

УДК 622.831.332

Article / Статья

© PNRPU / ПНИПУ, 2016

DEVELOPMENT OF MATHEMATICAL MODEL OF GAS-DYNAMIC PHENOMENA FORECASTING METHOD ACCORDING TO GEOLOGICAL DATA IN CONDITIONS OF VERKHNKAMSKOIE 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)

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)

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

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 of 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 Verkhnekamskoe 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 Verkhnekamskoe 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".

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».

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).

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).

Introduction

At the present time about 30 deposits of potassium and magnesium salts is developed by underground method. In the process of underground mining at all the developed deposits there is the problem of intense release of flammable and toxic gases, as well as the problem of various types of gas-dynamic phenomena (GDP). Intensive gas emission leads to mines gas contamination, flares and explosions of flammable gases, which are frequently accompanied by single and multiple accidents, including fatalities. Gas-dynamic phenomena in the form of sudden salt and gas outburst, the destruction of roof rocks and soil of mining, accompanied by gas evolution and the combined type effects by virtue of its suddenness and high power (up to 6.000 tons of shattered rock mass) threaten the lives of miners, incapacitate expensive mining equipment and disrupt the rhythm of the work of potash mines [1-4].

The problem of forecasting of gas-dynamic phenomena in underground mining of potash deposits is studied for a long time. As a rule, to solve the problem is used geological-mathematical methods of forecasting. A statistical approach in predicting the gas-dynamic danger allows significantly reduce the dependence of the final prediction results on subjective factors by introducing a model of informative strictly defined criteria. However, it should be noted that 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.

Research in the field of construction of mathematical models of gas-dynamic localization of the danger areas in the potash mines traditionally were limited to the use of parametric models based on linear discriminant functions. Studies have shown that the parametric mathematical model of forecasting method for zones dangerous because of GDP is a powerful statistical method, both in depth analysis of the data and the value of the results belongs to the most effective methods of statistical analysis. The model is sufficiently simple, concise, well interpreted, convenient to consider the geometric illustrations of separation areas of the array into the dangerous and non-dangerous GDP zones. The parametric mathematical model of forecasting method allows to select the most informative indicators, which makes it possible to

reduce the dimension of the original space of the features and drop uninformative variables. In the case of compliance with the terms of multidimensional data distribution normality and equality of covariance matrices 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. Thus, in dealing with applied problems of forecasting zones dangerous because of GDP, parametric mathematical model based on linear discriminant analysis using robust estimators is the most preferred among the other well-known parametric models.

Statement of the problem

The objective is basing on a set of geological parameters to determine zone as dangerous or non-dangerous because of GDP, the particular area of deposit or a mine field. According to the mathematical model of forecasting method is the decision rule, which is substituting with the values of geological parameters, allows the highest possible chance to assess considered area from the standpoint of the gas-dynamic danger.

The mathematical formulation obtaining a decision rule is reduced to the definition of a finite set of geological parameters, the intervals of their values, inherent to both dangerous and non-dangerous zones because of GDP, as well as finding the most efficient in terms of the probability of a correct forecasting of statistical procedures.

Formally, each record of available geological data represents a multi-dimensional observation of P -dimensional space with a finite set of attributes X . The set of observations belonging to a class of D_i , forms a "cloud" in the same space. Therefore for successful forecasting is necessary that the "cloud" of D_i was concentrated in a certain region R_i of the space X and a small part of D_j , according to another class, got into region R_i . In general, the construction of a decision rule for classifying the multidimensional observation to a particular class can be viewed as the problem of finding k disjoint regions R_i ($i = 1, k$), satisfying the above conditions. If the P -dimensional observation of space X Если P -мерное наблюдение пространства X falls in the region R_i , then we assume that the decision on an object belonging to D_i is accepted. By using such a forecasting rule in some cases, an object can be assigned to another class. So we denote the probability of misclassification of objects through $P(i/j)$, then the criterion of the average probability of an erroneous forecast is as follows:

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

where a_i – a priori probability of the object occurrence from the domain D_i .

The solution of this task is carried out at presence of a training set, which should consist of observations relating to both classes (dangerous and non-dangerous GDP zones). Therefore we consider that the training set for D_i class is a sampling of a general set of objects with the distribution density of $P_i(X)$. Forecasting rule that minimizes the loss function Q is as follows:

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

To find the probability meaning of the expression (2), we assume that $P(X) = \sum_{j=1}^k a_j P(X/j)$. The

value $P(j/X) = a_j P(X/j) \times [P(X)]^{-1}$ represents the conditional probability that the P -dimensional observation X belongs D_i . Therefore, the expression (2) is called forecasting rule on a maximum of conditional probability, as it gives the smallest value of the loss function (1). In reality, the exact location of the "clouds" of objects and their density $P_i(X)$ is unknown, and there are training set that give a rough idea about it. The training set is a set of multivariate observations obtained from research. The matrix of the original data is also included observations belonging to any class that is not clearly established.

At present it is known a large number of pattern recognition methods for constructing decision rules on the training set, or the so-called statistical classification, which are divided into two groups: parametric and nonparametric methods [5-11]. The parametric methods of obtaining the decision rules for predicting traditionally is used the assumption of homogeneity of training samples for each class, measuring all the components of the feature vectors, accessories conditional probability distributions for multivariate Gaussian family, independence of sample values. The most well known of parametric methods is the linear discriminant analysis based on linear discriminant function (LDF). Discriminant analysis based on LDF is characterized by rigorous mathematical justification, physical interpretability, visual presentation of the results. The disadvantages of this method are the requirements for the homogeneity of samples, belonging to a multivariate normal distribution and independence of the sample values.

Non-parametric methods of obtaining the decision rules for predicting ignore a priori assumptions about

the homogeneity of the samples of each class, the availability of measurements of all components of the feature vector, Gaussian conditional probabilities of features distribution. The most well-known non-parametric methods is " k -nearest neighbor algorithm" method and Rosenblatt-Parzen window method [12-14]. These methods are relatively easy, work well in predicting, but require constant memorization of training samples and are very sensitive to small amounts of training samples in which the risk of misclassification increases dramatically. At the same time robustness factor (sustainability), representing a relative increase of the probability of classification error, with the volume of training data for more than 20 non-parametric methods are 2-4 times greater than for linear discriminant analysis based on the LDF [14]. To obtain the decision rules for predicting zones dangerous because of GDP is used training samples, the volume of which far exceeds the 20, allowing give preference to a parametric model for obtaining decision rules for forecasting zones dangerous because of GDP.

The essence of constructing the decision rule based on the linear discriminant function is as follows. If we denote by X_{ij} the variable number i at the observation point number j , taken from the sample, characterizing zones dangerous because of GDP, we obtain the matrix V of order $p \times n_1$ of results of this sample observations:

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

Let X_{ij} as the result of measurement of the variable with index i at point with index j , taken from the sample characterizing the non-dangerous zone because of GDP. The result is a matrix W of the order $p \times n_2$:

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

where p – number of variables (indicators); n_1, n_2 – the volume of the indicator samples.

Then using these data for each group are calculated the average and sum of the cross products of deviations from the averages. The average is calculated from the formula.

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

where k – the number of groups ($k = 2$); n_k – the size of the sample in the group k ; $j = 1, 2, \dots, p$ – variables.

Then find sum of the cross products of the deviations from the average:

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

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

The next step is the calculation of the combined variance matrix

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

where q – the number of groups ($q = 2$).

Calculate the overall average for both groups:

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

where \bar{X}_{ij} – averages of variable j in group k ; n_k – the size of the sample in group k ; $j = 1, 2, \dots, p$ – variables.

Then for each group calculated inverse elements of the combined dispersion matrix D :

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

Discriminant function coefficients are calculated according to the formula

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

where $i = 1, 2, \dots, p$; k – the number of groups.

Then constant of discriminant functions are calculated:

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

For the case of two classes D_1 and D_2 (dangerous and non-dangerous zones because of GDP) we obtain two linear discriminant function of the form

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

Subtracting less discriminant function from a greater is reduced to a single linear discriminant function of the form

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

For the case of two classes the use of linear discriminant functions based on two assumptions. The first assumption is that the regions R_1 and R_2 , containing the most part of objects corresponding to D_1 and D_2 classes (non-dangerous and dangerous zones because of GDP) can be divided by $(p-1)$ – dimensional hyperplane. The equation of this hyperplane and is an expression of (13). Here a_1, a_2, \dots, a_p – coefficients characterizing the slope of the hyperplane to the coordinate axes, and b – parameter characterizing the distance of the hyperplane to the origin. The hyperplane (13) splits the p -dimensional space into two half- R_1 and R_2 , which are defined by inequalities

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

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

Consequently, inequality (14) gives us a decision rule for classifying the multidimensional vector of observations for one of the two zones – dangerous or non-dangerous because of GDP. Let us agree to refer the observation to D_1 , if $z(X) \leq b$, and to D_2 – otherwise. Equation (14) allows to switch from p -dimensional random variables to the classification of one-dimensional value $z(X)$. Denote through z_i random value $z(X)$, when $X \in D_i$.

The second assumption provides a method for determining the coefficients of the separating hyperplane such that separation was the best. Suppose that the division will be the better, the farther apart the average values of the random variables z_1 and z_2 , the distance is measured in terms of the dispersion:

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

where $m_i = E z_i$; D_z – the dispersion of z_i ($i = 1, 2$), assumed the same for both classes.

Vector \bar{A} , delivering maximum $h^2(A)$, given by the equation

$$\bar{A}_F = \left[\sum (\bar{\mu}_2 - \bar{\mu}_1) \right]^{-1}, \quad (16)$$

where μ – a mean vector for D_i ; Σ – covariance matrix considered to be the same for both classes.

The maximum value of $h^2(a)$ is called the Mahalanobis distance between classes of D_1 и D_2 and equals

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

Mahalanobis statistics can be used to check the significance of the discriminant function. If we assume a normal multivariate distribution, then D^2 -statistics is distributed as χ^2 with $p(q-1)$ degrees of freedom. Comparing the calculated value D^2 with χ^2 at the 5 % significance level, if $D^2 > \chi^2$, one can say that the resulting linear discriminant functions have meaning and can be used as decision rules in forecasting. At the stage of exploratory analysis establish that the objects from D_i ($i = 1, 2$) are distributed according to a multivariate normal distribution with identical covariance matrix. Therefore, the decision rules (14) based on linear discriminant functions (10) and (11) are the best that is delivering to the functional Q (1) the smallest possible value.

It is necessary to note one more very important aspect in finding the decision rules by linear discriminant functions. If at the stage of exploratory analysis found "clogging" of training samples by "emissions" and these anomalous observations could be of interest in studies, then it is advisable to use the distribution parameter estimation procedures that are not sensitive to the data structure. Such evaluation procedure called robust or stable [6, 15-22]. The following robust procedures could be used in the permutation discriminant analysis for obtaining of stable decision rules: winsorized estimators, truncated assessment, Huber weighted estimates, Hampel piecewise-linear M -assessment. If there is distortion in the training samples and using classical estimates of distributions parameters in the decision rules the probability an erroneous forecast is significantly increased, therefore it is necessary to use robust estimates.

If there are "emissions" in the training samples, you can use any of the following stable estimates. If the training sample is formed from distributions with "heavy-tailed" or different from the normality is better to apply the weighted estimates of Huber and Hampel. Huber robust iterative estimation is defined the following relations:

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

$\hat{x}_i, \hat{\Sigma}_i$ – estimation of parameters calculated in the previous step; x_1, \dots, x_{n_i} – training sample for the class i ; n_i – the volume of training sample for the class i . Classic estimates of the parameters are used as an initial approximation.

In general parametric mathematical model of forecasting method for zones dangerous because of GDP, based on the discriminant analysis is a powerful statistical method both in depth analysis of the data and the value of the results belongs to the most effective methods of statistical analysis. The model is sufficiently simple, concise, well interpreted, convenient to consider the geometric illustrations of separation areas of the array into the dangerous and non-dangerous GDP zones. The parametric mathematical model of forecasting method allows to select the most informative indicators, which makes it possible to reduce the dimension of the original space of the features and drop uninformative variables. In the case of compliance with the terms of multidimensional data distribution normality and equality of covariance matrices 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.

The procedure for obtaining decision rules

Taking into account the existing ideas about the formation's mechanism of the gas-dynamic phenomena's centers are the salt rock mass of Verkhnekamskoie potash salt deposit [23, 24]. Formation of the training sample was conducted the following way. All documented cases of GDP in mining AB formation in the southern part of Verkhnekamskoie potash salt deposit on the mine fields of Mine-2 and Mine-4 and their corresponding geological indicators was a priori assigned to zone dangerous because of GDP (Group 2). The rest of the total sample was related to zone non-dangerous because of GDP (Group 1). We investigated 658 observations, 478 are classified as non-dangerous because of GDP (Group 1) and 180 as dangerous (Group 2). To obtain the decision rules we used the following indicators: X_1 – formation thickness B, m; X_2 – the content in the formation B of potassium chloride (KCl), %; X_3 – the content in the formation B of magnesium chloride ($MgCl_2$), %; X_4 – the content in the formation B of sodium chloride (NaCl), %; X_5 – the content in the formation B of bromine (Br), %; X_6 – the content in the formation B of

calcium sulfate (CaSO_4), %; X_7 – the content in the formation B of insoluble residue, %; X_8 – occurrence depth of formation B roof, m; X_9 – absolute mark of formation B roof, m.

Exploratory analysis of baseline data showed heterogeneity of the total sample, the deviation from the normal distribution, and the presence in the sample of "heavy tails". Due to the fact that not all of used indicators is equally useful for the separation of the formation into non-dangerous and dangerous because of GDP zones, it was necessary to identify a subset of indicators by which one can build "the best" split between the two groups. Stepwise discriminant analysis (SDA) was implemented to solve this problem using the following estimates: classical estimation, Huber weighted estimates and truncated assessment with truncation level of 0.1. SDA used the F-statistic based on one-way ANOVA criteria for the selection of the most informative indicators. In fact, the SDA logic is: initially to determine indicator for which the average values of the two groups "the most different". For each indicator, $X_1, X_2, X_3 \dots X_9$ difference is measured using the F-statistic and one indicator, which corresponds to a larger value of F, is selected (included). At each step of the procedure is considered conditional distribution of each indicator not included in the subset for given including indicators. Then from among the non-included indicators the one for which the average values of the two groups "the most different" is determined once again. This difference is determined by using F-statistics. The process ends when none of indicators do contribute significantly to the group separation. In the process of implementing SDA were accepted standart value of F-inclusion minimum of variables not included in the sought subset and F-delete the selected variables at a significance level of 0.05. In step 8, the value of F-delete statistic was less than accepted minimum, equal to 3.0, hence "the best" discriminant equation are defined in step 7. Thus indicators $X_1, X_2, X_3, X_4, X_5, X_6$ and X_7 were included in the sought subset and indicators X_8 and X_9 – not included. Analysis of classification matrices in the implementation of the SDA procedure using classical estimation, Huber weighted estimates and truncated assessment with truncation level of 0.1 showed the following. In the application of classical assessments were correctly classified 62.3 % of the original grouped cases. In case of using Huber weighted estimates and truncated assessment with truncation level of 0.1 were correctly classified 61,7 % and 61,5 % of the original grouped cases.

The validity of the primary classification using both classical and robust procedure is virtually the same, and more than 60 %. So the final choice of the

decision rule was held after the reclassification. At the same time raw data were regrouped, and then reclassification procedure was carried out, which resulted in the refinement coefficients of linear discriminant functions for each of the groups, as well as the assessment of the correctness of the classification has been given. For each observation with the use of the obtained discriminant functions and D^2 -statistics of Mahalanobis posterior probabilities of belonging to a group dangerous and non-dangerous because of GDP were calculated. Based on these probabilities each observation it was classified as belonging to one of the groups and classification matrix was compiled. Table 1 shows the classification results using the classical procedures, using weighting Huber estimates, truncated estimates (truncation level of – 0.1).

Table 1

Classification matrix based on the results of the linear discriminant analysis

Group	Classification results				Total %
	non-dangerous because of GDP		dangerous because of GDP		
	abs.	%	abs.	abs.	
<i>Using classical estimation</i>					
GDP non-dangerous (1)	323	91.76	29	8.24	352
GDP dangerous (2)	18	5.88	288	94.12	306
Total	341	51.82	317	48.18	658
<i>Using Huber weighted estimates</i>					
GDP non-dangerous (1)	334	94.89	18	5.11	352
GDP dangerous (2)	8	2.61	298	97.39	306
Total	342	51.98	316	48.02	658
<i>Using truncated assessment with truncation level of 0.1</i>					
GDP non-dangerous (1)	319	90.63	33	9.37	352
GDP dangerous (2)	24	7.84	282	92.16	306
Total	343	52.13	315	47.87	658

As seen from Table 1 using classical statistical procedures of linear discriminant analysis was correctly classified 92.86 % of re-grouped observations.

By using linear discriminant analysis with weighted estimates of Huber the probability of correct classification is 96 %.

When used in linear discriminant analysis truncated assessment with truncation level of 0.1 the probability of correct classification is 91.34 %. By the criterion of the highest probability of correct classification we accept for forecasting zones dangerous because of GDP the decision rule obtained using weighting Huber estimates. Table 2 shows constants and coefficients of discriminant functions found using weighted estimates of Huber.

Since in this case there are two groups, two discriminant functions can be reduced to a single subtracting coefficients and constant of smaller

function from the larger. After subtracting we obtain a linear discriminant function of the form:

$$F_p = -19.98X_1 - 1.16X_2 + 31.18X_3 - 3.64X_4 - 747.02X_5 + 6.87X_6 - 4.57X_7 + 310.72 \geq 0, \quad (19)$$

where X_1 – formation thickness B, m; – the content in the formation B of potassium chloride

(KCl), %; X_3 – the content in the formation B of magnesium chloride ($MgCl_2$), %; X_4 – the content in the formation B of sodium chloride (NaCl), %; X_5 – the content in the formation B of bromine (Br), %; X_6 – the content in the formation B of calcium sulfate ($CaSO_4$), %; X_7 – the content in the formation B of insoluble residue, %.

Table 2

The constants and coefficients of the discriminant functions obtained using Huber weighted estimates

Group	Constants of the discriminant functions	Coefficients of the discriminant functions						
		X_1	X_2	X_3	X_4	X_5	X_6	X_7
GDP non-dangerous (1)	-385755.60	998.66	7678.37	16092.34	7734.02	-11769.6	7829.14	7715.20
GDP dangerous (2)	-385444.89	978.68	7677.20	16123.52	7730.39	-12516.62	7836.01	7710.63

Expression (19) is the desired decision rule for classifying AB formation areas to the zones, dangerous or non-dangerous because of GDP. When $F_p \geq 0$ observation point refers to the area dangerous because of GDP, otherwise to non-dangerous. For the probabilistic interpretation of the discriminant function, which is more convenient in constructing forecast maps, it is advisable to use the expression

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

where P_l – the probability corresponding to the largest discriminant function; q – the number of groups ($q = 2$); f_k, f_l – the value of the k -th and the largest discriminant function correspondingly.

Substituting the values of indicators in discriminant functions, which constants and coefficients are shown in Table. 2, we obtain some values of the discriminant functions. Then, substituting the obtained values of the discriminant functions in the expression (20), we find the probability of classifying the observation point to the zone, which value of the discriminant function larger. The value of P_l varies from 0 to 1 the boundary is a probability value of 0.5.

Assessing the adequacy of the decision rules and forecast maps of zones dangerous because of GDP

To assess the adequacy of obtained decision rules to practice of mining in the forecast maps of zones dangerous because of GDP were marked the places where under conduction of preparatory and sewage mining operations occurred gas-dynamic phenomena. In developing AB formation of Mine-1 field, 8 out of 98 GDP-cases are not included in forecasting zones dangerous because of GDP (7.8 % of the total). Thus,

the adequacy of the zones forecasting results dangerous because of GDP in the formation AB of sylvinitic mine

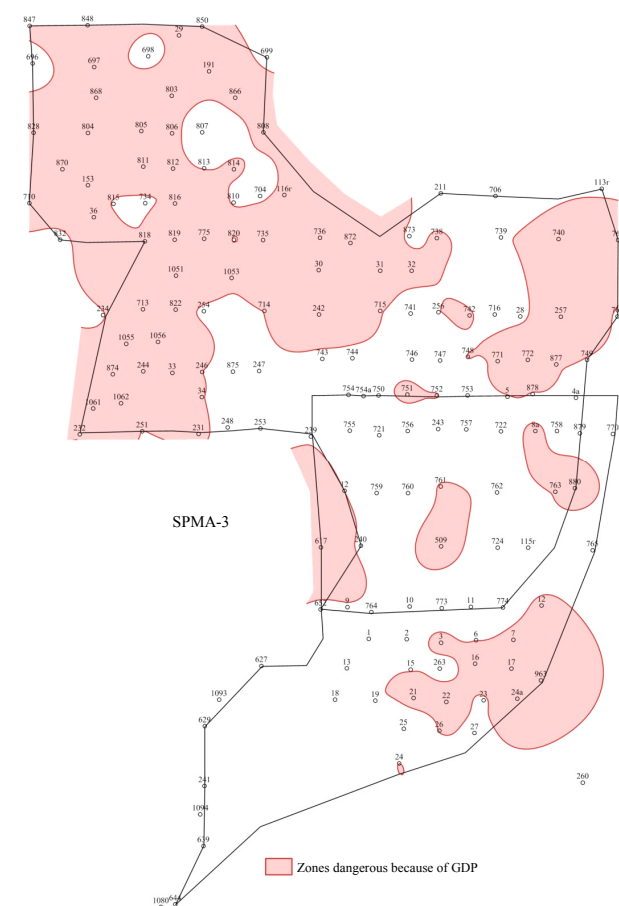


Fig. Map of zones dangerous because of GDP from a the roof of mine workings at mining AB formation in conditions of Polovodovsky sector of Verkhnekamskoiye potash salt deposit

field Mine-2 to practice of mining operations is 92.2 % correspondingly. To verify the performance obtained decision rules for forecasting zones dangerous because of GDP, the original sample was supplemented with

data on the mine fields of Mine-1 and Mine-4 initially not included in the original data. Then examination of new input data classification was carried out, which correctness for formation of AB was 96.7 %. Substituting the values of geological parameters obtained as a result of the wells drilling, detailed exploration from the surface, from underground exploration wells and furrow sampling data in underground mines, into decision rules allowed to construct forecast maps of zones dangerous because of gas-dynamic phenomena.

The figure shows a map of zones dangerous because of GDP from the roof of mine workings at mining AB formation for conditions of Polovodovskii area of Verkhnekamskoie potash salt deposit.

Conclusion

As a result of scientific research a mathematical model of the method forecasting zones dangerous because of gas-dynamic phenomena based on linear discriminant analysis using robust (resistant)

procedures, allowing to increase the reliability of multivariate forecasting was created. A method for forecasting of zones dangerous because of gas-dynamic phenomena in the exploration and development of AB sylvinite formation using in decision rules the following complex of geological parameters: formation thickness, the content of potassium chloride, the content of magnesium chloride, the content of sodium, the content of bromine, the content of calcium sulfate, the content of insoluble residue in formation rocks. The probability of a correct discrimination of AB sylvinite formation sections into zones dangerous and non-dangerous because of GDP is 0.96. Adequacy of zones forecasting results dangerous because of GDP to practice of mining operations on the formation of AB sylvinite formation is 92 % correspondingly. Forecast maps were created for zones dangerous because of GDP for the Ust-Yaivinskiimine field conditions on the base of PJSC "Uralkali" and Polovodovskii area of Verkhnekamskoie potash salt deposit.

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., Skripkin 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.
13. Epanenchikov V.A. Neparаметричeskaiia otsenka mnogomernoi plotnosti veroiatnostei [Non-parametric estimation of a multidimensional probability density]. *Teoriia veroiatnosti i ee primeneniie*, 1969, vol.14, no.1, pp.156-160.
14. Fukunaga K. Vvedenie 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 gualitative 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., Rausseau 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.

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.

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.

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., Ронchetti 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.

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

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