ГЕОЛОШКИ АНАЛИ БАЛКАНСКОГА ПОЛУОСТРВА ANNALES GÉOLOGIQUES DE LA PÉNINSULE BALKANIQUE	74	57–62	БЕОГРАД, децембар 2013 BELGRADE, December 2013

DOI: 10.2298/GABP1374057S

Application of factor analysis in identification of dominant hydrogeochemical processes of some nitrogenous groundwater of Serbia

JANA STOJKOVIĆ[1], PETAR PAPIĆ[1], MARINA ĆUK[1] & MAJA TODOROVIĆ[1]

Abstract. Multivariate statistical analyses are used for reducing large datasets to a smaller number of variables, which explain main hydrogeochemical processes that control water geochemistry. Factor analysis (FA) allows discovering intercorrelations inside the data matrix and grouping of similar variables, i.e. chemical parameters. In this way new variables are extracted, which are called factors, and each factor is explained by some hydrogeochemical process. Applying FA to a dataset that consists of 15 chemical parameters measured on 40 groundwater samples from Serbia, four factors were extracted, which explain 73.9% of total variance in the analyzed dataset. Interpretation of obtained factors indicated several hydrogeochemical processes: the impact of sea water intrusions and volatiles in previous geological periods, solutes diffusion from the marine clay, cation exchange and dissolution of carbonate and silicate minerals.

Key words: factor analysis, hydrogeochemical processes, groundwater, factor loadings, Serbia.

Апстракт. Мултиваријантне статистичке методе користе се у циљу свођења великог броја података на мањи број променљивих, које најбоље објашњавају доминантне хидрогеохемијске процесе одговорне за формирање састава подземних вода. Факторна анализа омогућава откривање интеркорелација унутар скупа података, тј. груписање параметара који су међусобно корелисани. На тај начин се издвајају тзв. фактори, при чему се сваки фактор објашњава одређеним хидрогеохемијским процесом. Применом факторне анализе на матрицу сачињену од 15 параметара хемијског састава одређиваних на 40 узорака подземних вода са територије Србије, издвојена су четири фактора, који објашњавају укупно 73,9% укупне варијансе података. Интерпретација добијених фактора указала је на следеће хидрогеохемијске процесе: утицај морске средине и вулканских испарења у геолошкој прошлости, истискивање везане воде из глина маринског порекла, катјонску измену и растварање карбонатних и силикатних минерала.

Кључне речи: факторна анализа, хидрогеохемијски процеси, подземне воде, факторски коефицијенти, Србија.

Introduction

Assessment of the results of chemical analyses of groundwater often involves a large number of data, rendering the interpretation and presentation of all the information available to the researcher rather challenging. Multivariate statistical methods are very useful tools in hydrogeochemical research, as they allow for the organization and simplification of large datasets. They are a significant contributor to the establishment of correlations between the analyzed chemical parameters, but also to the assessment of similarities between samples (i.e. groundwater occurrences). The goal of multivariate statistical methods is to identify the hydrogeochemical processes that govern the formation of groundwater composition. If the geological and hydrogeological characteristics of the aquifer are known, by applying these methods it is possible to determine the origin and circulation pathways of groundwater. Multivariate statistical methods are also used to define migration factors and the distribution of certain elements. They can point out certain anomalies in the chemical composition of groundwater, for example those of anthropogenic nature (HELENA *et al.* 1999; CLOUTIER *et al.* 2008; YIDANA *et al.* 2008; SUVEDHA *et al.* 2009).

[1] University of Belgrade, Faculty of Mining and Geology, Department of Hydrogeology, Djušina 7, 11000 Belgrade, Serbia. E-mail: janastojkovic@gmail.com

One of the methods often applied in hydrogeochemistry is **factor analysis (FA)**. It uncovers inter-correlations within datasets or allows for mutually-correlated variables to be grouped. The main goal of factor analysis is to isolate as few as possible new variables, which are called **factors**, in order to explain the variance of a large number of analytical data. Consequently, the purpose of this method is to reduce a large number of variables (measured chemical parameters) to the smallest possible number of factors, which are then subjected to interpretation (DREVER 1997; DAVIS 1986; CLOUTIER *et al.* 2008).

Study area

In this research 40 occurrences of Serbian groundwater (Fig. 1) were analyzed and a total of 15 chemical parameters were determined for each sample (macro and micro components, temperature and pH). Analyzed groundwaters are of nitrogenous composition, with a relatively low content of carbon dioxide (in most cases < 100 mg/L CO_2). Sampled groundwaters belong to different geological formations, comprised of igneous, sedimentary and metamorphic rocks, and the majority of these groundwater occurrences are located in Inner Dinarides (14 samples), Vardar Zone (20 samples) and Serbian-Macedonian Massif (six samples). Geological, structural and hydrogeological conditions in the area of investigated groundwaters are very complex. Different types of the Proterozoic to Paleozoic crystalline schists are present, and also varieties of Paleozoic and Mesozoic sediments, granitoide intrusions and the Tertiary volcanic rocks, and also characteristic oceanic elements (DIMITRIJEVIĆ 1995). Analyzed groundwaters are from different types of aquifers formed in these rocks, with the predominance of fracture aquifers.

Factor analysis was applied to this dataset to identify the dominant hydrogeochemical factors and processes that lead to the formation of the groundwater composition.

Methods

Factor analysis was applied to a set of hydrogeochemical data comprised of 15 measured chemical composition parameters of 40 groundwater samples collected in Serbia. The concentrations (in mg/L) of the following elements were analyzed: calcium, magnesium, sodium, potassium, chlorine, hydrocarbonate, sulfate, silicon, fluorine, boron, lithium, strontium and carbon dioxide, as was temperature (°C) and pH. IBM SPSS Statistics 19.0 software was used for statistical analysis.

Elementary statistical quantities (arithmetic mean, minimum and maximum values, median, etc.) were determined for the analyzed set of the hydrochemical data. All the variables were subjected to ln-transformation (computation of natural logarithm of all the analyzed data). The transformed data complied with the normal distribution criterion, corroborated by the Kolmogorov-Smirnov test.

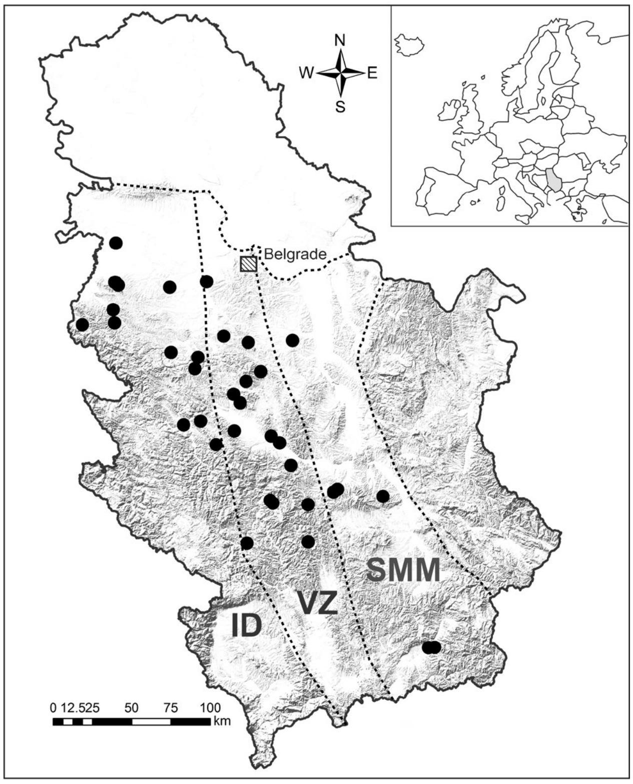

Fig. 1. Position of the study area, with the locations of analyzed groundwaters. Investigated geological-tectonic units of Serbia: **ID,** Inner Dinarides; **VZ,** Vardar Zone; **SMM,** Serbian-Macedonian masif.

The number of the extracted factors was determined based on the Kaiser criterion (KAISER 1960), according to which only those factors whose **eigenvalue** (characteristic value of the correlation matrix) is greater than one are taken into account. This was consistent with Cattell's scree plot, where factors constituted the X axis and their eigenvalues the Y axis. The curve was cut-off at the point of inflexion and the portion of the curve that exhibited a less steep decline was discarded (CATTELL 1966). To facilitate interpretation of the extracted factors, varimax orthogonal rotation was applied to enhance the contribution of significant variables and reduce that of less significant ones (HELENA *et al.* 1999; FIELD 2005).

Results

Based on the elementary statistical quantities shown in Table 1, it was concluded that the concentrations of most of the measured parameters did not follow normal distribution. Their distribution histograms were positively skewed, as indicated by distinctly positive coefficients of asymmetry (Table 1). For this reason ln-transformed data were used in factor analysis.

Table 1. Elementary statistical quantities for the 40 groundwater samples.

Parameters	Minimum	Maximum	Range	Mean	Median	Skewness
Temperature (°C)	13.10	83.20	70.10	28.36	21.10	1.72
pH	6.60	9.23	3.17	7.47	7.32	0.74
CO_2 (mg/L)	0.00	171.24	171.24	51.84	35.20	1.01
Ca^{2+} (mg/L)	0.00	130.26	130.26	51.21	46.89	0.48
Mg^{2+} (mg/L)	1.82	96.31	94.49	22.12	13.98	1.93
Na^+ (mg/L)	2.50	684.00	681.50	139.42	64.10	1.87
K^+ (mg/L)	0.20	51.20	51.00	5.80	2.60	3.90
Cl^- (mg/L)	2.00	223.34	221.34	44.20	21.98	2.04
HCO_3^- (mg/L)	124.44	1770.00	1645.56	504.04	400.00	2.35
So_4^{2-} (mg/L)	1.20	240.00	238.80	35.39	15.40	2.65
SiO_2 (mg/L)	9.39	91.60	82.21	32.59	25.10	1.27
F^- (mg/L)	0.05	13.00	12.95	1.62	0.70	3.12
B (mg/L)	0.00	32.60	32.60	2.02	0.29	5.09
Li^+ (mg/L)	0.003	4.78	4.777	0.31	0.10	5.21
Sr^{2+} (mg/L)	0.004	2.10	2.096	0.46	0.27	1.56

The application of factor analysis to the set of 15 variables (i.e. chemical parameters) determined for 40 groundwater samples produced four factors that together accounted for 73.9% of the total variance of the analyzed data. Table 2 shows the extracted factors, their factor loadings and the attributed percentage of the variance. **Factor loadings** represent coefficients of correlation between the variables and factors or, in other words, they indicate the relative contribution of a certain variable to each of the extracted factors (FIELD 2005). In this example, only the factor loadings whose absolute value was greater than 0.5 (bolded values in Table 2) were interpreted (STEVENS 1992). It was apparent that several variables exhibited high loadings on each factor, such that the 15 initial variables were classified into four groups, depending on their mutual similarity, to facilitate subsequent interpretation.

The first two factors accounted for nearly 50% of the variance, while the third and the fourth factors accounted for 13.3% and 11.6%, respectively. The first factor featured very high positive loadings of B, Na^+ and Cl^- (> 0.85), as well as high positive loadings of K^+, Li^+ and HCO_3^- (> 0.6). The relatively high loading of F^- (0.495) should also be noted. The second factor was characterized by high positive loadings of Ca^{2+}, Sr^{2+}, Mg^{2+}

Table 2. Factor loadings and percentage of variance explained by the four extracted factors, with varimax rotation (values in bold represent loadings with absolute values > 0.5).

Parameters	Factor 1	Factor 2	Factor 3	Factor 4
B	**0.901**	-0.190	0.085	0.045
Na^+	**0.891**	-0.270	0.164	0.158
Cl	**0.858**	0.037	0.121	-0.020
K^+	**0.687**	0.223	0.275	0.274
Li^+	**0.649**	-0.112	0.375	0.366
HCO_3	**0.632**	***0.473***	-0.430	-0.040
pH	-0.033	**-0.825**	-0.152	0.218
Ca^{2+}	-0.240	**0.808**	-0.210	-0.049
Sr^{2+}	0.207	**0.721**	0.036	0.050
Mg^{2+}	-0.260	**0.620**	-0.219	0.067
T	0.202	0.093	**0.862**	-0.025
SiO_2	0.205	-0.256	**0.742**	0.357
SO_4^2	0.191	-0.087	0.089	**0.830**
CO_2	-0.075	**0.537**	0.026	**0.579**
F	***0.495***	-0.415	0.362	**0.541**
% of variance	27.800	21.200	13.300	11.600
cumulative % of variance	27.800	49.000	62.300	**73.900**

and CO_2, as well as a high negative loading of pH, where the loading of HCO_3^- (0.473) should not be disregarded. All this is also shown in Fig. 2, where the factor loadings of all variables were plotted: the X axes represents factor 1 (left) and factor 3 (right), the Y axes represents factor 2 (left) and factor 4 (right). The variables that dominate each factor are apparent (marked by the ellipse).

The third and fourth factors accounted for the smaller portion of the variance. This was attributed to hydrogeochemical processes of a more local nature, which take place only in a certain number of groundwater occurrences (CLOUTIER *et al.* 2008). The third factor was characterized by high positive loadings of temperature and SiO_2. The fourth factor was dominated by SO_4^{2-}, but the factor loadings of CO_2 and F^- were also relatively high.

from the clays of marine origin (CLOUTIER *et al.* 2008; REIMANN & BIRKE 2010). Another possible process is cation exchange between Ca^{2+} and Mg^{2+} from the water and Na^+ from the aquifer matrix. Namely, as carbonate minerals dissolve, the groundwater becomes enriched with calcium, magnesium and hydrocarbonates, followed by the previously mentioned cation exchange, such that Ca^{2+} and Mg^{2+} concentrations in groundwater decrease while the Na concentration increases. This theory was supported by the negative factor loadings of Ca^{2+} and Mg^{2+}, and the positive factor loadings of Na^+ and HCO_3^- (GUO *et al.* 2007, CLOUTIER *et al.* 2008, SALIFU *et al.* 2011). The positive loadings for boron, potassium, lithium and fluorine of the first factor should also be noted, and they were attributed to paragenesis of these microelements and their similar hydrogeochemical behavior.

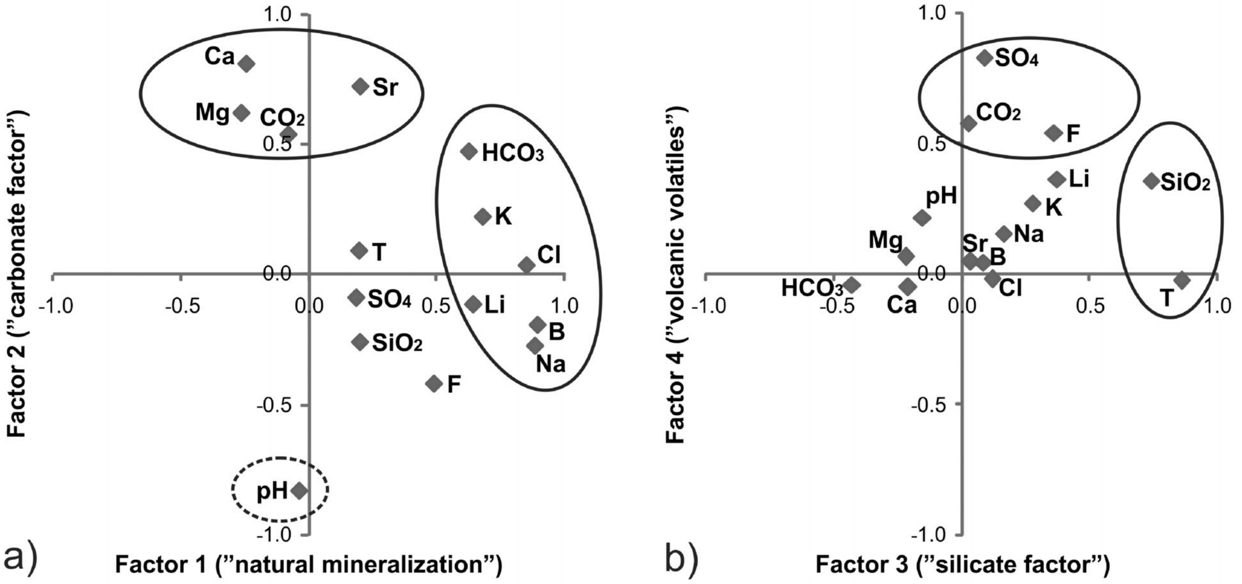

Fig. 2. Plot of factor loadings for the first and the second factors (**a**) and for the third and the fourth factors (**b**). The variables that dominate each factor are marked by the ellipse.

Discussion

If the extracted factors are viewed in a geological (primarily lithological) context, it is possible to gain insight into the main hydrogeochemical processes that lead to the formation of the chemical composition of the analyzed groundwater. In factor analysis, often all or at least the main factors are assigned conditional names, indicative of the variables that dominate the given factor. The first factor was dominated by B, Na^+, Cl^-, K^+, Li^+ and HCO_3^-, such that this factor could be called “natural mineralization” because it contains Na^+, Cl^-, K^+ and HCO_3^- that represent the ions of the basic chemical composition. Very high positive loadings of B, Na^+ and Cl^- (> 0.85) in the first factor were attributed to the groundwater mixing with seawater in the geological past, but also to the solutes diffusion

The second factor featured elevated positive loadings of Ca^{2+}, Sr^{2+}, Mg^{2+} and CO_2, and an elevated negative loading of pH. Here too, HCO_3^- needed to be taken into consideration. This factor can be called the “carbonate factor” because the dominant variables indicate the processes of dissolution of carbonate minerals. The presence of carbon-dioxide tends to render groundwater aggressive and enables the dissolution of calcite, dolomite etc., whereby Ca^{2+}, Mg^{2+} and HCO_3^- ions are released into the groundwater. This is consistent with the high positive loadings of Ca^{2+}, Mg^{2+}, CO_2 and HCO_3^-. The process takes place in an acidic environment, where the concentration of CO_2 and the pH level are inversely proportional, resulting in a negative factor loading of pH. The high positive factor loading of strontium was attributed to its paragenesis with Ca^{2+}. These two elements are chemically similar and Sr^{2+} is therefore a frequent ingredient of Ca^{2+} minerals (HITCHON 1999).

The third factor highlighted the loadings of temperature and SiO_2, attributed to the fact that the solubility of silicate minerals increases with increasing temperature (MATTHESS 1981), such that this factor could be called the “silicate factor”. The fourth factor featured elevated loadings of SO_4^{2-}, CO_2 and F^-. This association is indicative of the volatiles from volcanic activity in the geological past and the factor was given the name “volcanic volatiles”.

Conclusions

Factor analysis is an efficient tool for assessing hydrogeochemical data because of the high data variance caused by a series of geological, hydrogeological and other factors. It enables the identification of the correlations between the analyzed chemical parameters and also their grouping into factors based on similarity, which facilitates subsequent interpretation. In the present case study, factor analysis was applied to extract four dominant factors that accounted for most of the variance (73.9%) of the input dataset, which consisted of 15 chemical parameters measured on 40 groundwater samples from Serbia. The interpretation of obtained factors has indicated several hydrogeochemical processes: the effects of a marine environment and volcanic volatiles in the geological past, the solutes diffusion from the clays of marine origin, cation exchange, and the dissolution of carbonate and silicate minerals. The results uphold the significance of multivariate statistical analysis in the determination of groundwater genesis, or of the factors and processes that govern the formation of the chemical composition of groundwater.

Acknowledgments

Authors thank ALEXEY BENDEREV (Bulgarian Academy od Sciences) and TAQVEEM ALI KHAN (Department of Geology, A.M.U.Aligarh, India) for the critical comments and helpful suggestions This research was supported by the Serbian Ministry of Education, Science and Technological Development (Project No. 43004). .

References

CATTELL, R.B. 1966. The scree tests for the number of factors. *Multivariate Behavioral Research*, 1: 245–276.

CLOUTIER, V., LEFEBVRE, R., THERRIEN, R. & SAVARD, M.M. 2008. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology*, 353: 294–313.

DAVIS, J.C. 1986. *Statistics and Data Analysis in Geology.* 656 pp. John Wiley & Sons Inc., New York.

DIMITRIJEVIĆ, M.D. 1995. *Geologija Jugoslavije*. 267 pp. Geoinstitut, Beograd.

DREVER, J. 1997. *The Geochemistry of Natural Waters*. 436 pp. Prentice Hall, New Jersey.

FIELD, A. 2005. *Discovering Statistics Using SPSS.* 856 pp. SAGE Publications Ltd, London.

GUO, Q., WANG, Y., MA, T. & MA, R. 2007. Geochemical processes controlling the elevated fluoride concentrations in groundwaters of the Taiyuan Basin, Northern China. *Journal of Geochemical Exploration,* 93: 1–12.

HELENA, B.A., VEGA, M., BARRADO, E., PARDO, R. & FERNANDEZ, L. 1999. A case of hydrochemical characterization of an alluvial aquifer influenced by human activities, *Water, Air and Soil Pollution*, 112: 365–387.

HITCHON, B. 1999. *Introduction to Ground Water Geochemistry.* 310 pp. Geoscience Publishing Ltd., Alberta.

KAISER, H.F. 1960. The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20: 141–151.

MATTHESS, G. 1981. *The Properties of Groundwater*. 406 pp. Wiley-Interscience publications, New York.

REIMANN, C. & BIRKE, M. 2010. *Geochemistry of European Bottled Water.* 268 pp. Borntraeger Science Publishers, Stuttgart.

SALIFU, A., PETRUŠEVSKI, B., GHEBREMICHAEL, K., BUAMAH, R. & AMY, G. 2011. Fluoride occurrence in groundwater in the Northern region of Ghana, *Proceedings of IWA Specialist Groundwater Conference in Belgrade*, 267–275.

STEVENS, J.P. 1992. *Applied multivariate statistics for the social sciences.* 648 pp. Lawrence Erlbaum Associates, Inc., New Jersey.

SUVEDHA, M., GURUGNANAM, B., SUGANYA, M. & VASUDEVAN, S. 2009. Multivariate Statistical Analysis of Geochemical Data of Groundwater in Veeranam Catchment Area, Tamil Nadu. *Journal Geological Society of India*, 74: 573–578.

YIDANA, S.M., OPHORI, D. & BANOENG-YAKUBO, B. 2008. Hydrochemical evaluation of the Voltaian system – The Afram Plains area, Ghana. *Journal of Environmental Management*, 88: 697–707.

Резиме

Примена факторне анализе у циљу идентификације доминантних хидрогеохемијских процеса у неким азотним подземним водама Србије

Приликом статистичке обраде хидрохемијских података значајно место заузимају мултиваријантне статистичке методе. Оне олакшавају организовање и сагледавање великог броја аналитичких података, првенствено физичко-хемијских карактеристика подземних вода, а омогућавају и кла-

сификовање испитиваних узорака вода на основу већег броја одабраних параметара. Једна од метода које се често примењују у хидрохемији и хидрогеологији уопште јесте факторна анализа. Њеном употребом идентификују се и наглашавају статистичке релације између анализираних хидрохемијских параметара, уз накнадно тумачење успостављених релација са аспекта хидрогеохемијских процеса у подземним водама.

Употреба факторне анализе у овом раду омогућила је груписање хидрохемијских параметара који су међусобно корелисани и који се могу довести у везу са одређеним факторима и процесима формирања хемијског састава подземних вода. Применом ове статистичке методе на матрицу сачињену од 15 параметара хемијског састава, одређиваних на 40 узорака подземних вода са територије Србије, издвојена су четири фактора, који заједно објашњавају 73,9 % укупне варијансе података, од чега су прва два фактора одговорна за скоро 50 % укупне варијансе. Први фактор карактерише доминација B, Na, Cl, K, Li и HCO_3, па је условно назван „природна минерализација“, док код другог фактора доминирају Ca, Sr, Mg и CO_2, па му је додељен назив „карбонатни фактор“. Трећи и четврти фактор објашњавају мањи део укупне варијансе, па се њима приписују хидрогеохемијски процеси локалног карактера, који се јављају само код одређеног броја испитиваних појава подземних вода. Трећи фактор карактеришу температура и SiO_2 („силикатни фактор“), док су код четвртог фактора изражени SO_4, CO_2 и F („испарења вулкана“).

Сагледавањем издвојених фактора у геолошком, првенствено литолошком контексту, стиче се увид у главне хидрогеохемијске процесе који су значајни за формирање хемијског састава испитиваних подземих вода. Тако су издвојени следећи процеси: утицај морске средине и вулканских испарења у геолошкој прошлости, истискивање везане воде из глина маринског порекла, катјонска измена и растварање карбонатних и силикатних минерала. Добијени разултати указују на значај употребе факторне анализе, као и мултиваријантне статистичке анализе уопште, приликом утврђивања генезе подземних вода, то јест приликом дефинисања геохемијских и хидрогеолошких услова формирања тих вода.