

УДК 622.831.332

Статья / Article

© ПНИПУ / PNRPU, 2016

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ МЕТОДА ПРОГНОЗИРОВАНИЯ ГАЗОДИНАМИЧЕСКИХ ЯВЛЕНИЙ ПО ГЕОЛОГИЧЕСКИМ ДАННЫМ ДЛЯ УСЛОВИЙ ВЕРХНЕКАМСКОГО МЕСТОРОЖДЕНИЯ КАЛИЙНЫХ СОЛЕЙ

С.С. Андрейко

Горный институт Уральского отделения Российской академии наук (614007, Россия, г. Пермь, ул. Сибирская, 78а)

DEVELOPMENT OF MATHEMATICAL MODEL OF GAS-DYNAMIC PHENOMENA FORECASTING METHOD ACCORDING TO GEOLOGICAL DATA IN CONDITIONS OF VERKHNEKAMSKOIE POTASH SALT DEPOSIT

S.S. Andreiko

Mining Institute of the Ural Branch of the Russian Academy of Sciences (78a Sibirskaya str., Perm, 614007, Russian Federation)

Получена / Received: 23.11.2015. Принята / Accepted: 27.10.2016. Опубликовано / Published: 02.12.2016

Ключевые слова:

калийные рудники,
газодинамические явления,
геологические данные,
обучающая выборка,
математическая модель,
дискриминантный анализ,
статистические процедуры,
робастные оценки, решающие
правила, прогнозные карты.

Предложен статистический подход при прогнозировании газодинамической опасности в калийных рудниках, который позволяет существенно снизить зависимость конечных результатов прогнозирования от субъективных факторов путем введения в модель строго определенных информативных критериев. Получение наиболее адекватных действительности моделей прогнозирования возможно только с помощью комплексного подхода, отображающего факторы генерации, миграции, аккумуляции и сохранения в течение длительного геологического времени газовых скоплений и очагов газодинамических явлений в соляном породном массиве. При разработке математической модели метода прогнозирования газодинамических явлений по геологическим данным использовалась параметрическая математическая модель метода прогнозирования, основанная на дискриминантном анализе, которая является мощным статистическим методом и по глубине анализа данных и ценности получаемых результатов относится к самым эффективным методам статистического анализа. В случае соблюдения условий нормальности распределения многомерных данных и равенства ковариационных матриц параметрическая модель приводит к оптимальным результатам при прогнозировании. При невыполнении указанных ограничений применение в модели робастных оценок позволяет компенсировать рост вероятности ошибочного прогноза. Выполнен статистический анализ геологической информации в местах развития газодинамических явлений (ГДЯ) в калийных рудниках. Сформированы обучающие выборки, разработана математическая модель метода прогноза зон, опасных по ГДЯ, на основе многомерного дискриминантного анализа с использованием классических и робастных статистических процедур. Получены решающие правила для прогнозирования по геологическим данным зон, опасных по ГДЯ, для условий Верхнекамского месторождения калийных солей. Построены прогнозные карты зон, опасных по газодинамическим явлениям, для условий шахтного поля строящегося Усть-Яйвинского рудника ПАО «Уралкалий» и Половодовского участка Верхнекамского месторождения калийных солей. Дана оценка адекватности метода прогнозирования практике ведения горных работ на калийных пластах, опасных по ГДЯ, в условиях шахтных полей калийных рудников ПАО «Уралкалий».

Key words:

potash mines, gas-dynamic
phenomena, geological data,
learning sample, mathematical
model, discriminant analysis,
statistical procedures, robust
estimation, decision rules,
forecast maps.

Article proposes a statistical approach in predicting the gas-dynamic danger in potash mines, which can significantly reduce the dependence of the final prediction results on subjective factors by introducing a model of informative strictly defined criteria. Obtaining the most appropriate forecasting models of reality is only possible through an integrated approach, displaying factors of generation, migration, accumulation and preservation for a long period of geological time gas clusters and centers of gas-dynamic phenomena in the salt rock mass. In the mathematical model development for gas-dynamic phenomena's prediction method on geological data a parametric mathematical model of forecasting method was used based on discriminant analysis, which is a powerful statistical method and in depth of data analysis and the results significance is among the most effective methods of statistical analysis. In case of compliance with conditions of multidimensional data normal distribution and equality of covariance matrices the parametric model leads to optimal results in forecasting. In case of non-compliance with these restrictions in the use of robust evaluation model allows to compensate the incorrect prediction's probability growth. A statistical analysis of geological information in the field of gas-dynamic phenomena (GDP) in the potash mines was performed. Learning samples were formed, the mathematical model of forecasting method for zones dangerous because of GDP was developed based on multivariate discriminant analysis with classical and robust statistical procedures. Decision rules to predict the geological data of areas dangerous because of GDP were obtained for conditions of Verkhnekamskoie potash salt deposit. Forecast maps were created for zones dangerous because of GDP for the Ust-Yaivinskii mine field conditions on the base of PJSC "Uralkali" and Polovodovskii area of Verkhnekamskoie potash salt deposit. The adequacy estimation was given for the method of forecasting mining operations' practice on potash layers dangerous because of GDP in conditions of potash mine fields PJSC "Uralkali".

Андрейко Сергей Семенович – доктор технических наук, профессор, заведующий лабораторией геотехнологических процессов и рудничной газодинамики (тел.: +007 342 216 58 42, e-mail: ssa@mi-perm.ru).

Sergei S. Andreiko – Doctor of Technical Sciences, Professor, Head of the Laboratory of Geotechnological Processes and Mine Gas Dynamics (tel.: +007 342 216 58 42, e-mail: ssa@mi-perm.ru).

Введение

В настоящее время в мире подземным способом разрабатывается около 30 месторождений калийно-магниевых солей. В процессе подземной добычи на всех разрабатываемых месторождениях сталкиваются с проблемой интенсивных выделений горючих и ядовитых газов, а также с проблемой газодинамических явлений (ГДЯ) различного вида. Интенсивные газовыделения приводят к загазованию горных выработок, вспышкам и взрывам горючих газов, которые часто сопровождаются одиночными и групповыми несчастными случаями, в том числе и с летальным исходом. Газодинамические явления в виде внезапных выбросов соли и газа, разрушений пород кровли и почвы горных выработок, сопровождающихся газовыделениями, и явлений комбинированного типа в силу своей внезапности и большой мощности (до 6000 т разрушенной горной массы) угрожают жизни шахтеров, выводят из строя дорогостоящее горное оборудование и нарушают ритмичность работы калийных рудников [1–4].

Проблема прогнозирования газодинамических явлений при подземной разработке калийных месторождений исследуется достаточно давно. Как правило, для ее решения применялись геолого-математические методы прогнозирования. Статистический подход при прогнозировании газодинамической опасности позволяет существенно снизить зависимость конечных результатов прогнозирования от субъективных факторов путем введения в модель строго определенных информативных критериев. Однако следует отметить, что получение наиболее адекватных действительности моделей прогнозирования возможно только с помощью комплексного подхода, отображающего факторы генерации, миграции, аккумуляции и сохранения газовых скоплений и очагов ГДЯ в соляном породном массиве.

Исследования в области построения математических моделей локализации участков газодинамической опасности в калийных рудниках традиционно сводились к применению параметрических моделей, основанных на линейных дискриминантных функциях. Исследования показали, что параметрическая математическая модель метода прогнозирования зон, опасных по ГДЯ, основанная на дискриминантном анализе, является мощным статистическим методом и по глубине анализа данных и ценности получаемых результатов относится к самым эффективным методам статистического анализа. Модель отличается

достаточной простотой, лаконичностью, хорошей интерпретируемостью, удобна для рассмотрения геометрических иллюстраций разделения областей массива на опасные и неопасные по ГДЯ зоны. Параметрическая модель метода прогнозирования позволяет производить отбор наиболее информативных показателей, что дает возможность снизить размерность исходного пространства признаков, отбросить неинформативные и малоинформативные переменные. В случае соблюдения условий нормальности распределения многомерных данных и равенства ковариационных матриц параметрическая модель приводит к оптимальным результатам при прогнозировании. При невыполнении указанных ограничений применение в модели робастных оценок позволяет компенсировать рост вероятности ошибочного прогноза. Таким образом, при решении прикладных задач прогнозирования зон, опасных по ГДЯ, параметрическая математическая модель, основанная на линейном дискриминантном анализе с применением робастных оценок, является наиболее предпочтительной среди других известных параметрических моделей.

Постановка задачи

Поставленная задача заключается в отношении на основе комплекса геологических показателей к зоне, опасной или неопасной по газодинамическим явлениям, конкретного участка месторождения или шахтного поля. В соответствии с поставленной задачей математической моделью метода прогнозирования будет решающее правило, при подстановке значений геологических показателей в которое можно будет с максимально возможной вероятностью дать оценку рассматриваемому участку с позиций газодинамической опасности.

В математической постановке получение решающего правила сводится к определению конечного набора геологических показателей, интервалов их значений, присущих как опасным, так и неопасным по ГДЯ зонам, а также нахождению наиболее эффективных с точки зрения вероятности правильного прогнозирования статистических процедур.

Формально каждая запись имеющихся геологических данных представляет собой многомерное наблюдение в P -мерном пространстве с конечным набором признаков X . Совокупность наблюдений, относящихся к одному классу D_i , образует «облако» в этом же пространстве. Поэтому для успешного прогнозирования необходимо, чтобы «облако» из D_i было сконцентрировано в некоторой области R_i пространства X и в область R_i попала

незначительная часть D_j , соответствующая другому классу. В общем случае построение решающего правила для отнесения многомерного наблюдения к тому или иному классу можно рассматривать как задачу поиска k непересекающихся областей R_i ($i = 1, k$), удовлетворяющих указанным условиям.

Если P -мерное наблюдение пространства X попадает в область R_i , то будем считать, что принимается решение о принадлежности объекта к D_i . При использовании подобного правила прогнозирования в некоторых случаях объект может быть отнесен к другому классу. Поэтому обозначим вероятность неправильной классификации объектов через $P(i/j)$, тогда критерий средней вероятности ошибочного прогноза имеет следующий вид:

$$Q = \sum_{i=1}^{k-1} \sum_{j=1}^k a_i P(i/j), \quad (1)$$

где a_i – априорная вероятность появления объекта из области D_i .

Решение данной задачи проводится при наличии обучающей выборки, которая должна состоять из наблюдений, относящихся к обоим классам (к зонам, опасным и неопасным по ГДЯ). Поэтому считаем, что обучающая выборка для класса D_i представляет собой выборку из некоторой генеральной совокупности объектов с плотностью распределения $P_i(X)$. Правило прогнозирования, минимизирующее функцию потерь Q , будет следующим:

$$x \in D_i, \text{ если } \max_{1 \leq j \leq k} a_j P(X/j) = a P(X/i). \quad (2)$$

Чтобы выяснить вероятностный смысл выражения (2), предположим, что $P(X) = \sum_{i=1}^k a_i P(X/j)$. Величина $P(j/X) = a_j P(X/j) \times [P(X)]^{-1}$ представляет собой условную вероятность того, что P -мерное наблюдение X принадлежит D_i . Поэтому выражение (2) называется правилом прогнозирования по максимуму условной вероятности, так как дает наименьшую величину функции потерь (1). В реальной ситуации точное расположение «облаков» объектов и их плотности $P_i(X)$ неизвестны, а имеются обучающие выборки, которые дают об этом приближенное представление. Обучающая выборка представляет собой совокупность многомерных наблюдений, полученных в результате исследований. В матрицу исходных данных также включаются наблюдения, принадлежность к какому-либо классу которых точно не установлена.

В настоящее время известно большое число методов распознавания образов для построения решающих правил по обучающей выборке или так называемой статистической классификации, которые подразделяются на две группы: параметрические и непараметрические методы [5–11]. В параметрических методах получения решающих правил для прогнозирования традиционно используются предположения об однородности обучающих выборок для каждого класса, измерении всех компонент векторов признаков \vec{X} , принадлежности условных вероятностных распределений к многомерному гауссовскому семейству, независимости выборочных значений. Наиболее известным из параметрических методов является линейный дискриминантный анализ, основанный на линейных дискриминантных функциях (ЛДФ). Дискриминантный анализ на основе ЛДФ характеризуется строгим математическим обоснованием, физической интерпретируемостью, наглядным представлением результатов. Недостатками данного метода являются требования к однородности выборок, принадлежности к многомерному нормальному распределению и независимости выборочных значений.

Непараметрические методы получения решающих правил для прогнозирования игнорируют априорные исходные предположения об однородности выборки каждого класса, наличие измерений всех компонент вектора признаков, гауссовость условных вероятностей распределения признаков. Наиболее известными непараметрическими методами являются метод « k -ближайших соседей» и метод Розенблатта–Парзена [12–14]. Эти методы отличаются относительной простотой, хорошо работают при прогнозировании, но требуют постоянного запоминания обучающих выборок и характеризуются высокой чувствительностью к малым объемам обучающих выборок, при которых риск ошибочной классификации резко возрастает. При этом коэффициент робастности (устойчивости), представляющий собой относительное увеличение вероятности ошибки классификации, при объеме обучающей выборке больше 20 для непараметрических методов в 2–4 раза больше, чем для линейного дискриминантного анализа на основе ЛДФ [14]. Для получения решающих правил прогнозирования зон, опасных по ГДЯ, используются обучающие выборки, объемы которых намного превышают 20, что позволяет отдать при прогнозировании зон, опасных по ГДЯ, предпочтение параметрической модели получения решающих правил.

Суть построения решающего правила на основе линейных дискриминантных функций сводится к следующему. Обозначив через X_{ij} значение переменной с номером i в точке наблюдений с номером j , взятой из выборки, характеризующей опасные по ГДЯ зоны, получаем матрицу V порядка $p \times n_1$ результатов наблюдений над этой выборкой:

$$V = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n_1} \\ X_{21} & X_{22} & \dots & X_{2n_1} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn_1} \end{pmatrix}. \quad (3)$$

Обозначим через X_{ij} результат измерения переменной с номером i в точке с номером j , взятой из выборки, характеризующей неопасные по ГДЯ зоны. В результате получим матрицу W порядка $p \times n_2$:

$$W = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n_2} \\ X_{21} & X_{22} & \dots & X_{2n_2} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn_2} \end{pmatrix}, \quad (4)$$

где p – число переменных (показателей); n_1 , n_2 – объемы выборок показателей.

Затем, используя эти данные, для каждой группы вычисляются средние и суммы взаимных произведений отклонений от средних. Средние вычисляются по формуле

$$\bar{X}_{ik} = \left(\sum_{i=1}^{n_k} X_{ijk} \right) (n_k)^{-1}, \quad (5)$$

где k – число групп ($k = 2$); n_k – размер выборки в k -й группе; $j = 1, 2, \dots, p$ – переменные.

Далее находится сумма взаимных произведений отклонений от средних:

$$S_k = \sum (X_{ijk} - \bar{X}_{jk})(X_{ilk} - \bar{X}_{lk}), \quad (6)$$

где $j = 1, 2, \dots, p$; $l = 1, 2, \dots, p$.

Следующим этапом идет вычисление объединенной дисперсионной матрицы

$$D = \left(\sum_{k=1}^q S_k \right) \left(\sum_{k=1}^q n_k - q \right)^{-1}, \quad (7)$$

где q – число групп ($q = 2$).

Рассчитываются общие средние для обеих групп:

$$\bar{X}_j = \left(\sum_{k=1}^q n_k \cdot X_{jk} \right) \left(\sum_{k=1}^q n_k \right)^{-1}, \quad (8)$$

где \bar{X}_{ij} – средние j -й переменной в k -й группе; n_k – размер выборки в k -й группе; $j = 1, 2, \dots, p$ – переменные.

Затем для каждой группы вычисляются обратные элементы объединенной дисперсионной матрицы D :

$$[d_{ij}] = D^{-1}. \quad (9)$$

Коэффициенты дискриминантных функций находятся по формуле

$$a_{ik} = \sum_{j=1}^p d_{ij} \cdot \bar{X}_{jk}, \quad (10)$$

где $i = 1, 2, \dots, p$; k – число групп.

Затем вычисляются константы дискриминантных функций:

$$b_{ok} = -0,5 \sum_{j=1}^p \sum_{l=1}^p d_{jl} \cdot \bar{X}_{jk} \cdot \bar{X}_{lk}. \quad (11)$$

Для случая двух классов D_1 и D_2 (опасные и неопасные по ГДЯ зоны) получаем две линейные дискриминантные функции вида

$$\sum_{i=1}^p a_i \cdot X_i - b_{ok} = 0. \quad (12)$$

Вычитая меньшую дискриминантную функцию из большей, можем перейти к одной линейной дискриминантной функции вида

$$a_1 X_1 + a_2 X_2 + \dots + a_p X_p - b = 0. \quad (13)$$

Для случая двух классов использование линейных дискриминантных функций основывается на двух предположениях. Первое предположение состоит в том, что области R_1 и R_2 , где сконцентрирована основная часть объектов, соответствующих классам D_1 и D_2 (опасные и неопасные по ГДЯ зоны), могут быть разделены $(p-1)$ -мерной гиперплоскостью. Уравнение такой гиперплоскости и представляет собой выражение (13). Здесь a_1, a_2, \dots, a_p – коэффициенты, характеризующие наклон гиперплоскости к координатным осям, а b – параметр, характеризующий расстояние гиперплоскости до начала координат. Гиперплоскость (13) разбивает p -мерное пространство на два полупространства R_1 и R_2 , которые определяются неравенствами

$$R_1/z(X) < b \text{ и } R_2/z(X) > b, \quad (14)$$

где $z(X) = \sum_{i=1}^p a_i \cdot X_i$.

Следовательно, неравенство (14) дает нам решающее правило для отнесения многомерного вектора наблюдений к одной из двух зон – опасной или неопасной по ГДЯ.

Условимся относить наблюдение к D_1 , если $z(X) \leq b$, и к D_2 – в противном случае. Соотношение (14) позволяет перейти от p -мерных случайных величин к классификации одномерных величин $z(X)$. Обозначим через z_i случайную величину $z(X)$, когда $X \in D_i$.

Второе предположение дает метод определения коэффициентов разделяющей гиперплоскости таким образом, чтобы разделение было наилучшим. Предположим, что разделение будет тем лучше, чем дальше отстоят средние значения случайных величин z_1 и z_2 , причем расстояние измеряется в единицах дисперсии:

$$h_{(a)}^2 = (m_1 - m_2)D_z^{-1}, \quad (15)$$

где $m_i = E z_i$; D_z – дисперсия z_i ($i = 1, 2$), полагается одинаковой для обоих классов.

Вектор \bar{A} , доставляющий максимум $h^2(A)$, определяется уравнением

$$\bar{A}_F = \left[\sum (\bar{\mu}_2 - \mu_1) \right]^{-1}, \quad (16)$$

где μ – вектор средних значений для D_i ; Σ – ковариационная матрица, полагаемая одинаковой для обоих классов.

Максимальное значение $h^2(a)$ носит название расстояния Махалонобиса между классами D_1 и D_2 и равно

$$D^2 = \sum_{i=1}^p \sum_{j=1}^p d_{ij} \sum_{k=1}^k n_k (\bar{X}_{ik} - \bar{X}_i)(\bar{X}_{jk} - \bar{X}_j). \quad (17)$$

Статистика Махалонобиса может быть использована для проверки значимости полученной дискриминантной функции. Если многомерное распределение предположить нормальным, то D^2 -статистика распределена как χ^2 с $p(q-1)$ степенями свободы. Сравнивая расчетное значение D^2 с χ^2 при 5%-ном уровне значимости, если $D^2 > \chi^2$, можно говорить о том, что полученные линейные дискриминантные функции имеют смысл и могут использоваться как решающие правила при прогнозировании. На стадии разведочного анализа устанавливаем, что объекты из D_i ($i = 1, 2$) распределены согласно многомерному нормальному закону с одинаковой ковариационной матрицей. Следовательно, решающие правила (14) на основе линейных дискриминантных функций (10) и (11) являются наилучшими, т.е. доставляют функционалу Q (1) наименьшее возможное значение.

Необходимо отметить еще один весьма важный момент при нахождении решающих правил посредством линейных дискрими-

нантных функций. Если на стадии разведочного анализа установлено «засорение» обучающих выборок «выбросами» и эти аномальные наблюдения могут представлять определенный интерес при исследованиях, то целесообразно использовать процедуры оценки параметров распределения, не чувствительные к структуре данных. Такие процедуры оценивания называются робастными, или устойчивыми [6, 15–22]. Могут использоваться следующие робастные процедуры в подстановочном дискриминантном анализе при получении устойчивых решающих правил: винзоризованные оценки, усеченные оценки, весовые оценки Хьюбера, кусочно-линейные M -оценки Хампеля. При наличии искажений в обучающих выборках и использовании в решающих правилах классических оценок параметров распределений вероятность ошибочного прогноза значительно увеличивается, поэтому необходимо использовать устойчивые оценки.

Если в обучающей выборке присутствуют «выбросы», то можно использовать любую из приведенных устойчивых оценок. Если обучающая выборка сформирована из распределений с «тяжелыми хвостами» или отличных от нормального, то лучше применять весовые оценки Хьюбера или Хампеля. Робастные итерационные оценки Хьюбера определяются следующими соотношениями:

$$\begin{aligned} \hat{x}_i^* &= \sum_{j=1}^L x_j w_j / \sum_{j=1}^L w_j, \\ \hat{\Sigma}_i^* &= \sum_{j,k=1}^L w_j w_k \left(x_j - \hat{x}_j^* \right) \left(x_k - \hat{x}_k^* \right)^T / \sum_{j=1}^L w_j^2, \\ w_j &= \begin{cases} 1, & \text{если } d_j \leq d_0, \\ d_0 / d_j, & \text{если } d_j > d_0 \end{cases}, \quad d_0 = \sqrt{p} + 2\sqrt{2}, \quad (18) \\ d_j &= \left(x_j - \hat{x}_i^* \right) \hat{\Sigma}_i^{-1} \left(x_j - \hat{x}_i^* \right), \quad (j = \overline{1, n}; i = \overline{1, L}). \end{aligned}$$

Здесь $\hat{x}_i, \hat{\Sigma}_i$ – оценки параметров, вычисленные на предыдущем шаге; x_1, \dots, x_{n_i} – обучающая выборка для i -го класса; n_i – объем обучающей выборки для i -го класса. В качестве начального приближения используются классические оценки параметров.

В целом параметрическая математическая модель метода прогнозирования зон, опасных по ГДЯ, основанная на дискриминантном анализе, является мощным статистическим методом и по глубине анализа данных и ценности получаемых результатов относится к самым эффективным

методам статистического анализа. Модель отличается достаточной простотой, лаконичностью, хорошей интерпретируемостью, удобна для рассмотрения геометрических иллюстраций разделения областей массива на опасные и неопасные по ГДЯ зоны. Параметрическая модель метода прогнозирования позволяет производить отбор наиболее информативных показателей, что дает возможность снизить размерность исходного пространства признаков, отбросить неинформативные и малоинформативные переменные. В случае соблюдения условий нормальности распределения многомерных данных и равенства ковариационных матриц параметрическая модель приводит к оптимальным результатам при прогнозировании. При невыполнении указанных ограничений применение в модели робастных оценок позволяет компенсировать рост вероятности ошибочного прогноза.

Процедура получения решающих правил

В процессе получения решающих правил учитывались существующие представления о механизме образования очагов газодинамических явлений в соляном породном массиве Верхнекамского месторождения калийных солей [23, 24]. Формирование обучающей выборки проводилось следующим способом. Все зафиксированные случаи ГДЯ при отработке пласта АБ в южной части Верхнекамского месторождения калийных солей на шахтных полях рудников БКПРУ-2 и БКПРУ-4 и соответствующие им геологические показатели априори отнесены в опасную по ГДЯ зону (группа 2). Остальная часть общей выборки отнесена к зоне, неопасной по ГДЯ (группа 1). Всего исследовано 658 наблюдений, из них отнесенных к неопасным по ГДЯ (группа 1) 478, к опасным по ГДЯ (группа 2) – 180. Для получения решающих правил использовались следующие показатели: X_1 – мощность пласта Б, м; X_2 – содержание в пласте Б хлористого калия (KCl), %; X_3 – содержание в пласте Б хлористого магния ($MgCl_2$), %; X_4 – содержание в пласте Б хлористого натрия (NaCl), %; X_5 – содержание в пласте Б брома (Br), %; X_6 – содержание в пласте Б сульфата кальция ($CaSO_4$), %; X_7 – содержание в пласте Б нерастворимого остатка (н.о.), %; X_8 – глубина залегания кровли пласта Б, м; X_9 – абсолютная отметка кровли пласта Б, м.

Разведочный анализ исходных данных показал неоднородность общей выборки, отклонение распределения от нормального и наличие в выборке «тяжелых хвостов». В связи с тем что

не все используемые показатели в равной степени полезны для разделения пласта на опасные и неопасные по ГДЯ зоны, необходимо было выявить такое подмножество показателей, по которому можно построить «наилучшее» разделение между группами. Для решения этой задачи был реализован пошаговый дискриминантный анализ (ПДА) с использованием следующих оценок: классические оценки, весовые оценки Хьюбера и усеченные оценки с уровнем усечения 0,1. В ПДА для отбора наиболее информативных показателей использовалась F -статистика, основанная на критерии однофакторного дисперсионного анализа. В сущности, логика ПДА такова: вначале определяется показатель, для которого средние значения в двух группах «наиболее различны». Для каждого показателя $X_1, X_2, X_3, \dots, X_9$ различие измеряется с помощью F -статистики и выбирается (включается) тот показатель, которому соответствует большее значение F . На каждом шаге процедуры рассматривается условное распределение каждого показателя, не включенного в подмножество, при заданных включенных показателях. Затем из числа невключенных показателей вновь определяется показатель, для которого средние значения в двух группах «наиболее различны». Это различие определяется с помощью F -статистики. Процесс завершается, когда ни один из оставшихся показателей не вносит значимого вклада в разделение групп. В процессе реализации ПДА принимались стандартные значения минимума F -включения переменных, не вошедших в искомое подмножество, и F -удаления выбранных переменных при уровне значимости 0,05. На шаге 8 значение статистики F -удаления оказалось меньше принятого минимума, равного 3,0, следовательно, «наилучшие» дискриминантные уравнения задаются на шаге 7. Таким образом, показатели $X_1, X_2, X_3, X_4, X_5, X_6$ и X_7 вошли в искомое подмножество, а показатели X_8 и X_9 – нет. Анализ матриц классификации при реализации процедуры ПДА с использованием классических оценок, весовых оценок Хьюбера и усеченных оценок с уровнем усечения 0,1 показал следующее. При применении классических оценок правильно классифицировано 62,3 % исходных сгруппированных наблюдений. В случае использования весовых оценок Хьюбера и усеченных оценок с уровнем усечения 0,1 правильно классифицировано соответственно 61,7 и 61,5 % исходных сгруппированных наблюдений.

Правильность первичной классификации с использованием как классических, так и робастных процедур практически одинакова и превышает 60 %. Поэтому окончательный

выбор решающего правила был проведен после повторной классификации. При этом исходные данные были перегруппированы, а затем проводилась процедура повторной классификации, в результате которой уточнялись коэффициенты линейных дискриминантных функций для каждой из групп, а также давалась оценка правильности классификации. Для каждого наблюдения с использованием полученных дискриминантных функций и D^2 -статистики Махаллобиса вычислялись апостериорные вероятности его принадлежности к группе, опасной или неопасной по ГДЯ. На основании этих вероятностей каждое наблюдение классифицировалось как принадлежащее одной из групп и составлялась матрица классификации. В табл. 1 представлены результаты классификации с применением классических процедур, с использованием весовых оценок Хьюбера, усеченных оценок (уровень усечения $-0,1$).

Как видно из табл. 1, с использованием классических статистических процедур линейного дискриминантного анализа правильно классифицировано 92,86 % повторно сгруппированных наблюдений.

При использовании линейного дискриминантного анализа с весовыми оценками Хьюбера вероятность правильной классификации составляет 96 %.

Таблица 1
Матрица классификации по результатам линейного дискриминантного анализа

Группа	Результаты классификации				Всего
	неопасные по ГДЯ		Опасные по ГДЯ		
	абс.	%	абс.	%	
<i>С использованием классических статистических процедур</i>					
Неопасные по ГДЯ (1)	323	91,76	29	8,24	352
Опасные по ГДЯ (2)	18	5,88	288	94,12	306
Всего	341	51,82	317	48,18	658
<i>С использованием весовых оценок Хьюбера</i>					
Неопасные по ГДЯ (1)	334	94,89	18	5,11	352
Опасные по ГДЯ (2)	8	2,61	298	97,39	306
Всего	342	51,98	316	48,02	658
<i>С использованием усеченных оценок с уровнем усечения 0,1</i>					
Неопасные по ГДЯ (1)	319	90,63	33	9,37	352
Опасные по ГДЯ (2)	24	7,84	282	92,16	306
Всего	343	52,13	315	47,87	658

При применении в линейном дискриминантном анализе усеченных оценок с уровнем усечения 0,1 вероятность правильной классификации составляет 91,34 %. По критерию наибольшей вероятности правильной классификации принимаем для прогнозирования зон, опасных по ГДЯ, решающее правило, полученное с использованием весовых оценок Хьюбера. В табл. 2 представлены константы и коэффициенты дискриминантных функций, найденные с использованием весовых оценок Хьюбера.

Таблица 2

Константы и коэффициенты дискриминантных функций, полученные с использованием весовых оценок Хьюбера

Группа	Константы дискриминантных функций	Значения коэффициентов дискриминантных функций						
		X_1	X_2	X_3	X_4	X_5	X_6	X_7
Неопасные по ГДЯ (1)	-385755,60	998,66	7678,37	16092,34	7734,02	-11769,6	7829,14	7715,20
Опасные по ГДЯ (2)	-385444,89	978,68	7677,20	16123,52	7730,39	-12516,62	7836,01	7710,63

Так как в нашем случае имеется две группы, то можно перейти от двух дискриминантных функций к одной, вычитая коэффициенты и константы меньшей функции из большей. После вычитания получаем линейную дискриминантную функцию следующего вида:

$$F_p = -19,98X_1 - 1,16X_2 + 31,18X_3 - 3,64X_4 - 747,02X_5 + 6,87X_6 - 4,57X_7 + 310,72 \geq 0, \quad (19)$$

где X_1 – мощность пласта Б, м; X_2 – содержание KCl в породах пласта Б, %; X_3 – содержание MgCl₂ в пласте Б, %; X_4 – содержание NaCl в пласте Б, %; X_5 – содержание Br в пласте Б, %; X_6 – содержание CaSO₄ в пласте Б, %; X_7 – содержание нерастворимого остатка в пласте Б, %.

Выражение (19) является искомым решающим правилом для отнесения участков пласта АБ к зонам, опасным или неопасным по ГДЯ. При $F_p \geq 0$ точка наблюдения относится к зоне, опасной по ГДЯ, в противном случае – к неопасной. Для вероятностной интерпретации

полученных дискриминантных функций, которая более удобна при построении прогнозных карт, целесообразно использовать выражение

$$P_l = 1 / \sum_{k=1}^q \exp(f_k - f_l), \quad (20)$$

где P_l – вероятность, соответствующая наибольшей дискриминантной функции; q – число групп ($q = 2$); f_k, f_l – значение k -й и наибольшей дискриминантной функции соответственно.

Подставляя значения показателей в дискриминантные функции, константы и коэффициенты которых представлены в табл. 2, получим некоторые значения дискриминантных функций. Затем, подставляя полученные значения дискриминантных функций в выражение (20), найдем значение вероятности отнесения точки наблюдения к той зоне, значение дискриминантной функции для которой больше. Значение P_l изменяется от 0 до 1, граничным является значение вероятности, равное 0,5.

Оценка адекватности решающих правил и прогнозные карты зон, опасных по ГДЯ

Для оценки адекватности полученных решающих правил практике ведения горных работ на прогнозных картах зон, опасных по ГДЯ, были отмечены места, где при ведении подготовительных и очистных горных работ произошли газодинамические явления. При отработке пласта АБ шахтного поля рудника БКПРУ-2 из 98 случаев ГДЯ в прогнозируемые зоны, опасные по ГДЯ, не вошли 8 ГДЯ (7,8 % от их общего числа). Таким образом, адекватность результатов прогнозирования зон, опасных по газодинамическим явлениям, на сильвинитовом пласте АБ шахтного поля рудника БКПРУ-2 практике ведения горных работ составляет соответственно 92,2 %. С целью проверки работоспособности полученных решающих правил прогноза зон, опасных по ГДЯ, в исходную выборку были добавлены данные по шахтным полям рудников БКПРУ-1 и БКПРУ-4, первоначально не вошедшие в исходные данные. Далее была проведена экзаменационная классификация новых исходных данных, правильность которой для пласта АБ составила 96,7 %. Подстановка значений геологических показателей, полученных по результатам бурения скважин, детальной разведки с поверхности, подземных геолого-разведочных скважин и данных бороздowego опробования в подземных горных выработках, в решающие правила позволила построить прогнозные карты зон, опасных по газодинамическим явлениям.

На рисунке приведена карта зон, опасных по газодинамическим явлениям, из кровли горных выработок при отработке пласта АБ в условиях Половодовского участка Верхнекамского месторождения калийных солей.

Заключение

В результате проведенного научного исследования создана математическая модель метода прогнозирования зон, опасных по газодинамическим явлениям, на основе линейного дискриминантного анализа с использованием робастных (устойчивых) процедур, позволяющая повысить надежность многофакторного прогнозирования. Разработан метод прогнозирования зон, опасных по газодинамическим явлениям, при разведке и разработке сильвинитового пласта АБ, использующий в решающих правилах комплекс

Список литературы

1. Проскуряков Н.М. Внезапные выбросы породы и газа в калийных рудниках. – М.: Недра, 1980. – 264 с.
2. Ковалев О.В., Ливенский В.С., Былино Л.В. Особенности безопасной разработки калийных месторождений. – Минск: Полымя, 1982. – 96 с.

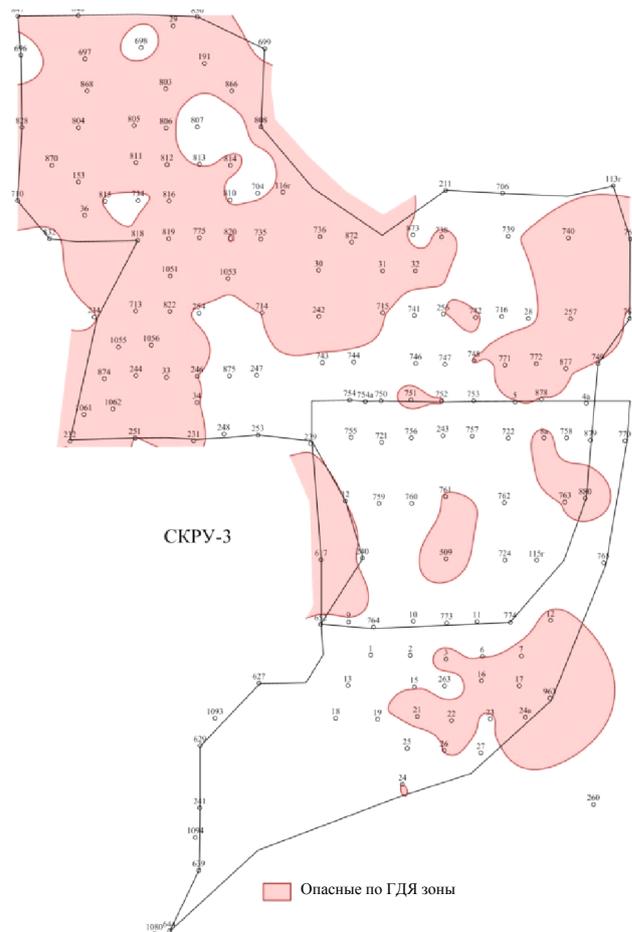


Рис. Карта зон, опасных по ГДЯ из кровли горных выработок при отработке пласта АБ в условиях Половодовского участка Верхнекамского месторождения калийных солей

следующих геологических параметров: мощность пласта, содержание хлористого калия, хлористого магния, хлористого натрия, брома, сульфата кальция и нерастворимого остатка в породах пласта. Вероятность правильной дискриминации участков сильвинитового пласта АБ на зоны, опасные и неопасные по ГДЯ, составляет 0,96. Адекватность результатов прогнозирования зон, опасных по газодинамическим явлениям, практике ведения горных работ на сильвинитовом пласте АБ составляет соответственно 92 %. Построены прогнозные карты зон, опасных по газодинамическим явлениям, сильвинитового пласта АБ для условий строящегося Усть-Яйвинского рудника ПАО «Уралкалий» и Половодовского участка Верхнекамского месторождения калийных солей.

3. Долгов П.В., Полянина Г.Д., Земсков А.Н. Методы прогноза и предотвращения газодинамических явлений в калийных рудниках. – Алма-Ата: Наука, 1987. – 176 с.
4. Проскуряков Н.М., Ковалев О.В., Мещеряков В.В. Управление газодинамическими процессами в пластах калийных руд. – М.: Недра, 1988. – 239 с.

5. Классификация и кластер. – М.: Мир, 1980. – 389 с.
6. Афифи А., Эйзен С. Статистический анализ: подход с использованием ЭВМ. – М.: Мир, 1982. – 488 с.
7. Горелик А.Л., Скрипкин В.А. Методы распознавания. – М.: Высшая школа, 1984. – 207 с.
8. Абусев Р.А., Лумельский Я.П. Статистическая групповая классификация. – Пермь, 1987. – 190 с.
9. Гирко В.Л. Многомерный статистический анализ. – Киев: Выща шк., 1988. – 320 с.
10. Фомин А.Я., Тарловский Г.Р. Статистическая теория распознавания образов. – М.: Радио и связь, 1988. – 264 с.
11. Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 606 с.
12. Parzen E. On estimation of a probability density function and mode // *Annals of Mathematical Statistics*. – 1962. – Vol. 33. – P. 1065–1076.
13. Епаненчиков В.А. Непараметрическая оценка многомерной плотности вероятностей // Теория вероятностей и ее применение. – М., 1969. – Т. 14, вып. 1. – С. 156–160.
14. Фукунага К. Введение в статистическую теорию распознавания образов. – М.: Наука, 1979. – 367 с.

References

1. Proskuriakov N.M. Vnezapnye vybrosy porody i gaza v kaliinykh rudnikakh [Sudden emissions of rock and gas in the potash mines]. Moscow: Nedra, 1980, 264 p.
2. Kovalev O.V., Livenskii V.S., Bylino L.V. Osobennosti bezopasnoi razrabotki kaliinykh mestorozhdenii [Features of safe development of potash deposits]. Minsk: Polymia, 1982, 96 p.
3. Dolgov P.V., Polianina G.D., Zemskov A.N. Metody prognoza i predotvrashcheniia gazodinamicheskikh iavlenii v kaliinykh rudnikakh [Method to forecast and prevent the gas-dynamic phenomena in potash mines]. Alma-Ata: Nauka, 1987, 176 p.
4. Proskuriakov N.M., Kovalev O.V., Meshcheriakov V.V. Upravlenie gazodinamicheskimi protsessami v plastakh kaliinykh rud [Management of gas-dynamic processes in potash ore seams]. Moscow: Nedra, 1988, 239 p.
5. Klassifikatsiia i klaster [Classification and cluster]. Moscow: Mir, 1980, 389 p.
6. Afifi A., Eizen S. Statisticheskii analiz: podkhod s ispol'zovaniem EVM [Statistical analysis: an approach using computers]. Moscow: Mir, 1982, 488 p.
7. Gorelik A.L., Skripkina V.A. Metody raspoznavaniia [Detection methods]. Moscow: Vysshiaia shkola, 1984, 207 p.
8. Abusev R.A., Lumel'skii Ia.P. Statisticheskaiia gruppovaia klassifikatsiia [Statistical group classification]. Perm', 1987, 190 p.
9. Girko V.L. Mnogomernyi statisticheskii analiz [Multivariate statistical analysis]. Kiev: Vysshiaia shkola, 320 p.
10. Fomin A.Ia., Tarlovskii G.R. Statisticheskaiia teoriia raspoznavaniia obrazov [Statistical theory of image recognition]. Moscow: Radio i sviaz', 1988, 264 p.
11. Aivazian S.A., Bukhshtaber V.M., Eniukov I.S., Meshalkin L.D. Klassifikatsiia i snizhenie razmernosti [Classification and reduction of dimension]. Moscow: Finansy i statistika, 1989, 606 p.
12. Parzen E. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 1962, vol. 33, pp.1065-1076.

15. Huber P.J. Robust statistics: a review // *Ann. Math. Statist.* – 1972. – Vol. 43. – P. 1041–1067.
16. Hampel F.R. A general qualitative definition of robustness // *Ann. Math. Statist.* – 1971. – Vol. 42, № 6. – P. 1887–1896.
17. Huber P.J. Robust statistics procedures. – Philadelphia: SIAM, 1977. – 56 p.
18. Huber P.J. Robust statistics. – N.Y.: Wiley. – 308 p.
19. Tiku M.L., Balakrishnan N.A. Robust test for testing the correlation coefficient // *Commun. Statist.* – 1986. – 15(4). – P. 946–971.
20. Харин Ю.С. Робастность в статистическом распознавании образов. – Минск: Университетское, 1992. – 232 с.
21. Титаренко Б.П., Харин Ю.С., Степанова М.Д. Робастное оценивание: методы, алгоритмы и программы. – М.: Финансы и статистика, 1992. – 256 с.
22. Хампель Ф., Рончетти Э., Рауссеу П. Робастность в статистике. – М.: Мир, 1989. – 512 с.
23. Андрейко С.С., Калугин П.А., Щерба В.Я. Газодинамические явления в калийных рудниках: генезис, прогноз и управление. – Минск: Высшая школа, 2000. – 355 с.
24. Андрейко С.С. Механизм образования очагов газодинамических явлений в соляном породном массиве. – Пермь: Изд-во Перм. гос. техн. ун-та, 2008. – 196 с.

13. Epanenchikov V.A. Neparametricheskaiia otsenka mnogomernoi plotnosti veroiatnoستي [Non-parametric estimation of a multidimensional probability density]. Teoriia veroiatnosti i ee primeneniie, 1969, vol.14, no.1, pp.156-160.
14. Fukunaga K. Vvedeniie v statisticheskuiu teoriuu raspoznavaniia obrazov [Introduction to statistical image recognition theory]. Moscow: Nauka, 1979, 367 p.
15. Huber P.J. Robust statistics: a review. *Ann. Math. Statist.*, 1972, vol. 43, pp.1041-1067.
16. Hampel F.R. A general qualitative definition of robustness. *Ann. Math. Statist.*, 1971, vol. 42, no.6, pp.1887-896.
17. Huber P.J. Robust statistics procedures. Philadelphia: SIAM, 1977, 56 p.
18. Huber P.J. Robust statistics. N.Y.: Wiley, 308 p.
19. Tiku M.L., Balakrishnan N.A. Robust test for testing the correlation coefficient. *Commun. Statist. Simulu.*, 1986, 15 (4), pp.946-971.
20. Kharin Iu.S. Robastnost' v statisticheskom raspoznavanii obrazov [Robustness in statistical image recognition]. Minsk: Universitetskoe, 1992, 232 p.
21. Titarenko B.P., Kharin Iu.S., Stepanova M.D. Robastnoe otsenivanie: metody, algoritmy i programmy [Robust estimation: methods, algorithms and programs]. Moscow: Finansy i statistika, 1992, 256 p.
22. Khampel' F., Ronchetti E., Rausseu P. Robastnost' v statistike [Robustness in statistics]. Moscow: Mir, 1989, 512 p.
23. Andreiko S.S., Kalugin P.A., Shcherba V.Ia. Gazodinamicheskie iavleniia v kaliinykh rudnikakh: genezis, prognoz i upravlenie [Gas-dynamic phenomena in potash mines: genesis, prognosis and management]. Minsk: Vysshiaia shkola, 2000, 355 p.
24. Andreiko S.S. Mekhanizm obrazovaniia ochagov gazodinamicheskikh iavlenii v solianom porodnom massive [Mechanism of the gas-dynamic phenomena centers' formation in salt rock mass]. Perm': Izd-vo Perm. gos. tekhn. un-ta, 2008, 196 p.

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:

Андрейко С.С. Разработка математической модели метода прогнозирования газодинамических явлений по геологическим данным для условий Верхнекамского месторождения калийных солей // Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело. – 2016. – Т.15, №21. – С.345–353. DOI: 10.15593/2224-9923/2016.21.6

Please cite this article in English as:

Andreiko S.S. Development of mathematical model of gas-dynamic phenomena forecasting method according to geological data in conditions of Verkhnekamskoe potash salt deposit. *Bulletin of PNRPU. Geology. Oil & Gas Engineering & Mining*, 2016, vol.15, no.21, pp.345–353. DOI: 10.15593/2224-9923/2016.21.6