Territories typification technique with use of statistical models

V I Galkin¹, A V Rastegaev¹, V V Seredin², A V Andrianov²

¹Perm National Research Polytechnic University, Komsomolsky Avenue, 29, Perm, 614990, Russia
²Perm State University, Bukireva St., 15, Perm, 614990, Russia,

E-mail: seredin@nedra.perm.ru

Abstract. Territories typification is required for solution of many problems. The results of geological zoning received by means of various methods do not always agree. That is why the main goal of the research given is to develop a technique of obtaining a multidimensional standard classified indicator for geological zoning. In the course of the research, the probabilistic approach was used. In order to increase the reliability of geological information classification, the authors suggest using complex multidimensional probabilistic indicator \( P_K \) as a criterion of the classification. The second criterion chosen is multidimensional standard classified indicator \( Z \). These can serve as characteristics of classification in geological-engineering zoning. Above mentioned indicators \( P_K \) and \( Z \) are in good correlation. Correlation coefficient values for the entire territory regardless of structural solidity equal \( r = 0.95 \) so each indicator can be used in geological-engineering zoning. The method suggested has been tested and the schematic map of zoning has been drawn.

1. Urgency of the research

Geological-engineering zoning is required for the solution of the range of problems. The results of zoning obtained by means of different techniques do not always correlate. It can be explained by various reasons and in the first place it is choosing and justification of the integral (generalized) indicator that is the most important classification characteristics during the procedure of zoning [1].

There is a range of techniques used for classification indicator assessment: geological [1], hydrobiological [2], rating [3], a normalization method and a probability statistical method [4, 5]. Currently geoinformational systems [6] are widely used when mapping. The most promising technique when determining the classification indicator is its probability and statistics assessment based on multidimensional models [7, 8]. This technique allows minimization of subjectivity of the research on the stage of choosing the classification indicator and the justification of taxa boundaries.

Thus the main aim of the research given is to develop the technique for calculation of the multidimensional standard classified indicator for geological-engineering zoning.

2. Research results and discussion

Geological-engineering zoning involves the following work stages:
- the purpose of zoning is stated;
- taking into the account the purpose, classification indicator \( (K_p) \) is chosen and justified;
- classification indicator \( (K_p) \) boundary values for taxa identification are justified;
the model of zoning is formed including classification indicator boundary values and taxa’ names;
- the zone under investigation is divided into sub-zones (equal by their areas) each of which is characterized by numeric values of a classified indicator;
- zoning is held and the taxa are described.

An integral indicator \([9]\) can serve as classification indicator \((K_p)\). And the opportunity to use the indicator of conditional complex probability \((P_k)\) and discriminator function \((Z)\) as an integral indicator was studied.

2.1. General characteristics of the subject of the research

The subject of the research was the territory of an active pipeline between points \(A\) and \(B\). The pipeline has been used for 35 years; it is 20.4 km long and its laying thickness is 2.8 m. The pipeline is made of steel pipe with the diameter of 89 m.

The pipeline condition analysis showed its satisfactory state \((G-1)\) between the pickets 25 and 35 while between the pickets 1 and 11 the condition of the pipeline is unsatisfactory \((G-2)\) as evidenced by annual accidental releases of oil onto the ground. So the task stated is to evaluate the pipeline condition according to geological-engineering data. The results are to be presented in the form of a schematic map of zoning.

To provide geological-engineering characteristics of the territory, a geotechnical investigation was conducted which included pipeline reconnaissance, well-drilling, core description, collection of samples and soil monolith at the areas between the pickets. The total number of wells drilled is 100. Laboratory and office operations were carried out and reported. Beyond that, the peculiarities of oil distribution over the earth cover section were studied which resulted in the development of mathematical models that allowed one to forecast the thickness of hydrocarbons penetration into soil mass \([10]\).

The results of the investigations revealed that geological structure consists of (from the top to the bottom) medium-grained sand with the thickness up to 8.9 m, very soft muckle with the thickness up to 5.3 m, very soft loam with the thickness up to 5.2 m and hard seamy sandstone with thickness up to 5 m.

2.2. Zoning technique with the use of multidimensional standard probability classified indicator \((P_k)\)

The foundation of the zoning technique under discussion is a probability statistical approach which operates according to the following algorithm.

At the first stage, the most informative geological indicators, which provide the most precise characteristics of the pipeline, are selected. These include: absolute height of the earth surface \((H)\), thickness of sand \((t_s)\), thickness of muckle \((t_m)\), thickness of loam \((t_l)\), absolute height of sandstone superface \((H_{sd})\) and overall thickness of the earth mass \((T_o)\).

At the second stage, the indicators under research are standardized in two variants. For the first variant, the values of \((dH_s)\), \((dH_{sd})\) and \((dT_o)\) are calculated:

\[
dH_s = \frac{(H_{ab} - H_{av})}{H_{av}};
\]

\[
dH_{sd} = \frac{(H_{ab1} - H_{av1})}{H_{av1}};
\]

\[
dT_o = \frac{(T_{ab} - T_{av})}{T_{av}}.
\]

where: \(H_{ab}\) and \(H_{ab1}\) are absolute height values of the earth surface and sandstone superface in the testing site;

Tab is overall thickness of the earth mass in the testing site;

\(H_{av}\) and \(H_{av1}\) are average values of absolute height of the earth surface and sandstone superface;

\(T_{av}\) is an average value of the thickness of the earth mass.

For the second variant, the following calculation are performed:

\[s_s = t_s/T_{av};\]
\begin{equation}
\begin{align*}
    s_m &= t_m/T_o; \\
    s_l &= t_l/T_o.
\end{align*}
\end{equation}

where: $S_s$, $S_m$, $S_l$ are standardized values of the thickness of sand, muckle and loam in the testing site; $T_o$ - an overall thickness of the earth mass in the drilling site; It should be noted that $S_s + S_m + S_l = 1$.

At the third stage, standard zones are established (classes G-1 and G-2). For doing so, the pickets where the pipeline is stable (G-1) and unstable (G-2) states are marked. The number of the pickets used for classes G-1 and G-2 is 50.

At the fourth stage, the assessment of how correctly the standard zones (classes G-1 and G-2) were identified happens. For this, standard values of geological characteristics of classes G-1 and G-2 are used and the comparison of their average values takes place. The results reveal that zones where constructions are in stable or unstable states differ in $dH_{sf}$, $dT_o$, $S_s$, $S_m$, $S_l$ and $dH_{sd}$.

It justifies the establishment of standard zones G-1 and G-2. To confirm this conclusion one can calculate the coefficient value of pair correlation ($r$) between the characteristics being studied for classes G-1 and G-2.

Correlation analysis shows that for the classes under study different types of correlation can frequently be observed. In some cases these correlation can differ even in sign. The most significant differences are those between $dH_{sf}$ and $S_s$, $S_m$, $S_l$ and $dH_{sd}$.

Obviously, correlation fields between $dH_{sf}$ and $S_m$ differ not only in statistical connection force but also in the values for stable (G-1) and unstable (G-2) zones.

Thus according to the criteria of average values of geological characteristics under investigation, correlation coefficients and the layout of testing sites on the scattergram, one can claimed that standard zones G-1 and G-2 have considerable differences and are well-established.

At the fifth stage, complex multidimensional probability indicator $P_k$ is proposed to use as a classification characteristic.

\begin{equation}
P_k = \frac{\prod_{i=1}^{n} P(W_i|X_j)}{\prod_{i=1}^{n} P(W_i|X_j) + \prod_{i=1}^{n} [1 - P(W_i|X_j)]}
\end{equation}

where $P(W_i|X_j)$ identifies probabilities: $P(dH_{sf})$, $P(dT_o)$, $P(S_s)$, $P(S_m)$, $P(S_l)$, $P(dH_{sd})$.

So the algorithm of multidimensional standard probability indicator $P_k$ calculation was proposed that can serve as classification characteristic $Kp$ during geological-engineering zoning.

At the sixth stage, the model of zoning (table) is developed.

| Taxon | State of object | Value $Z$ | Value $P_k$ |
|-------|----------------|-----------|-------------|
| Taxon1 | G-1           | Stable    | $Z>Z_o=0$   | $P_k>0.5$   |
| Taxon2 | G-2           | Unstable  | $Z<Z_o=0$   | $P_k<0.5$   |

At the seventh stage, numerical values of $P_k$ in each observation site are calculated, and with the use of the model of zoning (table) the pipeline route is zoned.

At the eighth stage, taxa are established and described. For the route under investigation, two taxa are identified; those matching the following geological conditions:

Taxon G-1 is represented by the chainage between the pickets 1-18 with the bedding of very soft muckle at the pipeline foundation;
taxon \( G-1 \) is represented by the chainage between the pickets 19-51 with the bedding of medium-grained sand at the pipeline foundation.

Thus, the technique of multidimensional standard classification indicator \( Kp \) was suggested, whose principle is to determine the probability to range testing sites to classes \( G-1 \) or \( G-2 \) according to the characteristics studied.

### 2.3. Zoning technique with the use of multidimensional standard classified indicator (Z)

The technique given is based on the discriminant approach which works according to the following algorithm.

At the first stage, the most informative geological indicators, which provide the most precise characteristics of the pipeline, are selected. These include: absolute height (\( H_s \)), thickness of sand (\( t_s \)), thickness of muckle (\( t_m \)), thickness of loam (\( t_l \)), absolute height of sandstone superface (\( H_{sd} \)) and overall thickness of the earth mass (\( T_o \)).

At the second stage, single (generalized) sampling population including values of geological indicators from classes \( G-1 \) and \( G-2 \) is made.

At the third stage, chi square test is applied to assess the reliability of standard zones \( G-1 \) and \( G-2 \) establishment.

At the fourth stage, according to the total sample the linear discriminant, function \( Z \) is worked out:

\[
Z = 0.988 \cdot H_s + 0.935 \cdot t_s + 0.23 \cdot t_m - 0.391 \cdot t_l - 100.189 \quad \text{when} \quad R=0.893
\]

At the fifth stage, the discriminant function is used to calculate numeric values of \( Z \) in each observation site. For class \( G-1 \), the average value is \( Z_1 = +1.973 \) and for \( G-2 \), it is \( Z_2 = -1.973 \).

At the sixth stage boundary value of discriminant function \( Z_o \) is determined which divides the data observed into classes \( G-1 \) and \( G-2 \). To calculate \( Z_o \), one applies geological characteristics average values; \( Z_o \) equals 0. If \( Z > Z_o \), the observation site falls into class \( G-1 \), and if \( Z < Z_o \) - to \( G-2 \).

At the seventh stage, the model of zoning represented in the table is developed.

At the eighth stage, numeric values of \( Z \) for each observation site and for the use of zoning model (see table) are calculated and the pipeline route is zoned.

At the ninth stage, taxa are determined and described. For the researched pipeline, two taxa are determined; these match the following geological conditions:

- taxon \( G-1 \) is represented by the chainage between pickets 1-18 with the bedding of very soft muckle at the pipeline foundation;
- taxon \( G-1 \) is represented by the chainage between picket 19-51 with the bedding of medium-grained sand at the pipeline foundation.

So the algorithm for the calculation of multidimensional standard indicator \( Z \) was suggested which can serve as classification characteristic \( Kp \) for zones with stable state of constructions \( r=0.89 \).

### 2.4. Assessed reliability of the zoning criteria proposed

The above mentioned criteria were used to assess the correctness of observation sites classification. In this regard, a distribution graph for classification indicators \( P_k \) and \( Z \) was plotted. It was determined that for stable territories \( G-1 \), values \( P_k < 0.5 \) are seen in one single case which means that the probability of the correct classification is 98%. As for \( Z \) values, the assessment proved wrong in two cases for stable territories \( G-1 \) and in one case for unstable \( G-2 \) ones; so the probability of the correct classification is 94%. Beyond that, there is a fairly good correlation between values of \( P_k \) and \( Z \), the correlation coefficient value for all zones regardless of the structural competence equals \( r=0.95 \); for zones with stable state of constructions \( r=0.89 \); for zones with stable state of constructions - \( r=0.59 \).

Thus, when performing geological-engineering zoning, one can use standard multidimensional probability criteria \( P_k \) and discriminant criterion \( Z \) as classified indicators.

### 3. Conclusion
In the course of the research, the algorithms for calculation of standard multidimensional probability criteria $P_K$ and discriminant criterion $Z$ were developed; these can serve as classification indicators when performing geological-engineering zoning. There is a good correlation between criteria $P_K$ and $Z$. The value of the correlation coefficient for the whole territory regardless of the structural competence equals $r=0.95$ so both of them $PK$ and $Z$ can be used in geological-engineering zoning.

The technique suggested was approved and the schematic map of zoning was developed. When $P_K$ was used as a classification indicator, the probability of the correct classification of observation sites for classes $G-1$ and $G-2$ is $98\%$, while it is $94\%$ when criterion was applied.

References

[1] Trofimov V T 1979. Theoretical issues of geological-engineering zoning. MSU Vestnik Series Geology, 1 94-106

[2] Yenikeyev F I 2015. Hydrobiological methods of mapping of hydrocryological environment in mountainous regions to the east of Lake Baikal. Geocology, 1 81-90

[3] Seredin V V, Pushkaryova M V, Leybovich L O, Bakharev A O, Tatkin A V and Filimonchikov A A 2014. Change of geological environment as a result of oil field exploitation in complex conditions. Neftyanoye Khozyaystvo, 12 153-5

[4] Lee S. and Pradhan B 2006. Probabilistic landslide risk mapping at Penang Island, Malaysia. J Earth Syst Sci 115(6) 661-72

[5] Pourghasemi H R, Moradi H R and Fatemi Aghda S M 2013. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69 749-79

[6] Chauhan S, Sharma M, Arora M K and Gupta N K 2010. Landslide susceptibility zonation through ratings derived from artificial neural network. Int J Appl Earth Observ Geoinform 12 340-50

[7] Nagibi S A and Pourghasemi H R 2015. Comparative evaluation between three models of machine learning and their comparison of efficiency using two-dimensional and multivariate statistical methods in mapping groundwater potential. Water Resources Management

[8] Youssef A M, Pourghasemi H R, Bosy El-Haddad A and Basem K. Dhahry, February 2016, Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. Bulletin of Engineering Geology and the Environment, 75(1) 63-87

[9] Ozdoyeva L I 1981. The use of integral indicator of geological-engineering conditions during large-scale geological-engineering zoning of city areas. Izvestya Vysshikh Uchebnikh Zavedeni. Geologiya i Razvedka, 8 70-4

[10] Seredin V V, Rastegayev A V, Galkin V I and Leonovith M F 2017 Influence of moisture content and soil type on depth of penetration of hydrocarbons into rock massif. 133 755-9