

UDC 622.276:550.832.92.05

Article / Статья

© PNRPU / ПНИПУ, 2019

USING DISCRIMINANT ANALYSIS FOR THE INTERPRETATION OF GAS LOGGING DATA ON THE EXAMPLE OF THE PAVLOV OIL FIELD

Mikhail N. Rastorguev

PITC «Geofizika» (16a, Petropavlovskaya st., Perm, 614000, Russian Federation)

ИСПОЛЬЗОВАНИЕ ДИСКРИМИНАНТНОГО АНАЛИЗА ДЛЯ ИНТЕРПРЕТАЦИИ ДАННЫХ ГАЗОВОГО КАРОТАЖА НА ПРИМЕРЕ ПАВЛОВСКОГО НЕФТЯНОГО МЕСТОРОЖДЕНИЯ

М.Н. Расторгуев

Пермский инженерно-технический центр «Геофизика» (614000, Россия, г. Пермь, ул. Петропавловская, 16а)

Received / Получена: 12.10.2018. Accepted / Принята: 14.01.2019. Published / Опубликовано: 29.03.2019

Key words:

field, oil saturation, reservoir, gas logging, geological and technological research, well logging, chromatography, luminescent-bitumen analysis, gamma-ray logging, neutron-neutron logging, hydrocarbon gases, drilling sludge, drillingmud, linear discriminant analysis, mathematical statistics, probability.

The article is devoted to the development of new methodology of gas-logging interpretation based on materials from drilled wells of the Pavlovsky oil field in the Lower-Middle-Visayan terrigenous oil and gas complex in the interval of the Tula terrigenous horizon C_{1td} , with using elements of mathematical statistics - stepwise discriminant analysis, for which the gas chromatographic data were used as variables - the total gas content of hydrocarbons in the drill mud and component composition of the gas-air mixture, as well as data of luminescent-bitumen analysis of drill cuttings. In addition, logging data were used - gamma-ray logging, neutron-neutron logging for thermal neutrons and the difference between the bit diameter and the well diameter. Based on the results of the analysis, the probability of attributing observations to oil-saturated sandstones, which takes into account gas logging and well logging, was calculated. In addition to oil-saturated sandstones, rocks occurring in this oil-gas-bearing complex were studied: mudstones, aleurolites, clayey sandstones and non-oil saturated sandstones. To visualize the results, a geological and geophysical diagram was plotted on which were shown: rock saturation according to well survey, rock saturation according to well logging and lithological column according to well survey and logging, as well as all variables that participated in the discriminant analysis. Based on the discriminatory analysis, three ranges of values were identified: 1) an area with observations related to oil-saturated rocks; 2) an area with observations related to non-oil-saturated rocks and 3) an area with observations of unclear saturation with indicators that are intermediate (the transition zone is possibly oil-saturated or washed reservoirs).

Ключевые слова:

месторождение, нефтенасыщенность, коллектор, газовый каротаж, геолого-технологические исследования, геофизические исследования скважин (ГИС), хроматография, люминесцентно-битуминологический анализ, гамма-каротаж, нейтрон-нейтронный каротаж, углеводородные газы, буровой шлам, буровой раствор, линейный дискриминантный анализ, математическая статистика, вероятность.

Статья посвящена разработке новой методики интерпретации газового каротажа по материалам пробуренных скважин Павловского нефтяного месторождения в нижне-средневизейском терригенном нефтегазоносном комплексе в интервале тульского терригенного горизонта C_{1td} с применением элементов математической статистики – пошагового дискриминантного анализа, для проведения которого в качестве переменных были использованы данные газового хроматографического анализа – суммарное газосодержание нефтенасыщенных углеводородов в буровом растворе и покомпонентный состав газозвушной смеси, а также данные люминесцентно-битуминологического анализа бурового шлама. Кроме того, были использованы данные геофизического исследования скважин (ГИС) – гамма-каротажа, нейтрон-нейтронного каротажа по тепловым нейтронам и разница между диаметром долота и диаметром скважины. По результатам анализа была рассчитана вероятность отнесения наблюдений к нефтенасыщенным песчанникам, которая учитывает показатели газового каротажа и ГИС. Также при разработке методики, кроме нефтенасыщенных песчанников, исследовались породы, попадающиеся в данном нефтегазоносном комплексе: аргиллиты, алевролиты, глинистые песчанники и нефтенасыщенные песчанники. Для визуализации результатов была построена геолого-геофизическая диаграмма, на которой были нанесены: насыщение пород согласно геолого-технологическим исследованиям (ГТИ), насыщение пород согласно ГИС и литологическая колонка по данным ГТИ и ГИС, а также все переменные, которые участвовали в дискриминантном анализе. На основании проведенного анализа были выделены три области значений: 1) область с наблюдениями, относящимися к нефтенасыщенным породам; 2) область с наблюдениями, относящимися к нефтенасыщенным породам и 3) область с наблюдениями неясного насыщения с показателями, занимающими промежуточное положение (переходная зона – возможно, нефтеводонасыщенные или промытые пласты).

Mikhail N. Rastorguev – interpreter engineer of geological and technological research group (tel.: +007 342 212 14 15, e-mail: mrastorguev@pitc.ru).

Расторгуев Михаил Николаевич – инженер-интерпретатор группы геолого-технологических исследований (тел.: +007 342 212 14 15, e-mail: mrastorguev@pitc.ru).

Introduction

In recent years, more and more forces have to be spent on the fields of the Perm Region to increase, and in some cases just to maintain hydrocarbon production at the current level. All the oil producing and oilfield services companies of the Perm Region are now concerned about this problem. The role of mud logging (ML) in ensuring of the wells drilling optimization is increasing. The range of tasks and requirements for the research quality, the speed of obtaining information, the filling of the ML complex is expanding.

Currently, ML is an integral part of geological and geophysical studies of oil and gas wells and is designed to monitor the well at all stages of its construction and commissioning to study the geological structure, to achieve high technical and economic indicators, and also ensure compliance with environmental requirements [1].

ML is carried out directly while drilling a well, without downtime in the work of the drilling crew and drilling equipment; solve a set of geological and technological tasks aimed at quickly identifying promising oil and gas reservoir formations, study their reservoir properties and saturation, optimize core sampling, express testing and selected objects logging, ensuring trouble-free wells drilling and drilling mode optimization. Gas logging is included in the ML complex and makes up a substantial part of it [2–3].

It is possible to divide all scientific and technical developments carried out in the field of mud logging into two groups: improving equipment performance (including sensors) and increasing the capacity of data interpretation. The development and introduction of new equipment is important, but they require a lot of material costs as compared to improving the method of interpreting the data collected by the equipment currently installed at ML stations. Therefore, it is proposed to improve the quality

of interpretation as a less costly process. In particular, to improve the efficiency of gas logging, it is possible to use methods of mathematical statistics - stepwise discriminant analysis, the methods and results of which are described in this article.

General geological characteristics of the Pavlov's field

The geological section of the Pavlov's field has been studied using materials from structural, exploration and production wells to a depth of 2,243 m and is represented by rocks from the Quaternary to the Upper Riphean age.

The unified stratigraphic scheme of the Russian platform, approved in 1988, is the basis for the dismemberment of the geological section. Unfortunately, not all of the drilled wells have sufficient core material with paleontological characteristics of the rocks, so the separation of stratigraphic horizons was carried out on the basis of a comparison of well sections of the Pavlov's field with sections of other areas (Krasnoyarsk-Kuedinskaya, Batyrbaiskaya, Tanypskaya and others).

In the modern regional tectonic plan, the territory of the Pavlov's field is confined to the area of the Chernushka rampart-like zone complicating the northern slope of the Bashkir arch (Fig. 1). The width of the rampart in the south is 25–30 km, in the north, in the area of the Tanyp Rise, 10–15 km. The rampart can be traced across all horizons of the Upper Paleozoic.

The Pavlov's field is associated with Upper Devonian reef massifs within the onboard zone of the Kama Kinel system of deflections that form the uplifts: Berezovskoye, Detkinskoe, Baranovskoye, Ulykskoe, Pavlovskoye, Yuzhno-Pavlovskoye, Grigorievskoye.

In Pavlov's field commercially oil bearing reservoirs are: Upper-Tournaisian carbonate, lower-srednevezysky terrigene, Oka-Serpukhpvsko-Bashkir carbonate, Moscow terrigenous-carbonate.

Techniques of the gas logging interpreting are described in [1]. The primary data obtained from the well are processed in the interpretation center. Interpreter, studying the relative composition of the gas and the absolute percentage of hydrocarbon gases in the gas-air mixture, coming together with the drilling fluid from the well, gives a conclusion about the layers saturation. At the same time, errors associated with differences in the approach to interpretation are possible,

i.e. human factor. It is proposed to use mathematical methods in order to create a uniform interpretation method, which will take into account all the necessary parameters for the reservoirs identification according to the type of fluid saturation and, thus, will mathematically explain the interpretation of gas logging, i.e. selection of oil-saturated rocks in the studied intervals. One possible tool for this purpose may be linear discriminant analysis.

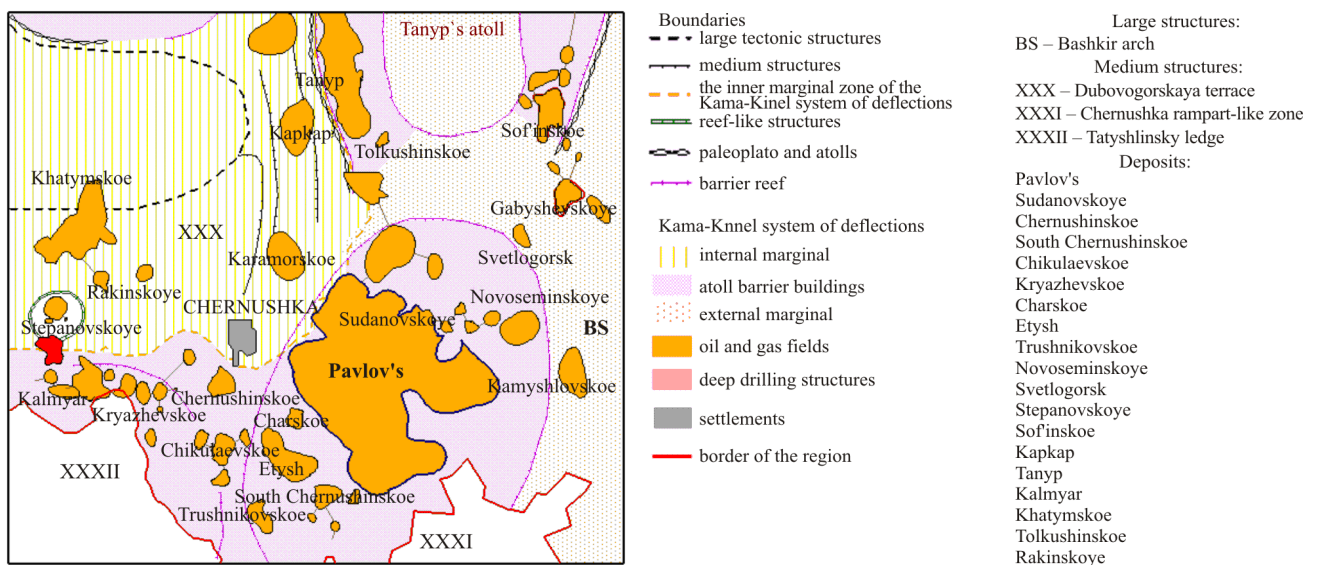


Fig. 1. Copying from the consolidated tectonic zoning scheme of the Perm region

The development of multidimensional models for oil-saturated sandstones allocation

Discriminant analysis is used to decide which variables distinguish (discriminate) two or more arising aggregates (groups). We first define the concept of separation (discrimination) and show how it differs from the concept of classification. Suppose that there are two groups of samples of shale, which are known in advance that they were formed in the freshwater and marine basins. This can be determined on the basis of a study of the remains of fossil organisms. Fossil residues in freshwater and marine basins

differ, which makes it easy to distinguish shale from each other, but in practice there are examples of shale without fossil residues, so it is necessary to find another criterion for the separation of these groups of shale. A number of geochemical characteristics were measured in the samples: the content of vanadium, boron, iron and other elements. The task is to find such a linear combination of these variables, which will give the maximum possible difference between the two previously defined groups. If we manage to find such a function, then we will be able to use it to assign new samples to one or another initial group. In other words, new samples of shale that do not contain diagnostic fossil residues

can be divided into marine and freshwater based on the linear discriminant function built on their geochemical components (This problem was considered by Potter, Shimp, and Witters [4–5]).

The classification task can be illustrated with a similar example. Suppose we collected a large collection of shale samples, each of which was subjected to geochemical analysis. Is it possible, based on the values of the measured variables, to divide the sample into relatively homogeneous groups (clusters) that are different from each other? Numerical methods for solving such problems are well developed and belong to a branch of science called taxonomy. There are several obvious differences between these methods and discriminant analysis methods. The classification is internally closed, that is, unlike the discriminant analysis, it does not depend on a priori information about the ratio between the samples. In the discriminant analysis, the number of groups is specified in advance, while the number of clusters that are obtained as a result of classification cannot be predetermined. Each sample from the original set in the discriminant analysis belongs to one of the specified groups. In most classification tasks, a sample can be included in any of the groups resulting from the classification. Other differences will become apparent when considering these two procedures. As a result of shale cluster analysis samples are distributed into groups. It is interesting to conduct geological comprehension thus found groups [5].

In the same way, the problem considered in this article can also be described: in order to determine which variables belong to certain intervals with oil-saturated rocks, data on the following variables were collected:

1. G_{sum} , abs. % – the total gas content of oil-saturated hydrocarbons in the drilling fluid;
2. C_1 , rel. % – Is the relative content of methane (CH_4) in the gas-air mixture;

3. C_2 , rel. % – Is the relative content of ethane (C_2H_6) in the gas-air mixture;

4. C_3 , rel. % – Is the relative content of propane (C_3H_8) in the gas-air mixture;

5. C_4 , rel. % – Is the relative content of butane (C_4H_{10}) in the gas-air mixture;

6. C_5 , rel. % – Is the relative content of pentane (C_5H_{12}) in the gas-air mixture;

7. LBA, points – data of the sludge luminescent-bitumen analysis;

8. GR, $\mu\text{R}/\text{h}$ – gamma ray logging readings;

9. NNT, rel. units – readings of neutron-neutron logging for thermal neutrons;

10. DS, mm – the difference between the diameter of the bit and the diameter of the well.

Log data (gamma logging, neutron-neutron logging for thermal neutrons and borehole diameter) were added to divide more precisely the sample into the five classes: argillite, aleurolite, clayey sandstone, non-oil saturated sandstone, oil-saturated sandstone.

Then you can use the discriminant analysis for the determination of such a decision rule (discriminant function), which would allow assigning a specific interval to a particular class.

For the study, data from the well gas logging of the Tula terrigenous horizon C_{1tl} in the Lower-Middle-Visayan terrigenous oil and gas complex of the Pavlov's field were taken. After collecting the data, the number of observations in different classes varied greatly, so a random sample was taken in order to equalize the number of observations in all classes (total number of observations $n = 1535$).

At the first stage of the classification attempt, it is necessary to compare the average values in the classes (Table 1).

This shows that the fluids from the oil-saturated intervals have a more "heavy" gas composition (the predominance of complex hydrocarbons compared to the other intervals). The highest values of G_{sum} (0.184) correspond to oil-saturated sandstone, lower values (0.100) – to clayey sandstone; most

likely, these readings are related to the transition zone between oil-saturated sandstone and argillites. The component composition of gas in oil-saturated sandstones indicates a high content of “heavy” hydrocarbon gases and a low methane content and inverse characteristics in other classes, for example, mudstone has the highest relative methane content and low relative gas content (from ethane to pentane). The highest average values of the DS variable are attributed to the

argillites, which is probably due to the outfalls from the borehole walls, the values of the DS in the oil-saturated and non-oil-saturated sandstones are the lowest, which can be explained by the mud cake formation.

Thus, for the allocation of oil-saturated intervals, all these variables should be used together. This problem was solved using elements of mathematical statistics – stepwise discriminant analysis, the results of which are described in this article.

Table 1

Average values of variables in classes $\left(\frac{x \pm \sigma}{\text{min} - \text{max}} \right)$

Indicator	Argillite	Aleurolite	Clayey sandstone	Non-oil saturated sandstone	Oil saturated sandstone
Gsum, abs. %	$\frac{0.040 \pm 0.074}{0.001 - 0.388}$	$\frac{0.050 \pm 0.098}{0.001 - 0.551}$	$\frac{0.100 \pm 0.375}{0.001 - 2.599}$	$\frac{0.039 \pm 0.079}{0.001 - 0.396}$	$\frac{0.184 \pm 0.359}{0.004 - 2.899}$
C ₁ , rel. %	$\frac{70.359 \pm 18.638}{25.698 - 100.000}$	$\frac{66.701 \pm 14.776}{34.712 - 94.820}$	$\frac{66.358 \pm 13.032}{36.034 - 96.748}$	$\frac{69.026 \pm 14.918}{28.974 - 100.00}$	$\frac{61.648 \pm 13.938}{23.155 - 91.704}$
C ₂ , rel. %	$\frac{15.881 \pm 11.685}{0.001 - 52.936}$	$\frac{17.439 \pm 8.831}{0.001 - 45.052}$	$\frac{18.348 \pm 8.795}{0.001 - 55.847}$	$\frac{16.969 \pm 9.415}{0.001 - 45.240}$	$\frac{18.146 \pm 6.863}{4.552 - 39.829}$
C ₃ , rel. %	$\frac{7.746 \pm 7.378}{0.001 - 29.060}$	$\frac{9.955 \pm 6.500}{0.001 - 27.323}$	$\frac{10.845 \pm 7.365}{0.001 - 37.057}$	$\frac{10.190 \pm 8.014}{0.001 - 41.480}$	$\frac{12.628 \pm 6.981}{0.001 - 41.287}$
C ₄ , rel. %	$\frac{4.309 \pm 4.843}{0.001 - 20.670}$	$\frac{4.029 \pm 3.393}{0.001 - 14.429}$	$\frac{3.253 \pm 3.258}{0.00 - 14.205}$	$\frac{2.684 \pm 3.195}{0.001 - 18.533}$	$\frac{5.158 \pm 4.608}{0.001 - 22.977}$
C ₅ , rel. %	$\frac{1.687 \pm 2.632}{0.001 - 13.265}$	$\frac{1.896 \pm 3.764}{0.001 - 41.186}$	$\frac{1.216 \pm 2.039}{0.001 - 9.792}$	$\frac{1.160 \pm 1.817}{0.001 - 10.325}$	$\frac{2.397 \pm 3.276}{0.001 - 15.194}$
NNT, rel. units	$\frac{2.397 \pm 2.216}{0.573 - 16.418}$	$\frac{7.540 \pm 2.552}{4.000 - 14.976}$	$\frac{9.751 \pm 5.120}{2.459 - 22.353}$	$\frac{8.243 \pm 4.902}{2.807 - 23.236}$	$\frac{8.835 \pm 5.606}{2.532 - 22.406}$
GR, $\mu\text{R/h}$	$\frac{10.620 \pm 3.793}{0.930 - 20.441}$	$\frac{14.200 \pm 2.475}{10.119 - 26.025}$	$\frac{7.014 \pm 1.597}{4.26 - 10.048}$	$\frac{2.836 \pm 0.812}{1.065 - 4.495}$	$\frac{2.807 \pm 0.791}{0.910 - 4.485}$
DS, mm	$\frac{17.697 \pm 24.779}{-8.385 - 151.745}$	$\frac{4.823 \pm 6.430}{-14.000 - 29.000}$	$\frac{1.111 \pm 4.442}{-7.400 - 39.635}$	$\frac{-1.267 \pm 2.631}{-7.148 - 26.000}$	$\frac{-1.428 \pm 2.986}{-13.000 - 30.000}$

Performed analysis of the average values of densities and distributions of variables studied showed that it is not possible to divide the above variables into any of the classes. Therefore, for complex (joint) use of the studied parameters, we will use linear discriminant analysis (LDA). Its capabilities for solving similar problems are given in [6–21].

Simple linear discriminant function converts original set of measurements included

in the sample into a single discriminant number. This number, or converted variable determines the position of the sample on the line defined by the discriminant function. Therefore, we can imagine the discriminant function as a way to transform a multidimensional problem into a one-dimensional problem [22, 23].

Discriminant analysis is based on finding a transform that gives the minimum ratio of the

difference between multidimensional means for a certain pair of groups and multidimensional dispersion within two groups. If we depict our two groups as sets of points in a multidimensional space, then it is easy to find a direction along which these sets are clearly separated and at the same time have the smallest convexity. We will show on the graph the possibility of identifying oil-saturated sandstones and mudstones by the two most informative parameters (Fig. 2). If we use the variables “GR, $\mu\text{R/h}$ ” and “LBA, points”, then we cannot achieve a satisfactory identification of groups A (oil sands) and B (mudstones). However, it is possible to find a direction along which the separation of the sets is obvious, and the convexity is minimal. The coordinates of points of this direction are given by the equation of a linear discriminant function. The overlaps of distributions for groups A and B along the axes “GR, $\mu\text{R/h}$ ” and “LBA, points” are indicated; projecting on the discriminant line allows to distinguish two groups [5].

Discriminant analysis has the following objectives:

1. Definition of discriminant functions or linear combinations of independent variables that best distinguish (discriminate) the categories (groups) of the dependent variable.
2. Check for the existence of significant differences between groups in terms of independent variables.
3. Identify predictors that contribute the most to intergroup differences.
4. The assignment of cases to one of the groups (classification) based on the values of predictors.
5. Evaluation of the data classification accuracy into groups [24–29].

Discriminant function is a linear combination of independent variables derived by discriminant analysis, with which one the categories of the dependent variable may be best distinguished (discriminated) [21].

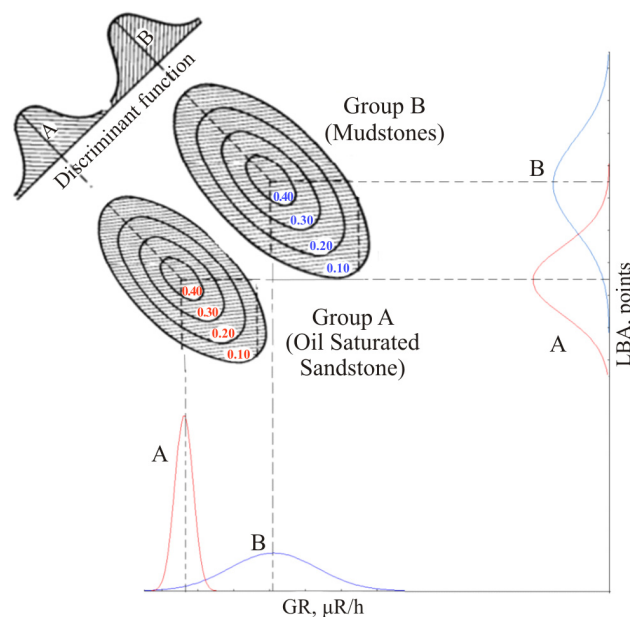


Fig. 2. Graphic representation of two two-dimensional distributions

The discriminant analysis method is described by the number of categories the dependent variable has. If it has two categories, the method is called two-group discriminant analysis. If three categories or more are analyzed, the method is called multiple discriminant analysis. The main difference between them is that if there are two groups, only one discriminant function can be derived. Using multiple discriminant analysis, several functions can be calculated [30, 31].

Using LDA, it is possible to build optimal surfaces (discriminant functions) Z in the feature space dividing the entire space into regions corresponding to objects of different classes. These surfaces (functions) serve as boundaries between regions and provide optimal separation of objects belonging to different classes. The classification rule in this case consists in determining, by the magnitude of the discriminant function, that the object belongs to one of the selected areas. Belonging to the corresponding area means belonging to the corresponding class [5].

The number of discriminant functions is defined as $K-1$ or $P-1$, where K is the number of

groups, and P is the number of variables. As a rule, the smallest of these numbers is chosen to determine the number of necessary discriminant functions [32–34]. This analysis is used 10 variables and 5 groups, i.e. discriminant functions should be four.

Performing discriminant analysis includes the following steps: problem formulation, calculation of the discriminant function coefficients, determination of significance, interpretation and reliability verification.

The first step in discriminant analysis is to formulate a problem by defining goals, dependent variables and independent variables. The dependent variable must consist of two or more mutually exclusive and mutually exhaustive categories.

If the dependent variable is measured by the interval or ratio scale, it should be first of all converted into a categorical status. The next step is to split the sample into two parts. One of them – the analysis sample – is used to calculate the discriminant function. The other part, the validation sample, is designed to test the discriminant function. This is called double cross validation [34–38].

Often the distribution of the number of cases in the analysis and validation samples is evident from the distribution in the total sample. For example, if the total sample contains 50 % of oil-saturated and 50 % non-oil-saturated intervals, then the analysis and validation samples should each contain 50 % of oil-saturated and 50 % non-oil-saturated intervals. In another case, if the sample contains 25 % of oil-saturated and 75 % non-oil-saturated intervals, you should select the analysis and validation samples in such a way that their distributions reflect a similar pattern (25 vs. 75 %).

Finally, the validation of the discriminant function is proposed to be performed repeatedly. Each time the sample should be divided into two parts: for analysis and

verification. Calculate the discriminant function and perform a model reliability analysis. Thus, the assessment of reliability is based on a number of tests.

After determining the analysis sample, we can calculate the discriminant function coefficients by using two methods. Direct method – calculation of the discriminant function with the simultaneous introduction of all predictors.

In this case, each independent variable is taken into account. Moreover, its discriminatory power is not taken into account. This method is more suited to the situation where the analyst, based on the results of a previous study or a theoretical model, wants all the predictors to be in the distinction basis.

In stepwise discriminant analysis, predictors are introduced sequentially, based on their ability to distinguish (discriminate) groups. This method is best used in situations where the researcher wants to select a subset of predictors for inclusion in the discriminative function. There is a two stepwise analysis of the varieties:

1. Stepwise analysis with inclusion. In a stepwise analysis of discriminant functions, the model of discrimination is built in steps. More precisely, at each step, all variables are viewed and the one that makes the greatest contribution to the difference between groups is found. This variable must be included in the model at this step, and the transition to the next step takes place.

2. Stepwise analysis with the exception. You can also move in the opposite direction; in this case, all variables will be included in the model first, and then at each step, variables that make a small contribution to the predictions will be eliminated. Then, as a result of successful analysis, only important variables in the model can be saved, i.e. those variables whose contribution to discrimination is more than others [39–40].

Mathematically, discriminant functions can be described by the following equation:

$$Z_{ik} = a + b_1 y_{i1} + b_2 y_{i2} + \dots + b_p y_{ip},$$

where a is a constant; b – the standardized coefficient (with average 0 and selective dispersion 1), allows to estimate

the contribution of each variable to this discriminant function [7].

Let's consider the possibility of using the linear discriminant function method on the example of a number of wells in the Pavlov's field in the interval of the Lower-Middle-Vissean terrigenous oil and gas complex. The values of gas logging parameters obtained at each point were grouped into five classes: argillite, aleurolite, sandstone, clayey sandstone, and oil-saturated sandstone. The grouping was carried out according to the mud log interpretation data.

Thus, a grouping variable (rock) and independent variables (gas logging parameters – G_{sum} , C_1 , C_2 , C_3 , C_4 , C_5 , LBA and log data – GR, NNT, and DS) will be used for discrimination. Since all these variables have different units of measurement, there are situations when a more significant parameter, with a small value of magnitude, overlaps with a less significant one, whose value is larger. Thus, we get information of poor quality, i.e. data may not be comparable. To eliminate such a situation, data was normalized – the transformation of formal parameters or criteria, expressed in general terms in different units, to a dimensionless form for the purpose of their comparison and comparative assessment. The following formula was used for normalization:

$$X'_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}.$$

Further, in order to follow up on what happens at each step of the discriminant analysis, a stepwise analysis was performed with the inclusion, allowing introducing variables into the model one by one, each

time choosing the one that makes the greatest contribution to discrimination. This analysis will be carried out until one of the following events occurs:

1. All variables are entered or discarded.
2. The maximum number of steps has been reached.
3. There are no other variables outside the model that have a larger F statistic value than the specified F-enable value, equal to 1, and when there are no other variables in the model that have a smaller F value than the F-delete value, equal to 0. For stepwise analysis with inclusion variables for inclusion are selected, giving the most significant single (additional) contribution to discrimination between groups, i.e. variables are selected with the largest F value (greater than the corresponding F-enable value, equal to 1). When executing steps with an exception, the least significant variables are selected for exclusion, i.e. variables with the smallest F value (smaller than the corresponding F-delete value, equal to 0).
4. Any variable in the next step has a tolerance value less than the tolerance value of 0.01. At each step, for each variable, the multiple correlation (R^2) with all other variables that were included in the model is calculated. Thus, the value of the tolerance of a variable is calculated as $1 - R^2$, therefore the value of tolerance is a measure of the redundancy of the variable. It should also be noted that when one or more included variable becomes too redundant, the matrix of dispersions/covariances for the variables included in the model may be irreversible and discriminant analysis cannot be performed. Consequently, the marginal value of tolerance was set to 0.01, because if the variable included in the model is reduced by 99 % with other variables, then its practical contribution to improving the quality of discrimination is very insignificant. More importantly, if you set

a significantly lower tolerance value, then rounding errors can lead to unstable results [4].

The results of the stepwise discriminant analysis are presented in Table. 2. Significance criteria were obtained, which are listed below.

Wilks' lambda criterion is the standard statistic used to denote the statistical significance of the discrimination power in the current model. Its value changes from 1 (there is no discrimination) to 0 (full discrimination).

Table 2

The results of the discriminant functions analysis

Variable	Wilks' lambda	Partial Lambda	<i>F-exclude</i> (2.447)	<i>p-level</i>
GR	0.390462	0.232090	1258.947	0.000000
LBA	0.118121	0.767203	115.457	0.000000
NNT	0.107498	0.843015	70.856	0.000000
DS	0.102912	0.880579	51.602	0.000000
C ₄	0.092164	0.983277	6.471	0.000037
C ₂	0.092219	0.982682	6.706	0.000024
C ₁	0.091740	0.987821	4.691	0.000915
G _{sum}	0.091702	0.988229	4.532	0.001216
C ₃	0.091327	0.992283	2.959	0.018916

Partial Wilks' lambda – is Wilks lambda statistics for a single contribution of the corresponding variable to discrimination between groups. This value can be considered as an analogue of the partial correlation coefficient. A lambda with a value of 0 means complete discrimination, therefore, the lower its value, the greater the single contribution of the corresponding variable to the degree of discrimination.

Thus, according to the results of the stepwise discriminant analysis, it can be concluded that the variable GR makes the greatest contribution to the overall discrimination – this is due to the identification of clay intervals from the reservoir intervals (the GR readings are increased in mudstones and lowered in sandstones). The second most important variable is the LBA – this parameter distinguishes the intervals of oil-saturated

sandstones, since the luminescence of the sludge during LBA is a direct indication of the presence of oil in the reservoir. The third important parameter is NNT – it is most likely associated with the separation of argillite from other classes. The fourth parameter in importance is the DS – it most likely also separates the intervals of mudstones from other classes, but also contributes to the identification of sandstones with reservoir properties. This can be seen if we compare the average values of DS in different classes (see Table 1) – sandstones with good reservoir properties during drilling are covered with a mudcake, and the well is narrowing. The remaining variables have virtually no effect on the data discrimination by classes, showing low values of contributions to discrimination.

To find out how the four variables separate different classes, find the actual discriminant function. Using canonical analysis, we calculate various independent (orthogonal) discriminating functions. Each subsequent discriminant function will contribute less and less to the overall discrimination. The maximum number of estimated functions is equal to the number of variables or the number of classes minus one, depending on which number is less. In our case, four discriminating functions are evaluated. First, we determine whether both discriminant functions (roots) are statistically significant (Table 3).

Table 3 shows a report on the stepwise criteria with inclusion for all canonical roots. The first line contains the criterion of significance for all the roots, the second - the data on the significance of the roots remaining after the removal of the first root, etc. Thus, this table shows how many canonical roots (discriminatory functions) should be interpreted. In our case, four statistically significant discriminant functions have been obtained that can be used to divide into classes.

Table 3

Criteria χ^2 consecutive roots

Excluded roots	Wilks' lambda	χ^2	p-level
0	0.090622	3666.409	0.000000
1	0.511245	1024.473	0.000000
2	0.771769	395.600	0.000000
3	0.981939	27.831	0.000101

In order to calculate the weights of the discriminant function and record the discriminant functions, we find out the initial coefficients for the canonical variables (Table 4).

Table 4

Initial coefficients for canonical variables

Parameter	Z_1	Z_2	Z_3	Z_4
GR	0.44205	0.141440	-0.03656	0.00967
LBA	-0.23902	0.320572	-0.64354	-0.09412
NNT	-0.03586	0.144381	0.05269	-0.11102
DS	-0.00210	-0.044725	-0.03096	-0.05547
C_4	0.05188	-0.105805	-0.07426	-0.11248
C_2	0.04584	-0.048629	0.01273	-0.17474
C_1	0.03131	-0.047732	0.02560	-0.15797
G_{sum}	-0.17723	0.206584	-0.89272	-0.52816
C_3	0.01530	-0.031519	0.04850	-0.14971
Constant	-5.99788	2.447964	-1.31175	16.69311
Cumulative distribution	0.85293	0.946577	0.99662	1.00000

The first discriminant function uses the most informative variables GR, LBA and G_{sum} (the highest initial coefficients, see Table 3). The second and third discriminant functions are weighted by the heaviest variables of the LBA and G_{sum} . The fourth function is weighted by the heaviest variable of G_{sum} . Other variables also contribute to these functions, but smaller.

$$Z_1 = -5.998 + 0.442 GR - 0.239 LBA - 0.036 NNT - 0.002 DS + 0.052 C_4 + 0.046 C_2 + 0.031 C_1 - 0.177 G_{sum} + 0.015 C_3;$$

$$Z_2 = -2.448 + 0.141 GR - 0.321 LBA - 0.144 NNT - 0.045 DS + 0.106 C_4 + 0.049 C_2 + 0.048 C_1 - 0.207 G_{sum} + 0.032 C_3;$$

$$Z_3 = -1.312 + 0.037 GR - 0.644 LBA - 0.053 NNT - 0.031 DS + 0.074 C_4 + 0.013 C_2 + 0.026 C_1 - 0.893 G_{sum} + 0.049 C_3;$$

$$Z_4 = 16.693 + 0.010 GR - 0.094 LBA - 0.111 NNT - 0.055 DS + 0.112 C_4 + 0.175 C_2 + 0.158 C_1 - 0.528 G_{sum} + 0.150 C_3.$$

Table 4 shows the eigenvalues (roots) for each discriminant function and the cumulative distribution of the explained variance accumulated by each function. As you can see, the first function is responsible for 85.29 % of the variance explained, i.e. 85.29 % of the total discriminating power is due to this function. Thus, it is clear that the first function is most important.

We now know which variables are involved in discrimination between different classes. The next task is to determine the nature of discrimination for each canonical root. The canonical averages of the first step are presented in Table 5.

Obviously, the first discriminant function separates mainly argillites and aleurolites (ie, dense clayey rocks) from other classes – sandstones and oil-saturated sandstones (average canonical variables vary greatly – from -2.52714 for oil-saturated sandstones to 3.10252 for aleurolites, while clayey sandstones are located approximately in the middle of this range with values of -0.35301, but still more to the sandstones). The second discriminant function seems to be designed to separate argillites and aleurolite; however, as was to be expected, based on the eigenvalues considered earlier, the quality is now slightly worse and will deteriorate with each new function. The third discriminant function most likely divides non-oil saturated sandstones and oil-saturated sandstones, and the fourth, apparently, serves to separate clayey sandstones.

Table 5
The canonical averages of the first step

Group	Z_1	Z_2	Z_3	Z_4
Argillites	1.75708	-1.20724	-0.320186	-0.039617
Aleurolite	3.10252	0.75046	0.118458	0.117952
Clayey sandstone	-0.35301	0.52841	0.124485	-0.248422
Sandstone	-1.99839	-0.36809	0.792809	0.083623
Oil-saturated sandstone	-2.52714	0.32116	-0.771621	0.086569

A quick way to visualize these results is to display the scatterplot of discriminant functions (Fig. 3).

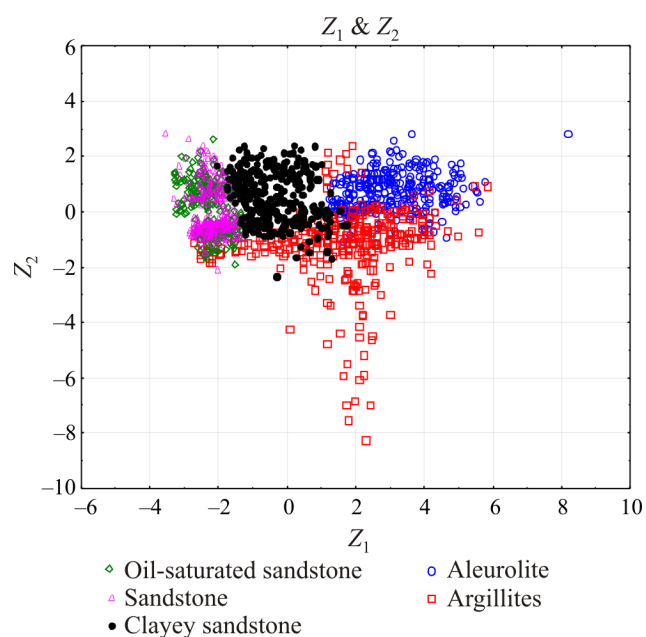


Fig. 3. Scatter diagrams for canonical values

From this it can be seen that the best separation is obtained by functions 1 and 2. It can be seen that non-oil saturated and oil saturated sandstones are shown in the diagram on the left, and they are separated from argillites and aleurolites by clayey sandstones. Therefore, the first discriminant function mainly divides non-oil saturated sandstones and oil-saturated sandstones from aleurolites and mudstones. There is also some discrimination between classes of argillites, aleurolite, and clayey sandstones. However, discrimination here is not as clear as for the first canonical function (root).

To summarize, we note that the clearest discrimination is possible using the first discriminant function. This function is marked

by negative coefficients for the presence of sludge luminescence at LBA, the total content of oil-saturated hydrocarbons in the drilling fluid and NNT indications and positive weights for the indications of GR, DS, and the relative content of methane, ethane, propane and butane. Thus, the greater the readings of the GR and DS and the lower the total content of oil-saturated hydrocarbons in the drilling fluid, and also if there is no luminescence of the sludge, the more likely it is argillites and aleurolite, i.e. non-collector rocks. After analyzing the coefficients of the third discriminatory function, it is clear that it can be used to separate sandstones and oil-saturated sandstones. This is due to the fact that the higher the total content of oil-saturated hydrocarbons in the drilling fluid and if there is a luminescence of the sludge, the more likely it is oil-saturated sandstones. A posteriori probabilities for each class were also obtained using discriminant analysis. These values indicate the probability that the observation belongs to a particular class. In order to graphically depict these probabilities, a geological-geophysical plate was created on which were applied: rock saturation according to mud logging, rock saturation according to well logging and lithological column according to mud and well logging, as well as all variables that participated in the discriminant analysis (Fig. 4).

In Fig. 4, the following picture can be seen: opposite the oil-saturated rocks, the probability curve is located on the right-hand side (values of 0.4–1.0), i.e. the probability of attributing this interval to the class of oil-saturated rocks is high. Opposite to dense rocks, the probability curve is on the left side (values are about 0), i.e. the probability of attributing this interval to the class of oil-saturated rocks is very low. Opposite to some intervals, the probability curve shows intermediate values, i.e. these are the intervals to which the interpreter should pay attention.

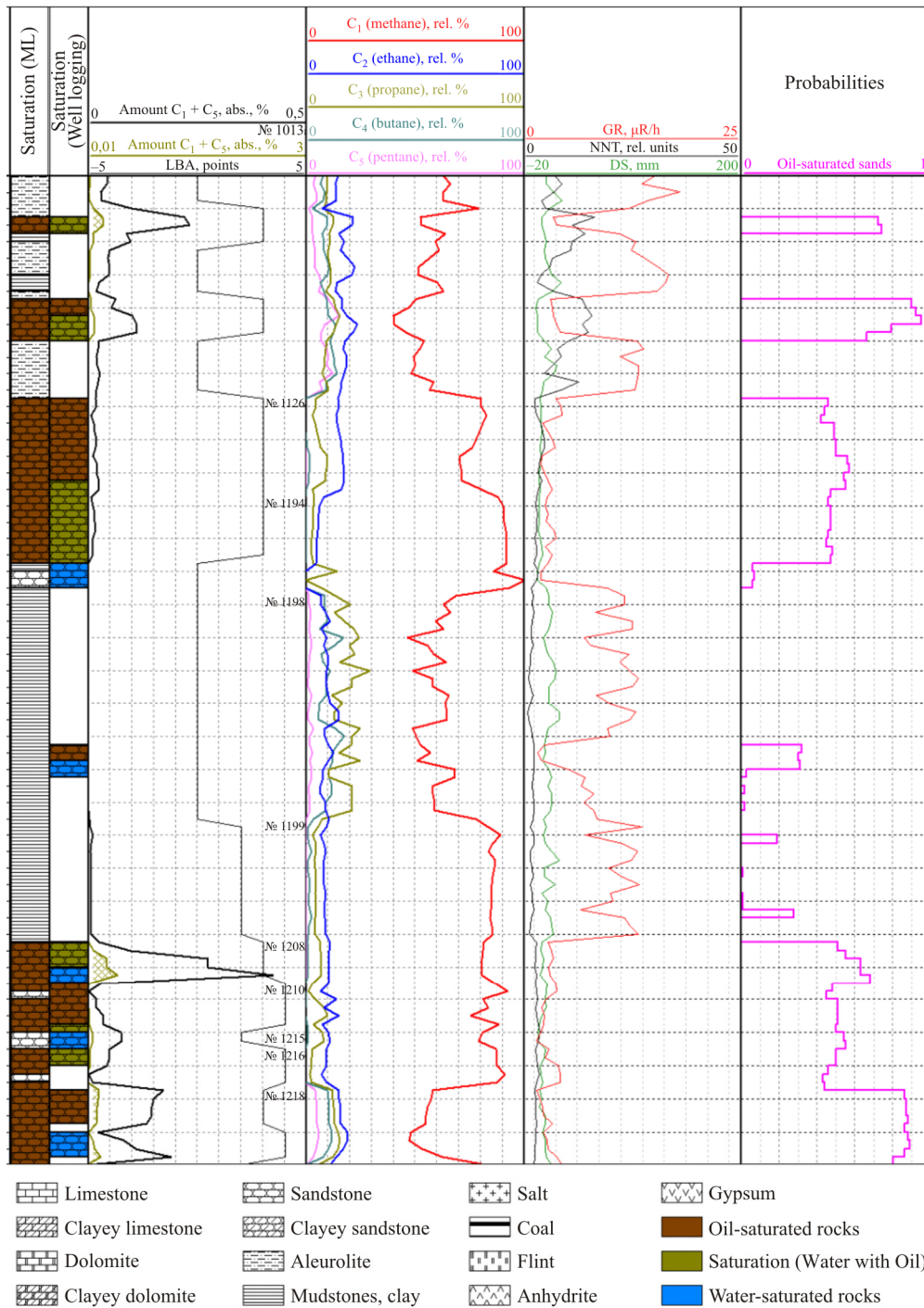


Fig. 4. The dependence of the oil layers presence probability on the geological and geophysical parameters

Conclusion

Thus, the use of stepwise discriminant analysis made it possible to determine the value of the probability of attributing observations to the rock class – argillites, aleurolites, clayey sandstones, non-oil-

saturated sandstones, and oil-saturated sandstones. An analysis of the distribution of geological and geophysical data, together with the calculated probability of assigning observations to the class of oil-saturated sandstones, allowed us to identify the following areas: the first zone with probability

values around zero (non-oil-saturated zone), the second zone with probability values from 0.4 to 1 (oil-saturated zone) sandstones) and the third zone – with intermediate values of probability (zone of unclear saturation). In this zone the interpreter must pay particular attention, since it may contain oil-saturated

sandstones, but did not fully manifest itself due to the lack of any data, for example, the luminescence of the sludge at the LBA. To interpret these obscure intervals, it is necessary to use additional data, such as well logging methods or the results of sampling and description of the core, if any.

References

1. Shmatchenko S.N. Geofizicheskie issledovaniia i raboty v skvazhinakh. Vol.7: Geologo-tekhnologicheskie issledovaniia v skvazhinakh [Geophysical surveys in wells. Vol.7: Geological and technological research in wells]. Ufa, Inform Reklama, 2010, 248 p.
2. GOST R 53375-2009. Skvazhiny nefianye i gazovye. Geologo-tekhnologicheskie issledovaniia. Obshchie trebovaniia [Oil and gas wells. Geological-technological logging. General requirements]. Moscow, Standartinform, 2009, 24 p.
3. RD 153-39.0-069-01. Tekhnicheskaiia instruktsiia po provedeniiu geologo-tekhnologicheskikh issledovaniia nefianykh i gazovykh skvazhin [Technical instruction for conducting geological and technological studies of oil and gas wells]. Moscow, 2001.
4. Potter P.E., Skimp N.F., Witters I. Trace elements in marine and fresh-water argillaceous sediments. *Geochimica et Cosmochimica*, 1963, vol.27, iss.6, pp.669-694. DOI: 10.1016/0016-7037(63)90019-X
5. Davis J.C. Statistics and data analysis in geology. 3rd ed. John Wiley & Sons, 2002. 656 p.
6. Statsoft, available at: URL: <http://statsoft.ru/home/textbook/default.htm> (accessed 08 May 2018).
7. Krivoshechekov S.N., Galkin V.I., Kozlova I.A. Determination of potentially oil bearing areas by behavioristical method by the example of Perm region (krai). *Perm Journal of Petroleum and Mining Engineering*, 2012, no.4, pp.7-15.
8. Matern Zh. Osnovy prikladnoi geostatistiki [Basics of applied geostatistics]. Moscow, 1968, 408 p.
9. Sosnin N.E. Development of statistical models for predicting oil-and-gas content (on the example of terrigenous devonian sediments of North Tatar arch). *Perm Journal of Petroleum and Mining Engineering*, 2012, no.5, pp.16-25.
10. Denisov S.V., Ismagilov R.N., Sidelnikov K.A. Rezultaty primeneniia metoda posledovatel'nogo prognozirovaniia veroiatnostei dlia diskriminantnogo analiza prostykh modelei plasta [The results of applying the method of successive prediction of probabilities for discriminant analysis of simple reservoir models]. *Izvestiia Tul'skogo gosudarstvennogo universiteta. estestvennye nauki*, 2009, pp.205-210.
11. Denisov S.V., Ismagilov R.N., Sidelnikov K.A. Rezultaty primeneniia metoda posledovatel'nogo prognozirovaniia veroiatnostei dlia diskriminantnogo analiza slozhnykh modelei plasta [The results of applying the method of successive prediction of probabilities for discriminant analysis of complicated reservoir models]. *Izvestiia Tul'skogo gosudarstvennogo universiteta. Estestvennye nauki*, 2009, pp.211-216.
12. Mikhalevich I.M., Primina S.P. Primenenie matematicheskikh metodov pri analize geologicheskoi informatsii (s ispolzovaniem kompiuternykh tekhnologii) [The use of mathematical methods in the analysis of geological information (using computer technology)]. Irkutsk, Irkutskii gosudarstvennyi universitet, 2006, part III, 115 p.
13. Porotnikov A.V., Popov M.P., Gorbunova N.P. Primenenie lineinogo diskriminantnogo analiza dlia avtomaticheskogo

opredeleniia proiskhozhdeniia izumruda po dannym rentgenofluorestsentnogo analiza [The use of linear discriminant analysis to automatically determine the origin of emerald according to X-ray fluorescence analysis]. *Trudy instituta geologii i geokhimii im. Akademika A.N. Zavaritskogo*. Ekaterinburg, Institut geologii i geokhimii im. Akademika A.N. Zavaritskogo, 2013, pp.353-355.

14. Kravtsov B.A., Miliutin L.I. Vozmozhnosti primeneniia mnogomernoi klassifikatsii pri izuchenii populiatsii drevesnykh rastenii [The possibility of using multidimensional classification in studying populations of woody plants]. *Prostranstvenno-vremennaiia struktura lesnykh biogeotsenozov*. Novosibirsk, Nauka, 1981, pp.47-65.

15. Urbakh V.Iu. Diskriminantnyi analiz i ego primeneniye v biologicheskoi sistematike i meditsinskoj diagnostike [Discriminant analysis and its application in biological systematic and medical diagnostics]. *Primeneniye matematicheskikh metodov v biologii*. Leningrad, 1964, iss.3, pp.67-87

16. Kagan E.S., Morozova I.S. Izuchenie faktorov optimizatsii poznavatelnoi deiatelnosti studentov s pomoshchiu metodov klasternogo i diskriminantnogo analizov [Studying the factors of optimization of students' cognitive activity using the methods of cluster and discriminant analyzes]. *Sibirskaiia psikhologiya segodnia*. Kemerovo: Kuzbassvuzizdat, 2002, pp.36-41.

17. Giniyatullin K.G., Valeeva A.A., Smirnova E.V. Ispolzovanie klasternogo i diskriminantnogo analizov dlia diagnostiki litologicheskoi neodnorodnosti pochvoobra-zuiushchei porody po granulometriches-
Ьkomu sostavu [Application of cluster and discriminant analyzes to diagnose lithological heterogeneity of the parent material according to its particle-size distribution]. Moscow, Nauka, 2017, pp.946-953. DOI: 10.7868/S0032180X17080044

18. Dat Thanh Tran, Moncef Gabbouj, Alexandros Iosifidis. Multilinear class-specific discriminant analysis. *Pattern Recognition Letters*, 2017, vol.100, iss.1, pp.131-136. DOI: 10.1016/j.patrec.2017.10.027

19. Yong Wang, Jian-Bin Xie, Yi Wu. Orthogonal discriminant analysis revisited. *Pattern Recognition Letters*, 2016, vol.84, pp.149-155. DOI: 10.1016/j.patrec.2016.09.010

20. Andrewartha H.G., Birch L.C. The distribution and abundance of animals. Chicago, Univ. Press, 1954, 782 p.

21. Cacoullos T. Discriminant analysis and applications. London, Academic press, Inc. Ltd, 1972, 434 p.

22. Kim Dzh.O., Miuller Ch.U, Klekka U.R. et al. Faktorny, diskriminantnyi i klasternyi analiz [Factor, discriminant and cluster analysis]. Moscow, Finansy i statistika, 1989, 215 p.

23. Gorban A.N. Obuchenie neironnykh setei [Neural Network Training]. Moscow, ParaGraph, 1990, 160 p.

24. Andersen T. Vvedeniye v mnogomernyi statisticheskii analiz [Introduction to multivariate statistical analysis]. Moscow, Fizmatgiz, 1963, 500 p.

25. Morrison D.F. Multivariate statistical methods. 2nd ed. New York, McGraw-Hill, Inc., 1976, 415 p.

26. Morrison D.F. Applied statistical methods. Prentice-Hall, Inc, Engle-wood Cliffs, N.J., 1983, 562 p.

27. Anderson T.W. An introduction to multivariate statistical analysis. New York, John Wiley and Sons, Inc., 1958, 374 p.

28. Cooley W.W., Lohnes P.R. Multivariate data analysis. New York, John Wiley and Sons, Inc., 1971, 364 p.

29. Fisher R.A. The precision of discriminant functions. *Annals of Human Genetics*, 1940, vol.10, iss.1, pp.422-429. DOI: 10.1111/j.1469-1809.1940.tb02264.x

30. Borovikov V.P. Statistika dlia studentov i inzhenerov [Statistica for engineering students]. Moscow, KompiuterPress, 2001, 301 p.

31. Gmurman V.S. Teoriia veroiatnostei i matematicheskaia statistika [Theory of probability and mathematical statistics]. Moscow, Vysshaya shkola, 1972, 368 p.
32. Aivazian S.A., Bezhaeva Z.I., Staroverov O.V. Klassifikatsiia mnogomernykh nabludenii [Classification of multidimensional observations]. Moscow, Statistika, 1974, 240 p.
33. Machine learning, neural and statistical classification. Ed. D. Mitchie et al. Ellis Horwood, Chichester, 1994, 304 p.
34. Li C.C. Introduction to experimental statistics. New York, McGraw-Hill, Inc., 1964, 460 p.
35. Marriot F.H.C. The interpretation of multiple observations. London, Academic Press, Inc., Ltd., 1974, 117 p.
36. Efroimson M.A. Multiple regression analysis. *Mathematical Methods for Digital Computers*, 1960, vol.1, pp.191-203.
37. Aivazian S.A., Stepanov V.S. Instrumenty statisticheskogo analiza dannykh [Statistical Data Analysis Tools]. *Mir PK*, 1997, no.8.
38. Aivazian S.A., Bushtaber V.M., Eniukov I.S., Meshalkin L.D. Prikladnaia statistika. klassifikatsiia i snizhenie razmernostei [Application statistics. Classification and reduction of dimensions]. Moscow, Finansy i statistika, 1989, 607 p.
39. Bikel P., Doksam K. Matematicheskaia statistika [Mathematical statistics]. Moscow, Finansy i statistika, 1983, iss.1, 278 p.; iss.2, 254 p.
40. Mikhalevich I.M., Alferova M.A., Rozhkova N.Iu. Osnovy prikladnoi statistiki [Basics of applied statistics]. Irkutsk, RIO gosudarstvennogo instituta usovershenstvovaniia vrachei, 2008, vol.3, 92 p.

Библиографический список

1. Геофизические исследования и работы в скважинах: в 7 т. Т. 7. Геолого-технологические исследования в скважинах / сост. С.Н. Шматченко. – Уфа: Информ реклама, 2010. – 248 с.
2. ГОСТ Р 53375-2009. Скважины нефтяные и газовые. Геолого-технологические исследования. Общие требования. М.: Стандартинформ, 2009. – 24 с.
3. РД 153-39.0-069-01. Техническая инструкция по проведению геолого-технологических исследований нефтяных и газовых скважин. – М., 2001.
4. Potter P.E., Skimp N.F., Witters I. Trace elements in marine and fresh-water argillaceous sediments // *Geochimica et Cosmochimica*. – 1963. – Vol 27, iss 6. – P. 669–694. DOI: 10.1016/0016-7037(63)90019-X
5. Davis J.C. Statistics and data analysis in geology. – 3rd ed. – John Wiley & Sons, 2002. – 656 p.
6. Statsoft [Электронный ресурс]: электрон. учеб. по статистике. – URL: <http://statsoft.ru/home/textbook/default.htm> (дата обращения: 08.05.2018).
7. Кривошеков С.Н., Галкин В.И., Козлова И.А. Определение перспективных участков геолого-разведочных работ на нефть вероятностно-статистическими методами на примере территории Пермского края // *Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело*. – 2012. – № 4. – С. 7–15.
8. Матерон Ж. Основы прикладной геостатистики. – М., 1968. – 408 с.
9. Соснин Н.Е. Разработка статистических моделей для прогноза нефтегазоносности (на примере терригенных девонских отложений Северо-Татарского свода) // *Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело*. – 2012. – № 5. – С. 16–25.
10. Денисов С.В., Исмагилов Р.Н., Сидельников К.А. Результаты применения

метода последовательного прогнозирования вероятностей для дискриминантного анализа простых моделей пласта // Известия Тульского государственного университета. Естественные науки. – 2009. – С. 205–210.

11. Денисов С.В., Исмагилов Р.Н., Сидельников К.А. Результаты применения метода последовательного прогнозирования вероятностей для дискриминантного анализа сложных моделей пласта // Известия Тульского государственного университета. Естественные науки. – 2009. – С. 211–216.

12. Применение математических методов при анализе геологической информации (с использованием компьютерных технологий): учеб. пособие / сост. И.М. Михалевич, С.П. Примина. – Иркутск: Иркут. гос. ун-т, 2006. – Ч. III. – 115 с.

13. Поротников А.В., Попов М.П., Горбунова Н.П. Применение линейного дискриминантного анализа для автоматического определения происхождения изумруда по данным рентгенофлюоресцентного анализа // Труды института геологии и геохимии им. академика А.Н. Заварицкого. – Екатеринбург: Институт геологии и геохимии им. академика А.Н. Заварицкого, 2013. – С. 353–355.

14. Кравцов Б.А., Милютин Л.И. Возможности применения многомерной классификации при изучении популяций древесных растений // Пространственно-временная структура лесных биогеоценозов. – Новосибирск: Наука, 1981. – С. 47–65.

15. Урбах В.Ю. Дискриминантный анализ и его применение в биологической систематике и медицинской диагностике // Применение математических методов в биологии. – Л., 1964. – Вып. 3. – С. 67–87.

16. Каган Е.С., Морозова И.С. Изучение факторов оптимизации познавательной деятельности студентов с помощью методов

кластерного и дискриминантного анализов // Сибирская психология сегодня. – Кемерово: Кузбассвуиздат, 2002. – С. 36–41.

17. Гиниятуллин К.Г., Валеева А.А., Смирнова Е.В. Использование кластерного и дискриминантного анализов для диагностики литологической неоднородности почвообразующей породы по гранулометрическому составу. – М.: Наука, 2017. – С. 946–953. DOI: 10.7868/S0032180X17080044

18. Dat Thanh Tran, Moncef Gabbouj, Alexandros Iosifidis. Multilinear class-specific discriminant analysis // Pattern Recognition Letters. – 2017. – Vol. 100, iss. 1. – P. 131–136. DOI: 10.1016/j.patrec.2017.10.027

19. Yong Wang, Jian-Bin Xie, Yi Wu. Orthogonal discriminant analysis revisited // Pattern Recognition Letters. – 2016. – Vol. 84. – P. 149–155. DOI: 10.1016/j.patrec.2016.09.010

20. Andrewartha H.G., Birch L.C. The distribution and abundance of animals. – Chicago: Univ. Press, 1954. – 782 p.

21. Cacoullos T. Discriminant analysis and applications. – London: Academic Press, Inc. ltd, 1972. – 434 p.

22. Факторный, дискриминантный и кластерный анализ / Дж.О. Ким, Ч.У. Мьюллер, У.Р. Клекка [и др.]. – М.: Финансы и статистика, 1989. – 215 с.

23. Горбань А.Н. Обучение нейронных сетей. – М.: ParaGraph, 1990. – 160 с.

24. Андерсен Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963. – 500 с.

25. Morrison D.F. Multivariate statistical methods. – 2nd ed. – New York: McGraw-Hill, Inc., 1976. – 415 p.

26. Morrison D.F. Applied statistical methods. – Prentice-Hall, Inc, Engle-wood Cliffs, N.J., 1983. – 562 p.

27. Anderson T.W. An introduction to multivariate statistical analysis. – New York: John Wiley and Sons, Inc., 1958. – 374 p.

28. Cooley W.W., Lohnes P.R. Multivariate data analysis. – New York: John Wiley and Sons, Inc., 1971. – 364 p.
29. Fisher R.A. The precision of discriminant function // *Annals of Human Genetics*. – 1940. – Vol. 10, iss. 1. – P. 422–429. DOI: 10.1111/j.1469-1809.1940.tb02264.x
30. Боровиков В.П. *Statistica для студентов и инженеров*. – М.: КомпьютерПресс, 2001. – 301 с.
31. Гмурман В.С. *Теория вероятностей и математическая статистика*. – М.: Высшая школа, 1972. – 368 с.
32. Айвазян С.А., Бежаева З.И., Староверов О.В. *Классификация многомерных наблюдений*. – М.: Статистика, 1974. – 240 с.
33. *Machine learning, neural and statistical classification* / ed. D. Mitchie [et al.]. – Ellis Horwood, Chichester, 1994. – 304 p.
34. Li C.C. *Introduction to experimental statistics*. – New York: McGraw-Hill, Inc., 1964. – 460 p.
35. Marriot F.H.C. *The interpretation of multiple observations*. – London Academic Press, Inc., Ltd., 1974. – 117 p.
36. Efron M.A. *Multiple regression analysis // Mathematical Methods for Digital Computers*. – 1960. – Vol. 1. – P. 191–203.
37. Айвазян С.А., Степанов В.С. *Инструменты статистического анализа данных // Мир ПК*. – 1997. – № 8.
38. *Прикладная статистика. Классификация и снижение размерностей* / С.А. Айвазян, В.М. Буштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.
39. Бикел П., Доксам К. *Математическая статистика*. – М.: Финансы и статистика, 1983. – Вып. 1. – 278 с.; Вып. 2. – 254 с.
40. Михалевич И.М., Алферова М.А., Рожкова Н.Ю. *Основы прикладной статистики: в 3 т.* – Иркутск: РИО Гос. ин-та усовершенствования врачей, 2008. – Т. 3. – 92 с.

Please cite this article in English as:

Rastorguev M.N. Using discriminant analysis for the interpretation of gas logging data on the example of the Pavlov oil field. *Perm Journal of Petroleum and Mining Engineering*, 2019, vol.19, no.1, pp.39-55. DOI: 10.15593/2224-9923/2019.1.4

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:

Расторгуев М.Н. Использование дискриминантного анализа для интерпретации данных газового каротажа на примере Павловского нефтяного месторождения // *Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело*. – 2019. – Т.19, №1. – С.39–55. DOI: 10.15593/2224-9923/2019.1.4