети скважин и перечня исследований в них. С одной стороны, это облегчает применение классификации, с другой — лишает характеристику степени изученности самостоятельности.

Характеристика степени пространственной локализации повторяет по существу номинацию объекта, к которому относятся запасы, лишь акцентируя то, что запасы высоких категорий селективно характеризуют части залежи.

Считается, что детальность геологоразведочных работ, а следовательно и плотность разведочной сети, определяется существующими требованиями к категориям запасов. По окончании каждой стадии разведки производится подсчет запасов определенной категории, и разведочная сеть этой стадии должна удовлетворять требованиям, предъявляемым к соответствующим категориям запасов. Каждая категория запасов требует определенной степени изученности, а степень изученности определяется (если опустить несущественные детали) в Инструкции плотностью сети. Образуется замкнутый круг и возникает вопрос, не являются ли категории запасов ненужным звеном, имеющим только номинальное значение. Далее возникает вопрос, есть ли необходимость вообще в подразделении на категории, т. е. в дискретном делении степени изученности. Ведь последняя, как и плотность сети, может принимать любые промежуточные «между категориями» значения. Ответить на эти вопросы можно, лишь приняв во внимание то, что изученность запасов должна рассматриваться не как самоцель, а служить критерием их пригодности к определенному производственному назначению. Освоение территорий и месторождений ведется стадиями, и на каждой стадии цель работ и производственное назначение запасов изменяются скачком, качественно. Следовательно, выделение «дискретных» категорий запасов необходимо. Но в таком случае число категорий и требования к их изученности должны строго соответствовать стадиям геологоразведочных работ и освоения месторождений. Классификация запасов при этом должна занимать подчиненное положение по отношению к схемам стадийности работ. В соответствии с принятой в настоящее время стадийностью работ на нефть (табл. 7.1.1) естественно вытекает подразделение запасов на следующие категории по производственному назначению:

- запасы, на основании которых может ставиться подготовка площадей к поисковому бурению (Д) — прогнозные ресурсы;
- запасы, на основании которых могут ставиться поиски залежей нефти — поисковое бурение (С₃) — перспективные ресурсы;
- запасы, на основании которых ставится разведка залежей и месторождений (С₂);
- запасы, на основании которых составляется технологическая схема разработки и осуществляется опытно-промышленная эксплуатация (С₁);
- запасы, на основании которых составляется проект разработки (В);
- запасы, обеспечивающие текущее планирование добычи (А).

Таблица 7.1.1

Схема стадийности геологоразведочных работ на нефть

Этапы	Стадии	Назначение работ	Виды и методы работ	Наименование и категории выявляемых (разведанных) запасов по Инструкции 1983 г.
Поисковый	Региональные геолого-геофизические работы	Изучение общего геологического строения, оценка перспектив нефтегазоносности, выделение зон нефтегазо-накопления и площадей, благоприятных для формирования ловушек	Региональные геофизические исследования, опорное и параметрическое бурение, геологическая съемка, научные тематические работы	Д ₂ — прогнозные ресурсы литолого-стратиграфических комплексов в пределах крупных региональных структур (промышленная нефтегазоносность не доказана); Д ₁ — то же (промышленная нефтегазоносность доказана)
	Подготовка площадей к поисковому бурению	Выявление перспективных на нефть и газ площадей (структур) и подготовка их к глубокому бурению	Детальные геофизические работы, структурное и параметрическое бурение, геологическая съемка	С ₃ — перспективные ресурсы подготовленных для глубокого бурения площадей
	Поиски залежей нефти и газа	Открытие залежей нефти и газа и их предварительная оценка	Поисковое бурение, ГИС	С ₂ — запасы залежи или ее части, подтвержденные лишь данными геологических и геофизических исследований (иногда С ₃)
Разведочный	Разведка нефтяных и газовых месторождений (подготовка месторождений к разработке)	Подсчет запасов по промышленным категориям, изучение эксплуатационных параметров. Подготовка данных к составлению технологической схемы разработки	Разведочное бурение ГИС, опытная эксплуатация	С ₁ — запасы залежи или ее части, подтвержденные промышленными притоками (в части скважин — испытателями пластов) и геолого-геофизическими данными (С ₂ — см. выше)
Промышленного освоения	Опытно-промышленная эксплуатация	Подготовка данных к составлению проекта разработки, добыча нефти	Эксплуатационное бурение в соответствии с технологической схемой	В — запасы залежи или ее части, изученные по данным ОПЭ
	Промышленная эксплуатация	Добыча нефти	Эксплуатационное бурение в соответствии с проектом разработки	А — запасы полностью изученной залежи или ее части

Характерно, что «эволюция» классификаций (1960—1983 гг.) и публикаций (Теория. . ., 1985) происходит именно в направлении конкретизации производственного назначения запасов. Но и в новом варианте категории запасов (последняя графа табл. 7.1.1) все еще неоднозначно связаны с их производственным назначением: для выполнения определенного производственного назначения предусматриваются различные категории запасов. Подготовка площадей к поисковому бурению планируется на основе ресурсов категорий Д₂ или Д₁. Поиски залежей как новых месторождений, так и в невоскрывших пластах разведанных месторождений осуществляются на основе перспективных ресурсов С₃ (здесь принцип определенного производственного назначения запасов выдерживается). Неоднозначны по производственному назначению запасы категории С₂. С одной стороны, это запасы открытых залежей, предназначенных для детальной разведки, с другой — неразведанных (недоразведанных) частей месторождений, передаваемых в промышленное освоение.

В классификациях указывается, что для передачи месторождения в промышленное освоение необходимо обеспечить заданный процент запасов определенной категории. Сравнение классификаций 1960 и 1971 гг. показывает, что если прежде этот процент изменялся для групп месторождений различной сложности и размера, а также в зависимости от освоенности района расположения, то в классификации 1983 г. для передачи месторождения в освоение требуются запасы категории С₁, а доля запасов категории С₂ при этом не должна превышать 20 %. Все это позволяет сделать уверенный вывод о том, что категории по действующей сейчас классификации являются, строго говоря, не категориями производственного назначения, а категориями изученности с параллельным указанием их приблизительного производственного назначения. Такое положение, создавая неопределенность, может служить причиной нерационального использования средств на разведку. Действительно, месторождение, согласно Инструкции, передается из одной стадии освоения в другую с запасами различных категорий (разной степени изученности), т. е. часть запасов оказывается изученной в степени, либо превышающей, либо недостаточной для выполнения производственного назначения данной стадии. А это ведет к «замораживанию» средств на более длительные, чем необходимо, сроки.

Этот анализ приводит к выводу, что в основу классификации, возможно, следовало бы положить не принцип степени изученности, а принцип производственного назначения. Производственное назначение (ПН) категорий запасов складывается из следующих элементов, которые приобретают различное содержание на каждой стадии геологоразведочных работ.

1. Обеспечение геолого-экономической оценки территорий, площадей, литолого-стратиграфических комплексов, объектов или их отдельных частей для правильного направления геолого-поисковых или разведочных работ, отбраковки бесперспективных территорий или нерентабельных месторождений.

2. Обеспечение народнохозяйственного планирования на разных уровнях и учета по данной сырьевой отрасли.

3. Обеспечение необходимой информацией, проектирование следующей стадии работ (стадии освоения).

Следует ли рассматривать выполнимость ПН запасов как альтернативу («выполнено—не выполнено») или как постепенное повышение степени выполнения («хуже—лучше»). От принятия решения по этому вопросу принципиально меняется подход к критериям и методам оптимизации разведки месторождений. Если принять первый вариант, необходимо обосновать тот минимальный уровень изученности, который позволяет выполнить ПН запасов. Если принять второй вариант, то подразделение запасов на категории по принципу производственного назначения окажется вообще невозможным предприятием. Современные тенденции в развитии стандартизации и метрологии склоняют к первому

варианту. Различные возможные подходы к роли категорий запасов в геолого-разведочном процессе изображены на рис. 7.1.1. Все указанные на рисунке варианты в принципе могут привести к решению основной задачи — разведать и освоить месторождение в кратчайшие сроки и с минимальными издержками, но организационно и методически различными путями. В первом варианте (рис. 7.1.1, а) анализ разведочных данных на каждой стадии разведки преследует цель определить, достигнута ли требуемая степень изученности. Этот путь требует количественного выражения степени изученности запасов, а классификация запасов, «приспособленная» к этому варианту, должна количественно регламентировать степень изученности по категориям. Кроме того, в этом случае классификация должна предписывать перечень категорий и их количественное соотношение для выполнения определенного производственного назначения. Этот вариант практически совпадает с существующим положением, с тем отличием, что требования к изученности выражаются скорее качественно, чем количественно. Недостаток этого варианта в том, что степень изученности запасов находится лишь в косвенной связи с их производственным назначением. Проблема оптимизации разведки разрывается при этом на части, находящиеся в компетенции различных организаций, что затрудняет ее осуществление.

Во втором варианте (рис. 7.1.1, б) анализ геологоразведочных данных направлен непосредственно на выяснение вопроса, удовлетворяют ли результаты разведки требованиям, предъявляемым к ним на следующей стадии работ. Преимущества этого варианта заключаются в неформальном творческом подходе к такому анализу, проводимому, как правило, совместно с исполнителями следующих стадий работ, с учетом всех местных и региональных условий, а также требований текущего момента. Недостатком этого варианта является большая субъективность в решениях, меньшая возможность контроля и стандартизации, а также то, что в результате под одноименными категориями будут скрываться запасы различной степени изученности. Количественный подход к оптимизации разведки в этом варианте вряд ли осуществим. Этот вариант весьма характерен в ситуации, когда разведка и разработка ведется одним и тем же объединением, и ГКЗ, учитывая это обстоятельство, не особенно придирчиво утверждает запасы.

В третьем варианте (рис. 7.1.1, в) трудность состоит в определении того, в какой степени ПН выполнено, и в обосновании допустимых «уровней» невыполнения. По-видимому, в этом варианте возможны большие спекуляции.

И наконец, в четвертом варианте (рис. 7.1.1, г) анализ проводится после каждого шага разведки, и разведку данной стадии останавливают и считают выполненной, если затраты на осуществление следующего шага превышают пользу от дополнительной информации, которую можно использовать при проектировании следующей стадии работ, т. е. останавливают в оптимальный момент. Преимущество такого подхода очевидно. Оно заключается в том, что квалификация запасов тесно связана с оптимизацией разведки — является ее непосредственным результатом. Необходимость категорий запасов в их прежнем понимании в этом варианте отпадает. Одновременно отпадает необходимость в специальных названиях и индексах. Именовывать запасы правильнее было бы по соответствующим завершаемым стадиям работ, например запасы «предварительно разведанных месторождений», «детально разведанных месторождений», запасы стадии «опытно-промышленной эксплуатации» и, наконец, «эксплуатационные» запасы. Но, как и во втором и третьем вариантах, под одними и теми же категориями будут фигурировать запасы, сильно отличающиеся по степени изученности.

Тенденции в развитии классификации запасов заключаются не только в более однозначном связывании категорий запасов со стадиями работ и производственным назначением запасов, но и в постепенном «ослаблении» требований к подготовленности разведанных месторождений для промышленного

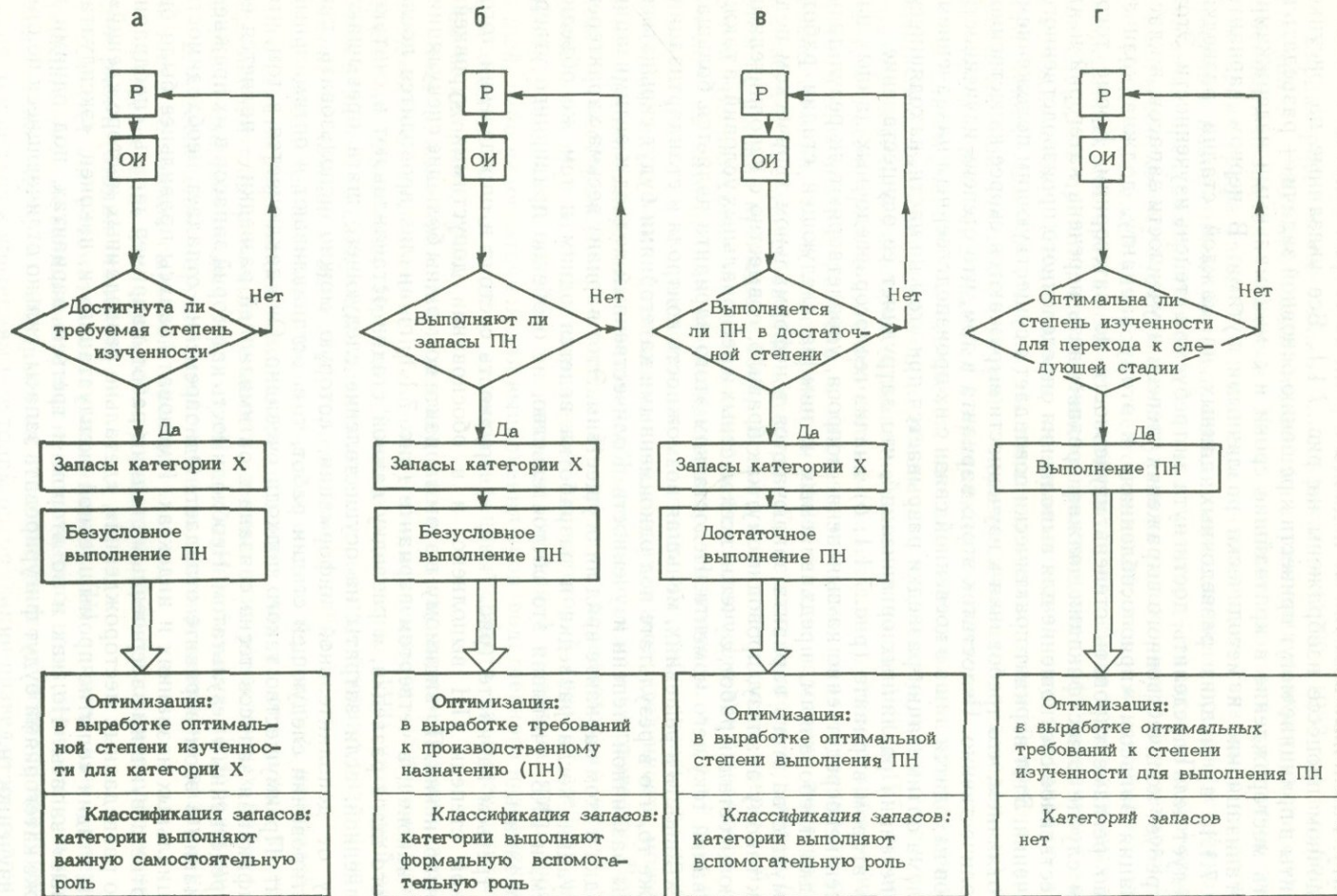


Рис. 7.1.1. Схемы, поясняющие изменение роли категорий запасов в зависимости от принципов их выделения.

Р — разведка, ОИ — обработка геологоразведочной информации, ПН — производственное назначение запасов; а, б, в, г — варианты.

освоения. Все авторы единодушно одобрили снижение доли запасов категории В, необходимых для передачи месторождения в проектирование разработки по классификации 1971 г. по сравнению с классификацией 1960 г.

Коллектив авторов Мингазпрома считает наилучшей основой категории запасов обоснованное *п л а н и р о в а н и е* добычи: 1) оперативное — $A+B+C_1$, 2) краткосрочное пятилетнее — $A+B+C_1$ плюс запасы категории C_2 на выявленных месторождениях, 3) долгосрочное — помимо перечисленных, запасы категории C_2 на новых площадях и прогнозные запасы D_1 и D_2 . Авторы приходят к выводу о необходимости подразделения запасов категории C_2 на подкатегории C_2^1 и C_2^2 на том основании, что в классификации 1971 г. под категорией C_2 объединены запасы, самые различные по степени изученности. Под C_2^1 предлагается объединить запасы уже открытых и разведанных месторождений (их подтверждаемость $\sim 35\%$), под C_2^2 — запасы на новых площадях (подтверждаемость $\sim 9\%$). Доразведку запасов C_2^1 рекомендуется проводить в процессе ОПЭ. Министерство геологии старается доразведать их своими силами, перевести в категории $B+C_1$ и утвердить в ГКЗ, пробуравив дополнительные (хотя и лишние при эксплуатации!) скважины. В классификации 1983 г. запасы C_3 соответствуют рекомендованным C_2^2 . В целом же уровень планирования сам по себе вряд ли можно считать основой классификации. Запасы по производственному назначению обеспечивают любой уровень планирования.

Месторождения даже одного и того же промышленного или генетического типа сильно отличаются по изменчивости изучаемых признаков, и плотность сети наблюдений для них должна быть различной. Все это неоднократно приводило многих авторов к мысли о включении в условия отнесения к категориям предельной погрешности оценки запасов. Еще В. М. Крейтер (1961) рекомендовал принять следующие предельные погрешности для категорий: А — 20%, В — 30%, C_1 — 60%, C_2 — 90%. Такая мысль оказалась привлекательной для многих исследователей, особенно склонных применять математические методы при оценке степени разведанности. Так, А. А. Петров (1967), в целом соглашаясь с В. М. Крейтером, считает, что предельные погрешности должны быть различными для разных полезных ископаемых и их следует ввести не в классификацию, а в инструкцию по ее применению к месторождениям различных полезных ископаемых.

С позиций достижения заданной точности в определении запасов к определению плотности сети подходит И. И. Пастушенко (1963 и др.). Рассматривая железорудные месторождения сложной формы, автор видит главный источник погрешностей в неточности проведения контуров и для оценки этих погрешностей вводит эмпирический коэффициент, зависящий от плотности сети и сложности контура. Предполагается, что на ранних стадиях разведки значение этого коэффициента может быть определено достаточно надежно. В качестве допустимых погрешностей автор принимает для запасов категории В — 25%, C_1 — 50% и на этом основании рекомендует плотность сети, отличающуюся от рекомендованной ГКЗ.

И. И. Гольдфельд и Г. И. Стадников (1964) методом разрежения сети определили, что на полиметаллических месторождениях лентовидной формы оптимальной плотностью сети будет 7—9 рудных профилей по залежи и 4—5 рудных пересечений в каждом из профилей. Погрешность определения запасов при такой плотности не превышает 13—16% при вероятности 0.85—0.90. При разрежении сети погрешности сильно возрастают, а при сгущении — почти не уменьшаются, чем и обосновывается плотность сети.

Э. Карлье (1966), критикуя старую французскую классификацию запасов («вскрытые» — оконтуренные выработками с четырех сторон, «вероятные» — с двух или трех сторон и «возможные» — с одной стороны), приводит мыслимый случай, когда вероятные запасы получаются точнее вскрытых. Новая французская классификация (резервы, ресурсы, перспективы) опирается исключи-

тельно на величину дисперсии оценки запасов. В. Б. Васильев (1980) считает, что в результате предварительной разведки должны быть подсчитаны запасы большей частью по категории C_1 с точностью параметров 10—15 %, в то время как по другим данным (Быков и др., 1978) фактическая точность по окончании детальной (!) разведки составляет для категории А — 10 %, В — 15 %, C_1 — 33 % и C_2 — 50 %.

Вместе с тем известно, что относительная точность запасов при прочих равных условиях в значительной степени (практически — прямо пропорционально) зависит от их величины: чем по большей площади подсчитываются запасы, тем меньше их погрешность, чем большее число блоков или залежей объединены для анализа подтверждаемости, тем она будет больше. В. И. Бирюков (1965, 1967) для этого явления вводит термин — регрессия категорий запасов. Он делает вывод, который можно сформулировать так: повышение категорий запасов прямо пропорционально густоте сети и обратно пропорционально размерам исследуемого участка. Для различных типов месторождений В. И. Бирюков приводит «кривые регрессии». И. И. Русаков и М. Н. Савосин (1973 г.) предлагают выражать требуемую достоверность разведки запасов предельно допустимой ошибкой подсчета и доверительной вероятностью. Для жильных золоторудных месторождений они рекомендуют цифры В. М. Крейтера, считая, что их можно принимать с вероятностью 0.9 для категорий А и В и с вероятностью 0.7 — для категорий C_1 и C_2 . Авторы приходят к выводу, что подсчитать запасы в отдельных блоках по категории А с рекомендуемой точностью и надежностью невозможно. В зависимости от степени изменчивости учитывать можно запасы по категориям: А — от 6 до 30 эксплуатационных блоков, В — от 3 до 11 блоков, C_1 — от 1 до 3 блоков. В фундаментальной работе П. А. Шехтман (1968) указывает, что запасы категории А следует относить к отдельному эксплуатационному блоку, В — к шахтному полю или его части, C_1 — в пределах нескольких шахтных полей, C_2 — в геологических контурах всего месторождения.

Из поистине обширной литературы, посвященной анализу подтверждаемости запасов, мы сделали небольшую намеренно случайную выборку для нефти с целью сравнить фактическую погрешность оценки запасов с вышеизложенными рекомендациями различных авторов (табл. 7.1.2). Такого сравнения, к сожалению, не получилось: коэффициент подтверждаемости вместо ожидаемого случайного колебания вокруг 100 % в среднем едва достигает 50 %. Единственное объяснение этому — субъективный фактор, направленный в сторону увеличения у геологов, выполняющих план прироста запасов, и в сторону уменьшения — у работников, план добычи которым устанавливается в зависимости от запасов. Это наносит существенный ущерб народному хозяйству и показывает, насколько важен вопрос об объективной оценке точности подсчитываемых запасов.

Однако допустимая погрешность оценки запасов по сравнению со степенью изученности представляется слишком односторонней характеристикой. На основании проведенного анализа мы пришли к выводу, что степень изученности можно связать с погрешностью оценки следующим образом: 1) повышение степени изученности требует увеличения числа (расширения перечня) параметров, требующих количественной оценки; 2) допустимая погрешность оценки по мере повышения степени изученности должна относиться ко все более мелким участкам (блокам запасов); 3) сопоставлять допустимые погрешности оценки параметров блоков разной степени изученности нельзя вследствие разных их размеров.

Но и при такой схеме категории запасов, выделенные по производственному назначению, нельзя, да и нет никакой необходимости, однозначно сопоставлять с определенной, регламентированной степенью изученности. Для каждого месторождения она может быть индивидуальной и должна устанавли-

Таблица подтверждаемости запасов при переходе с разведочного этапа на эксплуатационный

Источники информации	Объект анализа (регионы, месторождения)	Метод определения подтверждаемости	Степень подтверждаемости, %	Комментарии авторов
Ованесов, Фейгин, 1975	В целом по стране (1959—1972 гг.)	$\frac{A+B}{C_1}$	62.1	Средние цифры получены на основе данных по 21 региону СССР. Подтверждаемость запасов по регионам изменяется от 15 % (Киргизская ССР) до 82.4 % (Жоми АССР) при переводе из категории C_1 и от 3 % (Узбекская ССР) до 77 % (Удмуртская АССР) при переводе из категории C_2
		$\frac{A+B+C_1}{C_2}$	40	
Мелик-Пашаев и др., 1975	Средние цифры по объединениям (18 объединений)	$\frac{A+B}{C_1}$	56—71	Основной причиной невысокой подтверждаемости запасов категорий C_1 и C_2 является завышение двух параметров — площади нефтеносности и эффективной нефтенасыщенной мощности. Остальные параметры влияют незначительно
		$\frac{A+B+C_1}{C_2}$	43	
Крылова и др., 1975	Куйбышевская область	$\frac{A+B+C_1}{C_2}$	60	Изменение запасов по структурам связывается в основном с изменением площади нефтеносности и средней эффективной мощности. Отличие фактической площади от предполагаемой связано с невыясненным соотношением структурных планов опорных горизонтов и недостаточной точностью структурных карт
			(1959—1969 гг.)	
			58	
			(1959—1963 гг.)	
			70	
			(1963—1969 гг.)	
Долженков и др., 1978	Некоторые месторождения Татарии	Не указано	45.4—156.1	
Кочетов и др., 1974	Мегионское, Ватинское месторождения	$\frac{B}{C_1}$	15—36	Подтверждаемость запасов по исследованным залежам категории C_1 весьма высока. Запасы категории C_2 не подтверждаются. Основную роль в изменении запасов играют уточнение площади залежей и мощности пластов. Остальные параметры изменяются на 1—6 %
		$\frac{C_1+B}{C_1}$	67—114	
		$\frac{B+C_1+C_2}{C_2}$	0—4	
Борисенко, Соколов, 1975	Некоторые месторождения Ставропольского края	—	104	Запасы были уточнены вследствие изменения формы, размеров структуры и особенно очертания сводовых зон
		—	55	
Кузнецов и др., 1975	В целом по Башкирии	$\frac{A+B}{C_1}$	62	Низкая подтверждаемость запасов по категории C_2 на втором этапе обусловлена тем, что в бурение вводятся в основном небольшие структуры с часто встречающимся несоответствием структурных планов
		$\frac{C_1+A+B}{C_2}$	60	
		$\frac{C_1+A+B}{C_2}$	24	
		$\frac{C_2}{C_2}$		

Таблица 7.1.2 (продолжение)

Источники информации	Объект анализа (регионы, месторождения)	Метод определения подтверждения	Степень подтверждения, %	Комментарии авторов
Иванникова, Мандрыкин, 1972	Терско-Сунженская нефтегазоносная область	Подтверждаемость запасов — производство подтверждаемости отдельных параметров		Установлено, что подтверждаемость отдельных параметров и запасов может колебаться от 0 до 100 % (нет случаев, когда хотя бы по одному параметру подтверждаемость превысила 100 %)

ваться в процессе оптимизации плотности сети на данной стадии разведки. Количественная характеристика степени изученности необходима лишь как средство, используемое в процессе оптимизации. Категории запасов должны выделяться (и называться) по стадиям работ. Соответствующая стадия считается законченной, как только полученная информация (в том числе о количестве запасов) позволяет выполнить производственное назначение следующей стадии. Вопрос о допустимом соотношении категорий запасов при передаче месторождения в промышленное освоение при таком подходе отпадает.

Перечисленное может относиться не только к месторождению в целом, но и к его частям, представляющим самостоятельный промышленный интерес. При этом остальные его части могут находиться в других стадиях освоения.

§ 2. Аналитические методы оценивания точности запасов и других параметров, не связанные с их геометризацией

Оценивание погрешности параметров месторождений не является самоцелью, а может служить лишь вспомогательным средством при оптимизации детальности разведки. Последнюю можно расчленить на две проблемы: а) установление связи между детальностью разведочных работ и погрешностями определения основных параметров месторождений, б) установление оптимального значения допустимых погрешностей.

Общие тенденции развития стандартизации, выраженные в ГОСТах (ГОСТ 16263-70, ГОСТ 7.33-81, ГОСТ 8.0011-72, ГОСТ 8.207-76), требуют «с целью правильного и однозначного использования экспериментальных и других численных данных, оценки их достоверности и возможности воспроизведения» строгих стандартных форм представления точности результатов. Хотя сфера действия упомянутых ГОСТов не ограничена отдельными отраслями, в геологии и разведке месторождений (кроме маркшейдерской службы) их еще не придерживаются ввиду сложности расчета доверительных интервалов геологических параметров и еще большей сложности расчета границ неучтенных систематических погрешностей.

Некоторые материалы этого раздела частично пересекаются с материалами следующей главы, однако если здесь речь идет преимущественно о связи плотности сети с точностью оценки параметров, то в гл. 8 — об экономической оптимизации плотности сети.

Обобщение отечественной литературы раннего периода по данному вопросу можно найти, например, у Д. А. Зенкова (1950, 1951, 1957, 1961). Все существовавшие на тот период методы оценки точности геологических параметров Д. А. Зенков подразделил на следующие: 1) сравнения данных разведки с данными эксплуатации, 2) разрежения сети, 3) аналогии, 4) расчетно-аналитические, 5) экспериментальные. Первый метод почти всегда позволяет получать надежные данные и настойчиво рекомендуется автором, однако ретро-

спективные выводы не пригодны для управления разведкой. Метод разрежения также позволяет получить хорошие сравнительные данные, однако для его применения необходима заведомая переразведанность месторождения. Следовательно, первые два метода можно применить с пользой для дела лишь по аналогии на других месторождениях. Метод аналогии, по мнению Д. А. Зенкова, заслужено пользуется широким распространением в практике. На его основе составлено большинство действующих сейчас инструкций и руководств по разведке и подсчету запасов. В то же время Д. А. Зенков предостерегал от формального применения метода, так как в природе нет тождественных месторождений и рудных тел.

Под расчетно-аналитическими методами автор подразумевает методы математической статистики. Возражения против ее применения для решения рассматриваемого вопроса были к этому периоду достаточно хорошо известны из работ В. Г. Соловьева (1939), Д. А. Казаковского (1937, 1948, 1957, 1959), К. Л. Пожарицкого (1958) и других геологов и изложены Д. А. Зенковым сравнительно полно. Эти возражения заключаются в первую очередь в наличии зависимости между значениями параметра в точках, находящихся на близких расстояниях друг от друга.

Анализируя перечисленные Д. А. Зенковым методы, можно сделать вывод, что метод аналогии, выделенный в самостоятельный, является составной частью всех остальных. Действительно, в первом методе мы получаем интересные нас данные только после отработки месторождения, во втором — после его разведки (точнее — переразведки), тогда как они необходимы перед и в процессе разведки месторождений, т. е. результаты можно применить лишь на аналогичном, сходном объекте, причем сходство требуется устанавливать по данным, получаемым на ранних стадиях разведки. В четвертом методе применение аналогии и сравнения обнаруживается не сразу. В его основу кладется допустимая погрешность в определении параметров, а решение о плотности сети принимается в конечном счете на основании аналогии с месторождениями, имеющими близкое значение коэффициента вариации (или другого показателя изменчивости) изучаемого признака. В связи с проблемой оптимизации следует отметить, что с помощью перечисленных методов можно получить некоторые величины, характеризующие возможные отклонения данных разведки от истинных значений. При этом остается недоказанной допустимость этих отклонений, тем более — их оптимальность. Вопрос оптимизации только перемещается из одной сферы в другую: вместо того чтобы постулировать плотность сети, мы в явном или неявном виде постулируем вероятную погрешность определения параметров.

После рассмотренной статьи Д. А. Зенкова (1957) появилось значительное число вариантов расчетно-аналитического метода определения плотности сети, удовлетворяющей заданной погрешности, детальный анализ которых не входит в нашу задачу. Во многих из этих методов удается преодолеть принципиальные затруднения, препятствовавшие применению известной и до сих пор широко распространенной среди геологов формулы математической статистики:

$$n \geq t^2 \sigma^2 / m^2, \quad (7.2.1)$$

где n — необходимое число точек наблюдений, m — допустимая погрешность в определении параметра, σ — его наблюдаемое стандартное отклонение, t — коэффициент вероятности непревышения допустимой погрешности m . К ним относятся прежде всего методы, основанные на изучении первых, вторых и более высоких разностей между значениями показателя в соседних точках. Метод последовательных разностей впервые выдвинут Д. А. Казаковским (1948), но вначале не получил широкого распространения из-за того, что автор связал его с весьма частной моделью. Как теперь выясняется, метод

Д. А. Казаковского мог бы стать началом современной геостатистики, но не получил в свое время развития в должном направлении. Он заключается в определении «абсолютного показателя изменчивости» залежи

$$\mu = \sum_{i=1}^n \Delta''/n = \bar{\Delta}'', \quad (7.2.2)$$

где Δ'' представляют собой вторые разности (т. е. разности разностей значений изучаемой величины в соседних точках), и «относительного показателя изменчивости»

$$J = \mu/k_{cp}, \quad (7.2.3)$$

где k_{cp} — среднее значение изучаемой величины, а также в расчете «показателя разведанности»

$$R = n/(1000J), \quad (7.2.4)$$

где n — число точек наблюдений (скважин). Определяющим достоинством показателей μ , J и R является учет пространственной связи значений залежи. Рассматривая подход Д. А. Казаковского с позиций современной геостатистики, следует отметить, что средняя величина вторых разностей, характеризующая изменчивость первых разностей, фактически определяет значение коэффициента автокорреляции или полувариограммы. Переход от показателей изменчивости к «ошибкам аналогии», т. е. случайным погрешностям в определении параметров, Д. А. Казаковский видит в экспериментах на предметных (например, восковых) моделях. Автор считал зависимость между погрешностью аналогии и показателем разведанности не зависящей от модели. Такой эмпирический подход допустим, но он сводит на нет рациональное зерно метода, так как трудно подобрать модель, наиболее полно отображающую особенности данного месторождения. Если же такая модель будет подобрана, то по ней можно определить и сами погрешности, не привлекая промежуточные показатели.

Позже метод Д. А. Казаковского подвергался неоднократной модернизации. Например, Ю. М. Родченко (1965) предлагает строить так называемые диаграммы разведанности, связывающие предельные ошибки (объема) с вычисленным по Д. А. Казаковскому показателем разведанности, величиной разведанной площади и плотностью сети. Эти диаграммы разведанности, как и график Д. А. Казаковского, строятся эмпирическим методом разрежения сети и сравнения с данными эксплуатации. Их использование наталкивается на те же трудности: либо необходимо доказать аналогичность изучаемого месторождения тому, которое взято за эталон, либо предлагаемую диаграмму строить для изучаемого месторождения, а это возможно только после переразведки.

В. В. Шевелев (1967) при анализе плотности разведочной сети на месторождениях глин для аппроксимации распределения мощностей применил функцию Вейбулла и в последующем использовал ее для определения величин погрешностей среднего значения (при заданных вероятностях) в зависимости от числа точек наблюдений, а следовательно, и плотности сети. Вычисленные величины погрешностей, по его мнению, хорошо сходятся с данными Ю. М. Родченко для аналогичных сетей и разведанных площадей.

Рассмотрим очень интересный метод А. И. Осецкого (1953). В известных руководствах, например у Е. О. Погребницкого (1968), этот метод рассмотрен недостаточно глубоко и вне всякой связи с методом Д. А. Казаковского, хотя имеет с ним тесную логическую связь. Метод относится к анализу достоверности разведки квадратной сетью скважин. Результаты изображаются в виде объемной модели, в которой значения параметра залежи (мощность, содержание) откладываются по скважинам перпендикулярно плоскости, на которую проектируется залежь. При этом получают некоторую поверхность, объем под которой будет характеризовать запасы или, смотря по смыслу,

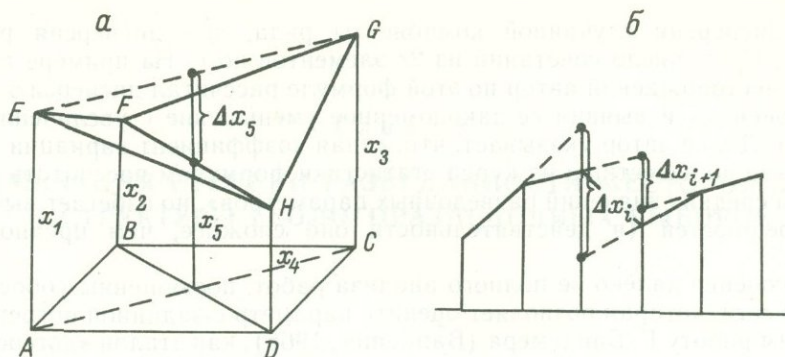


Рис. 7.2.1. Сравнение метода А. И. Осецкого (а) и его линейного варианта (б).

x_1, x_2, x_3, x_4 — содержания по скважинам; Δx_5 — разность содержаний в воображаемой центральной скважине, определенных различными вариантами интерполяции; $\Delta x_i, \Delta x_{i+1}$ — разность содержаний в пробах профиля, определенных с различными вариантами интерполяции, как показатель точности в интервале четырех проб.

среднее значение другого параметра. Каждую призму с основанием в виде квадрата-ячейки можно разбить вертикальной плоскостью на две трехгранные призмы, причем это можно сделать двумя различными способами. Объем в каждой ячейке можно подсчитать дважды в соответствии со способами разбиения. Разность между объемами характеризует, по замыслу А. И. Осецкого, возможную ошибку аналогии в данной ячейке. Задаваясь предельными значениями этих ошибок, так называемыми красными цифрами, необходимо сгущать разведочную сеть до тех пор, пока разность между объемами остается большей, чем «красные цифры». Метод А. И. Осецкого в отличие от многих других позволяет анализировать плотность сети не в целом по месторождению, а по отдельным ячейкам. Развивая (или, лучше сказать, — упрощая) метод А. И. Осецкого для линейного случая (профиля опробования, рис. 7.2.1), легко обнаружить его связь с методом вторых разностей Д. А. Казаковского. В течение двух десятилетий метод локальной характеристики степени разведанности, отличающийся простотой и наглядностью, был основательно забыт и лишь в семидесятых годах возрожден в модернизированном варианте В. А. Петровым (Иванов и др., 1981) и С. В. Шаклеиным (1979). В основу нового показателя положена все та же разность между интерполяционными значениями в точке, лежащей на пересечении линий интерполяции — разность перекрестной интерполяции. По существу речь идет о степени отклонения геологической поверхности от плоскости в пределах площади, ограниченной ближайшими четырьмя точками измерений. Недостаток метода Осецкого — Петрова — Шаклеина состоит в том, что их показатель трудно связать с величиной средней квадратической погрешности оценки. Именно в связи с методом Осецкого напрашивается мысль о количественном выражении точности разведки месторождений в виде среднего значения относительной средней квадратической погрешности оценки запасов или другого лимитирующего параметра как функции размера ячейки (блока запасов).

Подробного анализа заслуживает статья А. А. Петрова (1963), в которой для характеристики случайной компоненты изменчивости предлагается метод конечных разностей. Из математической статистики известно, что

$$s^2 = \frac{s_r^2}{C_{2r}^r} = \frac{\sum_{i=1}^{n-r} (\Delta_i^r)}{C_{2r}^r (n-r)}, \quad (7.2.5)$$

где s^2 — дисперсия случайной компоненты ряда, s_r^2 — дисперсия разностей порядка r , C_{2r}^r — число сочетаний из $2r$ элементов по r . На примере гравийно-песчаных месторождений автор по этой формуле рассчитал дисперсию «случайной компоненты» и выявил ее закономерное уменьшение с увеличением плотности сети. Далее автор указывает, что, «зная коэффициент вариации (дисперсию), можно по известным из курса статистики формулам рассчитать точность выведения средних значений разведочных параметров», но избегает вычисления этих погрешностей (в действительности оно сложнее, чем предполагается в статье).

В заключение далеко не полного анализа работ, посвященных обоснованию плотности сети, которая позволяет оценить параметр с заданной погрешностью, рассмотрим работу Г. Бандемера (Bandemer, 1966), как эталон «догестатистического» периода. Автор рассматривает два случая: а) функция распределения изучаемого параметра неизвестна, б) значения параметра распределены нормально. В первом случае по разведочным данным строится эмпирическая функция распределения изучаемой величины и по критерию А. Н. Колмогорова определяется полоса, в которой с заданной вероятностью расположена «истинная» функция распределения. Если бы мы имели способ связать погрешности в определении параметра с этой доверительной полосой, можно было бы определить необходимую плотность сети. По этому поводу Г. Бандемер пишет, что установление доверительной вероятности зависит от того, к какому последствию на практике приводит ошибка в определении $F(h)$ (истинной функции распределения параметра, — Ю. Т). Если это последствие можно выразить численно, то вместе с последующими выводами мы получим средство оптимизировать эту доверительную вероятность.

Для случая, когда значения изучаемой величины распределены нормально, идея метода в общем не нова и заключается в том, что отклонения оценки параметра от его истинного значения в единицах стандартного отклонения подчиняются нормированному нормальному распределению, и мы имеем возможность указать доверительный интервал (с заданной вероятностью), в котором находятся эти отклонения, или наоборот, задав этот доверительный интервал, можно рассчитать необходимое число наблюдений. Отличительная особенность решения этого вопроса Г. Бандемером заключается в том, что автор в целях недопущения переразведки прежде определяет n и h и n ю ю г р а н и ц у необходимого числа наблюдений, зависящую от нижней границы истинного значения дисперсии по ее оценке. Г. Бандемер исходит из полной независимости значений признака от их расположения, т. е. при всей кажущейся строгости решения допущения, которые положены в его основу, в большинстве случаев не выполняются и были опровергнуты более 30 лет назад (см., например, Соловьев, 1939). Плотность сети фактически можно изменять только скачкообразно в геометрической прогрессии со знаменателем, равным самое меньшее двум. Эти же замечания можно сделать к статье В. И. Раевского (1966), применившего для определения необходимого числа скважин при разведке разработанный А. Вальдом метод последовательного анализа. После каждой пробуренной скважины производится проверка гипотезы о том, заключено ли определяемое среднее содержание (или мощность) в заданном интервале.

**ОПТИМИЗАЦИЯ СТЕПЕНИ РАЗВЕДАННОСТИ МЕСТОРОЖДЕНИЙ
И СТРУКТУРЫ ГЕОЛОГОРАЗВЕДОЧНЫХ ВЫБОРОК**

В настоящее время слово оптимизация стало очень популярным и употребляется в широком смысле улучшения чего-либо. В этой главе термин оптимизация будет использоваться в своем первоначальном узком смысле. Пусть имеются некоторые параметры процесса (в данном случае разведки месторождения), которыми можно варьировать. Это плотность разведочной сети, параметры размещения скважин или других разведочных выработок, допустимые погрешности оценки запасов месторождения и т. д. Из экономических соображений можно определить так называемую целевую функцию, связывающую эти параметры с величиной полезного эффекта от разведки. Оптимизацией называется процедура нахождения таких значений параметров, для которых полезный эффект при заданных средствах на разведку достигает максимума или при заданном полезном эффекте затраты на его достижение достигают минимума.

Оптимизация разведки месторождений является в настоящее время более сложной проблемой по сравнению с оптимизацией других производственных или технологических процессов. Во-первых, параметры, определяющие этот процесс и его результаты, носят вероятностный характер. Во-вторых, целевая функция настолько сложна, что не имеет даже смысла говорить об ее аналитическом выражении. Полезный эффект можно исчислить лишь в результате выполнения сложного алгоритма, включающего в общем случае моделирование месторождения, его разведки, проектирования разработки и т. д. В-третьих, разведка преследует множество целей и производится в несколько стадий.

К настоящему времени уже накопилось большое количество работ по оптимизации разведки (Matheron, Formeri, 1963; Марголин, 1974; Оптимизация. . ., 1976; Быков и др., 1978; Аронов и др., 1982; Кноринг, 1980, и др.). Однако практически осуществимой методики еще нет, и контуры ее едва вырисовываются. Каждый из авторов развивает наиболее близкую ему одну сторону проблемы. В угоду высокому уровню абстракции и математическому изяществу зачастую игнорируется реальное положение вещей. При чтении работ бывает трудно установить, какую стадию разведки имеет в виду автор. Предлагаемые методы оптимизации, как правило, сложные в вычислительном отношении, обычно заимствованы из технических наук без учета особенностей организации геологического производства. Совершенно недостаточное внимание в публикациях было уделено анализу критериев оптимизации.

По нашему мнению, сейчас возникла острая необходимость в возврате к первоосновам и переосмыслении результатов начального периода создания методов оптимизации разведки месторождений с самых общих позиций.

§ 1. Аспекты оптимизации разведки месторождений

В соответствии со стадийностью геологоразведочных работ будем различать оптимизацию предварительной разведки и оптимизацию детальной разведки. В дальнейшем критерии и методы оптимизации разведки этих двух стадий будут рассматриваться раздельно. Всякая разведка месторождений сочетает два взаимопроникающих аспекта: а) открытие, выявление в пределах заданного участка полезного ископаемого; б) установление его количества. В зависимости от конкретных условий в разведке преобладает один из этих аспектов. Таким образом, мы будем отличать оптимизацию разведки «на прирост» запасов и оптимизацию разведки «на уточнение» (на повышение достоверности) запасов. В стадии детальной разведки преобладает аспект «на уточнение». В этом аспекте разведки мы предвидим необходимость различать еще два аспекта, или акцента. Действительно, все более настоятельно требуется оценка погрешностей разведки, т. е. изучение того, насколько точно разведка, другими словами — «разведка» самой разведки, или метаразведка. Этот аспект в явном виде никем не обсуждался, хотя попытки получить оценку точности запасов приводят к необходимости устанавливать такие характеристики изменчивости, которые с помощью «правильной» разведки установить невозможно (Ткачев, 1977) и на оценку которых необходимо отпускать часть разведочных средств. Какую именно — в этом заключается содержание оптимизации разведки (метаразведки), направленной на уточнение оценки точности.

Оценка экономической эффективности разведки также, с изложенных позиций, имеет свои аспекты. Например, эффективность разведки на прирост оценивается выявленными запасами на рубль затрат, на уточнение — отношением разности погрешностей оценки одной тонны запасов к затратам на это уточнение.

Далее, проблема оптимизации разведки, к какому бы аспекту она ни относилась, расчленяется на следующие основные части: а) определение необходимой и достаточной (оптимальной) степени разведанности месторождения; б) определение оптимальной методики разведки, т. е. методики, обеспечивающей заданную степень разведанности (заданную точность определения параметров) с минимальными затратами; в) выбор оптимальных технических средств для осуществления намеченной методики. Части (в) в данной книге мы касаться не будем, а из части (б) рассмотрим только оптимизацию структуры геологоразведочной выборки.

Изложенные представления условно изображены на рис. 8.1.1.

§ 2. Оптимизация объема геологоразведочных работ на прирост запасов

Разведка на прирост запасов наиболее характерна для стадии предварительной разведки, однако выводы этого раздела могут быть полезными на поисковом этапе геологоразведочных работ, в частности для стадии поисков залежей нефти и газа.

В пределах разведываемого месторождения разведка на прирост запасов может осуществляться на глубину и фланги, и если содержание или мощность в этих направлениях падают, оптимизация объема работ тесно переплетается с вопросом об оптимальных границах месторождения, за пределами которых разведку проводить нецелесообразно. В теоретическом аспекте этот вопрос детально обсуждался Ж. Матероном (1968), Ж. Матероном и Ф. Формери (Matheron, Forneri, 1963), А. М. Марголиным (1974) и др. Разведка на расширение может вестись и сгущением разведочной сети в пределах «первоначальных» границ месторождения.

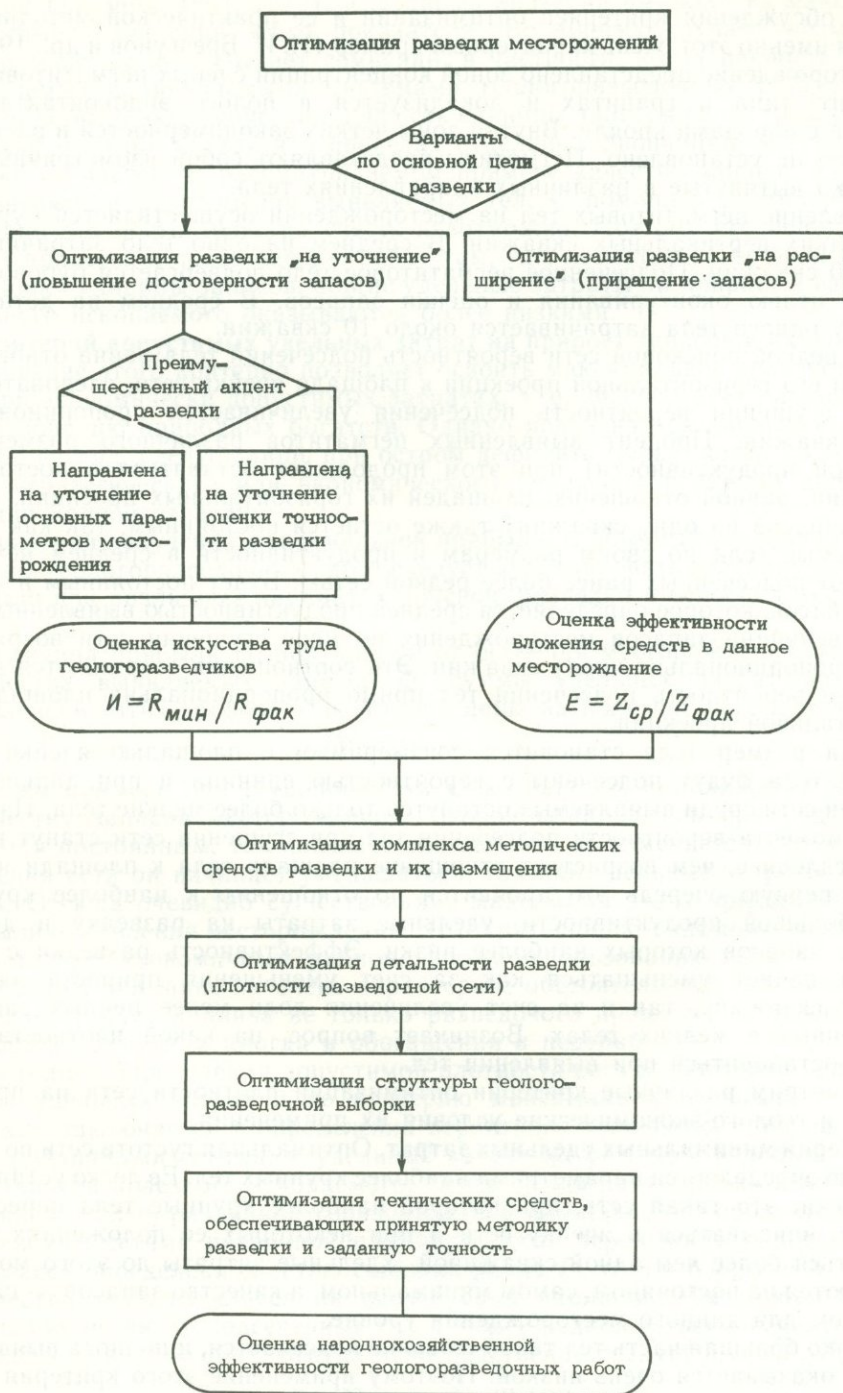


Рис. 8.1.1. Аспекты оптимизации в зависимости от основной цели разведки.

$R_{\text{фак}}$ — фактические затраты на разведку с заданной точностью; $R_{\text{мин}}$ — минимально возможные затраты на разведку с такой же точностью; $Z_{\text{фак}}$ — затраты на прирост тонны запасов данного месторождения; $Z_{\text{ср}}$ — то же, среднее по отрасли.

Для обсуждения критериев оптимизации и ее практической методики мы выбрали именно этот очень характерный пример (Н. И. Брезгунов и др., 1975 г.).

Месторождение представлено зоной концентрации слепых пегматитовых тел камерного типа в гранитах и локализуется в полосе эндоконтакта этих гранитов с породами кровли. Внутри зоны четких закономерностей в распределении тел не установлено. Пегматиты представляют собой изометричные или несколько вытянутые в различных направлениях тела.

Выявление пегматитовых тел на месторождении осуществляется бурением сети мелких вертикальных скважин. В среднем на одно тело затрачивается около 50 скважин. Подсеченное пегматитовое тело подвергается буровой разведке с целью оконтуривания и оценки запасов. В среднем на детальную разведку одного тела затрачивается около 10 скважин.

При редкой поисковой сети вероятность подсечения тела равна отношению площади его горизонтальной проекции к площади ячейки сети, следовательно, при ее сгущении вероятность подсечения увеличивается пропорционально числу скважин. Процент выявленных пегматитов различного размера (и различной продуктивности) при этом продолжает оставаться в постоянной пропорции, равной отношению площадей их горизонтальных проекций. Количество запасов на одну скважину также остается постоянным, так как вновь подсечаемые тела по своим размерам и продуктивности в среднем не отличаются от подсеченных ранее более редкой сетью. Будет постоянным и «качество» запасов, которое определяется средней продуктивностью выявленных тел. Общая величина запасов месторождения по мере сгущения сети возрастает прямо пропорционально числу скважин. Эти соотношения сохраняются до тех пор, пока вероятность подсечения тел прямо пропорциональна площади его горизонтальной проекции.

Когда размер тела становится соизмеримым с площадью ячейки сети, крупные тела будут подсечены с вероятностью единица и при дальнейшем сгущении сети среди выявляемых останутся только более мелкие тела. Начиная с этого момента вероятности подсечения тел при сгущении сети станут возрастать медленнее, чем возрастает отношение площади тела к площади ячейки сети. В первую очередь это проявится по отношению к наиболее крупным телам большой продуктивности, удельные затраты на разведку и добычу единицы запасов которых наиболее низки. Эффективность разведки с этого момента начнет уменьшаться как за счет уменьшения прироста запасов на одну скважину, так и за счет увеличения доли менее ценных запасов, заключенных в мелких телах. Возникает вопрос, на какой плотности сети следует остановиться при выявлении тел.

Рассмотрим различные критерии оптимизации плотности сети на прирост запасов и геолого-экономические условия их применения.

Критерий минимальных удельных затрат. Оптимальная густота сети по этому критерию определяется параметрами наиболее крупных тел. Ее легко установить графически: это такая сеть, при которой наиболее крупные тела перестанут свободно вписываться в ячейку сети и при некоторых ее положениях будут подсекаться более чем одной скважиной. Удельные затраты до этого момента сохраняются на постоянном, самом минимальном, а качество запасов — на максимальном для данного месторождения уровне.

Однако бóльшая часть тел такой сетью не подсекается, и полнота выявления запасов оказывается очень низкой. Поэтому применение этого критерия допустимо лишь при наличии неограниченного в пределах спроса на сырье резерва перспективных площадей. На практике такие условия встречаются крайне редко.

Критерий допустимых удельных затрат на выявление единицы запасов. Отраслевую экономическую эффективность геологоразведочных работ определяют отношением затрат на поисковые и (или) разведочные работы к приросту

запасов (Хрушов, 1967, 1978). Указанный критерий вытекает из максимально допустимого значения этого показателя. Во многих случаях возможность его применения не вызывает сомнений, в частности, когда ценность запасов, выявляемых сравниваемыми системами геологоразведочных работ, эквивалентна. Однако экономическая ценность единицы запасов существенно изменяется как по отрасли, так и в пределах отдельных провинций, месторождений и даже их частей. В рассматриваемом примере качество запасов изменяется в зависимости от размеров выявляемых тел. С помощью рассматриваемого критерия это обстоятельство учесть невозможно. Поэтому его использование может привести к прекращению работ там, где, несмотря на большие затраты на прирост единицы запасов, общие удельные затраты на прирост и извлечение полезного ископаемого оказываются более низкими.

Критерий допустимых удельных затрат на прирост запасов и их извлечение. Применение этого критерия позволяет освоить максимальное количество запасов при экономически допустимых затратах, т. е. способствует рациональному использованию природных богатств. Однако пользоваться данным критерием, на наш взгляд, следует лишь при остром дефиците сырья, когда по соображениям стратегического или экономического характера временно допускается дотация данной отрасли.

Критерий получения максимальной прибыли. В обычных условиях при оптимизации геологоразведочных работ следует анализировать величину ожидаемой прибыли при различных вариантах поисков, разведки и освоения месторождения.

Величина прибыли Π , получаемой от освоения месторождения, зависит от количества выявляемых запасов Q , суммы удельных затрат на прирост и извлечение этих запасов $У$ и оптовой цены на полезное ископаемое \mathcal{C} :

$$\Pi = Q(\mathcal{C} - У). \quad (8.2.1)$$

В нашем примере при сгущении сети Q возрастает, приближаясь к общему количеству запасов месторождения. Второй сомножитель при редких сетях является постоянным, так как качество запасов не изменяется; затем по мере сгущения сети он начинает уменьшаться, так как удельные затраты $У$ начнут расти. Отсюда очевидно существование максимума этого произведения, т. е. прибыли от освоения месторождения. Этот критерий, по-видимому, более других соответствует экономически рациональному использованию недр.

Обратим внимание на очень характерное обстоятельство, что оптимизация разведки требует анализа не только разведочного процесса, но и дальнейших процессов: добычи, а иногда и обогащения и переработки.

Методика определения допустимой плотности сети по критерию допустимой суммы удельных затрат и по критерию максимальной прибыли с единицы площади заключается в определении вероятного количества запасов, выявляемых различными сетями, и удельных затрат на разведку и добычу единицы этих запасов полезного ископаемого. Основой для расчета этих величин является число тел n_i различной горизонтальной площади S_i и продуктивности q_i , выявляемых сетями различной плотности. Если вероятность подсечения тела i -го класса определенной сетью скважин равна P_i , то $n_i = N_i P_i$, где N_i — общее количество тел i -го класса на исследуемой площади. Мы можем получить его оценку по числу подсеченных тел n'_i более редкой сетью и вероятностью P'_i их подсечения: $N_i = n'_i / P'_i$, откуда непосредственно следует $n_i = n'_i P_i / P'_i$. Запасы во всех выявляемых телах $Q = \sum_{i=1}^k n_i q_i = \sum_{i=1}^k n'_i P_i / P'_i \cdot q_i$, где q_i — запасы в теле i -го класса. Если имеется по одному телу на класс, то

$$Q = \sum_{i=1}^n P_i / P'_i \cdot q_i. \quad (8.2.2)$$

Решение располагается в области, где вероятности подсечения существенно отличаются от отношения площади тела к площади ячейки сети. В этом случае для прямоугольных сетей вероятности P_i и P'_i можно определить по таблицам И. Д. Савинского (1964), заменив реальные тела равновеликими эллипсами с такими же коэффициентами сжатия. Вычисленные вероятности даны в табл. 8.2.1. Они положены в основу расчета количества выявляемых запасов.

Таблица 8.2.1

Вероятности подсечения тел сетями различной плотности

Номера тел п./п.	Площади тел, м ²	Размеры ячейки поисковой сети					
		50×40	50×40ц	25×20	25×20ц	12.5×10	12.5×10ц
1	28	0.01	0.03	0.06	0.11	0.24	0.54
2	46	0.02	0.05	0.10	0.16	0.36	0.59
3	56	0.03	0.06	0.12	0.23	0.48	0.89
4	61	0.03	0.06	0.12	0.23	0.40	0.74
5	65	0.03	0.07	0.14	0.24	0.49	0.74
6	66	0.03	0.07	0.12	0.23	0.48	0.74
7	68	0.03	0.07	0.12	0.23	0.48	0.74
8	84	0.04	0.08	0.15	0.29	0.55	0.76
9	88	0.04	0.08	0.15	0.29	0.55	0.96
10	113	0.06	0.10	0.24	0.46	0.69	0.96
11	177	0.09	0.19	0.35	0.67	0.99	1.00
12	180	0.09	0.20	0.40	0.74	0.96	1.00
13	184	0.12	0.19	0.38	0.74	0.86	1.00
14	219	0.11	0.24	0.48	0.81	1.00	1.00
15	224	0.10	0.19	0.38	0.74	0.86	1.00
16	257	0.17	0.27	0.58	0.89	1.00	1.00
17	293	0.12	0.27	0.58	0.89	1.00	1.00
18	295	0.15	0.27	0.58	0.97	1.00	1.00
19	301	0.15	0.29	0.55	0.76	1.00	1.00
20	301	0.15	0.28	0.56	0.95	1.00	1.00
21	301	0.15	0.28	0.56	0.95	1.00	1.00
22	354	0.17	0.36	0.55	1.00	1.00	1.00
23	436	0.24	0.45	0.79	1.00	1.00	1.00
24	469	0.24	0.38	0.79	1.00	1.00	1.00
25	493	0.17	0.36	0.69	1.00	1.00	1.00
26	532	0.24	0.46	0.86	1.00	1.00	1.00

Примечание. 50×40ц означает сеть, образованную из сети 50×40 заданием скважин в центре каждой ячейки.

При стоимости одного погонного метра скважин a_1 , ее глубине l_1 и числе скважин N на единицу площади при данной плотности сети удельные затраты на поиски Y_n равны:

$$Y_n = a_1 l_1 N / Q, \quad (8.2.3)$$

где Q — количество запасов в выявленных телах.

Разведка каждого выявленного тела, необходимая для его вскрытия и отработки, осуществляется бурением m дополнительных скважин. При средней длине разведочных скважин l_2 удельные затраты на разведку единицы запасов Y_p составят

$$Y_p = \frac{1}{Q} a_1 l_2 m \sum_{i=1}^{n'} n_i. \quad (8.2.4)$$

Расчеты стоимости добычи произведены для варианта подземной разработки с раздельным вскрытием небольших тел шахтами. В стоимость добычи входят:

затраты на проходку шахты и вентиляционной скважины, прямо пропорциональные их глубине — a_2h ; постоянные затраты на сооружение околоствольного двора, подходного квершлага и восстающего — a_3 , затраты на выемку продуктивной массы, пропорциональные обрабатываемому объему — a_4V_i . Подлежащие выемке объемы и вероятные значения запасов для различных классов пегматитовых тел приведены в табл. 8.2.2. Удельные затраты на добычу единицы запасов сырья составят:

$$Y_x = 1/Q \left\{ (a_2h + a_3) \sum_{i=1}^{n'} n_i + a_4 \sum_{i=1}^{n'} n_i v_i \right\}. \quad (8.2.5)$$

Таблица 8.2.2

Расчетный объем подлежащих отработке пород V и вероятные запасы в телах q в зависимости от их горизонтальных размеров

Горизонтальная площадь тела, м ²	V	q	Горизонтальная площадь тела, м ²	V	q
28	111.0	6.0	219	373.0	27.5
46	235.0	8.9	224	506.0	27.9
56	253.5	10.3	297	435.0	30.8
61	251.5	11.0	293	437.0	34.0
65	323.5	11.4	295	393.5	34.2
66	255.0	11.6	301	500.5	34.7
68	258.0	11.8	301	412.0	34.7
84	305.0	13.8	301	412.0	34.7
88	309.5	14.3	354	538.5	38.5
113	352.5	17.0	436	485.0	43.6
177	345.0	23.4	469	485.0	45.4
180	375.0	23.7	493	575.5	46.6
184	506.0	24.1	532	646.5	48.3

Значения параметров, определяющих затраты на основные виды работ, и численные значения коэффициентов, полученные по справочнику укрупненных норм, приведены в табл. 8.2.3.

Таблица 8.2.3

Значения параметров и коэффициентов, принятые при определении затрат на основные виды работ

Параметры	Численные значения
Глубина поисковых скважин (l_1), м	50
Средняя глубина разведочных скважин (l_2), м	40
Средняя глубина шахты (h), м	30
Число разведочных скважин на одно тело (m)*, шт.	10
Затраты на проходку одного метра:	
а) скважины (a_1)* доли цены усл. ед. запасов	$2.2 \cdot 10^{-3}$
б) шахты и вентиляционной скважины (a_2), то же	3.1
в) околоствольного двора, подходного квершлага и восстающего (a_3), то же	7.9
Затраты на выемку 1 м ³ пород (a_4), то же	$2.8 \cdot 10^{-3}$

Результаты вычислений основных экономических показателей работ при различных поисковых сетях приведены в табл. 8.2.4 и графически представлены на рис. 8.2.1. Кривая a отражает характер увеличения выявляемых запасов с увеличением плотности сети. На этой кривой можно выделить участок, в пределах которого удвоение плотности сети ведет к удвоению выявляемых запасов. Этот участок простирается включительно до сети 50×40 ц (1 скважина на 1000 м²), которая является оптимальной, если в качестве критерия выбраны

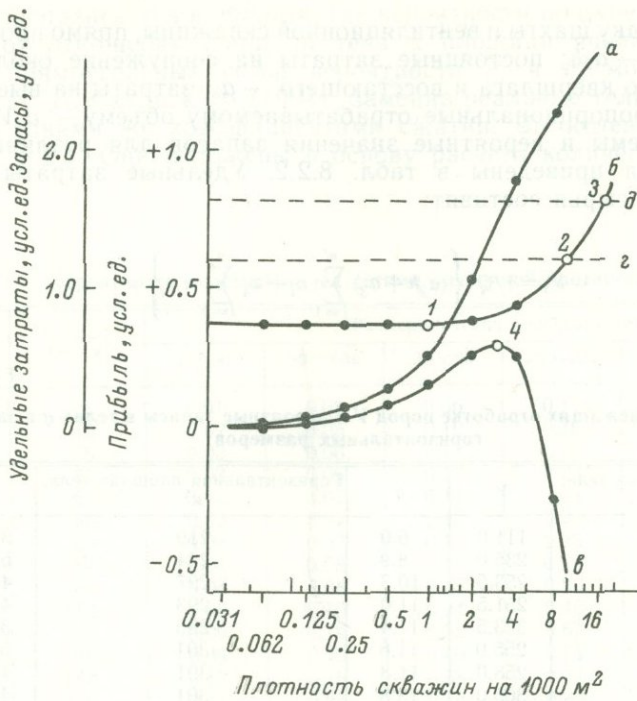


Рис. 8.2.1. Оптимизация плотности разведочной сети при разведке на «прирост запасов».

а — кривая выявляемых запасов, *б* — кривая удельных затрат на единицу выявляемых запасов, *в* — прибыль от разработки, *г* — среднеотраслевые удельные затраты, *д* — удельные затраты на замыкающих месторождениях. Оптимальная плотность сети: 1 — в условиях неограниченного резерва перспективных площадей, 2 — в условиях поддержания среднеотраслевых показателей, 3 — в условиях «сырьевого голода», 4 — по критерию максимальной прибыли.

Таблица 8.2.4

Основные расчетные показатели работ при различной конечной плотности поисковых сетей

Ячейка сети	Удельные затраты (в единицах оптовой цены сырья)				Количество выявляемых запасов (в усл. ед.)	Доля выявляемых запасов	Величина прибыли (+) или убытков (—) (в усл. ед.)
	на выявление тел	на детальную разведку	на добычу	общие			
50×40	0.239	0.042	0.461	0.742	288.9	9.6	75
50×40ц	0.243	0.042	0.464	0.749	566.9	18.8	142
25×20	0.254	0.043	0.469	0.766	1084.6	35.9	254
25×20ц*	0.314	0.045	0.493	0.852	1754.6	58.2	260 *
12.5×10	0.493	0.052	0.570	1.115	2235.1	74.1	— 257
12.5×10ц	0.820	0.061	0.652	1.533	2690.6	89.2	— 1433

* Оптимальная ячейка сети по критерию максимальной прибыли.

минимальные удельные затраты на прирост запасов (точка 1). Кривая *б* отражает изменение удельных затрат на прирост запасов и последующую добычу полезного ископаемого. При сгущении сети они сначала остаются постоянными, затем увеличиваются сперва медленно, затем быстрее. Пересечение кривой *б* с уровнем максимально допустимых удельных затрат на выявление и добычу (кривая *д*) определяет оптимальную плотность сети при условии применения данного критерия. Такой сетью является 12.5×10 ц (16 скважин на 1000 м², точка 3). Кривая *в*, отражающая изменение расчетной прибыли при освоении

месторождения, определяет как оптимальную более редкую сеть — 25×20 ц (точка 4). При этой плотности сети сумма ожидаемой расчетной прибыли достигает максимального значения.

Поскольку величина прибыли зависит от уровня оптовой цены, то при ее возрастании оптимальной будет более плотная сеть, при снижении — более редкая. Так как применение более дешевых и эффективных методов поисков, разведки и разработки позволяет получить большую прибыль, то максимальная прибыль будет достигаться при более плотных сетях, обеспечивающих повышение полноты извлечения заключенных в недрах запасов.

Проведенный анализ приводит к выводу о возможности применения различных критериев при определении оптимальной сети в зависимости от конкретных геолого-экономических условий. При неограниченном резерве перспективных площадей оптимальной является сеть, при которой удельные затраты на поиски все еще сохраняются на минимально возможном для данного месторождения низком уровне, т. е. обеспечивают максимальную эффективность геологоразведочных работ. В обычных условиях рекомендуется критерий максимума прибыли от освоения месторождения. При дефиците сырья допускаются приращивание запасов и добыча при удельных затратах выше уровня, при котором суммарная прибыль достигает максимума. Конкретная реализация критериев оптимизации разведки на прирост запасов других типов месторождений может существенно отличаться от описанного примера.

§ 3. Критерии оптимизации предварительной разведки (на уточнение)

Предварительная разведка должна выяснить, является ли месторождение промышленным, т. е. стоит ли на нем проводить детальную разведку или забраковать. Чтобы ответить на этот вопрос, запасы и другие горнотехнические параметры достаточно оценить в среднем по месторождению, без их пространственной детализации (поэтому предварительную разведку правильнее было бы называть общей в противоположность детальной).

Принципиальная возможность существования оптимальной степени разведанности месторождений подтверждается тем, что прирост положительного эффекта от разведки по мере ее уточнения убывает, а прирост затрат возрастает. Общий критерий оптимизации заключается в том, что разведка должна продолжаться до тех пор, пока стоимость следующего ее шага меньше положительного эффекта от него. Применительно к стадии предварительной разведки в реализации этого принципа необходимо рассмотреть два случая.

Первый случай. Данные разведки на момент анализа ее результатов указывают, что месторождение — непромышленное. Эта оценка, как и всякая оценка по выборке, содержит погрешность, и вероятность, что фактически месторождение является рентабельным, не равна нулю. Обозначим через $M(C_n)$ условное математическое ожидание прибыли от разработки такого месторождения, рассчитанное по данным n -го шага разведки. $M(C_n)$ по определению является математическим ожиданием усеченного слева (в точке θ_{\min}) распределения оценки прибыли от разработки месторождения. На рис. 8.3.1 $M(C_n)$ пропорционально заштрихованной площади под кривой $f_n(\theta)$, где θ — оценка параметра месторождения, определяющего прибыль. Тогда разведку следует прекратить в тот момент, когда будет выполнено неравенство

$$M(C_{n-1}) - M(C_n) = M(C_{n-1} - C_n) \leq R_n - R_{n-1}, \quad (8.3.1)$$

где R_n — затраты на разведку от ее начала до n -го (планируемого) шага включительно, т. е. разведку наиболее рационально прервать в тот момент,

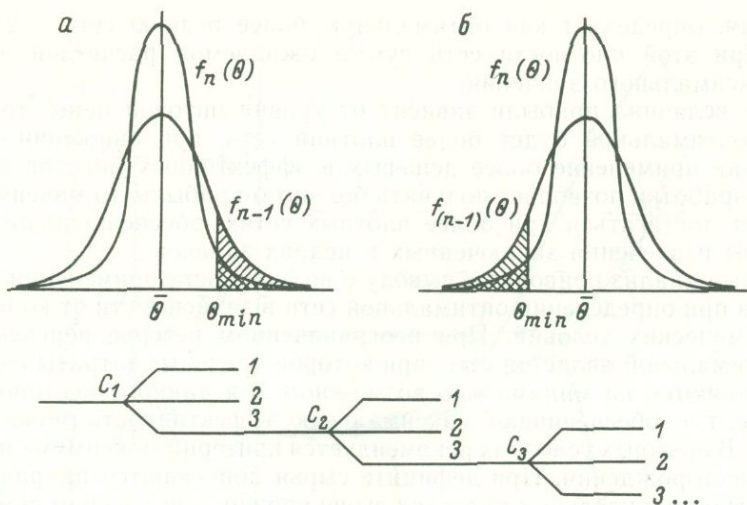


Рис. 8.3.1. К иллюстрации критериев оптимизации предварительной разведки.

Заштрихованы площади, соответствующие вероятности того, что: *a* — будет забраковано рентабельное месторождение, *б* — будет передано в детальную разведку нерентабельное месторождение; θ — рассматриваемый параметр; $f_n(\theta)$ — функция распределения оценки параметра; n — номера шагов разведки.

когда стоимость следующего шага разведки становится равной или больше разности вероятной упущенной прибыли из-за отбраковки рентабельного месторождения (рис. 8.3.1, а). Разведка прерывается с отрицательным результатом. Мотив прерывания на языке геологоразведчиков можно приблизительно сформулировать следующим образом: месторождение настолько нерентабельно, что шансы того, что по результатам следующего шага разведки оно окажется рентабельным, очень малы и не оправдывают средств на следующий шаг разведки.

Если неравенство (8.3.1) еще не выполняется, требуется проверить условие

$$M(C_{n-1}) < R_n - R_{n-1}. \quad (8.3.2)$$

Если это неравенство выполняется, разведку также следует прекратить, так как следующий шаг разведки требует больших затрат, чем ожидаемая прибыль от разработки месторождения. Практически этот вариант может реализоваться при разведке месторождения весьма незначительного масштаба, но сложного строения.

Если ни неравенство (8.3.1), ни (8.3.2) еще не выполняется, разведку необходимо продолжить. Цель ее заключается в испытании шансов того, что полученные новые данные будут более благоприятными и позволят дать общую положительную оценку месторождения. Если этого не произойдет, месторождение будет отбраковано, и вероятность, того, что оно в действительности рентабельно, будет сведена до разумного минимума. Общие потери (фактические и вероятные) составят

$$R_n + M(C_n). \quad (8.3.3)$$

Могут возразить, что подходы, основанные на вероятности, оптимальны в среднем, а к отдельному конкретному месторождению неприменимы. Но ведь ежегодно предварительной разведке подвергается огромное множество месторождений, и в масштабе всей страны или отрасли статистический подход оправдан.

Второй случай. Данные разведки на момент анализа результатов ее $(n-1)$ -го шага свидетельствуют в пользу того, что месторождение рентабельно. Обозна-

чив через P_n вероятность того, что месторождение может оказаться нерентабельным, а проектную стоимость последующей детальной разведки — через R_n , критерий оптимальности можно записать следующим образом: разведку наиболее рационально прекратить в тот момент, когда

$$R_n(P_{n-1} - P_n) \leq R_n - R_{n-1}, \quad (8.3.4)$$

т. е. когда затраты на следующий шаг разведки становятся равными или больше, чем уменьшение вероятных потерь от проведения (ненужного) детальной разведки (рис. 8.3.1, б). В такой формулировке критерия имеется определенный дефект, поскольку возможный ошибочный вывод о рентабельности месторождения может быть исправлен раньше, чем будет проведена детальная разведка в полном ее объеме. Однако предсказать заранее, когда это произойдет, невозможно, поэтому мы и предлагаем эту формулировку критерия.

На рис. 8.3.2 условно изображен процесс последовательного приближения к окончанию разведки с оптимальной плотностью сети. Рассмотрим следующий пример, имеющий чисто иллюстративное значение. Допустим, что: а) разведка месторождения производится отбором валовых проб из шурфов с поверхности; б) степень разведанности определяется общим их числом m ; в) распределение оценки среднего содержания, являющегося лимитирующим фактором оценки, близко к нормальному закону¹ с дисперсией $s_m^2 = s^2/m$, где s^2 — дисперсия содержания в месторождении ($s = 0.70$), коэффициент вариации содержания равен $V = s/C \cdot 100\% = 0.7/2.5 \cdot 100\% = 28\%$. Тогда рост затрат на разведку изобразится прямой 1 (рис. 8.3.3). Уменьшение вероятных убытков от неправильной отбраковки месторождения (кривая 2) пропорциональна величине $1 - F(t)$; $F(t)$ — функция вероятности нормального распределения, где

$$t = \frac{(c_{\min} - \bar{c}_n)}{s_m} = \frac{(3.25 - 2.5)}{0.7/\sqrt{m}} \%$$

Из рисунка видно, что при малом числе проб (1—15) очень велики вероятные убытки от отбраковки потенциального рентабельного месторождения. Они сильно уменьшаются при каждом шаге разведки (5 проб). При числе проб 15 и более темп уменьшения их замедляется и в сумме с затратами на разведку (кривая 3) они достигают минимума при $m = 20 \div 25$. При этом значении m начинает выполняться неравенство (8.3.1).

Наиболее значительный вклад в решение рассматриваемого вопроса внес Ж. Матерон (1968). В процессе разведки, как указывает он, наступает момент, когда дополнительно полученная информация перестает окупать затрачиваемые на нее средства. Между двумя крайними положениями — нулевые затраты на разведку, максимальный риск экономических потерь при отбраковке, полное отсутствие риска и бесконечные затраты на разведку — обязательно существует третья, соответствующее оптимальному соотношению между затратами и риском потерь. В качестве критерия при выборе решения принимается условие максимальной математического ожидания будущей прибыли. Последняя рассчитывается на основании предсказания результатов доразведки месторождения в зависимости от соотношения между стоимостью дополнительных разведочных работ и соответствующим уменьшением риска потерь. Таким образом, критерий Матерона является близким прототипом критерия (8.3.1).

Критерий (8.3.2) близок к идее экономического обоснования допустимых

¹ В работе Ю. А. Ткачева и Я. Э. Юдовича (1975, с. 86—91) и в других работах показано, что распределение оценки среднего значения уже при $n = 20$ очень близко к нормальному.

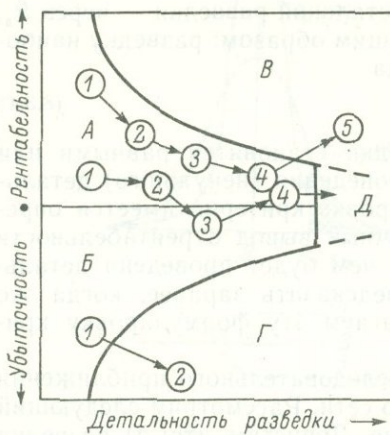


Рис. 8.3.2. Условная схема последовательного хода анализа необходимой степени разведанности на стадии предварительной разведки.

А — область продолжения разведки по критерию (8.3.4); Б — то же, по критерию (8.3.1); Д — область окончания разведки по критерию (8.3.2) с неопределенным эффектом; В — область окончания разведки с положительным эффектом; Г — область окончания разведки с отрицательным эффектом. Цепочка кружков — примеры разведки месторождений с различными исходами. Номера в кружках — шаги разведки.

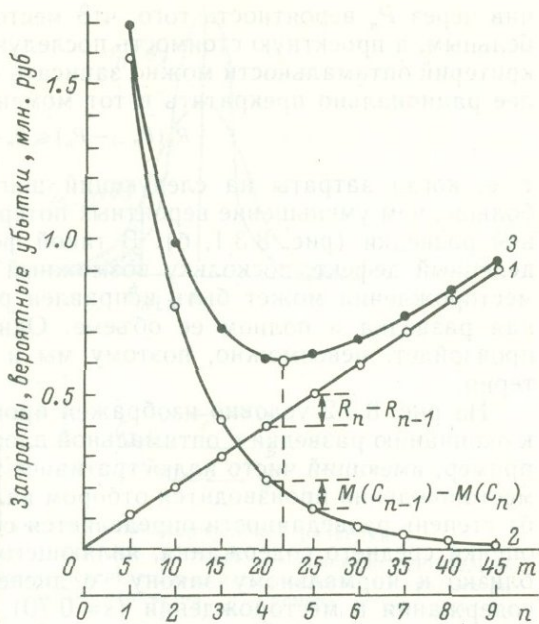


Рис. 8.3.3. Пример оптимизации предварительной разведки месторождений.

1 — затраты на разведку $R=10m$ (в тыс. руб.); 2 — кривая вероятных убытков вследствие ошибочной отборки рентабельного месторождения; 3 — кривая суммы затрат на разведку и вероятных убытков $M(C)=10(1-F(t))$ (в млн. руб.). Стрелка указывает оптимальное число проб (≈ 22).

затрат на разведку С. Р. Кондина и П. И. Штифанова (1965 г.). Из формулы уровня рентабельности

$$УР = \Pi \cdot 100 / ВС \%, \quad (8.3.5)$$

где Π — годовая прибыль от предприятия, определяемая допроектной экономической оценкой месторождения; $ВС$ — вложенные средства в месторождение, авторы предложили определять предельное значение $ВС$, исходя из нормального уровня рентабельности — НУР:

$$ВС = \Pi \cdot 100 / \text{НУР}. \quad (8.3.6)$$

Зная $ВС$, допустимые затраты на разведку P можно определить из формулы $P = BC - K - O$, где K — капиталовложения в строительство предприятия, O — необходимые ему оборотные средства.

Основной недостаток предлагаемого авторами метода состоит в следующем. При решении дилеммы — продолжать или окончить разведку — должны приниматься во внимание только предстоящие затраты.

Рассмотрим критерий Г. Бандемера (Bandemer, 1966, 1967a, 1967b). Автор разъясняет его следующим образом. Пусть μ_u — истинное значение некоторого параметра месторождения, а $\hat{\mu}_n$ — его оценка. Пусть экономисты в состоянии задать функцию потерь $R(\hat{\mu}_n - \mu_u)$. Вероятность отклонения ($\hat{\mu}_n - \mu_u$) зависит от числа проб n , поэтому можно говорить о математическом ожидании убытков $M(R)$, которые сопоставляются с ценой взятия n проб $K(n)$. Критерием оптимальности служит минимизация сумм

$$K(n) + M(R) = \min. \quad (8.3.7)$$

Стратегию разведки Г. Бандемер описывает как процесс, на каждом этапе которого производятся контроль по этой формуле и остановка при достижении минимума. В последующих разделах будет показано, что этот критерий не отличается от критерия С. П. Васильева для детальной разведки. Г. Бандемером не учитывается принципиальное изменение целей отдельных ее стадий, влекущее за собой смену критериев оптимальности.²

В заключение анализа критериев оптимизации предварительной разведки представим себе принятие решения о разработке или отбраковке месторождения как некоторую лотерею с математическим ожиданием выигрыша (положительным или отрицательным), равным прибыли от разработки месторождения. Отличительной особенностью такой лотереи является неравная нулю дисперсия выигрыша и возможность за определенную плату (затраты на разведку) уточнить оценку математического ожидания выигрыша (уменьшать ее дисперсию). Задачу оптимизации разведки можно сформулировать в этих терминах следующим образом: найти оптимальную степень уточнения условий лотереи, если известны затраты как функция этого уточнения.

Прежде рассмотрим следующий идеальный случай: время розыгрыша лотереи исчезающе мало, а играющий располагает неограниченными средствами повторять игру. Тогда все лотереи с одинаковыми математическими ожиданиями выигрышей одинаково предпочтительны независимо от дисперсии выигрыша. Действительно, увеличивая число игр любой из этих лотерей, можно обеспечить выигрыш, как угодно близкий к математическому ожиданию. Предпочтительность лотереи с меньшей дисперсией выигрыша обусловлена существенной продолжительностью одной игры и ограниченностью имеющихся средств. Именно отсутствие свободных средств заставляет предпочитать лотерею с более гарантированным выигрышем. В геологическом приложении продолжительность разведки велика, а ограничение касается как средств на разведку, так и фонда разведываемых месторождений.

Пусть t — длительность игры (т. е. суммарная длительность разведки и эксплуатации), а k — капиталовложения, которые необходимы для разведки. Чтобы увеличить гарантию выигрыша, можно играть n раз подряд и иметь дисперсию среднего выигрыша, равную σ_2^2/n . Для этого нужно произвести дополнительные затраты (по сравнению с одной игрой), равные в среднем

$$\Delta k = -k \sum_{i=1}^n (1+\alpha)^{it} - k(1+\alpha)^t = k(1+\alpha)^t \left\{ \sum_{i=2}^n (1+\alpha)^{it} - 1 \right\}, \quad (8.3.8)$$

где α — проценты за кредит. Эти затраты уменьшат выигрыш в расчете на одну игру на величину $\Delta k/n$, но зато дисперсия выигрыша уменьшится в n раз. Две лотереи с математическими ожиданиями выигрыша M_1 и $M_2 = M_1 - \Delta k/n$ и его дисперсиями σ_1^2 и $\sigma_2^2 = \sigma_1^2/n$ по условию одинаково предпочтительны. Двойное отношение $(M_1/M_2) : (\sigma_1^2/\sigma_2^2)$ одинаково предпочтительных лотерей является функцией двух аргументов $f(\sigma_1^2/\sigma_2^2, \alpha)$:

$$\frac{M_1}{M_2} : \frac{\sigma_1^2}{\sigma_2^2} = \frac{k(1+\alpha)^t}{k \sum_{i=1}^n (1+\alpha)^{it}} : \frac{\sigma_1^2}{n\sigma_1^2} = \frac{n}{\sum_{i=1}^n (1+\alpha)^i}. \quad (8.3.9)$$

Округляя до целого (n) отношение дисперсий, можно сравнивать предпочтительности лотерей: если двойное отношение для сравниваемых лотерей будет больше, чем вычисленное по (8.3.9), лотерея № 1 предпочтительнее; при равенстве обе лотереи одинаково предпочтительны. Если $\alpha=0$, т. е. неограни-

² Сходный критерий мы использовали практически при разведке месторождений с очень неравномерным характером распределения полезного ископаемого (Ткачев и др., 1970а, 1971).

ченный кредит предоставляется бесплатно, то независимо от отношения дисперсий выигрышей предпочтительнее лотерея с большим математическим ожиданием выигрыша. Действительно, при $\alpha=0$ $n/\sum_{i=1}^n (1+\alpha)^i = n/n = 1$.

В терминах изложенного примера предварительную разведку месторождения необходимо продолжать до тех пор, пока это будет соответствовать переходу к более предпочтительной лотерее. Как только продолжение разведки перестает улучшать предпочтительность лотереи, разведка должна быть закончена.

Для развития этой идеи в геологической разведке надо учесть, что фонд структур или рудопроявлений, которые могут быть подвергнуты разведке, ограничен, и при их отбраковке необходимо принимать во внимание стоимость поисков новых.

§ 4. Критерии и методы оптимизации плотности сети детальной разведки

В настоящее время окончание детальной разведки регламентируется инструкцией ГКЗ СССР, в которой указано, что в результате разведочных работ должно быть обеспечено получение необходимых исходных данных для количественной и качественной оценок запасов и обоснования проектирования разработки месторождения. Этот метод определения плотности сети детальной разведки можно считать простейшим инструктивным методом оптимизации. Он осуществляется в три стадии. На первой разведку относят к одной из установленных групп по изменчивости основных параметров на основании полученных геологоразведочных данных. На второй сравнивают фактические расстояния между разведочными выработками (точками наблюдения, пробами) и полноту их исследования с предусмотренными инструкцией для данной группы месторождений и устанавливают категорию запасов данного участка месторождения (А, В, С₁ или С₂). На третьей стадии сравнивают долю запасов каждой категории (А, А+В, А+В+С₁, А+В+С₁+С₂) с требуемой по инструкции, и если она не меньше, выносят суждение о достаточной степени изученности. В противном случае разведку месторождения продолжают. Основными недостатками этого способа являются невозможность свести все многообразие месторождений к группам по изменчивости и трудность надежного отнесения конкретного месторождения к определенной группе. Установленные инструкцией расстояния между выработками и предельные соотношения категорий запасов выведены из средних эмпирических данных и не могут варьировать в зависимости от индивидуальных особенностей разведываемых месторождений. В результате часть месторождений оказывается недоразведанной или переразведанной. И то и другое приводит к существенному экономическому ущербу. Достаточно напомнить, что по инструкции 1972 г. для месторождений нефти простого геологического строения (1-я группа) при передаче их в промышленное освоение по категории В должно быть разведано не менее 30 % запасов в новых районах и не менее 20 % — в районах с развитой нефтегазодобывающей промышленностью. Запасы месторождений сложного геологического строения достаточно разведать до категории С₁. В инструкции 1983 г. требуемые соотношения категорий уже не ставятся в зависимость от сложности и размера месторождения. По категории С₁ необходимо разведать не менее 80 % запасов (остальное — С₂). Таким образом, достижение цели разведки здесь устанавливается не непосредственно, а через достижение косвенной цели, а именно — разведать не менее чем заданную часть запасов до определенной категории, а остальную — по более низким категориям. Как бы ни была удачна классификация запасов и инструкция к ней,

все же появляется вероятность достижения косвенной цели при фактическом невыполнении прямой и — наоборот. Например, на более сложных месторождениях, требующих для проектирования, казалось бы, большего объема данных, требования к разведанности запасов значительно снижаются. Это еще раз подчеркивает противоречие между прямой и косвенной целью. В этих условиях ГКЗ, выполняя роль буфера между «поставщиками» и «потребителями» запасов, часто соглашается с предлагаемой геологами оценкой параметров и блокировкой запасов, если нет возражений со стороны представителей Миннефтепрома.

Неоднократно многие авторы предлагали использовать в качестве критерия окончания детальной разведки достижение требуемой точности. Однако это тоже подменяет прямую цель разведки косвенной, как и в предыдущем случае, хотя отдельные авторы (Оспанов и др., 1975, с. 69) считают, что «конечная цель разведки — достаточно точный подсчет запасов». Требуемая точность, как было показано в гл. 7, не может служить даже средством квалификации запасов, тем более как непосредственный критерий окончания разведки. Хотя расчет точности оценки параметров на месторождениях, вообще говоря, необходим.

Значительная доля специалистов предлагает считать основным критерием окончания разведки стабилизацию оценок главных параметров месторождений. В терминах теории информации это эквивалентно стабилизации информации или информационной энтропии (Дементьев и др., 1974). Действительно, при достижении такой стабилизации дальнейшее уточнение параметра ведет к стремительному увеличению затрат на разведку, однако совершенно не ясно, будет ли «стабилизированное» значение параметра удовлетворять по надежности основной цели разведки. Точку стабилизации некоторые авторы (Füst, 1980) предлагают определять по формуле $n_{\text{опт}} = k \cdot \sqrt{b}$, где b — параметр эмпирического уравнения зависимости погрешности оценки σ_x от числа скважин: $\sigma_x^2 = a + b/n$.

Рассмотренные предложения, строго говоря, не являются критериями оптимизации, поскольку определяемая ими процедура окончания разведки хотя и полезна, но не минимизирует какой-либо целевой функции.

Первой работой, в которой проблема оптимизации детальности разведки рассматривается с современных позиций, можно считать статью С. П. Васильева (1950). В ней автор говорит, что, разредив сеть скважин, можно добиться «удешевления» разведки, но это может повлечь за собой увеличение стоимости последующих горно-подготовительных и эксплуатационных работ. Для оптимального решения он предлагает два метода: метод так называемого шахтного анализа (на примере отработанных буроугольных шахтных полей определяются «лишние» разведочные выработки и их возможное влияние на последующую эксплуатацию) и метод экономической эквивалентности, который заключается в составлении проектов эксплуатации по текущим данным детальной разведки. По мере сгущения скважин проекты будут изменяться, все более и более приближаясь к оптимальному. На каждом шаге разведки сравнивается стоимость следующего шага с уменьшением лишних затрат при разработке (рис. 8.4.1). Эта методика была опробована на Широно-Сокольническом и Арцыбашевском месторождениях (было составлено по два проекта). Хотя в статье и нет строгой формулировки критерия оптимизации, в ней достаточно определенно намечена основная идея применительно к стадии детальной разведки и предложен хотя и громоздкий, но осуществимый эмпирический метод решения. Статья намного опередила свое время, и в 50-х годах не вызвала живого отклика среди геологов, а теперь почти забыта. Возрождение интереса к оптимизации производства на базе применения ЭВМ привело к идеям, очень близким идеям С. П. Васильева. Действительно, если «усреднить» мнения большого числа авторов, то полученный таким способом общий критерий

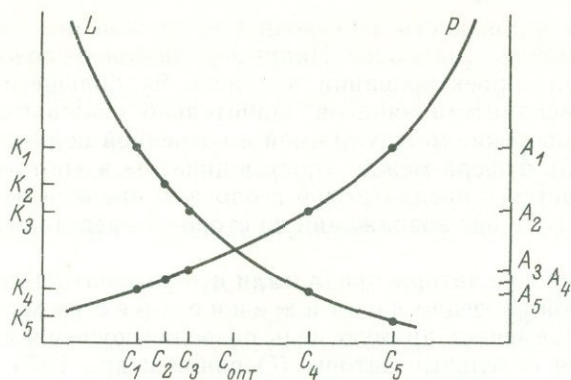


Рис. 8.4.1. Определение оптимальной разведанности на стадии детальной разведки по С. П. Васильеву.

L — кривая убытков от неоптимальности проекта; P — кривая затрат на разведку; $K_1, K_2 \dots$ — лишние затраты вследствие неоптимальности проекта эксплуатации; $A_1, A_2 \dots$ — затраты на разведку; $C_1, C_2 \dots$ — шаги детализации разведки.

заключается в минимизации суммы: а) затрат на разведку и б) вероятных потерь при разработке, а общая схема процесса оптимизации может быть изображена, как на рис. 8.4.2. Однако в такой общей формулировке критерий не конструктивен и допускает во второй части различное понимание.

Крупнейший французский специалист Ж. Матерон (1968) пишет, что ошибка в оценке запасов и содержания ведет к потере принятого проекта отработки своей оптимальности. В результате появляются экономические потери, вероятное значение которых может быть вычислено и которые должны быть включены в баланс вместе с затратами на дополнительные разведочные работы.

Группа сотрудников ВНИИнефть — Н. Е. Быков, Л. Д. Америка, А. В. Черницкий, Н. М. Николаевский, Н. А. Крылов, Е. Ф. Фролов, Р. А. Егоров, А. Я. Фурсов и др. (Быков и др., 1978; Оптимизация... , 1976) — предлагает сразу три варианта толкования «экономических потерь». В первом варианте — среднеотраслевые, среднерайонные или замыкающие затраты на подготовку (разведку) такого количества запасов, которое покроеет вероятный недобор от случайной погрешности их оценки на стадии разведки. Авторы исходят из посылки, что ошибка в запасах непременно приведет к недобору, тогда как равновероятен и «перебор», который авторы не принимают во внимание, дабы не разрушить предложенный критерий. Даже если принять, что в процессе оптимизации плотности детальной разведки необходимо учитывать только недобор, то все равно нет оснований ни планирующим органам, ни проектирующим организациям смотреть на разведанные запасы, как на детерминированную величину, которую в случае неподтверждения необходимо любой ценой компенсировать. Само планирование должно быть гибким с учетом статистических флуктуаций оценок планируемых величин. В частности, зная погрешность разведанных запасов по данному региону при планировании можно учитывать запасы, гарантированные с заданной степенью вероятности.

Второй вариант критерия тех же авторов заключается в толковании «экономических потерь» как среднеотраслевых, районных или замыкающих затрат на добычу количества нефти, «компенсирующего» ошибку запасов. Практически способ оптимизации, основанный на этом критерии, заключается в следующем (Оптимизация... , 1976, с. 107—113): 1) вычерчивают графическую модель месторождения в виде карт изолиний параметров; 2) строят по ней кривую зависимости средней квадратической погрешности оценки запасов от плотности разведочной сети методом последовательных разрежений послед-

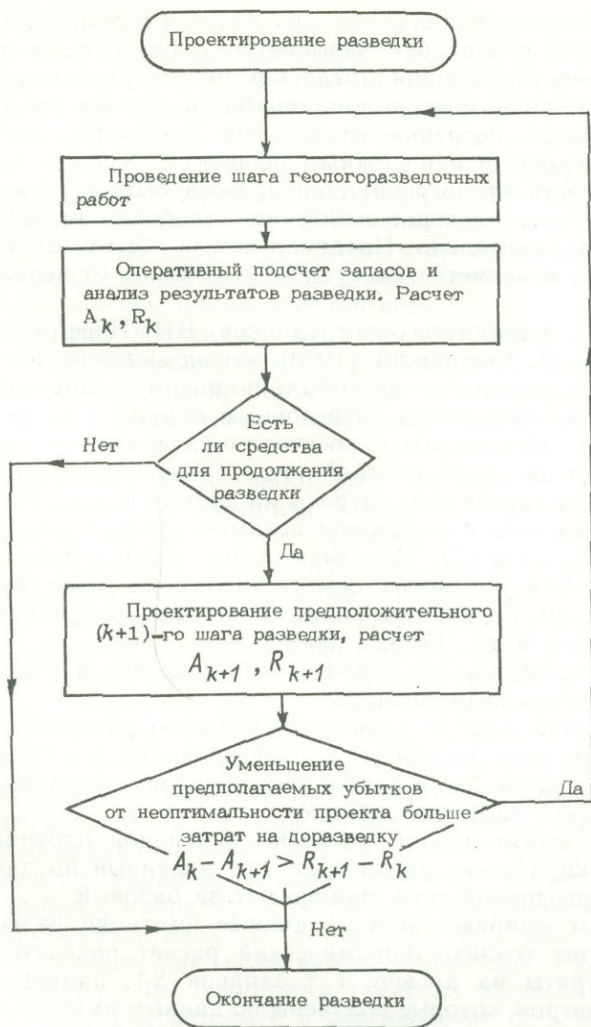


Рис. 8.4.2. Общая схема операции при оптимизации детальной разведки «на уточнение».

k — номера шагов разведки; A_k — вероятные убытки от неоптимальности проекта разработки месторождения, составленного по данным k шагов разведки; R_k — затраты на k шагов детальной разведки.

ней; 3) экстраполируют эту кривую в область еще не достигнутой более плотной разведочной сети и определяют предполагаемую среднюю квадратическую погрешность для нескольких значений плотности сети; 4) вычисляют нижнюю доверительную величину запасов $Q - \sigma_Q$, где Q — оценка запасов, σ_Q — их средняя квадратическая погрешность, и затраты на разведку 1 т этих запасов R ; 5) по эмпирической кривой, построенной на материалах анализа разработки ряда аналогичных месторождений и связывающей увеличение себестоимости добычи 1 т запасов от средней квадратической погрешности оценки запасов σ_Q , устанавливают себестоимость добычи 1 т запасов для нескольких предполагаемых значений σ_Q , соответствующих нескольким предполагаемым значениям плотности сети (по п. 3); 6) предполагаемую плотность сети, при которой сумма затрат на разведку и себестоимость добычи 1 т запасов минимальна, принимают за оптимальную.

Главный недостаток этого способа заключается в том, что изменение себестоимости добычи 1 т запасов связывается лишь с погрешностью оценки о б щ и х запасов месторождения, тогда как речь идет о детальной разведке, причем во внимание принимается лишь ошибка в сторону увеличения запасов. Экстраполяция кривой погрешности запасов в область значительно более плотных сетей приводит к ненадежным решениям. Кривая зависимости увеличения себестоимости от погрешностей должна быть получена как средняя для ряда аналогичных месторождений, но подобрать их для разведываемого месторождения затруднительно. Предполагаемый убыток от неоптимальности проекта разработки заменяется, по существу, убытком от неподтверждаемости запасов.

Наконец, часть упомянутых сотрудников ВНИИнефть — Н. Е. Быков, Л. Д. Америка, А. В. Черницкий (1978) «экономические потери» понимают как разность между затратами на добычу по оптимальному (безошибочному) проекту и реальному проекту по разведочным данным с погрешностями. Для этого составляют 3—10 вариантов проектов по различной степени разреженной сети, в том числе и по самой плотной (этот проект считается оптимальным). По каждому проекту определяют затраты на добычу и разность их с затратами по оптимальному, которая и составляет искомые потери (заметим: все — точно по С. П. Васильеву, 1950 г.!). Методика оптимизации в соответствии с таким толкованием критерия минимума суммы затрат на разведку и возможных экономических потерь от неоптимальности проекта разработки заключается в следующем (Быков и др., 1978, с. 86—88).

1. Составляют графическую модель месторождения в виде карт изолиний параметров и контуров нефтеносности.

2. На построенной модели проводят имитацию разведки для нескольких значений плотности разведочной сети, превышающих фактическую плотность на момент анализа разведки, а также расчет затрат на разведку 1 т запасов по каждому варианту плотности сети R_i .

3. По данным имитационной разведки различной плотности составляют проекты разработки. Проект разработки, составленный по данным имитации самой плотной разведочной сети, принимают за базовый.

4. По проектам, опирающимся на данные имитации разведки различной плотности, проводят технико-экономический расчет показателей разработки (приведенные затраты на добычу 1 т запасов S_i), принимая за истинные те значения параметров, которые получены по данным имитации самой плотной сети.

5. Определяют убытки от неоптимальности проекта разработки U_i как разность между приведенными затратами по i -му и по базовому проекту $U_i = S_i - S_0$.

6. Выбирают плотность сети, соответствующую минимальному значению суммы затрат на разведку и убытков от неоптимальности проекта $R_i + U_i$, и принимают ее за оптимальную.

В результате анализа и сопоставления всех предлагавшихся критериев мы пришли к выводу, что последний из рассмотренных критериев является как раз тем, что требуется для данного случая. Однако со способом его реализации согласиться трудно. Модель, построенная по данным редкой (фактически выполненной) разведочной сети, не включает и не может (по описанному способу) включать вероятные изменения параметров на расстояниях, меньших шага этой сети. Вследствие этого она является слишком гладкой и простой. Действительно, карты контуров и изолиний параметров при редких разведочных сетях всегда более просты, чем они должны быть на самом деле. Имитация разведки по таким картам плотными сетями дает результаты, мало отличающиеся от результатов по редким сетям. Тем самым создается искаженное впечатление хорошей разведанности редкими сетями. Искажение особенно

велико для очень плотных сетей, т. е. для базового варианта имитации разведки и проекта разработки. Во многих случаях плотность, соответствующая базовому варианту, фактически не дает надежных результатов разведки, а по приведенному способу вычисляемая оптимальная плотность разведочной сети всегда будет ниже базовой. Предложенный способ, таким образом, может применяться лишь ретроспективно и дает правильный результат только на заведомо переразведанных месторождениях. Недостатком способа является также практическая ограниченность числа анализируемых проектов для каждой плотности сети. Она не позволяет выявить среднюю тенденцию улучшения проектов по мере сгущения сети. По закону случая редкие сети могут дать более точную оценку, чем самая плотная сеть.

Одна из попыток обойти это затруднение заключается в нахождении оптимального числа скважин для многих разведочных месторождений и установлении эмпирической кривой «запасы—оптимальное число скважин» (Фурсов, 1979) или «площадь залежи (S)—оптимальное число скважин $n_{\text{опт}}$ » ($n_{\text{опт}}=0.32\sqrt{S}$).

Л. Д. Кноринг (1973, 1977, 1980, и др.) предлагает критерий оптимизации, основанный на той же идее минимизации суммы затрат на разведку и ущерба от неопределенности сведений о месторождении. Этот ущерб автором принимается прямо пропорциональным неопределенности оценки результатов разведки (т. е. погрешности оценки), а она в свою очередь в поле всей залежи характеризуется одним числом — некоторым функционалом, зависящим от дисперсионной матрицы оценок коэффициентов модели. По существу, автор применяет теорию оптимального планирования эксперимента. В гл. 5 на примере выбора эталонных образцов для построения градуировочного графика были проиллюстрированы лежащие в ее основе предположения. Если бы можно было знать наперед и выразить в достаточно простой аналитической (математической) форме процесс, приводящий к формированию месторождения, применение теории оптимального планирования эксперимента при разведке было бы эффективно. В реальных условиях априори нет достаточно веских оснований для выбора класса функций, аппроксимирующих геологический параметр. А если класс выбран, например алгебраические полиномы, то выбор степени полинома также субъективен. В методе Л. Д. Кноринга от степени полинома сильно зависит как величина функционала, так и «оптимальное» размещение скважин. При некоторых степенях полинома оно может быть весьма экзотичным и находящимся в противоречии со всем предшествующим опытом разведки (например, почти все скважины — на границах залежи нефти!). Заметим, что защищенный авторским свидетельством способ размещения (Габриэлянц, Пороскун, 1974) почти противоположен: равные запасы на скважину.

Рассмотренные и некоторые другие критерии и методы оптимизации плотности детальной разведки сведены в табл. 8.4.1. Ниже мы предлагаем метод оптимизации плотности детальной разведки, устраняющий недостатки оптимизации по Н. Е. Быкову и др.

Во-первых, вся работа по имитации разведки и проектированию разработки проводится на моделях, включающих вероятную изменчивость параметров месторождения. Если в известном методе моделью служит карта параметров, построенная по разведочным данным и отражающая только выявленные разведкой «закономерные» изменения, то в предлагаемом методе эта модель усложняется включением в нее случайных колебаний параметров, уровень и характер которых определяются статистическими характеристиками, такими как коэффициент вариации, автокорреляционная функция или полувариограмма.

Во-вторых, в отличие от известных методов, в которых для анализа может быть использована и используется только одна модель, в предлагаемом методе

Основные критерии и методы оптимизации плотности разведочной сети на стадии детальной разведки

Источники	Содержание метода
Дементьев и др., 1966; Ишаев, 1970; Гончаренко, 1969; Карлушин, 1979; Тихомиров и др., 1978; Огнев, 1979	Стабилизация значения подсчетного параметра. По мере детализации разведки значение какого-либо важного параметра (запасов) колеблется все меньше и практически стабилизируется. Варианты: стабилизация информации, энтропии Стабилизация погрешности оценки параметра. Кривая погрешностей от числа скважин (или других выработок) имеет гиперболический вид, и начиная с некоторого их числа погрешность практически не изменяется
Vandemer, 1966; Черницкий, 1977; Ковальчук, Предтеченская, 1976; Ковальчук, 1977; Емельянов, 1966; Васильев, Молчанова, 1977	Достижение заданной точности (допустимой погрешности) оценки параметра
Matheron, Formeri, 1963; Матерон, 1968; Добровольский, Володин и др., 1974; Сенюков, 1972; Фролов и др., 1976; Быков и др., 1972; Ткачев, 1971; Америка и др., 1976	Минимизация суммы затрат на разведку и потерь прибыли от разработки
Инструкции по применению. . . , 1961, 1972, 1984	Достижение заданного соотношения промышленных категорий запасов. Удовлетворение требованиям обоснованного выбора технологической схемы разработки месторождения
Головастов, Свищев; Габриэлянц и др., 1978а, 1978б, 1979	Число скважин устанавливается в зависимости от площади залежи. Является вариантом критерия стабилизации параметра. По данным многих залежей устанавливается зависимость числа скважин, стабилизирующих параметр на каждой залежи, от ее площади. По этой зависимости для вновь разведываемых залежей устанавливается оптимальное число скважин (Инструкцией ГКЗ рекомендуемое число скважин устанавливается в зависимости от категории запасов)
Васильев, Молчанова, 1977; Грудев, 1975; Фролов и др., 1976; Денисов, 1963, 1970; Козлов и др., 1970	Метод последовательного разрежения. Имеет много вариантов. По Васильеву строят графики изменения параметра от числа скважин, отмечают на них зону допустимых погрешностей. Если «текущее» значение параметра вошло в зону допустимых погрешностей, то скважин достаточно. По Грудеву эмпирическую кривую «погрешность—плотность сети» экстраполируют на область исходной сети, определяя ее погрешность. По Фролову аппроксимируют зависимость «погрешность—расстояние между скважинами» прямой, а для учета погрешности детальной карты параллельно переносят ее в начало координат

для получения статистически обоснованных решений по одной и той же схеме анализируется множество моделей, с одной стороны — удовлетворяющих разведочным данным, с другой — отличающихся друг от друга реализациями случайной компоненты модели.

В-третьих, на каждой модели имитация разведки проводится для двух значений плотности сети: фактически реализованной и планируемой для следующего шага уплотнения.

Применение модели, включающей вероятную изменчивость, исключает упрощение модели и максимально приближает ее по свойству «разведываемости» к изучаемому месторождению, что является самым важным условием при определении оптимальной плотности разведочной сети. Применение в предла-

гаемом способе ряда моделей, отличающихся друг от друга, но не противоречащих геологоразведочным данным на момент анализа разведки, нивелирует влияние случайного фактора. Многократная имитация разведки и проектирования на каждой модели с двумя значениями плотности сети, одна из которых соответствует фактически достигнутой, а другая — ближайшей планируемой, позволяет надежно установить среднюю неоптимальность проекта. Под оптимальным проектом мы понимаем проект, составленный на современном научно-техническом уровне с использованием истинных значений параметров месторождения в каждой точке залежи. При таком определении оптимальный проект по реальным данным разведки не может быть составлен, и это понятие можно применять только по отношению к известной модели. Оптимальному проекту противопоставляется разведочный проект, т. е. проект разработки, составленный по данным разведки.

По предлагаемому способу приведенные затраты на добычу 1 т запасов рассчитываются по результатам имитации разработки по разведочному проекту на модели (не на разведочной модели, а на исходной модели, по результатам имитации разведки которой составлен разведочный проект). Таким образом, предлагаемый способ заключается в следующем.

1. Составляют несколько моделей месторождения, удовлетворяющих геологоразведочным данным на момент анализа разведки, т. е. имеющих в точках наблюдений значения геологических величин, с точностью до погрешностей оценки совпадающих с наблюдаемыми, а в промежутках между точками наблюдений включающих вероятную изменчивость параметров.

2. На каждой модели несколько раз имитируют разведку с плотностью и по методике, которые соответствуют реализованным в натуре (на месторождении).

3) По каждой модели и каждой имитации разведки составляется проект (технологическая схема) разработки.

4. Каждый проект (технологическая схема) разработки применяют к модели и для каждой модели рассчитывают средние (из всех проектов по этой модели) приведенные затраты на добычу 1 т запасов (S_1).

5. Операции 2—4 повторяют для планируемой плотности сети, в результате чего определяют приведенные затраты на добычу 1 т запасов S_2 .

6. Оценивают затраты на планируемое уплотнение сети R .

7. Сравнивают затраты на уплотнение сети с разностью $S_1 - S_2$. Если они меньше ($R < S_1 - S_2$), планируемую плотность сети принимают для реализации в натуре. В противном случае разведку месторождения заканчивают либо планируют меньшее уплотнение сети и повторяют операции 2—7. Все затраты в расчетах приводят к единому (текущему) моменту времени.

Предлагаемый метод оптимизации можно реализовать только с применением ЭВМ, причем не для отдельных расчетов, а путем создания комплекса программ для реализации метода в целом. Блок-схема алгоритма такого комплекса приведена ниже.

Применение предложенного способа определения оптимальной степени разведанности месторождений приводит к значительному экономическому эффекту. Так, по разведанным месторождениям нефти Тимано-Печорской провинции средняя степень неоптимальности детальности разведки составляет 5—10 %. Предложенный способ по предварительным расчетам позволяет снизить ее до 0—5 %, т. е. в среднем на 5 % или 1.5—2 млн. руб. на каждом месторождении. Затраты по применению способа на одну процедуру анализа составляют ориентировочно 300—500 руб. (на оплату машинного времени и подготовку данных).

Блок-схема алгоритма оптимизации детальной разведки
(иллюстрационный вариант)

1

Исходные данные:

- 1) результаты разведки: значения параметров месторождения в точках наблюдений, координаты этих точек;
- 2) тип и параметры характеристики изменчивости месторождения (например, полувариограммы);
- 3) плотность реализованной (D) и планируемой (P) разведочной сети;
- 4) плотность, с которой считываются данные с модели;
- 5) затраты на реализацию одного разведочного элемента (например, затраты на скважину);
- 6) точность, с которой требуется установить степень неоптимальности проекта разработки по отношению к одной модели месторождения ($EPS1$);
- 7) точность, с которой требуется установить степень неоптимальности проекта разработки по отношению к серии моделей одного месторождения ($EPS2$)

2

Создание модели изучаемого месторождения (случайное моделирование геологического поля, удовлетворяющего заданной полувариограмме с плотностью M) DEP

3

Имитация натурной разведки месторождения (моделирование разведки на DEP с плотностью D наложением разведочной сети и интерполяцией между точками плотности M) с получением геологоразведочных данных — GRD

4

$k := 0$

5

$k := k + 1$

Из блока 21

6

Создание k -й модели месторождения по GRD (моделирование случайного поля, удовлетворяющего данным разведки, а также заданной полувариограмме — $MDEP_k$)

7

$i := 0$

8

$i := i + 1$

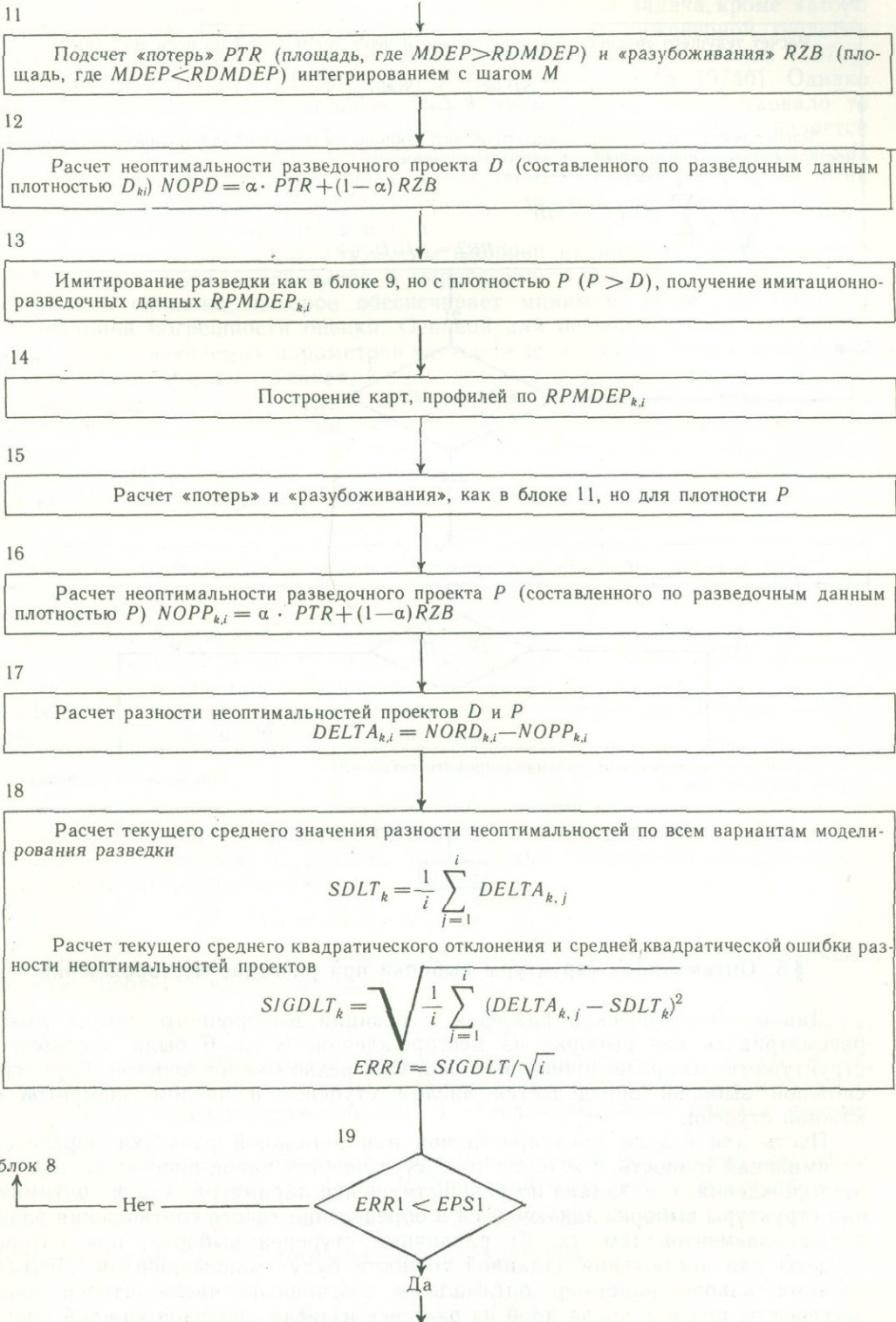
Из блока 19

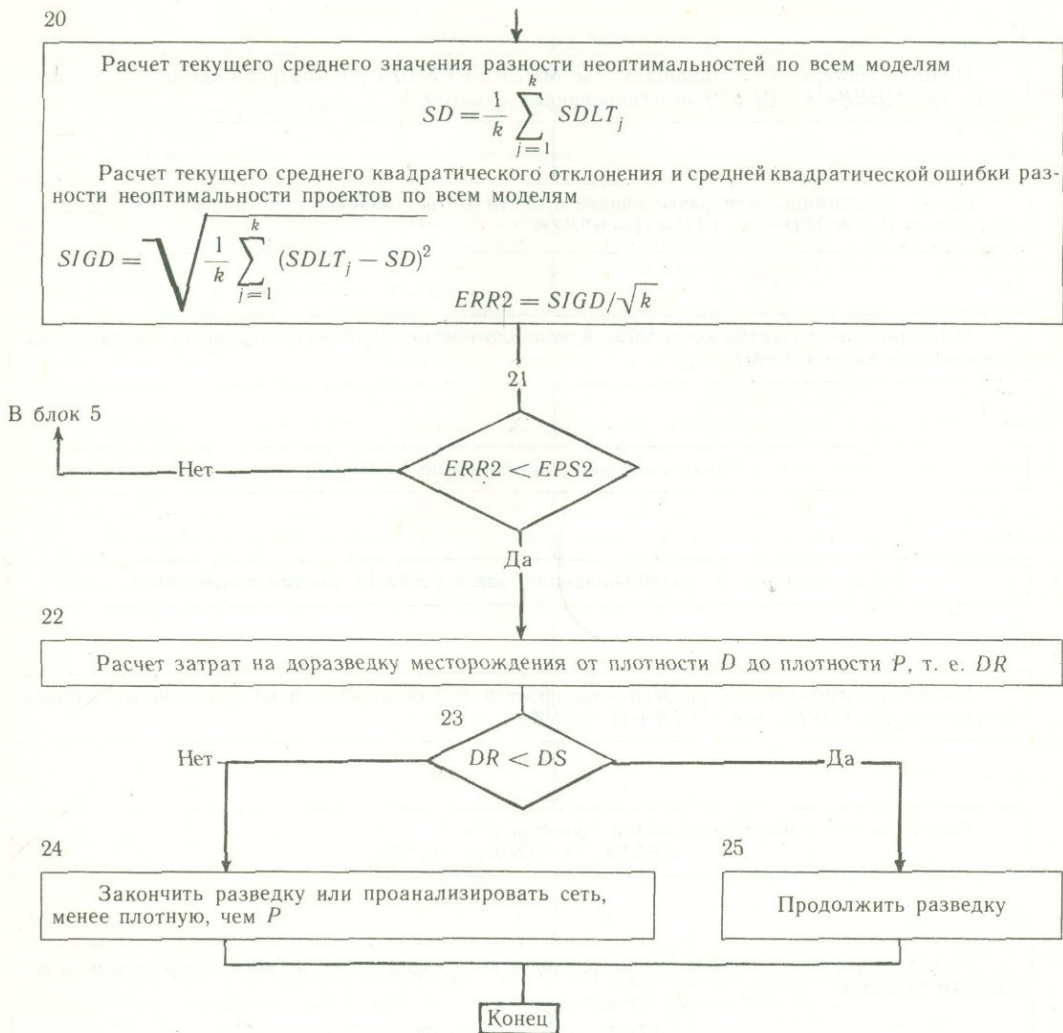
9

Имитация разведки с плотностью D модели $MDEP_k$ (случайный выбор начальной точки сети, считывание с $MDEP_k$ с шагом, определяемым плотностью D), получение имитационно-разведочных данных $RDMDEP_{k,i}$

10

Построение карт, профилей по $RDMDEP_{k,i}$





§ 5. Оптимизация структуры выборки при разведке месторождений

Данные геологической разведки с позиций выборочного метода можно рассматривать как выборку из месторождения. В гл. 6 была рассмотрена структура геологоразведочной выборки. Она редко бывает простой. Структура сложной выборки определяется числом ступеней и числом элементов на каждой ступени.

Пусть для стадии предварительной или детальной разведки определена оптимальная точность, с которой требуется оценить какой-либо из параметров месторождения, т. е. задана погрешность оценки параметра. Задача оптимизации структуры выборки заключается в определении такого соотношения разведочных элементов (см. гл. 6) различных ступеней выборки, при котором затраты для достижения заданной точности будут минимальными. Другими словами, каково, например, оптимальное соотношение числа штолен, числа расщечек из штолен, числа проб из расщечек и числа анализов каждой пробы, если затраты на «осуществление» разведочного элемента каждой ступени

известны. В применении к разведке месторождений эта задача, кроме автора (Ткачев, Кетрис, 1978), еще не рассматривалась. В инженерной геологии важные результаты были получены И. П. Кречетовой и М. В. Рацем (1973), И. П. Кречетовой (1974а, 1974б) и Б. Г. Слепцовым (1974а, 1974б). Однако они мало заинтересовали геологов. По-видимому, этому способствовало то обстоятельство, что: а) объектом инженерно-геологических работ является слой грунта на небольшую глубину; б) решение инженерно-геологических задач, как правило, нелинейно зависит от изучаемых параметров. Для геологии этот случай не типичен. Запасы, например, линейно зависят от площади, объема, содержания, пористости и т. д.

Рассмотрим оптимизацию структуры выборки на примере месторождений нефти. Задача здесь заключается в определении такого соотношения числа образцов и скважин, которое обеспечивает минимум затрат на разведку при заданной погрешности оценки. Основой для решения служат математические модели изучаемых параметров месторождений. Нами предложено решение для шести моделей (Ткачев, Кетрис, 1978).

Рассмотрим лишь одну модель, оказавшуюся по результатам практического применения наиболее типичной. Она заключается в следующем. Значения свойства в образцах в пределах скважины независимы, скважинные средние μ_j существенно отличаются, автокорреляции среднескважинных по площади залежи нет, т. е. значение свойства в i -м образце j -й скважины равно

$$x_{ij} = \mu_j + \xi_{ij}, \quad (8.5.1)$$

где ξ_{ij} — случайные величины с нулевыми средними и внутрискважинными дисперсиями σ_j^2 .

Критерием оптимизации является минимум дисперсии оценки параметра σ_0^2 при фиксированных затратах на разведку A , или, что то же самое, минимум затрат при фиксированной дисперсии.

Пусть a_1 — стоимость бурения скважины, a_2 — стоимость отбора и лабораторного исследования одного образца; тогда

$$A = a_1 N + a_2 K, \quad (8.5.2)$$

где N — число скважин, K — общее число образцов из них.

Дисперсия оценки параметра складывается из дисперсии оценки среднескважинного значения $\sigma_{\text{ср}}^2 = \sigma_x^2/k$, где k — число образцов из скважины, и дисперсии среднескважинных по залежи.

$$\sigma_0^2 = (\sigma_x^2/k + \sigma_{\text{пл}}^2)/N = \sigma_x^2/K + \sigma_{\text{пл}}^2/N. \quad (8.5.3)$$

Подставляя в (8.5.2) значение K из (8.5.3), получим нелинейную функцию, которую требуется минимизировать:

$$A = a_1 N + a_2 \sigma_x^2 / (\sigma_x^2 - \sigma_{\text{пл}}^2 / N) = a_1 N + a_2 \sigma_x^2 N / (\sigma_0^2 N - \sigma_{\text{пл}}^2) \quad (8.5.4)$$

при следующих ограничениях: 1) условия неотрицательности $N \geq 0$, $K > 0$; 2) предельно допустимое по смыслу соотношение числа образцов и скважин $K > N$; 3) слагаемые в выражении (8.5.3) — величины одного порядка, поэтому $N > \sigma_{\text{пл}}^2 / \sigma_0^2$.

Для определения оптимальной структуры выборки необходимо продифференцировать выражение (8.5.4) по N , приравнять производную к нулю:

$$A'_N = N^2 - 2\sigma_{\text{пл}}^2 / \sigma_0^2 N + (\sigma_{\text{пл}}^2 / \sigma_0^2)^2 - a_2 / a_1 \cdot \sigma_{\text{пл}}^2 \sigma_x^2 / \sigma_0^4 = 0.$$

Положительный корень этого квадратного уравнения

$$N = \left(\sigma_{\text{пл}}^2 + \sqrt{\frac{a_2}{a_1} \sigma_{\text{пл}}^2 \sigma_x^2} \right) / \sigma_0^2 \quad (8.5.5)$$

Подставляя значение N в (8.5.3), получаем

$$K = \left(\sqrt{\frac{a_1}{a_2} \sigma_{пл}^2 \sigma_x^2 + \sigma_x^2} \right) \sigma_\theta^2 \quad (8.5.6)$$

Оптимальное число образцов из одной скважины

$$k = K/N = \sqrt{a_1/a_2 \cdot \sigma_x / \sigma_{пл}} \quad (8.5.7)$$

Из геологических соображений очевидно, что полученное экстремальное значение может быть только минимумом, но это необходимо доказать математически. Исследуем вторую производную A_K'' :

$$A_K'' = \frac{a_1 \sigma_{пл}^2 \sigma_x^2 \cdot 2 (\sigma_\theta^2 K - \sigma_x^2) \sigma_\theta^2}{(\sigma_\theta^2 K - \sigma_x^2)^4}$$

Это выражение имеет смысл только тогда, когда $\sigma_\theta^2 K - \sigma_x^2 \neq 0$, а экстремум является минимумом, если $\sigma_\theta^2 K - \sigma_x^2 > 0$, т. е. $\sigma_\theta^2 > \sigma_x^2 / K$. Но из (8.5.3) известно, что $\sigma_\theta^2 = \sigma_x^2 / K + \sigma_{пл}^2 / N$. Оба слагаемых в первой части этого равенства положительны, поэтому действительно $\sigma_\theta^2 > \sigma_x^2 / K$. Следовательно, функция выпукла вниз при любых положительных значениях K , и найденное решение действительно является минимумом работы.

Таким образом, оптимальное число образцов из скважины пропорционально отношению скважинной и площадной средних квадратических отклонений свойства и обратному отношению стоимостей разведочных элементов. Действительно, чем дороже скважина по сравнению со стоимостью отбора образца, тем большее число образцов следует отбирать из каждой скважины. Число образцов следует также увеличивать, если увеличивается изменчивость свойства по разрезу скважины. Кажется парадоксальным, что оптимальное число образцов из скважины не зависит в данной модели от требуемой точности разведки, т. е. что увеличение точности следует проводить за счет увеличения числа скважин при постоянном числе образцов из каждой скважины. Другими словами, в процессе разведки структура выборки (соотношение числа образцов и скважин) должна сохраняться постоянной.

Табл. 8.5.1 наглядно иллюстрирует характер изменения числа образцов из скважины в зависимости от соотношения стоимости скважины и стоимости отбора и анализа образца a_1/a_2 , а также от соотношения квадратической изменчивости по разрезу и таковой по площади $\sigma_x^2/\sigma_{пл}^2$.

В применении к нефтяным месторождениям этой проблемы оптимизации двухступенчатой выборки касался А. В. Белов (1969, 1972), однако с примененными им принципами и полученными формулами мы не можем согласиться по причинам, рассмотренным в нашей статье (Ткачев, 1976). Здесь отметим только, что А. В. Белов применил принцип минимизации дисперсии оценки при фиксированном числе образцов, а для расчета дисперсии оценки предлагает формулу (в наших символах) $\sigma_\theta^2 = \sum_{i=1}^n \sigma_{xi}^2 / n_i / N^2 + \sigma_{пл}^2 / N$, ошибочность которой

можно показать на следующем примере. Пусть $\sigma_x^2 = 20$, $\sigma_{пл}^2 = 2$, $k_1 = 1$, $k_2 = 5$. Тогда, по методике А. В. Белова, дисперсия оценки одной скважиной, из которой отобрано 5 проб, равна $\sigma_\theta^2 = 20/5 + 2 = 6$, а двумя скважинами — $\sigma_\theta^2 = (20/1 + 20/5) / 2^2 + 2/2 = 7$, т. е. больше, чем одной скважиной! Правильный результат для двух скважин в соответствии с выводами гл. 6 следующий:

$$\sigma_\theta^2 = \frac{1}{\frac{1}{20/1+2} + \frac{1}{20/5+2}} = 4.7,$$

т. е. оценка двумя скважинами лучше, чем любой из них одной.

Оптимальное число образцов из скважины в зависимости от характера изменчивости свойства и стоимости видов работ

Отношение квадратической изменчивости свойства по разрезу скважины и по площади залежи ($\sigma_x/\sigma_{пл}$)	Оптимальное число образцов из скважины в зависимости от соотношения стоимости скважин и стоимости отбора и анализа образца (a_1/a_2)						
	1	10	100	1000	10 000	100 000	1 000 000
0.01	1	1	1	1	1	1	1
0.1	1	1	1	3*	10*	32*	100
1	1	3	10	32*	100*	316*	1000
10	10	32	100	316	1000	3160	10000
100	100	316	1000	3160	10000	31600	100000

Примечание. Звездочкой отмечены реальные соотношения.

Трехступенчатая выборка также широко распространена при разведке. На месторождениях нефти и газа выборку тоже можно считать трехступенчатой, если керновые образцы анализировать несколько раз и отдельный анализ считать разведочным элементом самой нижней ступени. Введем обозначения: σ_a^2 — дисперсия воспроизводимости анализа, l — число анализов одного образца, a_3 — стоимость анализа. Тогда «общая» итоговая дисперсия оценки параметра

$$\sigma_0^2 = [(\sigma_a^2/l + \sigma_x^2)/k + \sigma_{пл}^2]/N, \quad (8.5.8)$$

а общие затраты

$$A = a_3lkN + a_2kN + a_1N = a_3L + a_2K + a_1N, \quad (8.5.9)$$

где L — общее число анализов, N — число скважин.

Прямой путь минимизации функции A сопряжен с трудностями: требуется решение системы двух квадратных уравнений с двумя неизвестными:

$$\left. \begin{aligned} A'_k &= a_3\sigma_{пл}^2 l^2 k^2 + a_2\sigma_{пл}^2 lk - a_1\sigma_x^2 l - a_1\sigma_a^2 = 0 \\ A'_l &= a_3\sigma_{пл}^2 l^2 k^2 + a_3\sigma_x^2 lk - a_2\sigma_a^2 k - a_1\sigma_a^2 = 0 \end{aligned} \right\} \quad (8.5.10)$$

Поэтому применим принцип декомпозиции и будем рассматривать одну скважину как самостоятельный объект оценки двухступенчатой выборки: образцы — анализы. Соотношение «образцы — анализы» в соответствии с предыдущим результатом не зависит от требуемой точности оценки по скважине. Это позволяет провести отдельно оптимизацию числа анализов образца, а затем — числа образцов в скважине.

Применяя предыдущий результат (8.5.7), получим

$$l = \sqrt{a_2\sigma_a^2} / \sqrt{a_3\sigma_x^2}. \quad (8.5.11)$$

Учитывая, что каждый образец должен анализироваться l раз, а стоимость его отбора равна a_2 , стоимость отбора и анализа одного образца

$$a'_2 = la_3 + a_2 = \sqrt{\frac{a_2 a_3 \sigma_a^2}{\sigma_x^2}} + a_2. \quad (8.5.12)$$

Общая дисперсия свойства по разрезу скважины с учетом дисперсии воспроизводимости анализа при анализе каждого образца l раз

$$\hat{\sigma}_x^2 = \sigma_a^2/l + \sigma_x^2 = \sqrt{\frac{a_3}{a_2} \sigma_x^2 \sigma_a^2} + \sigma_x^2. \quad (8.5.13)$$

Подставляя в (8.5.7) вместо a_2 и σ_x^2 значения a'_2 и $\hat{\sigma}_x^2$, получим оптимальное значение числа образцов (элементов второй ступени) в трехступенчатой выборке:

$$k = \sqrt{a_1 \hat{\sigma}_x^2} / \sqrt{a'_2 \sigma_{пл}^2}. \quad (8.5.14)$$

Полученные результаты позволяют сделать следующие обобщения: а) решение для выборки с любым количеством ступеней имеет одинаковую структуру и одинаковые свойства; б) оптимальная структура выборки не зависит от требуемой точности разведки; для увеличения точности всегда необходимо увеличивать число элементов первой ступени, в данном случае — скважин; в) процесс оптимизации структуры выборки необходимо производить с последней ступени и двигаться к первой, анализируя последовательно по две смежные ступени: результаты оптимизации первой служат исходными данными для второй и т. д.; г) оптимальное число элементов n -й ступени в каждом элементе $(n+1)$ -й ступени определяется по формуле

$$k = \sqrt{a_{n-1} / a_n \cdot \sigma_n / \sigma_{n-1}}, \quad (8.5.15)$$

где a_{n-1} — стоимость элемента $(n-1)$ -й ступени; a_n — суммарная стоимость элемента n -й ступени и всех входящих в него элементов всех последующих ступеней; σ_{n-1}^2 — дисперсия средних значений свойства между элементами $(n-1)$ -й ступени; σ_n^2 — средняя дисперсия оценки значений свойства в элементах n -й ступени. Обозначим

$$C_n = a_n / \sigma_n^2, \quad C_{n-1} = a_{n-1} / \sigma_{n-1}^2. \quad (8.5.16)$$

Тогда (8.5.15) можно записать в виде

$$k = \sqrt{C_{n-1} / C_n}, \quad (8.5.17)$$

и интерпретировать следующим образом: число элементов n -й ступени, приходящееся на один элемент $(n-1)$ -й ступени, равно корню квадратному из отношения стоимости единицы дисперсии $(n-1)$ -й и n -й ступеней.

При переменной мощности исследуемого тела кажется очевидным, что число образцов должно быть пропорционально мощности. Определим дисперсию оценки среднего по залежи при различной ее мощности. Расчет должен производиться по формуле

$$\sigma_0^2 = \sum_{i=1}^N p_i^2 (\sigma_{\bar{x}_{скв_i}}^2 + \sigma_{пл}^2), \quad (8.5.18)$$

где $p_i = m_i / M$, m_i — мощность в i -й скважине, M — сумма мощностей по всем скважинам. Примем для проверки $p_i = \text{const} = p = 1/N$, $\sigma_{\bar{x}_{скв_i}}^2 = \text{const} = \sigma_{\bar{x}_{скв}}^2$; тогда $\sigma_0^2 = (1/N)^2 N (\sigma_{\bar{x}_{скв}}^2 + \sigma_{пл}^2) = 1/N (\sigma_{\bar{x}_{скв}}^2 + \sigma_{пл}^2)$, что совпадает с (8.5.3). Подставив вместо $\sigma_{\bar{x}_{скв}}^2$ его значение, равное σ_x^2 / k_i , получим

$$\sigma_0^2 = \sum_{i=1}^N p_i^2 \left(\frac{\sigma_x^2}{k_i} + \sigma_{пл}^2 \right) = \sigma_x^2 \sum_{i=1}^N \frac{p_i^2}{k_i} + \sigma_{пл}^2 \sum_{i=1}^N p_i^2. \quad (8.5.19)$$

Необходимо определить условия (т. е. значения k_i), при которых это выражение достигает минимума. Поскольку принцип распределения образцов в зависимости от мощности должен быть одинаковым для любого числа скважин, рассмотрим для определенности две скважины: $\sigma_0^2 = \sigma_x^2 (p_1^2 / k_1 + p_2^2 / k_2) + \sigma_{пл}^2 (p_1^2 + p_2^2)$; так как $\sum p_i = 1$, то $p_2 = 1 - p_1$ и $\sigma^2 = \sigma_x^2 \left[\frac{p_1^2}{k_1} + \frac{(1-p_1)^2}{k_2} \right] + \sigma_{пл}^2 \times$

$\times [p_1^2 + (1 - p_1)^2]$. Возьмем производную по k_1 и приравняем ее к нулю:

$$(\sigma_x^2)_{k_1}' = -\frac{\sigma_x^2 p_1^2}{k_1^2} - \frac{\sigma_x^2 (1 - p_1)^2}{(K - k_1)^2} = 0,$$

так как $\sigma_x^2 \neq 0$, то $\frac{p_1^2 (K - k_1)^2 + (1 - p_1)^2 k_1^2}{k_1^2 (K - k_1)^2} = 0$. Но $k_1 = 0$, $K - k_1 = 0$; тогда $p_1^2 (K - k_1)^2 + (1 - p_1)^2 k_1^2 = 0$. После преобразования получаем квадратное уравнение $k_1^2 (1 - 2p_1) + k_1 2kp_1^2 - K^2 p_1^2 = 0$, положительный корень которого

$$k_1 = \frac{K p_1 (1 - 2p_1)}{1 - 2p_1} = K p_1, \quad (8.5.20)$$

где K — общее число образцов. Обобщая этот вывод на N скважин, можем записать $k_i = K p_i$, т. е. число образцов из скважины должно быть пропорционально мощности изучаемых отложений в этой скважине. Методика определения числа образцов из каждой скважины в случае различной мощности состоит из следующих операций: а) определяется значение k , как если бы залежь имела одинаковую мощность по скважинам m , равную средней мощности \bar{m} ; б) определяется средняя мощность по скважинам \bar{m} ; в) составляется пропорция $k/m = k_i/m_i$, откуда $k_i = k m_i / \bar{m}$.

Методика оптимизации структуры выборки на многозалежных месторождениях более громоздка, и мы отсылаем читателей к работе Ю. А. Ткачева и М. П. Кетрис (1978). Для многозалежных месторождений оказалось существенным, разведуются ли они одной и той же сетью скважин или верхние залежи требуют большего числа скважин. В первом случае «корректировка» решения заключается в разделении стоимости скважины в \sqrt{n} раз, где n — число залежей. Во втором случае стоимость скважины для каждой нижележащей залежи берется по разности между стоимостью скважины до глубины анализируемой залежи и до глубины вышележащей залежи.

Условия разведки месторождений накладывают ряд ограничений на изложенную теоретическую схему. Так, на месторождениях нефти анализ образцов керна, особенно терригенных коллекторов, перестал быть основным способом оценки пористости. Чаще он используется лишь для эталонировки результатов геофизических исследований.

Недостаточный вынос керна ставит существенную техническую границу верхнему числу образцов из пласта по скважине. Наконец, скважина используется для геофизических исследований, для промысловых испытаний, а в дальнейшем и как эксплуатационная или нагнетательная. Ясно, что результаты оптимизации, предусматривающие меньшее число скважин, чем необходимо для решения других задач разведки, должны пересматриваться в сторону их увеличения.

Наконец, имеется важное обстоятельство, присущее всем ситуациям опробования и разведки. При решении задачи предполагалось, что число образцов и число скважин меняются непрерывно. Фактически N и k могут принимать только целочисленные значения. Если требования к точности оценки изучаемого параметра не определены, число образцов из каждой скважины необходимо определять по изложенной здесь методике с округлением полученного k в большую сторону. Число скважин определяется при этом другими соображениями. Если требования к точности оценки среднего значения изучаемого параметра (пористости, проницаемости, нефтенасыщенности) установлены (в виде предельного значения дисперсии σ^2), то поступают следующим образом.

1. Случай первый — число скважин, требуемое для решения других задач разведки, превышает то число скважин, которое требуется для оценки изучаемого параметра с заданной точностью. При этом условии число образцов из скважины может быть уменьшено по сравнению с оптимальным. Оно определяется по формуле

$$k = \sigma_x^2 / (\sigma_\theta^2 N - \sigma_{пл}^2), \quad (8.5.21)$$

где σ^2 — максимально допустимая дисперсия оценки параметра, N — число скважин на месторождении.

2. Случай второй — из всех задач разведки оценка изучаемого параметра с заданной точностью требует наибольшего числа скважин. Это условие проверяется следующим образом. По изложенной методике определяем оптимальное число образцов из скважины и, задаваясь предельной дисперсией σ^2 , — оптимальное число скважин. Берем ближайшие целые значения N с избытком (N_1) и недостатком (N_2). Подставляем их в формулу (8.5.21) и находим соответствующие требуемые при этом числа образцов k_1 и k_2 :

$$k_1 = \frac{\sigma_x^2}{\sigma_\theta^2 N_1 - \sigma_{пл}^2}, \quad k_2 = \frac{\sigma_x^2}{\sigma_\theta^2 N_2 - \sigma_{пл}^2}.$$

Берем их с округлением в большую сторону. Затем находим затраты A_1 (при N_1 и k_1) и A_2 (при N_2 и k_2) по формулам $A_1 = a_1 N_1 + a_2 k_1 N$, $A_2 = a_1 N_2 + a_2 k_2 N$. Выбираем ту пару N и k , которая дает меньшие затраты.

В табл. 8.5.2 дано сравнение оптимального «непрерывного» и целочисленного вариантов по среднедевонской залежи Усинского нефтяного месторождения. Из таблицы видно, что в подавляющем большинстве случаев оптимальный целочисленный вариант получается из «непрерывного» округлением числа скважин в большую сторону. Число образцов из скважины при этом уменьшается. Лишь в редких случаях, когда оптимальное число скважин на 0.1—0.3 превышает целое число, округление может производиться в меньшую сторону. При этом необходимое число образцов резко увеличивается и становится обычно технически не осуществимым. Таким образом, вывод о постоянстве структуры выборки независимо от требуемой точности, справедливый для «непрерывного» случая (приблизительно — при большом числе образцов и скважин), здесь имеет относительный характер: оптимальное число образцов интерпретируется в большинстве случаев как верхняя граница. Для расчета оптимального числа образцов из скважины и числа скважин как для непрерывного, так и для целочисленного случаев в Институте геологии Коми филиала АН СССР составлены алгоритмы и программы.

Структура реальной выборки разведанного месторождения никогда не бывает оптимальной. Для измерения степени отклонения ее от оптимальной мы предлагаем следующий показатель неоптимальности:

$$\Pi = \frac{A - A_0}{A_0}, \quad (8.5.22)$$

где A — затраты на разведку при фактической структуре выборки; A_0 — затраты при оптимальной структуре выборки, которая дала бы ту же точность, что и фактическая.

По разработанной методике были проведены расчеты по среднедевонской залежи Усинского месторождения. В результате расчетов установлено, что оптимальное среднее число образцов из каждой скважины равно 150 против 100 фактических. Показатель неоптимальности составил 8—13%. Обычно это значение для разведанных месторождений больше. Но и здесь оптимальная стратегия позволила бы достигнуть той же точности с затратами, на 1.2 млн.

Сравнение оптимального числа образцов и скважин по «непрерывному» и целочисленному оптимальным вариантам

Предельно допустимая квадратическая погрешность оценки пористости, абс. %	Число образцов из скважины	Число скважин	Затраты, усл. ед.
0.1	155 (151)	755.8 (756)	304.0 (304)
0.12	<u>155 (174)</u>	<u>534.4 (534)</u>	215.4 (215.5)
0.14	155 (151)	377.9 (378)	152.3 (152.4)
0.17	<u>155 (174)</u>	<u>267.2 (267)</u>	107.7 (107.7)
0.2	<u>155 (151)</u>	<u>188.9 (189)</u>	76.2 (76.2)
0.24	155 (113)	133.6 (134)	53.8 (53.9)
0.28	155 (91)	94.5 (95)	38.1 (38.2)
0.33	155 (113)	66.8 (67)	26.9 (27.0)
0.4	<u>155 (456)</u>	<u>47.2 (47)</u>	19.0 (19.2)
0.47	<u>155 (47)</u>	<u>33.4 (34)</u>	13.5 (13.6)
0.57	155 (51)	23.6 (24)	9.5 (9.6)
0.67	155 (47)	16.7 (17)	6.7 (6.8)
0.8	155 (51)	11.8 (12)	4.7 (4.8)
0.95	155 (14)	8.3 (9)	3.3 (3.6)
1.1	155 (51)	5.9 (6)	2.3 (2.4)
1.3	155 (6)	4.2 (5)	1.7 (2.0)

Примечание. Число перед скобкой — оптимальный вариант, в скобках — оптимальный целочисленный вариант. Число образцов округлено до целого. Подчеркнуты случаи уменьшения числа скважин целочисленного варианта против непрерывного.

руб. меньшими фактических (методика расчета экономической эффективности от использования НИР изложена в работе автора (Ткачев, 1985)). Практически это происходит за счет того, что совсем небольшие затраты на отбор и анализ дополнительно 50 образцов из каждой скважины позволили бы достигнуть фактической точности подсчета запасов сетью из 12 скважин вместо 13.

В результате определились направления дальнейших исследований по оптимизации структуры выборки. Важнейшие из них следующие.

1. Учет многоцелевого назначения скважин. Фактически скважина используется не только для определения пористости, проницаемости и нефтенасыщенности по керну, но и для геофизических и других исследований, а также для промысловых испытаний.

2. Учет автокорреляции изучаемых параметров по разрезу скважин и конечного характера геологических совокупностей. В простейшем варианте учет первого фактора может быть произведен с помощью коэффициента $[1-r^2(h)]$, где $r(h)$ — коэффициент автокорреляции изучаемого свойства в зависимости от расстояния h . Для аппроксимации зависимости r от h можно рекомендовать экспоненциальную функцию $r(h) = e^{-ah}$. Учет второго фактора можно произвести с помощью коэффициента отбора $f = 1 - l_0 k / L$, где l_0 — длина единичного образца, k — их число, L — длина оцениваемого интервала по скважине. Равенство (8.5.3) после этого примет вид

$$v_0^2 = \left[\frac{\sigma_x^2 (1-r^2(h))}{k} \left(1 - \frac{l_0 k}{L} \right) + \sigma_{пл}^2 \right] / N.$$

При частом расположении скважин и большом числе образцов из них решение может быть получено лишь с применением геостатистических характеристик изменчивости по площади и разрезу.

3. Учет при решении взаимной корреляции изучаемых свойств (пористости, проницаемости, нефтенасыщенности), позволяющий уточнить оценку при тех же затратах.

4. Учет необходимости статистической увязки значений коллекторских свойств по керну и по геофизическим данным и получения комплексной оценки.

5. Учет ограничений в максимальном числе представительных образцов из скважины, связанных с недостаточным выносом керна и его избирательным размывом.

6. Учет многозалежности месторождений в том случае, когда верхняя залежь требует разведки бóльшим числом скважин, чем нижняя.

§ 6. К оптимизации разведки на уточнение оценки точности

Один из аспектов оптимизации разведки заключается в выборе способа размещения разведочных элементов, например скважин, по объему или площади месторождения. Ему посвящена обширная литература, в которой рекомендуются различные системы размещения разведочных горных выработок или скважин, оптимальные, по мнению авторов, либо для минимизации погрешности оценки запасов, либо по вероятности подсечения рудных тел, либо для наилучшего использования разведочных горных выработок при последующей разработке месторождения. В этой работе мы не касаемся перечисленных сторон оптимизации размещения выработок, а ставим вопрос, какое их размещение оптимально для оценки погрешности оценки параметров. Равномерная сеть на статистически однородном объекте дает максимум информации о значении параметра, однако, как выяснилось, она не является лучшей для оценки точности параметра.

Еще в 1910 г. С. Ю. Доборжинский для оценки точности содержания и запасов золоторудных месторождений предложил формулу

$$s_{\theta}^2 = s^2/n, \quad (8.6.1)$$

где s_{θ}^2 — дисперсия параметра, n — число наблюдений. Для указанных месторождений эта формула не потеряла своего значения и теперь. Что касается большинства месторождений других полезных ископаемых с более равномерным и «закономерным» в пространстве распределением компонентов, то решение по формуле (8.6.1) дает для них почти всегда сильно завышенный результат, т. е. фактическая погрешность оценки меньше, чем вычисленная. Это явление впервые было детально рассмотрено С. Траскоттом (цит. по: Раевский, Медведев, 1976), Г. А. Мельниковым (1934) и в целом ряде работ В. Г. Соловьева (например, 1938). Их основной вывод, выраженный в современной терминологии, заключается в том, что изучаемый объект не подходит под определение совокупности случайных величин, а разведочные данные не являются случайной выборкой из нее. Свойства месторождения являются пространственными переменными с закономерной тенденцией изменения. Это утверждение повторяется в большинстве работ, подвергающих критике применение в геологической разведке методов математической статистики.

Обозначим через σ_{θ}^2 истинную (фактическую) дисперсию оценки параметра θ , тогда рассматриваемое утверждение можно записать в виде неравенства

$$\sigma_{\theta}^2 < s^2/n, \quad (8.6.2)$$

где s^2 — дисперсия изучаемой величины по данным разведки. Однако условия, при которых выполняется это неравенство, четко не сформулированы. Действительно, всегда ли вычисленная среднеквадратическая погрешность превышает

фактическую? Может ли быть обратное соотношение и в каких случаях? Можно ли разместить скважины или пробы так, чтобы, выбрав соответствующий метод расчета среднего, получить фактическую дисперсию оценки среднего, соответствующую формуле (8.6.1)? Легко показать, что для выполнения последнего условия достаточно: а) разместить скважины случайным образом и б) расчеты вести методом среднего арифметического. Под случайным размещением при этом следует понимать такое, при котором расположение каждой последующей скважины не зависит от расположения предыдущих.

Реальное размещение скважин может отклоняться от случайного в двух «направлениях» — быть более равномерным и менее равномерным, чем случайное, когда скважины сгущены на отдельных участках месторождения. При разведочной сети, более равномерной, чем случайная, формула (8.6.1) дает завышенную оценку дисперсии; при менее равномерной сети — заниженную. Если первое соотношение освещено в литературе подробно, то о последнем соотношении, т. е. недооценке погрешности, к сожалению, даже не упоминается. Условия, при которых фактическая дисперсия больше, меньше или равна рассчитанной по формуле (8.6.1), сведены в табл. 8.6.1. Случайный способ размещения разведочных элементов, обеспечивающий правильность оценки точности по формуле (8.6.1), неэффективен, с точки зрения повышения этой точности, и тем больше, чем сильнее выражена закономерная составляющая изменчивости. Однако означает ли это, что оптимальной является равномерная сеть? При ответе на данный вопрос следует учесть следующее: а) по результатам разведки нам известно не точное значение параметра θ , а лишь его выборочная оценка $\hat{\theta}$; б) эта выборочная оценка параметра характеризуется некоторой среднеквадратической погрешностью $\sigma_{\hat{\theta}}$, точное значение которой нам также неизвестно, а известна лишь ее выборочная оценка $s_{\hat{\theta}}$; в) величина $s_{\hat{\theta}}$ характеризуется квадратической погрешностью $\sigma_{s_{\hat{\theta}}}$, которую мы также не знаем, но по результатам разведки имеем ее выборочную оценку $s_{s_{\hat{\theta}}}$. При разведке необходимо получить оценки всех трех величин: $\hat{\theta}$, $s_{\hat{\theta}}$, $s_{s_{\hat{\theta}}}$. Необходимость оценки изучаемого параметра доказательства не требует, так как составляет самую суть разведки. В последнее время все острее ощущается необходимость сопровождать эту оценку характеристической погрешности $s_{\hat{\theta}}$. Труднее всего осознается необходимость иметь количественное представление о том, насколько корректна характеристика погрешности, т. е. насколько мала величина $s_{s_{\hat{\theta}}}$.

Таблица 8.6.1

Условия различного соотношения расчетной и фактической дисперсии оценки параметров

Размещение разведочных элементов	Способ расчета среднего	Изменение параметра в пространстве	
		закономерное	случайное
Менее равномерное, чем случайное	Среднеарифметический	$M \left\{ \frac{S_x^2}{n} \right\} < \sigma_{\theta}^2$	$M \left\{ \frac{S_x^2}{n} \right\} = \sigma_{\theta}^2$
	Средневзвешенный		
Случайное	Среднеарифметический	$M \left\{ \frac{S_x^2}{n} \right\} \approx \sigma_{\theta}^2$	
	Средневзвешенный	$M \left\{ \frac{S_x^2}{n} \right\} > \sigma_{\theta}^2$	
Равномерное	Среднеарифметический	$M \left\{ \frac{S_x^2}{n} \right\} > \sigma_{\theta}^2$	
	Средневзвешенный		

Геологов-практиков в настоящее время в первую очередь интересует получение правильной, более точной оценки самого параметра. Оценка же точности не является еще обязательным атрибутом разведки. Кроме того, возможность получения наиболее точной оценки параметра находится в противоречии с возможностью получить наиболее точную оценку ее точности. Поэтому разведчики размещают свои выработки равномерно, насколько позволяет ситуация, и поступают в основном правильно в соответствии с крейтеровским принципом равной разведанности. Позицию же практиков можно выразить так: не бурить же нам скважины намеренно неравномерно (например, по случайной схеме) только для того, чтобы теоретики могли правильно посчитать оценку точности. Согласно принципу равной разведанности, каждая часть месторождения должна быть изучена с одинаковой детальностью. «В основе этого принципа лежит положение о том, что природные тела полезных ископаемых характеризуются изменчивостью форм и качества, уловить которую можно лишь при более или менее равномерном распределении разведочных выработок и пунктов опробования в пределах пространства, занимаемого месторождением» (Крейтер, 1961, с. 26).

В геологоразведочной науке важным моментом была публикация статьи о статистической связанности рудных тел (Грудев, 1961), в которой автор предложил использовать для расчета точности корреляцию значений параметра в соседних по разрезу пробах (r) путем подстановки в формулу (8.6.1) вместо дисперсии признака s_x^2 «исправленную» дисперсию $s_x^2(1-r^2)$. В дальнейшем выяснилось, что для правильного расчета $s_{\bar{r}}$ необходимо знание среднего значения \bar{r} из коэффициентов корреляции, между значениями свойства на расстояниях от нуля до h , где h — шаг опробования. Как ведет себя функция $r(h)$ при значениях аргумента меньше шага опробования, можно узнать, лишь проведя дополнительное опробование, отбирая пробы на расстояниях, меньших основного шага. С позиций современной геостатистики для расчета дисперсии оценки необходимо знать среднее значение полувариограммы $\gamma(h)$ для расстояний от $l=0$ до $l=h$. При равномерной сети можно определить значения $\gamma(l)$ только для $l=h, 2h, 3h, \dots, kh$, т. е. для той области, для которой это не нужно.

Таким образом, противоречие между точностью и ее оценкой носит принципиальный характер, и его нельзя обойти применением какого-нибудь способа оценки точности. Это позволило Ж. Матерону (1968) заявить, что «теоретически невозможно по одним лишь экспериментальным данным рассчитать одновременно и оценку количества металла, и дисперсию этой оценки» (с. 39). Автор геостатистики имел в виду то обстоятельство, что для получения среднего значения полувариограммы на отрезке $[0, h]$ необходима экстраполяция, которая не может быть произведена без априорных предположений о виде функции, которая подходит для описания полувариограммы. Отказавшись от равномерной сети и сгустив ее в некоторых случайно выбранных участках, можно получить сведения о поведении полувариограммы на расстояниях, меньших основного шага опробования, но при этом часть выработок будет мало полезной для уточнения среднего значения параметра. Нельзя одновременно получить а) наилучшую для заданного числа скважин оценку количества металла, т. е. минимальное значение $s_{\bar{r}}$, и б) наилучшую оценку точности, т. е. наименьшее значение $s_{s_{\bar{r}}}$. Для уменьшения необходима как можно более равномерная сеть, для уменьшения $s_{s_{\bar{r}}}$ необходим «спектр» расстояний между точками опробования. Это можно осуществить, если на базе основного шага опробования часть шагов уплотнить в n_1 раз, на этих сгущенных участках некоторую часть уплотнить в n_2 раз и т. д., и т. д. Такой способ обеспечивает равномерность опробования на уровне основного шага, равномерность пунктов сгущений первого, второго и т. д. порядка, большее число точек

на расстоянии первого сгущения, меньшее — второго, еще меньшее — третьего и т. д.

Проведенный анализ убеждает в том, что равномерная разведочная сеть не является оптимальной, если принять во внимание, что она не позволяет давать правильную оценку точности, с которой определены изучаемые свойства месторождения. Возникает проблема определения доли точек опробования от общего их числа, которое надо располагать на расстояниях, меньших основного шага опробования, т. е. оптимизации числа и степени местных сгущений или оптимизации соотношения между дисперсией оценки параметра месторождения и дисперсией оценки этой дисперсии. Действительно, точность оценки в значительной степени обесценивается, если эта точность является «вещью в себе», если мы не подозреваем, что являемся ее обладателями. Если ограничить постановку задачи вариантом сгущений вдвое, т. е. $n_1 = n_2 = \dots = n_n = 2$, то она сведется к определению оптимальной последовательности чисел $m_1, m_2, m_3, \dots, m_k$, к которым должно стремиться число разведочных элементов первого, второго, \dots , k -го уровня сгущений.

Выбрать критерий оптимальности для этой задачи нелегко. Он должен обеспечивать в некотором смысле наилучшее соотношение между самой точностью и точностью ее оценки при заданном числе скважин, т. е. при фиксированных затратах. В первом приближении такой критерий можно выбрать исходя из следующих соображений. Точность запасов можно охарактеризовать доверительным интервалом с заданной вероятностью. При разведке желательно уменьшение доверительного интервала, поэтому критерием предлагается выбрать минимум суммы

$$s_{\bar{0}}^2 + s_{s\bar{0}}^2 \rightarrow \min, \quad (8.6.3)$$

а процесс оптимизации нам представляется лишь в самых общих чертах, преимущественно путем перебора вариантов.

ЛИТЕРАТУРА

- Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Основы моделирования и первичная обработка данных. М., 1983. 471 с. (Справочное издание).
- Америка Л. Д., Николаевский Н. М. К вопросу оптимизации промышленной разведки нефтяных месторождений на стадии проектирования // Экон. нефт. пром-сти. 1972. № 2. С. 3—9.
- Америка Л. Д., Николаевский Н. М., Кочетова Л. М. Учет некоторых геолого-географических факторов при обосновании минимально допустимых объемов разведки нефтяного месторождения // Экон. нефт. пром-сти. 1976. № 6. С. 3—8.
- Америка Л. Д., Черницкий А. В. К вопросу определения рациональных объемов разведочных работ при допустимой точности подсчета запасов нефти // Экон. нефт. пром-сти. 1971. № 8. С. 5—8.
- Аронов В. И., Аракелян В. А., Жуков Н. И., Френкель М. А. Оптимизация размещения скважин при разведке залежей нефти и газа на основе использования ЭВМ. М., 1982. 50 с. (ВИЭМС).
- Барышев Н. В. Контроль опробования // Материалы ВКЗ по методике разведки и подсчета запасов. М., 1948. Вып. 2. С. 88—105.
- Барышев Н. В., Красников В. Н. К методике экспериментального исследования вопроса о рациональном сокращении проб // Тр. МГРИ, 1936. Т. 2. С. 22—37.
- Батурин Ю. Н. О некоторых особенностях расчета ошибок среднего значения пористости при подсчете запасов нефти объемным методом // Оценка точности определения параметров залежей нефти и газа. М., 1965а. С. 114—119.
- Батурин Ю. Н. Расчет средней величины пористости пласта по совокупности керновых и промыслово-геофизических данных // Оценка точности определения параметров залежей нефти и газа. 1965б. С. 122—128.
- Белов А. В. О выборе способа осреднения исходных данных при оценке коллекторских свойств неоднородных пластов по керну // Нефтегазовая геология и геофизика. 1969. № 4. С. 43—49.
- Белов А. В. Оценка погрешности одного из способов определения средней пористости пласта в условиях значимой корреляции с эффективной мощностью / Тр. н.-и. и проект. ин-та «Гипрогаз», 1972. Вып. 30. С. 76—83.
- Бирюков В. И. Регрессия категорий запасов // Изв. Вуз. Геол. и разв. 1965. № 4. С. 86—93.
- Бирюков В. И. Методика разведки полезных ископаемых за 50 лет // Сборник научных статей Всесоюзного научно-исследовательского института экономики минерального сырья и геолого-разведочных работ. М., 1967. С. 19—45.
- Борисенко З. Г., Сосон М. Н. Оценка подтверждаемости запасов промышленных категорий // Геология нефти и газа. 1975. № 3. С. 69—72.
- Бортницкая В. М. Об оценке емкости коллекторов лабораторными методами: Особенности формирования пород-коллекторов впадин запада Русской платформы. Минск, 1977. С. 135—136.
- Быков Н. Е., Америка Л. Д., Черницкий А. В. Поваринтное проектирование разведки многопластных нефтяных месторождений. М., 1978. 184 с.
- Быков Н. Е., Егоров Р. А., Николаевский М. Н. и др. О точности определения параметров залежей при экономической оценке разведки нефтегазовых месторождений // Геология нефти и газа. 1972. № 11. С. 16—23.
- Васильев В. Б. Управление процессом разведки нефтяных и газовых месторождений. М., 1980. 175 с.
- Васильев В. Б., Молчанова Н. Г. К вопросу минимизации количества скважин, необходимых для оценки параметров нефтяной залежи / Тр. ЗапСибНИГНИ, Тюмень, 1977. Вып. 123. С. 19—27.
- Васильев С. П. Метод экономической эквивалентности в определении частоты буровых скважин на уголь в Подмосковном бассейне // Разведка недр. 1950. № 1. С. 29—34.
- Воробьев В. Я., Элланский М. М. Построение иерархических систем геологических объектов // Математические методы в геологии. Саратов, 1979. № 2. С. 17—24.

Габриэлянц Г. А. Пути оптимизации разведки нефтяных и газовых месторождений // Тр. ВНИГРИ, 1979. № 213. С. 3—16.

Габриэлянц Г. А., Карпушин В. З., Мясников В. И., Сердюков В. В. Методика определения оптимального количества разведочных скважин // Геология нефти и газа. 1976. № 3. С. 22—27.

Габриэлянц Г. А., Карпушин В. З., Пороскун В. И. Принципы изучения контакта газ—вода (нефть—вода) при разведке массивных залежей // Методика разведки массивных залежей нефти и газа. М., 1978. С. 115—124. (Тр. ВНИГНИ; Вып. 197).

Габриэлянц Г. А., Пороскун В. И. Методика определения рациональной сетки скважин при разведке массивных залежей // Геология нефти и газа. 1974. № 8. С. 61—66.

Гавришин А. И. Оценка и контроль качества геохимической информации. М., 1980. 287 с.

Гавришин А. И., Вострокнутов Г. А. Исследование качества приближенно-количественного спектрального анализа по данным междублабораторного контроля // Геохимические методы при поисках и разведках рудных месторождений. М., 1971. Вып. 2. С. 88—122.

Головастов Д. С., Свищев М. В. О влиянии площади поднятия на плотность сетки разведочных скважин // Геология нефти и газа. 1972. № 6. С. 36—41.

Гольдфельд А. И., Стадников Г. П. Определение оптимальной плотности сети при предварительной разведке месторождений лентовидной формы // Разведка и охрана недр. 1964. № 8. С. 14—24.

Гончаренко Э. А. Математический анализ разведочной сети нефтяных месторождений // Методика геометризации нефтяных и газовых залежей при их разведке и разработке с применением современных методов и средств вычислительной техники. Пермь, 1969. (Тр. Пермского совещания).

Грудев А. П. Статистическая связанность рудных тел // Изв. вузов. Геология и разведка. 1961. № 1. С. 78—82.

Гудошников С. С., Югин Л. Г. Некоторые методические особенности подсчета запасов нефти в зонально и вертикально неоднородных пластах. Куйбышев, 1977. Вып. 29. С. 123—127.

Гусинский М. Н., Заводская Н. Н., Макулов Н. А. Исследование возможности повышения чувствительности и правильности результатов атомно-абсорбционного пламенного спектрального анализа с помощью методов математической статистики // VIII Сибирское совещание по спектроскопии: Атомная спектроскопия и спектральный анализ. Иркутск, 1972. С. 43—44. (Тез. докл.).

Давтян Ю. А. Принципы выделения объектов оценки для количественного описания изменчивости оруденения // Изв. вузов. Геология и разведка. 1978. № 11. С. 96—99.

Дементьев Л. Ф., Азаматов В. И. Об определении средней пористости и нефтенасыщенности при подсчете запасов и проектировании разработки // Новости нефтяной и газовой техники. М., 1961. № 1. С. 22—25.

Дементьев Л. Ф., Хитров Е. А., Дурьшанова Г. А. О предельных объемах исследований при изучении нефтяных пластов // Геология и разработка нефтяных месторождений. Пермь, 1966. С. 24—34.

Дементьев Л. Ф., Хитров Е. А., Шурубор Ю. В. Применение информационных мер в нефтепромысловой геологии // Тр. ПермНИПИнефть, Пермь, 1974. Вып. 10. 155 с.

Демин В. И., Конторович А. Э. Применение математических методов и ЭВМ при подсчете запасов нефти и газа промышленных категорий // Геология и геофизика. 1973. № 10. С. 3—13.

Денисов М. Н. Методика определения необходимой и достаточной плотности разведочной сети на примере пластообразных месторождений // Изв. вузов. Геология и разведка. 1963. № 7. С. 68—77.

Денисов М. Н., Куличихин С. Н., Павлов В. К. Штокверковые месторождения и рациональная методика их предварительной разведки. М., 1970. 110 с. (Обзор ВИЭМС. Сер. Геология и методы поисков и разведки месторождений цветных, редких и благородных металлов).

Добровольский М. Б., Володин И. Н. Применение математических методов определения рациональной степени разведанности нефтяных и газовых месторождений при передаче из разведки в разработку. М., 1974. 62 с. (Тем. науч.-техн. обзор / ВНИИОЭНГ. Сер. Нефтегазовая геология и геофизика).

Долженков В. Н., Ахметов Н. Г., Лиходедова С. А. О подтверждении величины и структуры запасов нефти небольших месторождений Татарии по данным эксплуатационного бурения // Труды Татарского научно-исследовательского и проектного института нефтяной промышленности, 1978. № 38. С. 73—76.

Дубов Р. И. Дисперсионный анализ ошибок измерений // Спектральный анализ элементов-примесей в горных породах. М., 1972. С. 57—68.

Дубов Р. И., Рабинович И. М., Тонкопий М. С., Фурсов В. В. Оценка погрешностей измерений при анализе геохимических проб // Разведка и охрана недр. 1970. № 7. С. 21—25.

Дубов Р. И., Фурсов В. В. Соотношение между параметрами статистического распределения химических элементов в коренных горных породах и мелкой фракции элювия-делювия // Математические методы геохимических исследований. М., 1966. С. 82—91.

Емельянов Н. Н. Методика определения числа и рационального размещения скважин и объемов исследований, необходимых для изучения параметров при промышленной разведке / Автореф. дис. ... канд. геол.-минер. наук. М., 1966.

Емельянов Н. Н., Кочетов М. Н. К вопросу обоснования необходимого количества образцов при исследовании пористости пород по кернам в разрезах скважин / Тр. ВНИИ. М., 1966. Вып. 45. С. 123—133.

Жданов М. А., Лисунов В. Р., Гришин Ф. А. Методика и практика подсчета запасов нефти и газа. М., 1967. 403 с.

Закс Л. Статистическое оценивание. М., 1976. 598 с.

Зенков Д. А. Об основных свойствах разведочных сеток // Разведка недр. 1950. № 4. С. 16—21.

Зенков Д. А. Развитие разведочной сети // Разведка недр. 1951. № 1. С. 15—22.

Зенков Д. А. Методы определения плотности разведочной сети // Сов. геология. 1957. № 61. С. 130—143.

Зенков Д. А. Изменение статистических показателей содержаний компонентов в связи с методами исследования при разведке // Разведка и охрана недр. 1961. № 9. С. 13—20.

Иванникова Л. С., Мандрыкин И. Н. Методика анализа подтверждаемости запасов нефти и газа низших категорий при переводе их в высшие // Геология нефти и газа. 1972. № 7. С. 72—77.

Иванов В. В. О системе приближенных оценок средних содержаний химических элементов в геологических образованиях // Математические методы в геологии. М., 1968. С. 48—53.

Иванов М. П., Петров В. А., Стражевич А. И. Количественная оценка сложности строения рудных объектов и разведанности запасов. М., 1981. 46 с. (Геол. методы поисков и разведки м-ний метал. полез. ископаемых. Обзор / ВНИИ экон. минер. сырья и геол.-развед. работ. ВИЭМС).

Иванова Т. И. Изучение вида распределения случайных погрешностей спектрального анализа // Геология и полезные ископаемые Северо-Востока европейской части СССР: (Ежегодник-1975). Сыктывкар, 1976. С. 136—141.

Иванова Т. И., Ткачев Ю. А. Построение градуировочных графиков в спектральном анализе с учетом априорной информации // Прикладная спектроскопия. Минск, 1982. Т. 37. Вып. 4. С. 558—564.

Инструкция по применению классификации запасов к месторождениям нефти и горючих газов (1960 года). М., 1961. 63 с.

Инструкция по применению классификации запасов к месторождениям нефти и горючих газов (1971 года). М., 1972. 64 с.

Инструкция по применению классификации запасов месторождений, перспективных и прогнозных ресурсов нефти и горючих газов (1983 года). М., 1984. 65 с.

Иоффе О. П. Вычисление среднего значения пористости продуктивного пласта методом статистических испытаний // Тр. ВНИИ, М., 1968. Т. 52. С. 134—136.

Ишаев У. Г. К вопросу о рациональной методике промышленной разведки нефтяных месторождений в Шаимском районе Западной Сибири. Тюмень, 1970. Вып. 10. С. 3—10.

Каган Л. М. О математической обработке данных спектрального анализа с учетом его точности // Методы анализа вещественного состава горных пород и вод при геохимических исследованиях. Минск, 1978а. С. 96—105.

Каган Л. М. Оценка случайной ошибки анализа с учетом вклада градуировки // Материалы геологического изучения земной коры Белоруссии. Минск, 1978б. С. 40—43.

Казаковский Д. А. К вопросу о влиянии интервала между скважинами и способа их расположения на величину ошибки в подсчете запасов / Тр. ЦНИМБ, 1937. № 6.

Казаковский Д. А. Оценка точности результатов в связи с геометризацией и подсчетом запасов месторождений. М., 1948.

Казаковский Д. А. О применении формул статистики к оценке точности подсчета запасов месторождений // Исследования по вопросам горного и маркшейдерского дела. М., 1957, сб. 31. С. 139—148.

Казаковский Д. А. О характеристике изменчивости залежи полезного ископаемого. Колыма, 1959, № 5. С. 32—34.

Калабин Ю. Я., Канов Г. В., Кирсанов А. Н., Кирсанова Н. С. К методике расчета средних значений пористости для построения карт: (на примере Медвежьего месторождения) / Тр. ВНИИ экон., организ. пр-ва и техн.-экон. информ. в газ. промышленности, 1975. С. 32—36.

Каллистов П. Л. По поводу статьи К. Л. Пожарицкого «К вопросу о рационализации обработки проб» // Зав. лаб. 1950. № 4.

Каневский М. П., Кочев Н. И. Опыт применения методов математической статистики для обоснования оптимального объема опробования пород. М., 1976, с. 290—296. (Тезисы 3-го межведомс. совещ. по вопросу прогноза гидрогеол., инж.-геол. и почв.-мелиор. условий; Вып. 4).

Карелина А. А. О методах оценки среднего значения пористости // Оценка точности определения параметров залежей нефти и газа. М., 1965. С. 120—121.

Карлье Э. Количественная оценка месторождений урана. М., 1966. 351 с.

Карпушин В. З. Оценка оптимальных объемов отбора керна в терригенных продуктивных разрезах большой мощности / Тр. ВНИГРИ, 1979. № 213. С. 113—125.

Кельин А. М., Михайлович М. П. Способ вычисления систематической ошибки химических анализов // Разведка и охрана недр. 1969. № 7. С. 23—25.

К методике расчета средних значений пористости для построения карт: (на примере Медвежьего месторождения) / Ю. Я. Калабин, Г. В. Канов, А. Н. Кирсанов, Н. С. Кирсанова / Тр. ВНИИ. Экон., организ. произ-ва и техн.-экон. информ. в газовой промышленности, 1975. С. 32—36.

Кноринг Л. Д., Деч В. Н. Оптимизация объемов работ при разведке на нефть и газ // Сов. геология. 1973. № 6. С. 92—101.

Кноринг Л. Д. Метод поиска оптимальных объемов работ при разведке пластовых сводовых залежей газа и нефти / Тр. ВНИГРИ, 1977. № 395. С. 16—23.

Кноринг Л. Д. Основы теории оптимизации разведки нефтяных месторождений: (в аспекте оценки месторождений разведочными скважинами). Л., 1980. 304 с.

Ковальчук Н. Р. Методы оптимизации процесса подготовки нефтяных и газовых месторождений к промышленному освоению / Автореф. дис. ... д-ра геол.-минер. наук. Львов, 1977. 38 с.

Ковальчук Н. Р., Предтеченская Н. С. Определение рационального количества скважин в процессе разведки нефтяных и газовых месторождений // Геология нефти и газа. 1976. № 3. С. 27—32.

Козлов А. Л., Соломахин В. И., Фриман Ю. М. Влияние числа и расположения скважин на точность определения геологопромысловых параметров // Газовая промышленность. 1970. № 7. С. 10—13.

Кокрен У. Методы выборочного исследования. М., 1976. 440 с.

Комаров В. Л. Определение средней пористости объекта по данным выборочного метода / Тр. УФНИИ, 1961а. Вып. 8. С. 218—236.

Комаров В. Л. Статистическая оценка коллекторских параметров песчаных пластов / Там же, 1961б. С. 237—249.

Кондин С. Р., Штифанов П. И. Экономическое обоснование допустимых размеров затрат на разведку месторождений / Тр. САИГИМС, 1965. Вып. 6. С. 157—161.

Королев В. П. Об обработке результатов анализов внешнего контроля разведочных проб // Разведка и охрана недр. 1969. № 7. С. 22—23.

Котляр Н. А. О теоретических обоснованиях в деле взятия и сокращения проб полезных ископаемых / Тр. I Всесоюз. горн. науч.-техн. съезда, 1926. Т. 8.

Кочетов М. Н., Власенко В. В., Молотова Н. А. Анализ причин изменения цифр запасов нефти при пересчетах на примере залежей Мегионского и Ватинского месторождений // Нефтепромысловая геология. М., 1974. С. 154—167.

Краснов Д. А. Закономерности сокращения и определение предельного веса проб // Материалы по методам разведки полезных ископаемых. М., 1962. С. 241—243.

Краснов Д. А. Теоретическое определение минимального веса химических проб // Зап. ЛГИ, 1963. Т. 42, вып. 3. С. 121—127.

Краснов Д. А. Теоретические основы и расчетные формулы определения веса проб. М., 1969. 126 с.

Крейтер В. М. Поиски и разведка месторождений полезных ископаемых 1961. Ч. 2. М., 386 с.

Крейтер В. М. Поиски и разведка месторождений полезных ископаемых. М., 1961. Ч. 2. 386 с.

Кречетова И. П. Некоторые задачи теории площадного опробования. М., 1974а.

Вып. 26. С. 35—46. (Тр. Проект. и НИИ по инж. изыск. в стр-ве Госстроя СССР).

Кречетова И. П. Применение минимаксного принципа планирования в теории опробования / Там же, 1975б. С. 23—34.

Кречетова И. П., Рац М. В. Решение задач площадного опробования слоистой толщи // Бюл. МОИП. Отд. геол. 1973. Т. 48, № 1. С. 141—142.

Крылова Г. Г., Казьмин Л. Л., Алексеев Г. И. Подтверждаемость запасов категории C_2 при переводе в высшие категории: (на примере Куйбышевской области) // В кн.: Проблемы геологии нефти. М., 1975. Вып. 5. С. 259—262.

Кудряшов П. И., Кузьмин В. И. Геометризация и учет запасов месторождений твердых полезных ископаемых. М., 1981. 275 с.

Кузнецов Ю. И., Зайцев Б. В., Муртазин Т. Т., Цоцур В. С. Геолого-экономическая оценка ресурсов и подтверждаемость запасов нефти и газа категорий C_1 и C_2 в Башкирии // Геология и разработка нефтяных месторождений востока Волго-Уральской провинции. М., 1975. С. 124—126.

Кузьмин В. И. Условия применения формул средневзвешенного при подсчете запасов // Разведка и охрана недр. 1966. № 2. С. 11—19.

Кузьмин В. И. Геометризация и подсчет запасов месторождений твердых полезных ископаемых. М., 1967. 242 с.

Кузьмин В. И. К теории эффективного применения формул среднего арифметического и среднего взвешенного при подсчете запасов твердых полезных ископаемых // Применение математических методов в геологии, 1968. С. 224—228.

Кузьмин В. И. Конструирование и исследование двумерных цифровых моделей на ЭВМ // Изв. вузов. Горн. журн. 1972а. № 12. С. 55—58.

Кузьмин В. И. Построение на ЭВМ вероятностных номограмм для эффективного применения средневзвешенных и среднеарифметических оценок подсчета запасов // Изв. вузов. Горн. журн. 1972б. № 8. С. 39—45.

Кузьмин В. И. Эффективное применение средневзвешенных и среднеарифметических оценок при подсчете запасов и оценке геологоразведочной информации // Геология и геофизика. 1972в. № 7. С. 74—81.

Кузьмин В. И. Некоторые вопросы теории и методики оценок параметров запасов // Применение математических методов и ЭВМ при решении типовых геологических задач. Новосибирск, 1976. С. 68—83.

Лончих С. В. К оценке точности спектро-физического анализа металлотрических проб // Науч. тр. ИРГиредмет, 1965. Вып. 12. С. 3—13.

Макулов Н. А. Оптимальное построение калибровочной кривой и оценка случайной погрешности анализа // Прикладная спектроскопия. Минск, 1974. С. 277—290.

Макулов Н. А., Славный В. А., Кузнецов Ю. Н. Статистическая интерпре-

тация результатов спектроскопических измерений // Прикладная спектроскопия. Минск, 1974. С. 291—300.

Митрофанов В. П., Тульбович Б. И. Дифференцированный подход к оценке коллекторских свойств карбонатных пород // Геология нефти и газа. 1974. № 2.

Мамедов Т. Г., Мамедова К. Д., Мадатова Э. А. О точности определения коллекторских свойств пород по каротажу // Тематический сборник научных трудов Азербайджанского научно-исследовательского и проектного института нефтяной промышленности, 1979, № 48. С. 115—122.

Марголин А. М. Методика учета систематических погрешностей пробоотбора и химических анализов при геометризации и подсчете запасов месторождений полезных ископаемых // Совершенствование методов маркшейдерских работ и геометризации недр. М., 1972. С. 279—294.

Марголин А. М. Оценка запасов минерального сырья: Математические методы. М., 1974. 264 с.

Марсал Д. Выделение ошибок из технических данных о резервуаре // Исследования по математической геологии. Л., 1978. С. 99—104.

Матерон Ж. Основы прикладной геостатистики. М., 1968. 407 с.

Мелик-Пашаев В. С., Бреев В. А., Воронова Г. Т., Кочетов М. Н. Анализ подтверждаемости запасов нефти категорий C_1 и C_2 при переводе их в высшие категории: Нефтегазоносность, геология и геофизика / Научно-технический сборник, 1975. № 6. С. 28—31.

Мельников Г. А. Опыт определения густоты бети опробования методами вариационной статистики // Разведка недр. 1934. № 4.

Методика определения рациональных пределов разведки залежей нефти и газа / А. С. Гончаренко, Э. А. Гончаренко, А. А. Гончаренко, В. В. Малышева // Геология, поиски и разведка месторождений горючих полезных ископаемых. Пермь, 1975. Вып. 1. С. 219—224.

Мончак Л. С. Определение необходимого и достаточного количества анализов пористости для подсчета запасов нефти и газа // Разведка и разработка нефтяных и газовых месторождений. М., 1972. Вып. 9. С. 34—36.

Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. М., 1982. 272 с.

Налимов В. В. Применение математической статистики при анализе вещества. М., 1960. 430 с.

Нелепченко А. Е., Горелова Т. Л. Оценка коллекторских свойств при изучении геологической неоднородности: (на примере Медвежьего месторождения) // Геология и разработка и бурение газовых и газоконденсатных месторождений Сибири. М., 1975. Вып. 8. С. 19—22.

Ованесов Г. П., Фейгин М. В. К вопросу о подтверждаемости запасов нефти категорий C_1 и C_2 // Геология нефти и газа. 1975. № 8. С. 7—14.

Огнев А. Ф. Обоснование оптимальных пределов изучения месторождений нефти широтного Приобья // Тезисы докладов IV научно-технической конференции молодых ученых и специалистов ЗапСибНИГНИ. Тюмень, 1979. С. 60—61.

Оптимизация разведки нефтяных месторождений / Е. Ф. Фролов, Н. Е. Быков, Р. А. Егоров, А. Я. Фурсов. М., 1976. 303 с.

Орлов А. Г. Методы расчета в количественном спектральном анализе. 1977. 223 с.

Орлов А. Г. О сравнении экспериментальных данных двух литературных источников статистическими методами // Заводская лаборатория. 1978. 44, № 7. С. 852—854.

Осецкий А. И. Показатель соответствия густоты разведочных точек характеру разведваемой залежи / Автореф. дис. ... канд. техн. наук. Днепрпетровск, 1953.

Оспанов Е. С., Аралбаев Ю. Б., Яременко А. Д. Определение плотности сети буровых скважин математическим методом: (на примере стратиформного месторождения Шалкия) // Изв. АН КазССР. Сер. геол. 1975. № 3. С. 69—74.

Пастушенко И. И. Об оптимальной плотности разведочной сети на железорудных месторождениях сложной формы // Разведка и охрана недр. 1963. № 7. С. 16—22.

Петров А. А. Применение математической статистики для решения основных вопросов разведки месторождений / Сов. геология. 1963. № 9. С. 132—141.

Погребницкий Е. О. Поиски и разведка месторождений полезных ископаемых. М., 1968. 460 с.

Пожарицкий К. Л. Теория и практика сокращения проб // Цветные металлы. 1935. № 6.

Пожарицкий К. Л. Пути увеличения эффективности геологоразведочных работ и борьба с излишествами в них // Сов. геология. 1958. № 1. С. 127—144.

Прокофьев А. П. Приемы обработки контрольных химических анализов // Разведка и охрана недр. 1955. № 3. С. 25—29.

Прокофьев А. П. О внутреннем и внешнем контроле анализов разведочных проб // Разведка недр. 1962. № 6. С. 6—10.

Раевский В. И. Применение последовательного анализа при разведке месторождений полезных ископаемых / Тр. Пермск. политех. ин-та, 1966, сб. 20. С. 281—286.

Раевский В. И., Медведев Н. С. Методы изменчивости свойств пластовых месторождений. Л., 1976. 100 с.

Раевский В. И., Щурубор Ю. В. Обработка данных внешнего контроля химических анализов геологических проб // Изв. вузов. Геол. и разв. 1958. № 11. С. 63—69.

Рац М. В. Неоднородность горных пород и их физических свойств. М., 1968. 107 с.

Родионов Д. А. О внешнем контроле результатов анализов при подсчете запасов рудных месторождений // Разведка и охрана недр. 1964. № 5. С. 14—17.

Родионов Д. А., Иванов В. В. Статистические оценки средних содержаний по со-

вокупности разной представительности // Геохимия. 1967. № 1. С. 109—117.

Родченко Ю. М. Анализ плотности разведочной сети // Сов. геология. 1964. № 7. С. 52—62.

Родыгин В. Р., Сливков В. И. Пространственное размещение и формирование пород-коллекторов Вуктыльского газоконденсатного месторождения // Особенности формирования пород-коллекторов впадин запада Русской платформы. Минск, 1977. С. 114—115.

Рошин Ю. В. Об использовании уравнения регрессии с целью выявления систематических ошибок при сопоставлении рядов основных и контрольных измерений // Вопросы рудной радиометрии. М., 1968а. С. 108—119.

Рошин Ю. В. Способы конъюнктного анализа и выявление систематических ошибок по сопоставлению рядов измерений // Вопросы рудной радиометрии. М., 1968б. С. 120—132.

Рошин Ю. В. Основные направления развития прикладной геостатистики к решению геологоразведочных и горнотехнологических задач. М., 1983. 105 с. (МГРИ. Инф. вып.; № 56 (248)).

Саввин В. А., Бадоев Т. И., Коростышевский М. Н. Об оптимальном объеме керновой информации для достоверной оценки пористости при подсчете запасов нефти и газа // Геология нефти и газа. 1975. № 6. С. 66—69.

Савинский И. Д. Таблица вероятностей подсечения эллиптических объектов прямоугольной сетью наблюдений. М., 1964. 86 с. (ВИЭМС).

Сараева Г. Д. К методике оценки погрешности измерения открытой пористости терригенных пород // Вопросы газонасыщенности европейской части СССР и среднеазиатских районов. М., 1976. С. 74—80.

Сенюков Р. В. О модели минимизации стоимости разведочных работ при определении запасов нефтяных и газовых месторождений // Экономика нефт. пром-сти. 1972. № 5. С. 10—15.

Сенюков Р. В. Применение математических методов и ЭВМ для оценки перспектив нефтегазодосных регионов. М., 1982. Ч. II. 98 с. (МИНХ и ГП).

Слепцов Б. Г. Двухэтапное оценивание средних значений показателей свойств грунтов с учетом концентрации проб в выработках. М., 1974а. Вып. 26. С. 68—77. (Тр. Проект. и НИИ по инж. изыск. в стр-ве Госстроя СССР).

Слепцов Б. Г. Методы последовательного оценивания при определении необходимого числа наблюдений // Бюл. МОИП. Отд. геол. 1974б. т. 49, № 1. С. 151—152.

Смирнов В. И. и др. Подсчет запасов месторождений полезных ископаемых. М., 1960. 672 с.

Смирнова Е. В., Лувсандорж М., Райхбаум Я. Д., Видинеева Н. В. К теории градуировочных графиков при спектральном анализе. Иркутск, 1973. С. 455—460.

Соловьев В. Г. Вариационная статистика в применении к разведке и подсчету запасов полезных ископаемых // Разведка недр. 1938. № 1.

Соловьев В. Г. Методы вариационной статистики в применении к разведке и подсчету запасов месторождений полезных ископаемых. М., 1939.

Теория и практика разведки месторождений нефти и газа / И. И. Нестеров, В. Б. Васильев, А. М. Волков и др. М., 1985. 215 с.

Тихомиров Ю. П., Огнев А. Ф., Стовбун Ю. А. Оптимальные пределы разведанности нефтяных и газоконденсатно-нефтяных месторождений Западной Сибири / Тр. Зап. СибНИГНИ, Тюмень, 1978. Вып. 134. С. 115—132.

Ткачев Ю. А. Оценка экономической эффективности использования результатов научно-исследовательских работ в геологии / Сер. препринтов-сообщений «Новые научные методики». Сыктывкар.

Ткачев Ю. А. Оценка параметров и их погрешностей при подсчете запасов нефти и газа / Сер. препринтов-сообщений «Научные рекомендации — народному хозяйству». Сыктывкар. 1979. Вып. 18. 54 с.

Ткачев Ю. А. Статистическая обработка и интерпретация геохимических данных в условиях логнормального распределения погрешностей анализа // Физические методы и математическая обработка анализов горных пород и минералов. Сыктывкар, 1981. С. 59—73. (Тр. ИГ Коми фил. АН СССР; Вып. 36).

Ткачев Ю. А. Повышение надежности оценок средних в геологии без увеличения выборки. Сыктывкар, 1973. С. 177—183.

Ткачев Ю. А. Анализ сравнительной эффективности оценок средних при подсчете запасов нефти // Геология и нефтегазоносность Тимано-Печорской провинции. Сыктывкар, 1975. С. 114—123.

Ткачев Ю. А. Подсчет запасов минерального сырья на месторождениях с гнездовым распределением полезного ископаемого / КФАН СССР. Сер. препринтов-сообщений «Научные рекомендации — народному хозяйству», 1976. Вып. 10. 65 с.

Ткачев Ю. А. Оптимальна ли равномерная сеть опробования при разведке месторождений // Геология и полезные ископаемые Северо-Востока европейской части СССР: Ежегодник-1976. Сыктывкар, 1977. С. 149—156.

Ткачев Ю. А. Применение дисперсионного анализа при обработке геологоразведочных данных и подсчете запасов нефтяных и газовых месторождений // Геология, разведка и разработка газовых и газоконденсатных месторождений Тимано-Печорской провинции. Сыктывкар, 1978. С. 74—86.

Ткачев П. В. Анализ изменчивости коллекторских свойств продуктивных горизонтов крупных месторождений северных районов Урало-Поволжья / Тр. ВНИГРИ, 1976. Вып. 191. С. 79—84.

Ткачев Ю. А. Вывод формулы сокращения проб // Изв. АН Кирг. ССР. Сер. естеств. и тех. наук. 1960, Т. 2, вып. 8. С. 127—140.

Ткачев Ю. А. Составление надежных схем обработки проб // Изв. АН Кирг. ССР. Сер. естеств. и тех. наук. 1962. Т. 4, вып. 3. С. 113—131.

Ткачев Ю. А. Экспериментальное подтверждение формулы надежного сокращения проб // Изв. АН Кирг. ССР. Сер. естеств. и тех. наук. 1964. Т. 6, вып. 3. С. 99—108.

Ткачев Ю. А. Исследования вопросов достоверности отбора и обработки проб / Тр. САИГИМС, 1965. Вып. 6. С. 98—105.

Ткачев Ю. А. Влияние деятельности разведки на цену запасов в недрах // Определение стоимости оценки запасов минерального сырья в недрах. Берлин, 1971. (Матер. науч.-техн. семинара СЭВ), 1985. Вып. 14. 39 с.

Ткачев Ю. А., Брезгунов Н. И., Брезгунов А. Г. Д. О сравнительной эффективности валового и объемно-газодового способов подсчета запасов на месторождениях с гнездовым оборудованием // Разведка и охрана недр. 1970а. № 9. С. 12—16.

Ткачев Ю. А., Брезгунов Н. И., Брезгунов А. Г. Д. Определение объема валового опробования, обеспечивающего минимальную сумму возможных экономических потерь, связанных с разведкой и риском неправильной оценки месторождений / Тр. ВНИИСИМС, 1970б. Т. 13. С. 113—119.

Ткачев Ю. А., Кетрис М. П. Методика оценки погрешностей средней пористости и проницаемости и обоснование обеспеченности керном при разведке нефтяных и газовых месторождений / Тр. ИГ Коми фил. АН СССР. Сыктывкар, 1978. № 27. С. 118—124.

Ткачев Ю. А., Морозов В. А., Лобанов В. К. Подсчет запасов на некоторых типах жильных месторождений горного хрусталя с использованием корреляции между параметрами // Изв. вузов. Геол. и разв. 1971. № 4. С. 59—67.

Ткачев Ю. А., Шейн А. А. Обработка проб полезных ископаемых. М., 1987.

Ткачев Ю. А., Юдович Э. Э. Статистическая обработка геохимических данных. Л., 1975. 236 с.

Урбах В. Ю. Математическая статистика для биологов и медиков. М., 1963. 323 с.

Усиков Ю. Т. Математические методы оценки точности измерений в геологии. М., 1983. 43 с. (ВИЭМС).

Фурсов А. Я. Ошибки значений пористости и нефтенасыщенности, полученных с применением классификации терригенных коллекторов // Тр. ВНИИ. М., 1968. Т. 52. С. 99—105.

Фурсов А. Я. Методика обоснования рациональных объемов бурения при разведке пластовых залежей нефти // Геология нефти и газа. 1979. № 5. С. 42—45.

Хрущов Н. А. Основные показатели экономической эффективности затрат на геологоразведочные работы и подход к определению цены разведанных запасов минерального сырья в недрах // Сов. геология. 1967. № 8. С. 52—60.

Хрущов Н. А. Основные этапы развития и результаты научных исследований в области методики разведки и геолого-экономи-

ческой оценки месторождений полезных ископаемых // Методика разведки и оценки месторождений твердых полезных ископаемых. М., 1978. С. 3—11.

Цирульников В. П. Анализ погрешностей определения геофизических параметров по результатам многократных замеров в скважине // Разведочная геофизика. 1980. № 88. С. 143—148.

Черницкий А. В. Опыт применения количественных требований при анализе разведанности нефтяных месторождений Западной Сибири // Геометризация месторождений полезных ископаемых. М., 1977. С. 330—334.

Четвериков Л. И. Теоретические основы моделирования тел твердых полезных ископаемых. Воронеж, 1968. 152 с.

Чечотт Г. О. Опробование и испытание полезных ископаемых. М., 1932.

Шаκлеин С. В. Оценка достоверности разведанных запасов угля // Изв. вузов. Горн. журн. 1977. № 10. С. 46—47.

Шаκлеин С. В. Уравнение сети замеров мощности на угольных месторождениях // Изв. вузов. Горн. журн. 1979. № 4. С. 39—40.

Шарапов И. П. О контрольных анализах геологических проб // Разведка и охрана недр. 1954. № 1. С. 16—27.

Шашкин В. Л. О контрольном опробовании и контрольных анализах геологических проб / Тр. Ин-та геол. АН Кирг. ССР, 1956. Вып. 7. С. 111—124.

Шевелев В. В. О плотности разведочной сети при детальной разведке глиен центральной части Иркутского угольного бассейна // Изв. вузов. Геол. и разв. 1967. № 3. С. 77—84.

Шехтман П. А. Принципы и методы разведки эндогенных месторождений. Ташкент, 1968. 94 с.

Шишигин С. И. Ошибки определения запасов нефти объемным методом, обусловленные погрешностями подсчетных параметров коллекторских свойств / Тр. Зап.-Сиб. НИГНИ, 1975. Вып. 90. С. 113—115.

Шустеф И. Н., Тульбович Б. И., Митрофанов В. П., Викторин В. Д. Учет структурно-генетических особенностей карбонатных коллекторов при подсчете запасов нефти // Геология нефти и газа. 1973. № 12. С. 23—25.

Юфа Б. Я. Учет систематических ошибок при подсчете запасов // Разведка недр. 1951. № 6. С. 23—28.

Юфа Б. Я. Замечания по статье Шаропова И. П. «О контрольных анализах геологических проб (Разведка и охрана недр, 1954, № 1)» // Разведка и охрана недр. 1954. № 5. С. 60—63.

Юфа Б. Я. Об ошибках В. Л. Шашкина в решении некоторых вопросов контроля результатов опробования / Тр. Ин-та геол. АН Кирг. ССР, 1958. Вып. 10. С. 193—200.

Юфа Б. Я. Метрологическое обеспечение качества работ при региональных геохимических исследованиях: (Временные методические рекомендации). Л., 1979. 54 с. (ВСЕГЕИ).

Ящук Г. Н. К вопросу о представительности выборки, обеспечивающей необходимую точность оценки пористости коллектора / Тр. ВНИГРИ, 1974. Вып. 156. С. 107—113.

Vandemer H. Die für eine vorgegebene Genauigkeit notwendige Bohrlochzahl bei quantitativer Auswertung der Ergebnisse // Zeitsch. angew. Geol. 1966. Bd 12, H. 3. S. 122—126.

Vandemer H. Gedanken zur Schätzung von Lagerstättenparametern // Zeitsch. angew. Geol. 1967a. Bd 13, H. 3. S. 120—123.

Vandemer H. Zur Optimierung der Erkundung einer Lagerstätte // Zeitsch. angew. Geol. 1967b. Bd 13, H. 5. S. 253—255.

Bartlett M. S. Fitting a straight line when both variables are subject to error // Biometrics. 1948. Vol. 5. P. 207—212.

Brooker P. I. Optimal block estimation by kriging // Austr. IMM. Proc. 1975. N 253. P. 15—20.

Brooker P. I. Block selection according to deposit variability // Austr. IMM. Proc. 1976. N 257. P. 33—36.

Brooker P. I. Geostatistical calculations using true and estimated values from a simulated deposit // Austr. IMM. Proc. 1978. N 268. P. 63—70.

Clark I. Practical kriging in three dimensions // Comp. and Geosc. 1977. 3 (1). P. 173—180.

David M., Dagbert M. Case studies in geostatistical ore reserve estimation of uranium deposits // Uranium evaluation and mining techniques. IAEA. Vienna, 1980. P. 111—137.

Diamond C., Halferdahl R. Mechanical sampling of Ore // Eng. and Min. Journ. Press. 1922. Vol. 114. No 7.

Füst A. Optimale Erkundung ungarischer Bauseitlagerstätten // Neue Bergbau-techn. 1980. Bd 10, N 1. S. 27—31.

Journel A. G., Huijbregts Ch. J. Mining geostatistics. London; N. Y.; San Francisco, 1978. 600 p.

Lindley D. V. Regression lines and the linear functional relationship // Journ. Royal Stat. Soc., London. Ser. B. 1947. Vol. 9. P. 218—244.

Lindley D. V. Estimation of a functional relationship // Biometrika. 1953. Vol. 40. P. 47—49.

Matheron G., Formeri Ph. Recherche d'optimum dans la reconnaissance et la mise en exploitation des gisements miniers // Ann. des mines. 1963. Vol. 6, N 5.

Rendu J. M. An Introduction to Geostatistical Methods of Mineral Evaluation. South Afr. Inst. Min. and Met. Johannesburg, 1978. 84 p.

Stamberg R. F. Theoretische Grundlagen der Bemusterung von Lagerstätten fester mineralischer Rohstoffe. Akademie — verlag. Berlin, 1965. 120 p.

ОГЛАВЛЕНИЕ

Предисловие	3
Глава 1. Введение	4
Глава 2. Корреляционно-регрессионные методы повышения достоверности оценки параметров месторождений	6
§ 1. Анализ и классификация ситуаций, возникающих при изучении и использовании статистических зависимостей в геологии и разведке месторождений	7
§ 2. Оценка геологических переменных по истинному уравнению регрессии	9
§ 3. Оценка геологических переменных по выборочному уравнению регрессии	14
§ 4. Непосредственная оценка по регрессии средних значений параметров геологических совокупностей	19
§ 5. Некоторые свойства реальных геологоразведочных выборок	26
§ 6. Особенности оценки средних в ограниченных совокупностях с учетом погрешности измерения	33
§ 7. Оценка среднего в совокупности через индивидуальную оценку y по зависимости в каждом элементе сфинкс-выборки	36
Глава 3. Оценивание среднего значения произведения нескольких величин с применением корреляционно-регрессионного анализа	50
§ 1. Общие замечания	50
§ 2. Тождества, связывающие среднее значение произведения нескольких величин с произведением их средних	51
§ 3. Анализ условий, при которых произведение средних предпочтительнее среднего значения произведения	52
§ 4. Повышение точности оценки произведения нескольких параметров без увеличения объема выборки	59
Глава 4. Контроль погрешностей и оптимизация обработки проб при разведке месторождений и подсчете запасов	65
§ 1. Общие замечания	65
§ 2. Обзор методов и формул контроля обработки проб полезных ископаемых	66
§ 3. Вывод формулы сокращения смеси «рудных» и «нерудных» зерен правильной формы	73
§ 4. Оценка погрешности сокращения при многоэтапной обработке проб	75
§ 5. Изменение контрастности в процессе обработки пробы	76
§ 6. Методы оптимизации обработки проб	81
§ 7. Составление надежных и экономически оптимальных схем обработки проб	84
Глава 5. Погрешности измерений и анализов в разведке месторождений, их контроль и учет при подсчете запасов	99
§ 1. Общие замечания	99
§ 2. Характеристики качества анализа	99
§ 3. Характеристика воспроизводимости анализа, когда она зависит от содержания	104
§ 4. Способы выражения и расчета воспроизводимости при логнормальном распределении погрешностей	107
§ 5. Внешний контроль анализов. Выявление систематических погрешностей	113
§ 6. Сопоставление измерений пористости по керну и скважинно-геофизическими методами. Комплексная оценка пористости	118
§ 7. Градуировочные графики при измерениях и анализах в геологии и разведке месторождений	129

Глава 6. Оценка параметров составных и неоднородных геологических объектов при подсчете запасов	139
§ 1. Иерархия объектов изучения в геологии и разведке месторождений. Структура геологоразведочной выборки	140
§ 2. Оценка среднего по данным нескольких серий измерений	143
§ 3. Оценка среднего по объединенным данным независимых систем опробования	146
§ 4. Оценка среднего в объекте по данным, относящимся к составным частям этого объекта	148
§ 5. Оценка среднего в объекте по данным, относящимся к опробованным частям этого объекта	153
§ 6. Оценка среднего в совокупности объектов, по средним в отдельных группах объектов, выделенных случайным образом	156
§ 7. Анализ применяющихся способов оценки средних значений коллекторских свойств нефтяных месторождений	159
Глава 7. Способы выражения степени разведанности и точности оценки параметров месторождений	165
§ 1. Анализ категорий запасов как показателя степени изученности и точности оценки параметров	165
§ 2. Аналитические методы оценивания точности запасов и других параметров, не связанные с их геометризацией	174
Глава 8. Оптимизация степени разведанности месторождений и структуры геологоразведочных выборок	179
§ 1. Аспекты оптимизации разведки месторождений	180
§ 2. Оптимизация объема геологоразведочных работ на прирост запасов	180
§ 3. Критерии оптимизации предварительной разведки (на уточнение)	187
§ 4. Критерии и методы оптимизации плотности сети детальной разведки	192
§ 5. Оптимизация структуры выборки при разведке месторождений	202
§ 6. К оптимизации разведки на уточнение оценки точности	210
Литература	214

Научное издание

Юрий Андреевич Ткачев

**ТЕОРИЯ И МЕТОДЫ ПОДСЧЕТА ЗАПАСОВ
ПОЛЕЗНЫХ ИСКОПАЕМЫХ**

*Утверждено к печати
Коми филиалом АН СССР*

Редактор издательства *Т. П. Жукова*
Художник *О. М. Разуевич*
Технический редактор *Р. А. Кондратьева*
Корректоры *Г. А. Александрова, И. Н. Вакуленко*
и *М. В. Орлова*

ИБ № 21545

Сдано в набор 18.05.87. Подписано к печати 16.03.88.
М-21153. Формат 70×100¹/₁₆. Бумага офсетная № 1.
Гарнитура литературная. Печать офсетная. Фото-
набор. Усл. печ. л. 18.2. Усл. кр.-отт. 18.36.
Уч.-изд. л. 19.91. Тираж 1200. Тип. зак. 1587.
Цена 3 р.

Ордена Трудового Красного Знамени
издательство «Наука». Ленинградское отделение.
199034, Ленинград, В-34, Менделеевская линия, 1.

Ордена Трудового Красного Знамени
Первая типография издательства «Наука».
199034, Ленинград, В-34, 9 линия, 12.

3р.

5029



ИЗДАТЕЛЬСТВО „НАУКА“
Ленинградское отделение