

A multivariate Statistical Analysis of Groundwater Chemistry Data

Belkhiri, L.^{1*}, Boudoukha, A.² and Mouni, L.³

¹Department of hydraulics, University Hadj Lakhdar 05000 Batna, Algeria

²Research Laboratory in Applied Hydraulics, University Hadj Lakhdar 05000 Batna, Algeria

³Departement de Génie des Procèdes, Faculté de Technologie, Université de Bejaia, Targa, Ouzemour 06000, Algeria

Received 6 Feb. 2010;

Revised 10 Aug. 2010;

Accepted 17 Aug. 2010

ABSTRACT: Q-mode hierarchical cluster (HCA) and principal component analysis (PCA) were simultaneously applied to groundwater hydrochemical data from the three times in 2004: June, September, and December, along the Ain Azel aquifer, Algeria, to extract principal factors corresponding to the different sources of variation in the hydrochemistry, with the objective of defining the main controls on the hydrochemistry at the aquifer scale. Hydrochemical data for 54 groundwater samples were subjected to Q-mode hierarchical cluster and principal component analysis. The study finds, from Q-mode HCA that there are three main hydrochemical facies namely the less saline water (group 1: Ca-Mg-HCO₃), mixed water (group 2: Mg-Ca-HCO₃-Cl) and blended water (group 3: Mg-Ca-Cl-HCO₃). In principal component analysis, the first 4 factors explain 72.14% of the total variance, their loadings allowing the interpretation of hydrochemical processes that take place in the area. The results of this study clearly demonstrate the usefulness of multivariate statistical analysis in hydrochemical.

Key words: Hydrochemistry, Multivariate techniques, Q-mode hierarchical cluster analysis, Principal component analysis, Ain Azel plain, Algeria

INTRODUCTION

Water is very vital for nature and can be a limiting resource to men and other living beings. Without a well functioning water supply, it is difficult to imagine productive human activity be it agriculture or livestock. The quality of water is of almost importance to quantity in any water supply planning. Water quality is influenced by natural and anthropogenic effects including local climate, geology and irrigation practices. The chemical character of any groundwater determines its quality and utilization. The quality is a function of the physical, chemical and biological parameters, and could be subjective, since it depends on a particular intended use.

Literature shows the always-increasing potentiality of chemometric methods in obtaining useful information from environmental data, which hardly could be otherwise correlated and interpreted. In particular many examples can be found of the application of multivariate analysis to sets of variables collected for surface and ground waters.

Multivariate statistical techniques, cluster analysis (CA) and factor analysis (FA), are effective means of manipulating, interpreting and representing data concerning groundwater pollutants and geochemistry. They are frequently employed to characterize the quality of groundwater. Liu *et al.* (2003) adopted FA to evaluate groundwater quality in a blackfoot disease area, Taiwan. Two factors, seawater salinization and arsenic pollutant, were considered and their factor scores were mapped. Love *et al.* (2004) applied FA to distinguish several groundwater signatures, including uncontaminated groundwater, agricultural activities, mining activities and sewage pollution. Olmez *et al.* (1994), Mahknecht *et al.* (2003) and Farnham *et al.* (2003) used FA to discuss geochemical evolution and mineralization and groundwater contamination. Additionally, the CA was also used to interpret the hydrochemical data based on factor scores, see, for example, Suk and Lee (1999), Reghunath *et al.* (2002) and Kim *et al.* (2005).

The main objectives of this paper are: (1) to assess the chemistry of groundwater and (2) to identify the

*Corresponding author E-mail: BELKHIRI_Laz@yahoo.fr

geologic factors that presently affect the water chemistry in the region by using multivariate statistical techniques.

The studied area is located in the east of Algeria. Most of inhabitants (more than 30000 inhabitants) are centered on the town of Ain Azel.

The principal activity in this area is the production of cereals (barley and corn) and its climate is semi-arid. The amount of rainfall varies from place to place and from season to season, ranging from 346 mm/yr in the plains (altitude < 1000 m) to 534 mm/yr in the mountains (altitude > 1000 m). The annual maximum and minimum temperatures are 21.8 and 8.5°C, respectively (Belkhir, 2005).

According to many authors (Savornin, 1920; Galcon, 1967; Guiraud, 1973; Vila, 1980) the area of concern is distinguished by two geological sets. In the South, autochthonous Jurassic and Cretaceous carbonate lithologies are mainly observed in Djebels Boutaleb, Djebel Hadjar Labiod and Fourhal (Fig. 1). In the North, allochthonous Jurassic and Cretaceous lithologies are observed in Djebels Kalaoun and Sekrine. A lithostratigraphic study of Ain Azel area identifies the following two aquifer formations. The first is a carbonate and sandy formation of about 700 m thick presenting fracture porosity and constitutes the Barremian formation. The second is an alluvial formation of about 250 m thick corresponding to a Mio-Plio-Quaternary formation with interstitial porosity (Boutaleb, 2001; Belkhir, 2005). A mineralogical study shows that the metalliferous minerals, particularly the sphalerite, the gangue minerals and the dolomite contain variable quantities of trace metals such as Zn, Pb, Cd, Fe and Cu (Boutaleb, 2001). The studied area is situated in the alluvial plain of the Mio-Plio-Quaternary (Fig. 1) showing a heterogeneous continental detrital sedimentation (Boutaleb, 2001; Belkhir, 2005; Attoucheik, 2006).

MATERIALS & METHODS

For chemical analysis, a total of 54 groundwater samples from the Ain Azel plain (from 18 wells) were obtained during 2004 from three separate sampling campaigns (June, September and December) at the sampling points shown in Fig. 1. The samples were collected after pumping for 10 min. This was done to remove groundwater stored in the well. These samples were collected using 4-l acid-washed polypropylene containers. Each sample was immediately filtered on site through 0.45 µm filters on acetate cellulose. Filtrate for metals analyses were transferred into 100-cm³ polyethylene bottles and immediately acidified to pH < 2 by the addition of Merck™ ultrapure nitric acid (5 ml 6N HNO₃). Samples for anions analyses were collected into 250-cm³ polyethylene bottles without

preservation. All the samples were stored in an ice chest at a temperature of <4°C and later, transferred to the laboratory and stored in a refrigerator at a temperature of <4°C until analyzed (within 1 week). Immediately after sampling, pH, and electrical conductivity (EC) were measured in the field using a multi-parameter WTW (P3 MultiLine pH/LF-SET).

Subsequently, the samples were analyzed in the laboratory for their chemical constituents such as calcium, magnesium, sodium, potassium, chloride, bicarbonate, sulfate, lead, iron, zinc and copper. This was achieved using standard methods as suggested by the American Public Health Association (APHA 1989, 1995a, 1995b). Ca, Mg, HCO₃ and Cl were analyzed by volumetric titrations. Concentrations of Ca and Mg were estimated titrimetrically using 0.05N EDTA and 0.01N and those of HCO₃ and Cl by H₂SO₄ and AgNO₃ titration, respectively. Concentrations of Na and K were measured using a flame photometer (Model: Systronics Flame Photometer 128) and that of sulfate by turbidimetric method (Clesceri *et al.*, 1998). Standard solutions for the above analysis were prepared from the respective salts of analytical reagents grades. Trace metals were determined by Graphite Furnace Atomic Absorption Spectrophotometer (Perkin-Elmer AAnalyst 700) using multi element Perkin-Elmer standard solutions. The accuracy of the chemical analysis was verified by calculating ion-balance errors where the errors were generally around 10%.

All mathematical and statistical computations were made using Excel 2003 (Microsoft Office[®]) and STATISTICA 6 (StatSoft, Inc.[®]). Multivariate analysis of the lake groundwater quality data set was performed through cluster and principal component analysis techniques (Liu *et al.*, 2003).

Cluster analysis was used to determine if the samples can be grouped into statistically distinct hydrochemical groups that may be significant in the geologic context. A number of studies used this technique to successfully classify water samples (Alther, 1979; Williams, 1982; Farnham *et al.*, 2000; Alberto *et al.*, 2001; Meng and Maynard, 2001). Comparisons based on multiple parameters from different samples were made and the samples were grouped according to their 'similarity' to each other. Classifications of samples according to their parameters are known as Q-mode classifications. In the present study Q-mode hierarchical cluster analysis (HCA) was used to classify the samples into distinct hydrochemical groups. The Ward's linkage method (Ward, 1963) was used in this analysis. A classification scheme using Euclidean distance (straight line distance between two points in *c*-dimensional space defined by *c* variables) for similarity measurement, together with

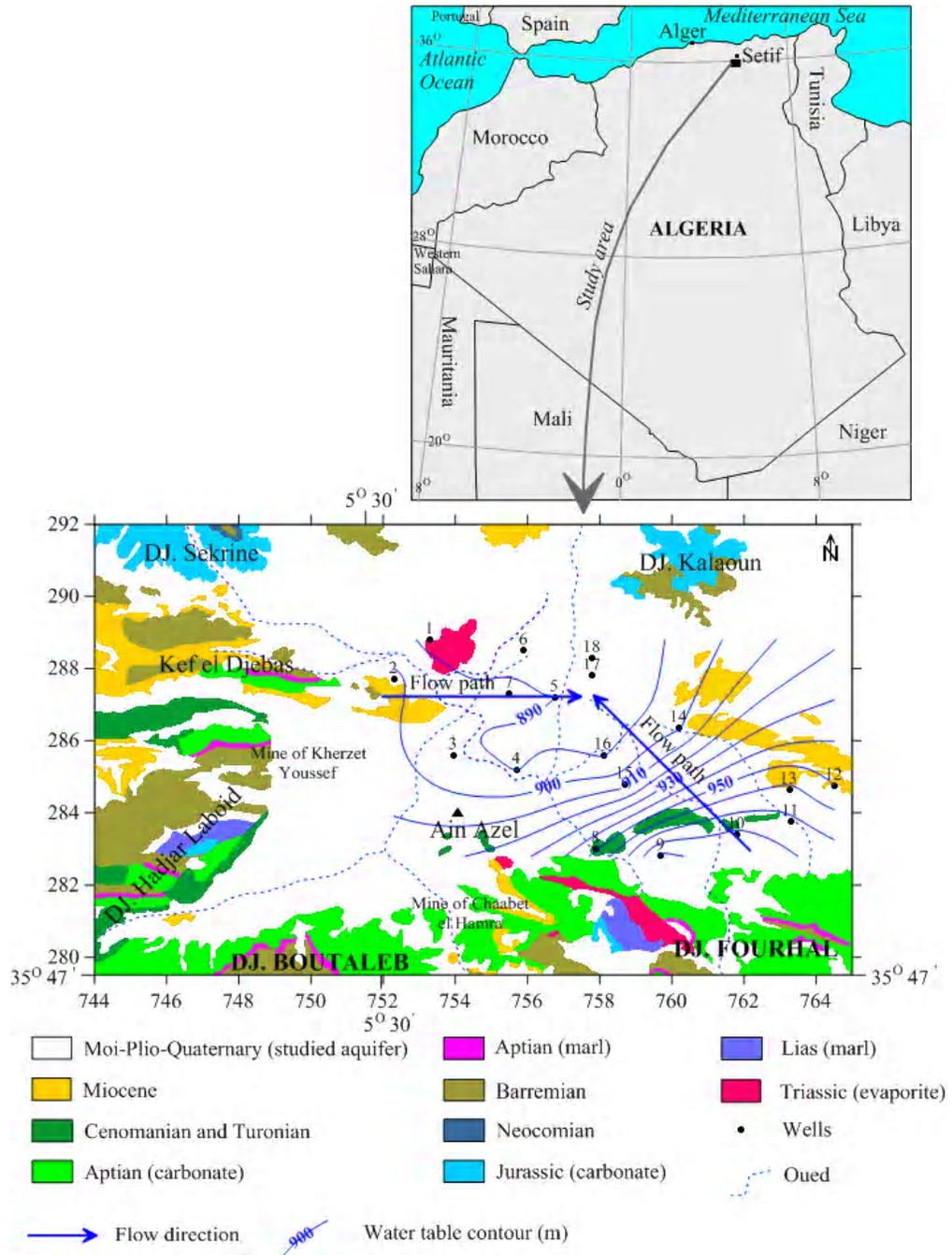


Fig. 1. Map showing water sampling locations and geology of the studied area

Ward's method for linkage, produces the most distinctive groups where each member within the group is more similar to its fellow members than to any member outside the group (Güler *et al.*, 2002). All 13 hydrochemical variables measured (consisting of EC, pH, Ca, Mg, Na, K, Cl, SO_4 , HCO_3 , Pb, Fe, Zn and Cu) were utilized in this analysis. For statistical analysis, all the variables were log-transformed and more closely correspond to normally distributed data. Subsequently, they were standardized to their standard scores (z-scores) as described by Güler *et al.* (2002).

Factor analysis is a multivariate analytical technique, which derives a subset of uncorrelated variables called factors that explain the variance observed in the original dataset (Anazawa and Ohmori, 2005; Brown, 1998). Factor analysis is used to uncover the latent structure of a set of variables. In technical terms, common factor analysis represents the common variance of variables, excluding unique variance, and is thus a correlation-focused approach seeking to reproduce the inter-correlation among the variables.

On the other hand, components (from PCA) reflect both common and unique variance of the variables and may be seen as a variance-focused approach that reproduces both the total variable variance with all components as well as the correlations. PCA is far more commonly used than principal factor analysis (PFA). However, it is common to use “factors” interchangeably with “components” in multivariate analysis. Factor analysis can be performed on any kind of scientific data to establish a pattern of variation among variables or reduce large data sets into factors for easy handling and interpretation. The total number of factors generated from a typical factor analysis indicates the total number of possible sources of variation in the data. Factors are ranked in order of merit. The first factor or component has the highest eigenvector sum and represents the most important source of variation in the data. The last factor is the least important process contributing to the chemical variation. Factor loadings on the factor loadings tables are interpreted as correlation coefficients between the variables and the factors. In this research, PCA was applied to chemical data from the Ain Azel plain to extract the principal factors corresponding to the different sources of variation in the data. Here, PCA was selected for the reasons stated above. In order to maximize the variation among the variables under each factor, the factor axes were subsequently varimax rotated.

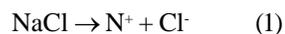
RESULTS & DISCUSSION

A statistical summary of chemical parameters measured in the groundwater samples is presented in Table 1. The groundwater samples of the study area have pH values ranging from 6.7 to 7.9, which indicate that the groundwater is slightly alkaline. The electrical conductivity (EC) values ranged 470 to 2,578 μ S/cm. The order of abundance of the major cations is Ca²⁺ Mg²⁺ > Na⁺ > K⁺ and all samples exceeded the desirable limit of Ca for drinking water (75 mg/L) except sample 10, but only 78% of them exceed that of Mg (50 mg/L). The abundance of the major anions is HCO₃⁻ Cl⁻ > SO₄²⁻ and almost 28 % of the samples exceeded the desirable limit of Cl (200 mg/L), but the sulfate concentrations are all below health guidelines (WHO, 1993). The concentration of lead for the three campaigns ranges from 0.02 to 1.83 mg/L and all samples exceeded the WHO guideline limit of 0.01 mg/L. In the case of iron, the concentration in many of the samples was higher than the WHO permitted limit of 0.3 mg/L and the percent samples above the WHO limit ranges from 60 to 83 % for the three campaigns. Since in this region, except for agricultural activity no other anthropogenic activities were being carried out in this region and hence the high Fe concentration in these waters could not be assigned to the

anthropogenic activities but rather the water-rock interaction should be the reason for the higher value of Fe observed in the plain. Zinc and copper values ranging from 0.045 to 0.304 mg/L, and 0.067 to 0.431 mg/L respectively, all of the studied samples Zn and Cu values are lesser than the prescribed WHO standard (WHO, 1993). It is clearly observed that the concentrations of the major elements do not undergo an appreciable change during the three campaigns. For this reason the three campaigns were grouped as one.

The data were classified in a simple and direct manner with results presented as dendrograms (Fig. 2). Based on an imaginary horizontal line (phenon line) on the cluster scale three distinct groups or clusters. The first group of waters, group 1, has low salinity (mean EC = 782 μ S/cm) and abundance orders (meq/l) Ca²⁺ Mg²⁺ > Na⁺ > K⁺ and HCO₃⁻ > Cl⁻ > SO₄²⁻ (Fig. 3). This group is basically bicarbonate dominated, however; calcium and magnesium are also present (Ca-Mg-HCO₃). Among the trace elements identified, the higher concentrations of the lead and iron (Table 2). The water type 2 (group 2) is bicarbonate and chloride dominated and it also has low concentrations of sulfate (Fig. 3). The electrical conductivity for this group is 1,062 μ S/cm which is the characteristic of mixed water (Mg-Ca-HCO₃-Cl). Among trace elements, lead and iron concentrations are highest (Table 2). The water type 3 (group 3) is chloride and bicarbonate dominated but also contains low concentrations of sulfate. The magnesium and calcium are the dominant elements. The EC is 1,956 μ S/cm which is the characteristic of blended water (Mg-Ca-Cl-HCO₃). As far as trace elements are concerned, this water has the lowest concentrations of most of the trace elements compared to other waters of the study area. This water type has the highest level of lead (Table 2).

Correlation coefficient is commonly used to measure and establish the relationship between two variables. It is a simplified statistical tool to show the degree of dependency of one variable to the other. The correlation matrix of ten variables has been presented in Table 3. The EC values exhibit high positive correlation with Ca, Mg, Na, Cl and HCO₃⁻. Calcium and magnesium presented a strong positive correlation (0.76), indicating a common source. There are two outliers, which may be related to different soil parent materials or geology of the sampling sites. Cl and Na possess a very good positive correlation (0.91) between each other. The high Na and Cl contents detected in certain samples may suggest the dissolution of chloride salts. The dissolution of halite in water release equal concentrations of Na and Cl into the solution:



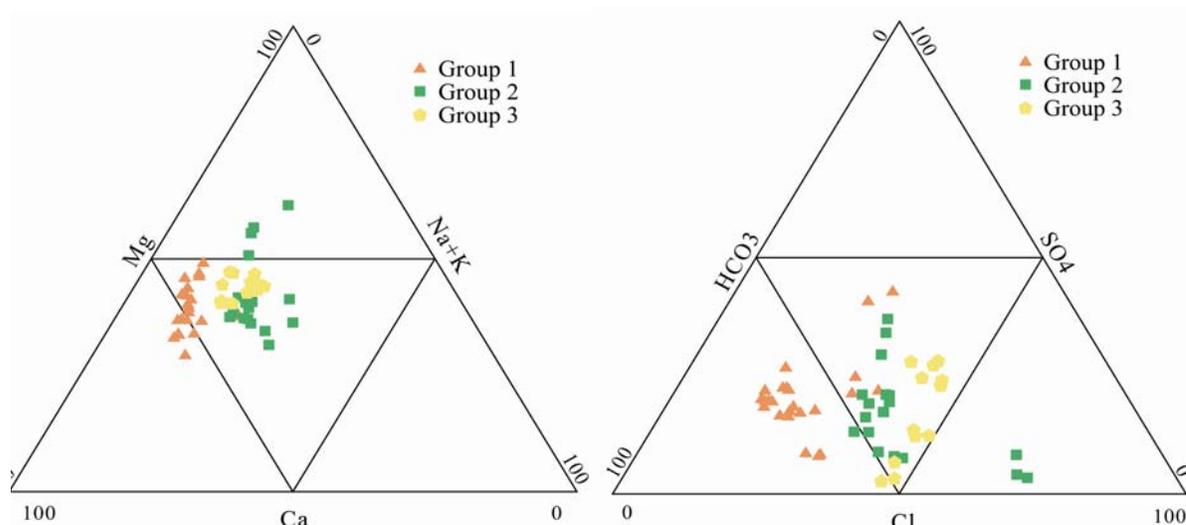


Fig. 3. Ca-Mg-(Na+K) and HCO₃-SO₄-Cl trilinear diagrams for groundwater samples

Table 2. Mean parameter values of the three principal water groups

Groups	EC	pH	Ca	Mg	Na	K	Cl	SO ₄	HCO ₃	Pb	Fe	Zn	Cu
1	782	7.0	108.93	54.98	26.3	10.81	82.83	98	352.03	0.45	0.319	0.125	0.213
2	1062	7.1	120.08	89.44	78.0	12.37	198.05	127	342.62	0.54	0.336	0.192	0.310
3	1956	7.3	195.22	139.25	105.4	15.43	284.01	164	653.41	0.70	0.185	0.173	0.231

Group 1: Wells: 3, 4, 5, 6, 7, 10, 12 and 13.
 Group 2: Wells: 1, 8, 11, 15, 16 and 17.
 Group 3: Wells: 2, 9, 14 and 18.

pH (standard units), EC (μ.Siemens/cm) and mean concentrations (mg/l).

Table 3. Correlation coefficient among groundwater quality parameters

Variables	EC	pH	Ca	Mg	Na	K	Cl	SO ₄	HCO ₃	Pb	Fe	Zn	Cu
EC	1												
pH	0.34	1											
Ca	0.76	-0.01	1										
Mg	0.76	0.13	0.76	1									
Na	0.70	0.11	0.70	0.78	1								
K	0.38	0.26	0.34	0.33	0.48	1							
Cl	0.76	0.21	0.66	0.83	0.91	0.51	1						
SO ₄	0.16	-0.19	0.19	0.39	0.33	0.36	0.42	1					
HCO ₃	0.57	0.51	0.30	0.29	0.06	0.16	0.12	-0.14	1				
Pb	0.09	0.37	0.09	0.28	0.08	-0.02	0.10	-0.11	0.35	1			
Fe	-0.19	-0.08	-0.19	-0.02	-0.16	-0.31	-0.23	-0.11	-0.00	0.16	1		
Zn	0.23	0.3	0.01	0.09	0.23	0.18	0.14	-0.25	0.42	0.21	-0.03	1	
Cu	0.04	0.02	0.07	0.08	0.15	0.05	0.06	0.16	0.07	0.02	0.01	0.34	1

The Kaiser criterion (Kaiser, 1960) was applied to determine the total number of factors for each dataset in this analysis. Under this criterion, only factors with eigenvalues greater than or equal to 1 will be accepted as possible sources of variance in the data, with the highest priority ascribed to the factor that has the highest eigenvector sum. The rationale for choosing 1 is that a factor must have a variance at least as large as

that of a single standardized original variable to be acceptable.

Four principal components (PC) were extracted and rotated using the varimax normalization (Kaiser, 1960). An initial run using the Kaiser criterion (Kaiser, 1960) resulted in five principal components. However, it was observed that the fifth factor would not constitute a unique source of variance in the hydrochemistry since

it had only one loading greater than 0.50. It was therefore dropped and four factors were chosen for varimax rotation. The results show that the four PC account for more than 72.14% of the total variance (Table 4), which is quite good and can be relied upon to identify the main sources of variation in the hydrochemistry. PC1 represents about 36.60% of the variance and has high absolute loadings for EC, Ca, Mg, Na, K and Cl and probably shows the result of mineral water reactions in the area. PC 2, which accounts for 16.68% of the total variance, contains high loadings for pH, SO₄, HCO₃, Pb and Zn. PC1 and PC2 are assumed to be indicative of the natural processes and water-rock interaction. PC3 and PC4, respectively, show high loadings for Fe and Cu, respectively. PC3 and PC4, which, respectively, account for 9.72% and 9.15% of the variation in the hydrochemistry.

Fig. 4 show the projection of the two PC scores (PC1 vs. PC2) for datasets in a 2D scatter-plot. The distribution of PC scores for datasets suggests a continuous variation of the chemical and physical properties of some of the samples. In Fig. 4, for datasets, groups 1, 2 and 3 samples are well separated in the PC space and completely consistent with the HCA derived groupings. Compact PC distributions for majority of the water samples within groups suggest that all the water samples in their respective groups have similar chemistries hence similar flow paths or sources. If distribution of the samples in the PC space is broad it may indicate changes in the water chemistry due to processes such as a source of contamination, dilution or abrupt changes in vertical-horizontal connectivity of the aquifer.

Table 4. Varimax rotation PCA loading matrix

	PC1	PC2	PC3	F4
EC	0.89	0.13	0.10	-0.14
pH	0.31	0.67	-0.14	-0.28
Ca	0.81	-0.16	0.23	-0.04
Mg	0.88	-0.10	0.32	0.13
Na	0.88	-0.21	0.00	0.13
K	0.58	-0.11	-0.45	-0.20
Cl	0.91	-0.21	0.02	-0.01
SO ₄	0.36	-0.58	-0.11	0.22
HCO ₃	0.41	0.70	0.04	-0.12
Pb	0.21	0.56	0.38	0.13
Fe	-0.24	0.18	0.59	0.50
Zn	0.27	0.61	-0.41	0.30
Cu	0.15	0.11	-0.44	0.77
Eigenvalue	4.76	2.17	1.26	1.19
Variance				
(%)	36.60	16.68	9.72	9.15
Cumulative variance %	36.60	53.27	62.99	72.14

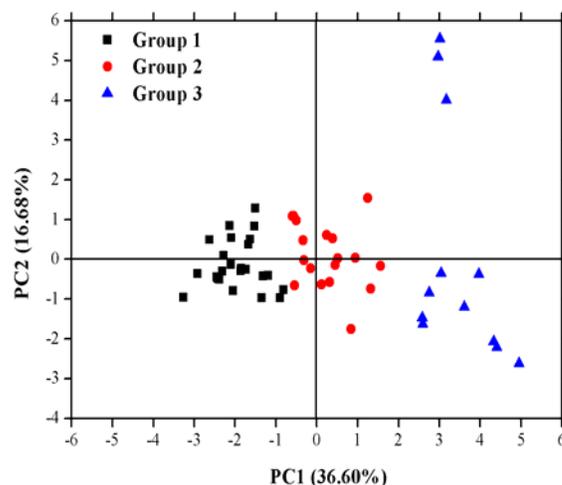


Fig. 4. Plots of PC scores for PC1 versus PC2

CONCLUSION

In this study, different multivariate statistical techniques were used to evaluate variations in groundwater quality of the Ain Azel plain, Algeria. Interpretation of analytical data showed that the abundance of the major ions is as follows: Ca²⁺ Mg²⁺ Na⁺ K⁺ and HCO₃⁻ Cl⁻ > SO₄²⁻. The EC values exhibit high positive correlation with Ca, Mg, Na, Cl and HCO₃. Three major water facies are suggested by the Q-mode HCA analysis. The samples from the area were classified as less saline water (group 1: Ca-Mg-HCO₃), mixed water (group 2: Mg-Ca-HCO₃-Cl) and blended water (group 3: Mg-Ca-Cl-HCO₃). In principal component analysis, the first 4 factors explain 72.14% of the total variance, their loadings allowing the interpretation of hydrochemical processes that take place in the area. The results of this study clearly demonstrate the usefulness of multivariate statistical analysis in hydrochemical.

REFERENCES

Alberto, W. D., Del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A. and De Los Angeles, B. M. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Cordoba-Argentina). *Water Res*, **35**, 2881-2894.

Alther, G. A. (1979). A simplified statistical sequence applied to routine water quality analysis: a case history. *Ground Water*, **17**, 556-561.

Anazawa, K. and Ohmori, H. (2005). The hydrochemistry of surface waters in Andesitic Volcanic area, Norikura volcano, central Japan. *Chemosphere*, **59**, 605-615.

APHA, (1995a). American Public Health Association), AWWA (American Water Works Association), and WPCF (Water Pollution Control Federation). Standard methods for the examination of water and waste water (19th ed.). New York, USA.

- APHA, (1995b). American Public Health Association, Standard methods for the examination of water and wastewater, 19th edn. American public Health Association, Washington, DC.
- APHA, (1989). American Public Health Association, Standard methods for examination of water and wastewater, 17th edn. American Public Health Association, Washington, DC.
- Attoucheik, L. (2006). Geochemical study of discharges from mining complex of Kherzet Youssef (SETIF) and its environmental impact, Thesis, IST. USTHB, Algeria, 200 p. (in Algeria).
- Belkhiri, L. (2005). Hydrogeologic study and problem of the groundwater quality in Ain Azel plain, Sétif, East Algerian. Thesis, University of Batna, Algeria, 170 p. (Algeria).
- Boutaleb, A. (2001). Mineralizations with Pb-Zn of the Sétifien-Hodna field: Gtology, petrography of dolomites, microthermometry and implications metallogenic. Ph.D. Thesis statement in mining geology, IST. USTHB, Algeria, 260 p. (Algeria).
- Brown, C. E. (1998). Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer, New York.
- Clesceri, L. S., Greenberg, A. E. and Eaton, A. D. (1998). Standard methods for the examination of water and wastewater, 20th edn. American Public Health Association, American Water Works Association, Water Environment Federation, Washington.
- Farnham, I. M., Johannesson, K. H., Singh, A. K., Hodge, V. F. and Stetzenbach, K. J. (2003). Factor analytical approaches for evaluating groundwater trace element chemistry data. *Analytica Chimica Acta*, **490** (1-2), 123-138.
- Farnham, I. M., Stetzenbach, K. J., Singh, A. K. and Johannesson, K. H. (2000). Deciphering groundwater flow systems in Oasis Valley, Nevada, Using trace element chemistry, multivariate statistics, and Geographical Information System. *Math. Geol.*, **32**, 943-968.
- Galcon, J. (1967). Research on the geology and the metalliferous lodgings of Sétifien Such, Doct Thesis Sc. Nat. Publ. Serv. Geol. From Algeria, 751 p. (Algeria).
- Guiraud, R. (1973). Post-triassic evolution of before country of the Alpine chain of Algeria, according to the study of the El Eulma basin and close areas, Thesis Sc. Nat. Nice, France, 270 p. (France).
- Güler, C., Thyne, G. D., McCray, J. E. and Turner, A. K. (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*, **10**, 455-474.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20**, 141-151.
- Kim, J. H., Kim, R. H., Lee, J. H., Cheong, T. J., Yum B. W. and Chang, H. W. (2005). Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrological Processes*, **19** (6), 1261-1276.
- Liu, C. W., Lin, K. H. and Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *The Science of the Total Environment*, **313**, 77-89.
- Love, D., Hallbauer, D., Amos, A. and Hranova, R. (2004). Factor analysis as a tool in groundwater quality management: two southern African case studies. *Physics and Chemistry of the Earth*, **29** (15-18), 1135-1143.
- Mahlknecht, J., Steinich, B. and Navarro de Leon, I. (2003). Groundwater chemistry and mass transfers in the Independence aquifer, central Mexico, by using multivariate statistics and mass-balance models.
- Meng, S. X. and Maynard, J. B. (2001). Use of statistical analysis to formulate conceptual models of geochemical behavior: water chemical data from Butucatu aquifer in Sao Paulo State, Brazil. *Journal of Hydrology*, **250**, 78-97.
- Olmez, I., Beal, J. W. and Villaume, J. F. (1994). New approach to understanding multiple-source groundwater contamination: factor analysis and chemical mass balances. *Water Research*, **28** (5), 1095-1101.
- Reghunath, R., Murthy, T. R. S. and Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research*, **36** (10), 2437-2442.
- Savornin, J. (1920). Etude géologique du Hodna et du plateau Sétifien, Thèse Sc. Nat. Lyon. France.
- STATISTICA® 5.0 for Windows. (1998). StatSoft, Inc., Tulsa OK. USDA, Natural Resources Conservation Services, 1999. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. *Agriculture Handbook No. 436*, p. 871.
- Suk, H.J. and Lee, K.K. (1999). Characterization of a ground water hydrochemical system through multivariate analysis: clustering into ground water zones. *Ground Water*, **37**(3), 358-366.
- Vila, J. M. (1980). The Alpine chain of Eastern Algeria and the Algerian-Tunisian border, Thesis of doctorate be-Sc. Nat. Paris VI, France, 665 p. (Algeria).
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*, **69**, 236-244.
- WHO, (1993). World Health Organization, Guidelines for drinking water quality, vol.1, Recommendations (2nd edn.), Geneva, WHO, p. 130.
- Williams, R. E. (1982). Statistical identification of hydraulic connections between the surface of a mountain and internal mineralized sources. *Ground Water*, **20**, 466-478.