The Methodological Aspect of Development and Application Multivariate Classification G-Mode for Analyses Geochemical Trend

A.I. Gavrishin*

South-Russian State Polytechnic University named after M.I. Platov, Novocherkassk, Russia

Abstract: This article is unique G-mode of multidimensional classification method and its application in the analysis of hydrogeochemistry Donetsk basin. He has the following main advantages over other methods of classification: does not require a priori information for classification of observations; earmarks homogeneous observing classes and subclasses; evaluates information weight of each indicator, determines the distance between the homogeneous taxon's; assessment of the descriptiveness of the sign of the classification and others. G-method is widely used to analyze geochemical, environmental, kosmochemical, distance and other types of information. The method successfully used in examining objects, phenomena and processes on Earth, Moon, Mars, Saturn, comets, asteroids and deep space. The results of the use of G-method in analysis of hydrogeological data for the Donetsk basin identified direct and inverse geochemical zonality. This shows that in the region can be discovered oil and gas accumulations.

Keywords: Methodology, geochemical investigation, modeling, classification of G-mode, Donbass, hydrogeochemical zonality.

Quantitative nature of modern geochemical information creates opportunities to use mathematical techniques and computer technology for processing of raw data and justification of conclusions about the patterns of distribution of chemical elements in rocks and underground water. However, there are serious difficulties due to the complexity of the matters dealt with in geochemistry (in geology) objects and processes, the shortcomings of the existing classifications, disadvantages of many geological concepts, etc.

MATHEMATICAL MODELING

Application of mathematics in geochemistry, according to most researchers, should be considered primarily as an application of the method of mathematical modeling. This means that it is necessary to establish similar object (process) and the model. The model is designed for the study of the object and should be similar to the signs of that object (the other properties-featured model is different from the object; otherwise it ceases to be just a model).

Under the mathematical modeling in geochemistry is understood this way of discovering the laws of space-time distribution and migration of chemical elements in natural objects, according to certain rules when a mathematical description of some of the geochemical properties of the object or process, and further: 1) on the basis of a study of this description are improved and expanded (often projected) geochemical knowledge on the same object or process; 2) on the basis of a similarity of the new object known object on mathematical description of geochemical properties, prediction of other properties for the new object [2.6].

In each case, can be used by different mathematical models, but most commonly in the geochemistry of applied probability, as the concentration of a chemical element geological area at this point is influenced by many factors, which are practically impossible to deterministically. Probabilistic-statistical model consists of the following main parts: 1) deterministic part that describes the changes in the chemical composition of natural objects and processes under the influence of the leading factors; 2) random part of the describing the actual random changes in the chemical composition by secondary factors; 3) random part, showing the observed changes in composition due to errors.

From the point of view of the geochemistry is the most interesting study of the first two parts of the model (deterministic and random), and the third part due to errors, distorts the actual distribution of the concentrations of chemical elements and prevents the researcher. Under this model, the concentration of chemical elements in some geological space can be written as follows:

$$C(x, y, z, t) = (\mu)(x, y, z, t) + (\nu)(x, y, z, t) + \upsilon(x, y, z, t),$$

where C (x, y, z, t) is the chemical element concentration observed at some point geological space

*Address correspondence to this author at the 346428 Novocherkassk, Prosveschenia 132, SRSPU, Russia; Tel: 8-863-52-55352;
E-mail: agavrishin@rambler.ru

E-ISSN: 2409-5761/14

with coordinates x, y, z, and the coordinate time t;(μ)(x, y, z, t) is the concentration of the element at coordinates x, y, z, t, due to major natural factors;(ν)(x, y, z, t) is a real element of concentration deviation(μ)due to occasional minor factors; υ (x, y, z, t)-deviation due to errors by the geological space with coordinates x, y, z, t, we get more í as compact expression:

$$(C)_{(i)} = (\mu)_{(i)} + (\nu)_i + \upsilon_{(i)}.$$

Specialist seeks within the studied object, set the first legitimate change of $(\mu)_i$, for example, uses a trend-analysis, moving average, factor analysis, etc. the most successfully used classification procedures with homogenous geological parts of space, i.e. the selection of parts with permanent(μ)Now the model can be written as follows:

$$C_i = \mu + \nu_i + \upsilon_{(i)}.$$

Thus, geologist, homogeneous parts of an object gets a reasonable statistical method of their geochemical characteristics and objective comparison of homogeneous parts-natural ways to identify changes for the object.

Statistical analysis of geochemical data to draw conclusions of the specified reliability. It includes the following main stages: planning, data processing management of observation, evaluation and quality control of primary geochemical information, verification of the homogeneity of the sample selection models (laws) distribution of components in natural objects, estimate the parameters of distributions, distributions, comparison study of relationships between content components, application of multidimensional data processing techniques (various modifications of the classification methods, factorial analysis, allocation of homogeneous statistical aggregates, pattern recognition, etc.).

G–MODE MULTIDIMENSIONAL CLASSIFICATION METHOD OF OBSERVATION

When studying rocks, minerals, water, and other natural objects and processes one of the most important is the task of building their classifications that can be kept to determine the taxonomic structure of the geological area. For selection in the space-time coordinates of similar taxa in geology are various multidimensional mathematical methods. These methods are based on a variety of statistical and heuristic principles, each of which has specific advantages and disadvantages.

Multidimensional classification method of geological observations [1.2], the following general requirements are to ensure that:

- construction of classification in the absence of a priori information about the taxonomic structure of the observations;
- the use of dependent indicators and observations;
- allocation of taxonomic structures of various levels;
- unlimited ratio the number of indicators (M) and the number of observations (N);
- use of the information on changing average values, and the relationship between signs;
- assessment of the descriptiveness of the sign of the classification;
- evaluation of similarity-difference between homogeneous taxonomy;
- classification of new observations about the taxonomic structure.

A detailed analysis of the different classification methods and their problems can be found in the literature, here we will focus on the new (G)-multidimensional classification method, developed by work [1, 2, 6, 7].

The basis for the classification procedure, called G-method based on the criterion of Gavrishin - Z^2 with distribution of quasi- χ^2 [1]:

$$Z^2 = \frac{M}{\sum_{SK} r_{SK}^2} \sum_{ij} Z_{ij}^2 = K \sum_{ij} Z_{ij}^2,$$

$$Z_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i},$$

where x_{ij} -value indicator i (1.2, ..., M), j (1.2, ..., N); μ_i and $\sigma_{(i)}$ -average value and standard deviation (i) in a homogeneous classroom observations; r_{SK} -correlation coefficients between indicator s and k.

The hypothesis of one or more observations to this homogeneous class is rejected, if the computed

$Z^2 > \chi^2_{q,f}$, where $\chi^2_{q,f}$ -critical distribution χ^2, when the significance level q and the number of degrees of freedom (f):

$$f = N \cdot M \cdot K \,.$$

To simplify the procedures for the use of the distribution χ^2 his conversion was applied in normal parameters (0.1):

$$G = \sqrt{2z^2} - \sqrt{2f - 1}.$$

If (G)>G_q the surveillance data do not belong to the homogeneous taxon, where G_q -critical values for level of significance q.

Classification procedure (G-mode) is reduced to the following main operations: selecting a coordinate system in which the transformation of the multidimensional space of attributive to the distribution of Z^2, and find the center of the first homogeneous taxon; the transformation of the coordinate system and finding all the observations of the first homogeneous taxon; repeat these operations for the observations, which were not included in previous similar taxa; evaluation of similarity-difference between homogeneous taxonomy in each and all indications simultaneously; estimation of informatively of the taxonomic structure; repeat all steps for the different levels of reliability allocation of similar taxa.

Largest G all observations found that belong to this homogeneous taxon. Modifying the critical radius of a homogeneous taxon (G_q) you can get different levels of detail and classification of varying degrees of homogeneity of the sub-family G lower values. G_q the higher the homogeneity of the taxa, more detail of the classification, but a lower reliability validity of differences between taxonomy.

Evaluation of similarity-difference between homogeneous taxonomy is also based on criteria Z^2 and is determining the international taxonomic variance for each account and all the indicators together. Media weight characteristics can be estimated as the sum of the difference between the uniform taxonomy [2].

Largest Z^2 or G can be evaluated any number of new observations belonging to the fusion of units, that is, to produce a classification of observations. G-method is implemented in the form of computer technology AGAT to automatically build a multidimensional classification of observations at different levels of detail, and successfully apply to natural and natural-human systems on Earth, Moon, Mars, asteroids, and comets in deep space for astrophysical, kosmochemical, remote, hydrogeochemical, geological and other data [1, 2, 6, 7]. This paper considers the technique and results of application G-method based on the analysis of hydrogeochemistry of Donbass.

GEOCHEMICAL ZONING OF GROUNDWATER

The most important tributaries in the coal mines of Donetsk basin form the water of Carboniferous, Cretaceous, Paleogene and Neogene deposits. To identify and quantify the vertical geochemical zonality groundwater systems listed have been used chemical analysis of samples of water taken from wells [4]. The application of a consistent classification modeling using G-method identified two main trends in the chemical composition of the groundwater, which are similar to water the aquifer complexes; below I'll elaborate on the geochemical zonation of water in Carboniferous deposits (C). Since changing contents of many of the features in depth is curved, the specifications of the subject was offered the following function, which allows you to describe the positive, negative, positive in the negative (and vice versa), and the periodic function:

$$Y = \sum_i a_i \cdot 10^{-\frac{(\lg x - b_i)^2}{c_i}},$$

where $a_{(i)}$ - factors characterizing the modal (top) the value of the function; $b_{(i)}$ -the coefficients describing the location of modal values on the x axis; $c_{(i)}$ - factors characterizing the slope of the regression line; i-number of the modal value.

Water Carboniferous deposits on the territory are very diverse chemical composition: water varies from hydrocarbonate calcium to chloride sodium, salinity ranges from 0.2 to 57.2 g/l, Cl^- -0.012-35.6, Na^+ -0,002-17.6 g/l, etc. (Table **1**). Distribution of the contents of components does not correspond to the normal model and correlation may be curvilinear. For example, for SO_4^{2-} the transition is clearly positive due to the depth (H) of mineralization and (M) the negative. A very strong correlation with salinity (r > 0.95) discover Na^+, Cl^-, Ca^{2+}, Mg^{2+}, with a depth of strongly connected M, Cl^- and Na^+. However, you can see that at significant depths (more than 400-500 m) and salinity (M > 50 g/l), and fresh water (M-2-3 g/l).

Table 1: Groundwater Chemical Composition of Carboniferous Deposits

Component	X_m	Me	X_{min}	X_{max}	S
pH	7.7	7.7	6.4.	8.6	0.5
HCO_3^-	358	352	77	947	135
SO_4^{2-}	485	400	15	1427	405
Cl^-	2366	243	12	35636	7086
Ca^{2+}	337	153	20	4084	716
Mg^{2+}	140	82	4	11145	227
Na^+	1237	252	2	17582	3316
(M)	4729	1611	178	57419	11149
(H)	124	75.5	3	922	163

Note: in tables X_m-arithmetic means, Me - median, x_{min} and x_{max} -the minimum and maximum value, S-standard deviation (in mg/l, H-depth in meters).

With the help of G-method of computer technology AGAT in the waters of Carboniferous deposits by chemical composition has been allocated 11 similar geochemical types plus 1 anomalous (Table **2**). Interesting patterns identified in examining detailed water classification by type in the coordinates of "the depth (H)-content component-mineralization (M)". Surely there are two main trends in geochemical

Table 2: Geochemical Composition of Carboniferous Deposits of Groundwater (mg/l and %-mole)

Tendention	View	H	Ph	Components						(M)
				HCO_3^-	SO_4^{2-}	Cl^-	Ca^{2+}	Mg^{2+}	Na^+	
1	1.6	40	7.1.	404	355	46	111	47	134	900
				44	48	8	37	25	38	
	1.2	79	7.7	351	425	213	158	58	183	1200
				28	43	29	38	24	38	
	1.3	82	7.7	395	790	315	185	99	333	1900
				20	52	28	29	26	45	
	2.2	102	8.0	380	810	1180	275	168	660	3300
				11	30	59	24	25	51	
	2.1	106	8.0	387	1160	786	316	177	506	3010
				12	46	42	30	28	42	
	3.1	333	8.3	293	86	14700	1300	532	6920	24000
				1	0.4	98.6	18	11	71	
	A. 1	730	8.4	290	73	33930	3500	1100	16000	55000
				0.5	0.2	99.3	18	10	72	
2	1.4	28	6.8	247	65	18	71	16	25	320
				68	24	8	59	22	19	
	1.1	50	7.4	382	114	57	65	41	82	550
				61	23	16	32	33	35	
	1.5	73	7.2	310	176	56	90	34	68	580
				49	36	15	44	28	28	
	1.7	100	7.6	372	450	173	98	118	128	1153
				30	46	24	24	48	28	
	4.1	380	7.8	548	418	1250	46	38	1100	3100
				17	16	67	4	6	90	

composition of groundwater of Carboniferous deposits by depth [3], which reflects the direct and inverse vertical geochemical zonality groundwater. In Table **2** geochemical types are positioned as the depth of water and geochemical trends.

The first trend is a typical representative of direct vertical geochemical zonality and, despite the high heterogeneity of the composition of water, it is characterized by a natural transition from low mineralized hydrocarbonate and hydrocarbonate-sulphate mixed cation composition of waters to the chloride-sulphate and further mineralized chloride sodium (Table **2**) with increasing depth.

A central role in the formation of the chemical composition of the waters of Carboniferous deposits on the first trends are Cl^- and Na^+, the contents of which increases with depth, grows so does the amount of water pH and decreases SO^{2-}_4. The coefficients are reduced dramatically-relationships $rHCO^{3-}/rCa^{2+} + rMg^{2+}$ from 0.8 to 0.02 and rSO_4^{2-}/rCl^- from 6 up to 0.002. Clearly the patterns changes in the composition of the waters with depth are visible on the equations (Table **3**). Increase of SO_4^{2-} turns out (H_{max} = 90 m), the pair correlation coefficient of no significant, while curvilinear correlation factor (r_1 = 0.66) and the regression equation is as follows (Table **3**):

$$SO_4 = 790 \cdot 10^{-\frac{(\lg H - 1{,}96)^2}{0.6}}.$$

On the genesis of the first geochemical trends, you can quite confidently say that from a depth of 150-200m starting to wane of infiltration in the formation of the chemical composition of groundwater and the increasing role of the sedimentation. This has an impact on reducing the contents in the waters SO_4^{2-} and the HCO_3^{2-} and increasing Cl^- and Na^+; water of II type by O.A. Alekin go III contents; of the J typically 5-10 mg/l, Br -20-30 mg/l. In the open part of the Eastern Donbass to mineralized water chloride sodium occurs at depths of about 1 km in the outlying parts of the basin depth of mineralized waters is much closer to the surface.

The second trend reflects a backward vertical geochemical zonality of the underground water of Carboniferous deposits, where a slight increase in salinity with depth gives way to its reduction and the formation water soda type. These trends are described well curved exponential function (Table **3**) with high correlation coefficients (r_i) parametric equations. The regression equation editing, the following patterns are clearly visible: the maximum mineralization is achieved at depths of 250-300 m; the HCO_3^- increases with depth and a maximum is reached where reliably predict fails; the content of SO_4^{2-} and Cl^- max at depths of 200-350 m and deeper decline; the content of Ca^{2+} and

Table 3: The Parameters of the Regression Curved Equations of Components Contents in Groundwater Carboniferous Deposits of Donbass on Depth

Trend	Component	$(a)_{(i)}$	$(b)_{(i)}$	$c_{(i)}$	r_i	$(H)_{max}$
1	HCO_3^-	372	2.06	8.23	0.3	115
	SO_4^{2-}	790	1.96	0.6	0.66	90
	Cl^-	32200	3.05	0.89	0.90	>1000
	Ca^{2+}	4550	3.5	1.97	0.82	>3000
	Mg^{2+}	1280	3.5	2.6	0.82	>3000
	Na^+	15950	3.1	1.0	0.92	>1000
	(M)	54000	3.13	1.16	0.90	>1000
2	HCO_3^-	525	3	8.2	0.46	>1000
	SO_4^{2-}	491	2.32	0.63	0.52	210
	Cl^-	1400	2.38	0.3	0.91	240
	Ca^{2+}	83	1.44	5.0	0.34	30
	Mg^{2+}	75	2.14.	0.7	0.62	140
	Na^+	1150	2.55	0.3	0.93	350
	(M)	3300	2.45	0.45	0.89	280

Note:$(a)_{(i)}$ – X maximum in mg/l, $(b)_{(i)}$ -the location of the maximum depth H (lgH) , c_i -the slope of the regression line, r_i - the coefficient of correlation curvilinear, $(H)_{max}$ -maximum depth in m.

Mg^{2+} is observed at depths of 100-250 m and decreased with depth.

The second type of vertical geochemical zonality of the water depth moving away from hydrocarbonate calcium to a sulphate hydrocarbonate and hydrocarbonate sulphate mixed cation composition and chloride sodium (soda) with a salinity of 2-3 g/l; a second type of water gives way to the first with high content of HCO_3^- and very low - Ca^{2+} and Mg^{2+} . The attitude now $rHCO_3^-/rCa^{2+}+rMg^{2+}$ increases to 1.7, and attitude rSO_4^{2-}/rCl^- reduced only to 0.2 for the extrapolation of equations (Table **3**) suggests that at depths of more than 1 km are hydrocarbonate sodium waters with a salinity of less than 1 g/l.

On the education of hydrogeochemical inversions and the emergence of the mineralized waters depths there are a variety of hypotheses, of which the most popular are: infiltration, juvenile, degidratation and evaporation-condensation. There are currently a large number of photographs clearly show the presence of a not much mineralized water depths associated with oil and gas deposits [5], which refer to the soda type and which are very close to the chemical appearance to the water as described above. The author believe that soda fresh water second geochemical trends in Donbass is most likely associated with the processes of condensation of water vapour from the gas phase [3].

These waters are found at different depths in the Carboniferous, Cretaceous, Paleogene and Neogene sediments and often confined to zones of vertical tectonic fracturing. Now, taking the hypothesis of evaporation-condensation of the genesis soda water with a high content of HCO_3^- and very low - Ca^{2+} and Mg^{2+} you must acknowledge within Donetsk basin in certain traps oil and gas accumulations, as is the adjacent geological structures (Dnieper-Donetsk, Donetsk-Don, the Azov-Kuban basin, etc.).

Vertical geochemical zonality regularities of groundwater had a significant impact on the formation of mine water, forming a different direction.

Thus, this article is a unique(G-mode)-multidimensional classification method and its application in the analysis of hydrogeochemistry Donetsk basin classification methods are the primary means of knowledge of the world around us. To know means to categorize.

Today, the constantly increasing number of studied objects, processes and phenomena, which are characterized by plenty of signs (chemical, physical, geological, biological, environmental, medical, sociological, economical etc.). We have developed a unique G-classification method of multivariate observations, which lets you describe the new spatial and temporal patterns of forming objects, phenomena and processes, make discoveries, to formulate new laws, describing the genesis of, establish the diagnosis. It has the following main advantages over other methods of classification:

- construction of classification in the absence of a priori information about the taxonomic structure of the observations;
- the use of dependent traits and observations;
- allocation of taxonomic structures of various levels;
- unlimited ratio characteristics (M) and the number of observations (N);
- use of the information on changing average values, and the relationship between signs;
- score information of individual traits in the classification;
- evaluation of similarity-difference between homogeneous taxonomy's;
- classification of new observations about the taxonomic structure.

(G-mode)-the method is widely used to analyze geochemical, environmental, kosmohemical, distance and other types of information. the method successfully used in examining objects, phenomena and processes on Earth, Moon, Mars, Saturn, comets, asteroids and deep space [6-8]. It can be successfully applied in studies in geology, geochemistry, geography, economics, biology, physics, chemistry and other fields of knowledge. The results of the use of G-method in the analysis of hydrogeological data for the Donetsk basin identified direct and inverse geochemical zonality. This shows that in the region can be discovered oil and gas accumulations.

REFERENCES

[1] Gavrishin AI. Hydrogeochemical studies using mathematical statistics and computing. M.: Nedra 1974; p. 146.

[2] Gavrishin AI, Coradini A. Multidimensional classification method and its application in the study of natural objects. M.: Nedra 1994; p. 92.

[3] Gavrishin AI, Coradini A. Origins and patterns of forming of underground and mine waters in the Eastern Donbass//water resources 2009; 36(5): 564-574.

[4] Hydrogeology of the USSR. Donbass. /Red. D.I. Schegolev, M.: Nedra 1970; p. 480.

[5] Fresh water aquifers deep oil and gas provinces. Ed. V.V. Kolodiy. Kiev: Nauk. Dumka 1985; p. 280.

[6] Gavrishin AI. Multivariate classification method as a methodological basis for natural object simulation/ Hydrological, Chemical and BrologicalProcess of Transformation and Transport of Contaminats/IAHS Publ 1994; 219: 337-341.

[7] Gavrishin AI, Coradini A, Cerroni P. Multivariate classification method in planetary sciences//Earch. Moon and Planets 1992; 59: 141-152.

[8] Coradini A, Tossi F, Gavrishin AI. Identification of spectral units in Phoebe//ICARUS 2008; 193: 233-251. http://dx.doi.org/10.1016/j.icarus.2007.07.023

Received on 28-06-2014 Accepted on 31-07-2014 Published on 10-10-2014

DOI: http://dx.doi.org/10.15377/2409-5761.2014.01.01.4