GEOLOGICA BALCANICA, 21. 5, Sofia, Oct. 1991, p. 77-85.

Hydrochemical classification of ground waters in Northeast Bulgaria

D. Dimitrov[1], *M. Machkova*[1], *B. Velikov*[2]

[1]*Institute of Meteorology and Hydrology, Bulgarian Academy of Sciences, 1184 Sofia*
[2]*University of Mining and Geology, 1156 Sofia*

(Received 12. 12. 1989; accepted 21. 12. 1989)

Д. Димитров, М. Мачкова, Б. Великов — Гидрохимическая классификация подземных вод в Северо-Восточной Болгарии. Работа посвещена региональной классификации гидрохимической информации 341 водной пробы разных водоносных горизонтов в Северо-Восточной Болгарии. Анализ проб проведен в полевых условиях (ϰ, pH, T, °C, CO_2) или в лаборатории стандартными аналитическими методами для остальных химических ингредиентов. Использована современная методика, основанная на многомерных методах для статистического (кластерного и дискриминантного) анализа и с применением микрокомпьютерной техники.

Abstract. The present paper deals with the regional classification of hydrochemical information on 341 water samples from different aquifers in Northeastern Bulgaria, the analysis of which is carried out in situ (pH, T, °C, CO_2, ϰ) or by standard analytical methods for the other chemical ingredients. Modern methodology, based on multivaried methods for statistical (cluster and discriminant) analysis applying microcomputer technique, is used.

Introduction

One of the main tasks, when treating the empirical hydrochemical data, is the creation of initial hypothesis. The choice of the methods for further processing and analysis and, to a certain extent, the type and nature of the results obtained and conclusions very often depend on the postulation of these hypotheses. One of the most widely used methods for initial processing of data, aiming at the creation of initial hypotheses and analysis is the cluster analysis. In this sense, the aim of the present paper is to apply the latter method for classification of the authors' data on the chemistry of ground waters in Northeast Bulgaria on the basis of statistical criteria for distinction and similarity of the individual observations for different hydrochemical parameters.

Methods

The cluster analysis allows to draw conclusions, free from hypotheses on the empirical probability distribution; it allows to classify a practically unlimited number of observations; it allows to apply different metrics in the phase space of variables (similarity criteria); it allows the use of a large number of techniques for observation standardization, etc. (Афифи, Эйзен, 1982). The Euclidean metric was used for calculation of the distances between the observations and the centres of the clusters, whereas the observations were standardized by the standard deviations, calculated for each variable along the whole data set.

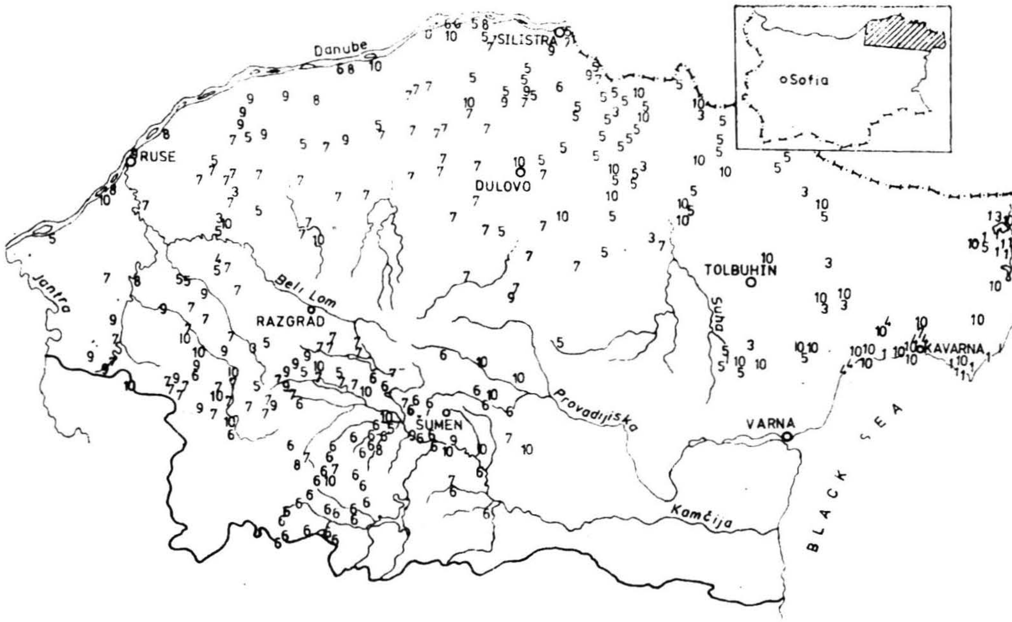


Fig. 1. Map of the region under investigation with cluster membership location

For this purpose the programmes "1D" add "KM" were used (D i x o n, 1981). Software package 6 was used for drawing the graphic applications (S t a t g r a p h i c s, 1985).

The methods used for sampling, field and laboratory hydrochemical analysis are described in В е л и к о в, 1986; В е л и к о в и др., 1986. The region studied (Fig. 1) comprises almost the whole territory of the Varna and Razgrad districts and the area is about 13 000 km^2.

This part of the Moesian bedded platform consists of deposits of Malm-Valanginian, Hauterivian-Barremian-Aptian, Cenomanian-Senonian and Miocene (predominantly Sarmatian) sediments, as well those of the Quaternary. The latter are presented by alluvium in the terrace of the Danube and the terraces of some other rivers (Cherni and Beli Lom rivers, Topchiiska river, the tributaries of the Kamchia river, etc.) of the internal hydrographic network, and also by loess.

The main rock-forming minerals are the carbonates in the clayey, marly, sandy and dolomitizised limestones, in the pure limestones and dolomites of the Malm-Valanginian, Hauterivian-Barremian-Aptian, Upper Cretaceous, Sarmatian and Pliocene, as well as the alumosilicates — for alluvium, loess, sandstones in the base of the Sarmatian and in the Pliocene, partially sandstones for the Aptian and Upper Cretaceous.

Except for part of the Malm-Valanginian, the other aquifers, due to the lack of safe and areally defined aquitards, form joint water complexes of unconfined waters (А н т о н о в, Д а н ч е в, 1980). The main direction of the groundwater flow is northeast-east and the waters are drained by the Danube, the Black Sea and partially by the internal hydrographic network.

The appurtenance of the water sources, from which samples were taken for studies in the region up to now, was determined on the basis of geologic-hydrogeologic consideration (А н т о н о в, Д а н ч е в, 1980; V e l i k o v et al., 1989). In the present paper this preliminary information on the ground waters is also taken into consideration, assuming that the total amount of the available 341 samples could be grouped in 10 clusters (mostly concerning the existing aquifers). Owing to this, when treating the results obtained such a condition on the number of the hydrochemical taxons was layed down.

Table 1

Mean values of the hydrochemical parameters by clusters

Cluster	T	pH	$\varkappa$	CO_2	HCO_3^-	SO_4^{2-}	Cl^-	Ca^{2+}	Mg^{2+}	Na^+	NO_3^-	μ	ß	M	C_s^c	C_s^d	I_{ank}
	[°C]		[μS/cm]	[bar]	[mmol/dm³]									[mg/dm³]			
1	17.72	7.408	1736.5	0.006	6.95	0.661	10.16	1.29	2.21	11.13	0.700	21.22	1.96	1231.6	0.154	0.045	—5.70
2	14.60	7.100	4050.0	0.033	7.73	1.800	27.08	1.60	4.02	25.63	0.700	41.91	4.29	2407.0	0.042	0.004	—6.46
3	12.88	7.144	1409.8	0.028	7.85	0.548	2.50	2.23	4.38	3.86	5.950	20.86	4.01	913.5	0.079	0.017	—6.32
4	27.67	7.350	585.4	0.005	4.91	0.337	1.12	0.97	1.20	2.20	0.015	9.03	1.23	495.4	0.506	0.547	—5.19
5	13.11	7.135	779.9	0.026	7.62	0.248	0.66	1.29	2.21	2.29	0.764	11.59	3.53	676.0	0.092	0.017	—6.54
6	13.12	7.228	574.8	0.026	4.34	0.542	0.61	1.89	0.77	1.02	0.357	9.26	1.92	459.0	0.245	0.039	—6.19
7	13.30	7.027	686.9	0.032	6.92	0.291	0.52	2.10	1.41	1.34	0.497	11.24	3.86	621.5	0.101	0.009	—6.53
8	18.69	7.855	411.0	0.007	3.58	0.537	0.44	1.32	0.52	1.47	0.217	7.52	0.38	388.9	0.935	0.495	—4.71
9	13.25	6.999	988.7	0.039	7.97	0.812	1.24	2.71	2.34	2.19	1.567	15.90	4.80	828.8	0.085	0.007	—6.51
10	14.67	7.356	679.1	0.015	6.51	0.287	0.89	1.41	1.70	1.99	0.559	10.73	1.96	604.9	0.188	0.067	—5.96
Total	14.15	7.188	802.5	0.025	6.59	0.410	1.40	1.77	1.75	2.40	0.864	12.25	3.02	665.3	0.169	0.052	—6.25

Results and discussions

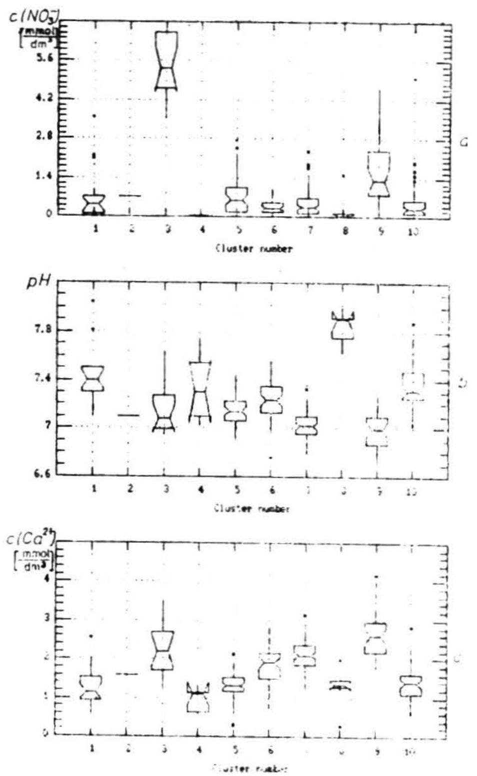
From Table 1 we can see that for equal or similar hydrochemical characteristics the samples fall into one and the same cluster, almost regardless of the aquifer. In the 12-metric space the defining of clusters is performed by taking into account all these characteristics simultaneously. That is why at longer distances from the centre of the "cloud" consisting of 341 units, the grouping in taxons is more definite and in such more clearly defined clusters (with respect to origin) falls a small number of samples. For example cluster 2 consists of only one sample, which is characterized by the highest electric conductivity $\varkappa$ — 4050 μS/cm, correspondingly mineralization M — 2407 mg/dm^3, highest concentrations of $Na^+(+K^+)$ — 25.63 mmol/dm^3, Cl^- — 27.08 mmol/dm^3, SO_4^{2-} — 1.80 mmol/dm^3, an increased content of magnesium — 4.02 mmol/dm^3 (Tables 1 and 2). Going back to the input data we could see that this sample is from a coastal water complex of the Sarmatian (the village of Krapets) whose hydrochemical type is strongly metamorphized as a result of sea water intrusion.

In the first cluster 19 samples are classified, which are characterized at one and the same time by a high $\varkappa$ — 1736.5μS/cm, high ionic strength — 21.22 mmol/dm^3, content of $Na^+(+K^+)$ — 11.13 mmol/dm^3, Cl^- — 10.16 mmol/dm^3 (Table 1, Figs 3c, 4b, 5a). From these 19 samples 13 are surely polluted as a result of the sea water intrusion in the Sarmatian aquifer along the north Black Sea coast. The composition of the other 6 samples (Table 2), obviously also samples 25 and 33, which belong to the Sarmatian, is influenced by the interaction with sea water. There is a communal-living contamination in the case with sample 25 (the village of Durankulak) (the concentration of NO_3^- is 3.58 mmol/dm^3). Sample 35 also comes under this cluster, which probably means that the well, from which the sample is taken (the village of Balgarevo), has wrongly been regarded as pumping water from the loess, instead of the Sarmatian affected by the intrusion. An interesting point is also the inclusion of three samples of the Valanginian (30, 34 and 42) in the first cluster. They are from three deep wells in the region of the village of Vaklino, Russalka and the town of Balchik and are of a chloride-sodium type (the concentrations of Cl^- and $Na^+(+K^+)$ are of the order of 12-13 mmol/dm^3 and in a ratio close to 1, while the mineralization increases — 1 to 1.3 g/dm^3). Since their composition could have hardly been formed as a result of a recent intrusion of sea water, we could rather suggest, that in this case the process of substitution of the syngenetic sea water with infiltration water, although being near to its end, is not over jet.

Table 2

Number of samples by clusters and apriori suggested groups (aquifers)

Apriori suggested groups	Cluster numbers									
	1	2	3	4	5	6	7	8	9	10
$N_2 — Cr_2$	—	—	—	—	6	6	2	—	—	2
River's	—	—	—	—	—	—	—	5	—	—
Q^{al}	—	—	—	—	1	15	13	—	9	5
Q^{ls}	1	—	4	—	9	2	5	—	5	3
N_1^{sm}	2	—	8	2	29	—	—	—	—	20
Cr_1^{apt}	—	—	—	—	—	8	5	1	2	1
Cr_1^{b}	—	—	1	—	9	1	35	2	13	5
Cr_1^{h}	—	—	—	—	3	7	6	—	1	3
Cr_1^{v}	3	—	1	5	6	12	20	2	—	15
Intrusion	13	1	1	—	—	—	1	—	—	4
Total	19	1	15	7	63	51	87	10	30	58



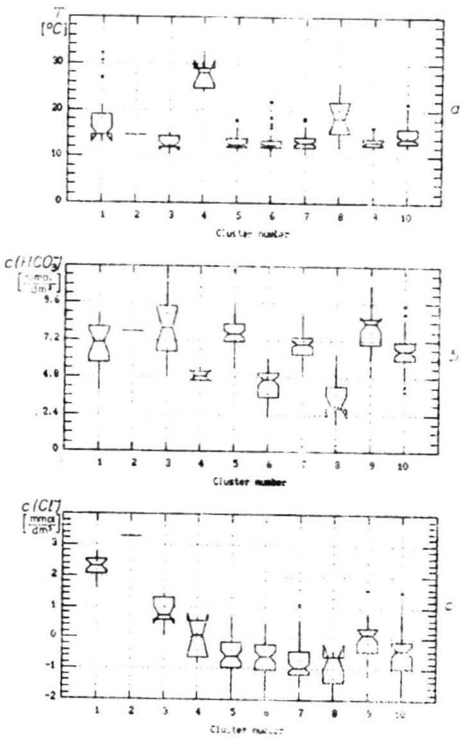


Fig. 2. Box-and-whisker plots of the concentrations of nitrates (*a*) and calcium (*c*) and of pH (*b*)
Fig. 3. Box-and-whisker plots of the temperature (*a*) and of the concentrations of hydrocarbonates (*b*) and chlorides (*c*)

The ground waters of the Valanginian, formed under similar conditions of occurrence, but where the refreshment is in a more advanced phase, belong to the fourth cluster (7 samples). The last one is characterized by the highest temperature of the water (27.7°C) and by the lowest values for the concentrations of Ca^{2+} (0.97 mmol/dm^3), CO_2 (0.0057 bar) and NO_3^- (below 1 mg/dm^3). Under this cluster come also samples 812 and 814 (the village of Gurkovo and Albena), which were wrongly classified when taken, i. e. the Sarmatian (Tables 1 and 2, Figs 2a, 2c, 3a, 4c).

The analysis of the samples (a total of 15), grouped on the basis of mathematical-statistical criteria in cluster 3 shows (Table 1) that here the main criteria are the highest average content of nitrates (369 mg/dm^3 or 5.95 mmol/dm^3) and the lowest temperature (12.9°C) (Figs. 2a, 3a) regardless of the appurtenance of the ground waters to a certain aquifer. Nevertheless the largest number of samples belongs to the Sarmatian (8, and if sample 27 from the well near the village of Durankulak is considered — 9) and to the loess — 4. These are samples taken mainly from wells and drill-holes in towns and villages, where the contaminating sources are of a community-living and an agriculture nature. The fact, however, that in the third cluster are included one sample both from the Barremian and from the Valanginian (Table 2), shows that the front part of contamination (particularly when isolation between the aquifers is missing and also an active microbiological activity has not been performed) advances in depth or that in these cases there is a local contamination. The enhanced content of nitrates is related also to the increase of aggressiveness of the ground waters with respect to water-bearing rocks, building materials and concrete. Thus for the samples of cluster 3 the content of Ca^{2+}, Mg^{2+} and HCO_3^- is also increased (Figs 2c, 3b, 4a). The coefficients of saturation with respect to calcite — C_s^c and dolomite — C_s^d, calculated by the programme WATROCK (V e l i k o v, 1985), have rather low values—respectively 0.079 and 0.017. The distributions of C_s^c and of $\ln C_s^d$, as well as of the index of transformation (anortite-kaolinite) — I_{an-k} (V e l i k o v et al., 1989), are schematically presented in Fig. 6.

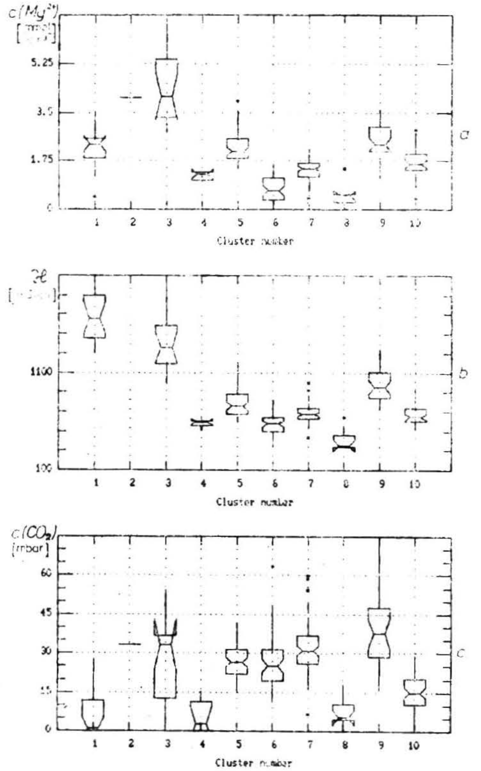
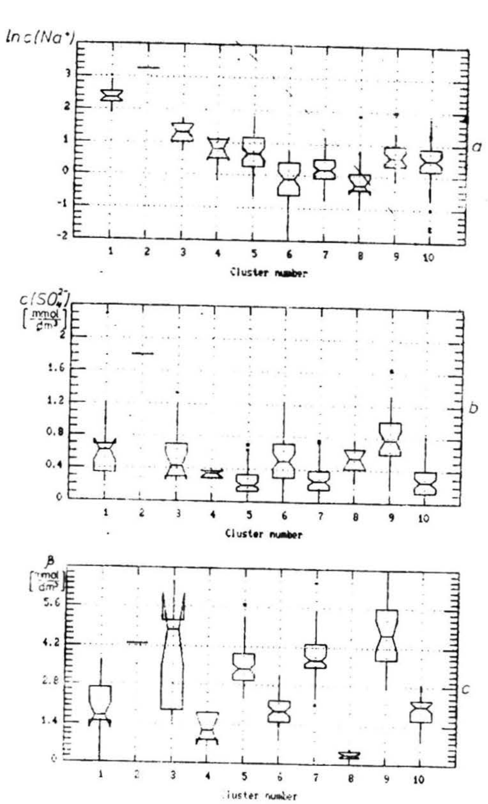


Fig. 4. Box-and-whisker plots of the concentrations of magnesium (*a*) and dissolved inorganic carbon (*c*) and of the electric conductivity (*b*)

Almost the same is the case with cluster 9 (30 samples, 13 belonging to the Barremian, 9 — to the Sarmatian). Here also the concentrations of nitrates are enhanced (an average of 97.2 mg/dm³ or 1.57 mmol/dm³) due to contamination. And thence the aggressiveness and the content of HCO_3^- and Ca^{2+} are increased, pH remaining below 7 (Figs 2a, 2b, 2c, 3b), and C_s^c and C_s^d have also very low values — correspondingly 0.085 and 0.0073.

The waters with the lowest mineralization, respectively concentrations of HCO_3^-, Cl^-, Mg^{2+} and with highest pH — 7.85 are differentiated in cluster 8 (Table 1, Figs 2b, 3b, 3c, 4a). Turning back to the input information we can see that here we have samples from different river waters (702, 705, 715,728, 739), as well as ground waters, whose composition is formed as a result of the direct infusion of surface waters (samples 206 andr 303) or, due to an active water-exchange, theihydrochemical parameters are close to those of the surface waters (samples 313,332,416). It is interesting that all samples of this cluster are characterized by maximal values of the saturation coefficients and of the transformation index (Fig. 6) but with a minimum value of the buffer capacity β — 0.38 mmol/dm³ (Table 1, Fig. 5c).

The most numerous are the samples in the fifth, sixth, seventh and tenth cluster and considerably slightly differing in composition. In the fifth cluster (a total of 63 samples) are included mainly samples of ground waters (mostly from the Sarmatian — 29, loess — 9, the Pliocene and Upper Cretaceous — 6), whose composition is formed not only at the expense of the interaction with carbonates but also with silicate rocks. This is supported by the high value of the ratio Na/Cl — 3.4. The samples of cluster 5 have a lower content of SO_4^{2-}, very low content of Cl^{-2} and an enhanced content of Na^+, Mg^{2+} and often of nitrates (Figs 4a, 5a, 5b, 3c). To this cluster also belong water samples of the part of the Valanginian, which, when isolating layers are missing, is connected, to a certain extent, either to the overlying aquifers or to the earth surface. Probably due to the same reasons here come also 9 samples of the Barremian. The situation is also similar in the sixth cluster; the samples (a total of 51) are also defined with respect to territory — in the southern part of the territory studied (Fig. 1) where the ground waters are accumulated in the limestone of the Valanginian, in the upper weathered part of the marl-calcareous and marl deposits, and also in alluvial deposits of the rivers which are tributaries of the Kamchia



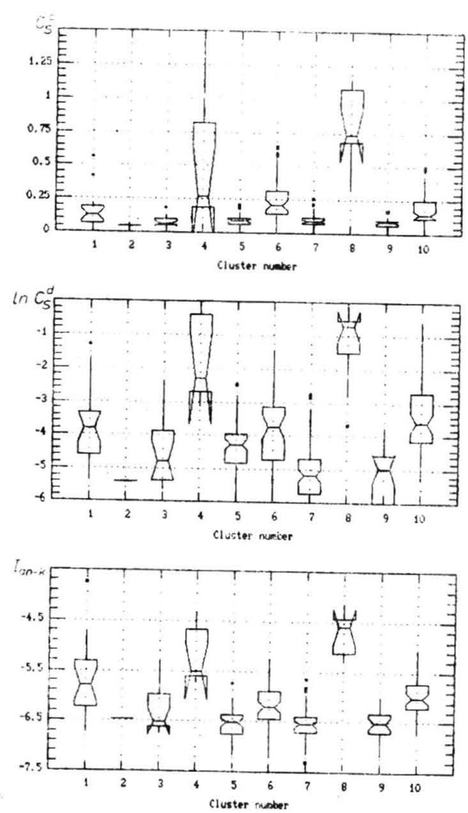


Fig. 5. Box-and-whisker plots of the logarithmus of the sodium concentration (*a*), of the concentration of sulfates (*b*) and of the buffer capacity (*c*)

Fig. 6. Box-and-whisker plots of the saturation coefficient with respect to calcite (*a*), of the logarithmus of the saturation coefficient with respect to dolomite (*b*) and of the transformation index anortite-kaolinite (*c*)

River. These tributaries in the period of summer water deficiency form their out-flow entirely at the expence of the ground waters of the region. The samples in cluster 6 are with a ratio of Na/Cl — 1.7 and have very low concentrations of HCO_3^- (4.34 mmol/dm^3), Na^+ (1.02 mmol/dm^3), Mg^{2+} (0.77 mmol/dm^3), as well as one of the lowest values of specific electric conductivity (Figs 3b, 4a, 4b, 5a).

In the seventh cluster (a total of 87 samples, from which 35 belong to the Barremian and 20 to the Valanginian — Table 2) are grouped the ground waters with a composition formed mainly by their interaction with typical carbonate rocks and Lower Cretaceous marl-calcareous sediments from the zone of an active weathering. The appurtenance of the second largest group of alluvial waters to this taxon may be easily explained by the period of sample-taking, i. e. the summer water deficiency. The content of chloride (0.52 mmol/dm^3), sulphate (0.29 mmol/dm^3) and sodium (1.34 mmol/dm^3) is very low, that of calcium is high (2.11 mmol/dm^3), and of hydrocarbonates is also above the average for the region (6.92 mmol/dm^3) — Figs 2c, 3c, 5a, 5b. The coefficients of saturation with respect to calcite and dolomite (0.1 and 0.009 respectively) allow that the waters in this group are aggressive with respect to the main rock-forming minerals (Figs 6a, 6b).

The tenth cluster — Table 1 — practically unites all the other samples (58 in number),

Table 3

Statistical parameters of the discriminant analysis steps

Step number	Variable entered	F-value to enter or remove	U-statistics	Approximate F-statistic
1	4 Cl	287.7417	0.1133	287.742
2	12 NO_3	84.4695	0.0343	161.304
3	8 pH	50.7480	0.0144	116.635
4	3 HCO_3	40.1181	0.0068	95.132
5	1 T°C	29.4339	0.0038	80.797
6	2 Ca	28.6215	0.0021	72.470
7	5 SO_4	10.7942	0.0016	62.031
8	14 CO_2	9.2579	0.0013	54.534
9	6 Na	8.2275	0.0011	48.909

Table 4

Comparison of classifications performed by cluster and discriminant analysis

Group (Cluster)	Percentage of similarity	Number of cases classified into group									
		A (3)	B (10)	C (5)	D (1)	E (4)	F (7)	G (8)	H (9)	I (6)	J (2)
A (3)	100.0	15	0	0	0	0	0	0	0	0	0
B (10)	89.7	1	52	1	0	1	1	1	0	1	0
C (5)	96.8	0	1	61	0	0	1	0	0	0	0
D (1)	100.0	0	0	0	19	0	0	0	0	0	0
E (4)	100.0	0	0	0	0	7	0	0	0	0	0
F (7)	96.6	0	1	0	0	0	84	0	0	2	0
G (8)	100.0	0	0	0	0	0	0	10	0	0	0
H (9)	100.0	0	0	0	0	0	0	0	30	0	0
J (6)	98.0	0	0	0	0	0	1	0	0	50	0
J (2)	100.0	0	0	0	0	0	0	0	0	0	1
Total	96.5	16	54	62	19	8	87	11	30	53	1

for which the hydrochemical characteristics almost do not differ from the average ones for the region (Figs 2a, 3c, 4b, 5a).

From the results mentioned above it is obvious that there is a good possibility for classification of data using the cluster analysis. One of the problems which, however, cannot be solved by this method and is important for the practical analysis of data, is the classification of observations into groups which have already been defined. With respect to such studies this question comprises also the case of a possible prediction on the appurtenance of a measurement to a certain group, that measurement not being included in the determination of the groups. As far as this has a defined practical value, one of the methods for solving it is the discriminant analysis on the basis of the groups formed by the cluster analysis. This was performed by the programme "7M2" (Dixon, 1981). With the step discriminant analysis for determination of the parameters of the multivariate standard distribution, the different variables are successively included in the analysis, depending on their statistical importance. In Table 3 the variables used and the corresponding Fisher criteria are given. The correspondences between the classification made by the cluster analysis and by the discriminant analysis are presented in Table 4. It can be seen that the results of both methods are almost identical. The differences may be explained on the account of a deviation of the empirical distribution from the standard multivariate one. As a result of the use of the discriminant analysis it is possible to assign the new observationg. with a certain probability, to one of the 10 groups described by simple calculations — e. ,s with a calculator calculating linear polynomials, whose coefficients are the determined discrimination functions (a table with their values is not given here due to their large size and impossibility for clear presentation).

Conclusion

As a conclusion it may be noted that the use of cluster analysis gives a good possibility for grouping of ground waters depending on the processes of formation of their composition and taking into account all hydrochemical characteristics. For example, the fact that the ground waters belonging to clusters 5, 6, 7 and 10 do not differ substantially with respect to chemical composition (regardless of their geological-hydrogeological appurtenance) in our opinion proves the lack of sufficiently strong horizontally and vertically waterproof layers. On the other hand, a most clearly expressed different chemical composition have the waters of the covered part of the Valanginian, included in the 4th cluster, the 8th cluster (mainly river waters), the waters from the 1st and 2nd cluster, to which belong the intrusive waters, as well as the partially altered syngenetic ones of the 3rd and partially of the 9th cluster, in which samples of an increased content of nitrates are included. Thus the cluster analysis allows the early diagnosis of regions and places, where changes in the hydrochemical composition of the waters occur and thus accelerates taking precautions when necessary.

References

Dixon, W. J. 1981. *BMDP Statistical Software — User's Manual.* Univ. of California Press, Los Angeles; 537 p.

Stratigraphics, 1985. *Statistical Graphic System.* STSC Inc., Toronto; 256 p.

Velikov, B. L. 1985. Estimation hydrochimique quantitative des interactions eau-roches carbonatées. — *XXI-ème congrés international de technique hydrothermal,* Varna—Albena, Recueil des rapports, *2*; 170—179.

Velikov, B. L., Machkova, M. S., Dimitrov, D. I. 1989. Estudio hidroquimico de las aguas subterraneas de la region noreste de Bulgaria. — *Hidrogeologia, 4*; 13-24.

Антонов, Хр., Данчев, Д. 1980. *Подземни води на НРБ.* Техника, С.; 340 с.

Афифи, А., Эйзен, С. 1982. *Статистический анализ.* Мир, М.; 488 с.

Великов, Б. Л. 1986. *Хидрохимия на подземните води.* МНП—ВМГИ, С.; 280 с.

Великов, Б. Л., Мачкова, М. С., Димитров, Д. И. 1986. Хидрохимични изследвания на подземните води в Североизточна България. — *Годишн. ВМГИ, 32,* 6; 133—148.