APPLICATION OF MULTIVARIATE STATISTICAL METHODS TO SURFACE WATER AND GROUNDWATER OF THE SOUMMAM BASIN, ALGERIA

LAZHAR BELKHIRI (1), LOTFI MOUNI (2)

(1): Hydraulics Department, University of Hadj Lakhdar Batna, Algeria
(2): Technical Sciences Department, Institute of Sciences, University of Akli Mohand Oulhadj, Bouira, Algeria

ABSTRACT

Multivariate statistical methods, i.e., cluster analysis (CA), discriminant analysis (DA) and analysis of variance (ANOVA), were used to assess spatial variation in the water quality of the Soummam basin, Algeria. The application of hierarchical cluster analysis, based on all possible combinations of classification method, showed three main groups of samples. The group 1 samples are exclusively composed of surface water. Groups 2 and 3 samples are consisted of groundwater. Discriminant analysis (DA) was assigned about 98.6% of the cases grouped by CA. All groups are super-saturation with Ca-montmorillonite, dolomite, gibbsite, K-mica, kaolinite and quartz, and all these groups are under-saturation with albite, anhydrite, anorthite, $CO_{2(g)}$, gypsum, halite, melanterite and smithsonite. The ANOVA results indicate that the saturation indices of each of the mineral phases are significantly except chalcedony and quartz $(p > 0.05)$.

Keywords: Cluster analysis; Discriminant analysis; Analysis of variance; Surface water and groundwater; Soummam basin; Algeria.

1. INTRODUCTION

Water has a profound influence on human health. At a very basic level, a minimum amount of water is required for consumption on a daily basis for survival and therefore access to some form of water is essential for life. However, water has much broader influences on health and wellbeing and issues such as the quantity and quality of the water supplied are important in determining the health of individuals and whole communities [1,2]. The water demand is continuously increasing mainly due to population growth and raising needs in agriculture, industrial uses and domestic services. Integrated water management has a strong impact on long-term protection and sustainability [3-6]. The multivariate statistical techniques such as analysis (CA), discriminant analysis (DA) and analysis of variance (ANOVA) have widely been used as unbiased methods in analysis of water quality data for drawing meaningful conclusions. The multivariate analysis is widely used to characterized and evaluate groundwater quality and it is useful for evidencing spatial variation caused by natural and anthropogenic processes [7-11]. The objective of the present study is to identify water quality variables for spatial dissimilarity, and to ascertain the influence of the pollution sources on the water quality parameters.

2. Description of the study area

The Soummam basin is situated in the northeastern of Algeria and extends from 36°-36° 45' of latitude north to 3° 40'-5° 45' of longitude east (Fig. 1). The Soummam basin area is bounded by the Djurdjura mounts on the western, by the Babor mounts on the east and southeastern and by the Mediterranean Sea on the northern side. The climate of the Soummam basin shows a series of transition between humid climate in the mountains near the Mediterranean and semi-arid climate of high plains. The mean minimum and maximum temperatures are 11.8 and 24.8°C, respectively. The mean annual rainfall of the basin is about 700 mm. Rocks and unconsolidated deposits in the area can be divided into four geologic formations: (1) Oligocene; (2) Cretaceous; (3) Lower Miocene; (4) Alluvial terraces.

3. MATERIALS AND METHODS

3.1. Chemical data

Surface water ($n = 4$) and Groundwater samples ($n = 20$) from the Soummam basin were collected in April 2007 (Fig. 1). Water samples were collected in stopper-fitted polyethylene bottles and refrigerated at 4°C in order to be analyzed as soon as possible [12]. Conductivity, temperature and pH were measured in situ using a portable water tester. The water samples were analyzed for major and trace elements. Anions analyzed include sulfate, chloride, bicarbonate and nitrate; cations include calcium, magnesium, sodium and potassium; trace elements include Fe, Zn, SiO_2 and Al. Ca, Mg, HCO_3 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 HCO_3 and Cl by H_2SO_4 and $AgNO_3$ titration, respectively. Concentrations of Na and K were measured using a flame photometer (Model: Systronics Flame Photometer 128) and that of sulfate (SO_4) by turbidimetric method. Nitrate (NO_3) was analyzed by colorimetry with a UV-visible spectrophotometer. Trace elements were determined by Graphite Furnace Atomic Absorption Spectrophotometer (Perkin-Elmer AAnalyst 700) using multi element Perkin-Elmer standard solutions. Standard solutions for the above analysis were prepared from the respective salts of analytical reagents grades.

3.2. Data treatment and multivariate statistical methods

Water quality datasets were subjected to four multivariate techniques: cluster analysis (CA) and discriminant analysis (DA) and analysis of variance (ANOVA). CA, DA and ANOVA were applied to experimental data, standardized through z-scale transformation to avoid misclassifications arising from the different orders of magnitude of both numerical values and variance of the parameters analyzed [13,14]. All mathematical and statistical computations were made using STATISTICA 8.

4. RESULTS AND DISCUSSION

4.1. Cluster analysis

Q-mode hierarchical cluster analysis (HCA) was used to classify the samples into distinct hydrochemical groups. The hydrochemical data having complete analysis of the variables were classified by HCA in 12-dimensional space (Ca, Mg, Na, K, Cl, HCO_3, SO_4, NO_3, Fe, Zn, SiO_2, Al) and the result is presented as a dendrogram (Fig. 2).

Three preliminary groups are selected based on visual examination of the dendogram each representing a hydrochemical facies. The choice of number of clusters is subjective and choosing the optimal number of groups depends on the researcher since there is no test to determine the optimum number of groups in the dataset. However, the large linkage distance between groups 1, 2 and 3 (Fig. 2) suggests that three groups exist for the dataset. Detailed evaluation of the data (comparing Fig. 2 and Table 1) revealed that group 1 samples are exclusively composed of surface water. Groups 2 and 3 samples are consisted of groundwater.

Group 1: The electrical conductivity in this cluster (surface water) ranges from 2,950 to 5,480μS/cm with a mean value of 3,610μS/cm. The EC show strong positive correlation with Ca, Mg, Na, Cl, SO_4, SiO_2 and Al. The mean value of pH is 8.03. This shows that the surface water of the study area is mainly alkaline in nature. The abundance of the major ions in surface water (group 1) is in the following order: Mg > Na > Ca > K and SO_4 > HCO_3 > Cl > NO_3, respectively. The desirable limit for Mg for drinking water is specified as 50 mg/l. It is observed that all the surface water samples from the study area are within the permissible limit as per WHO standard [15]. The desirable limit of SO_4 for drinking water is specified as 200 mg/l. All samples exceed the desirable limit as per WHO [15]. SO_4 was higher in this water indicating breaking of organic substances from topsoil/water, leachable sulfate preset in fertilizer and other human influences. The desirable limit of iron for drinking water is specified as 0.3 mg/l [15]; most of the samples from the study area exceed the desirable limit.

Group 2: The second group represents the groundwater. The EC varies from 540 to 2,380μS/cm with a mean value of 1,395μS/cm. The general dominance of anions was in the order of Cl > SO_4 > HCO_3 > NO_3. The concentration of Cl ranges from 35.5 to 319.5 mg/l with a mean of 142.53 mg/l that was higher than those from other groups. The general dominance of cations was in the order of Mg > Ca > Na > K. Magnesium is the most abundant cation, with concentration ranges from 66.20 to 120 mg/l with a mean of 79.72 mg/l. Calcium ion concentration ranges from 96 to 172 mg/l, with a mean of 124.2 mg/l. Both Ca and Mg concentrations in the groundwater are quite consistent and account for creating major water types in this group. The samples of this water type are found in the south-western part of the study area (Fig. 1).

Group 3: The last group (groundwater) is further classified into two subgroups representing different residence time, recharge, circulation, water-rock interactions and anthropogenic influences.

Group 3-I: The concentrations of HCO_3, Mg and Ca in waters from this group were higher than those from group 2 and lower than those from group 3-II. This group is located at the center of the area (Fig. 1).

Group 3-II: This subgroup of waters is found in the north-east part of the study area, which is the discharge area of the regional groundwater system (Fig. 1). As a result, they have Mg-Ca-HCO_3-Cl type water with high EC values (806 < EC < 3,440 μS/cm; Table 2), indicating pollution.

4.2. Discriminant Analysis

DA is a method of analyzing dependence that is a special case of canonical correlation, and one of its objectives is to determine the significance of different variables, which can allow the

separation of two or more naturally occurring groups. DA operates on original data, and the method constructs a discriminant function for each group as follows Eq(1):

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} \cdot p_{ij} \tag{1}$$

Where i is the number of groups (G), k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, w_j is the weight coefficient, assigned by DA to a given parameter (p_j).

DA was performed on the raw dataset comprised of 15 parameters. The objectives of DA in this study were (1) to test the significance of discriminant functions and (2) to determine the most significant variable associated with the differences between the groups. As shown in Table 2, the values of Wilks' lambda and chi-square for each discriminant function were small and rather high, respectively, which suggested that the DA in this study was valid and effective. Discriminant functions (DFs) and classification matrices (CMs) obtained from the forward stepwise mode of DA are shown in Tables 3 and 4. In forward stepwise mode, variables were included step-by-step, beginning with the most significant, until no significant changes were obtained. Only eight out of 15 parameters analyzed were proven important in discriminating the location groups. The seven parameters which were not shown to be important in differentiating the water groups were T, pH, Ca, K, HCO_3, Zn and Al. The first tow discriminant functions accounted 98.6% of the variation in group separation. The scatter plot of DFs 1 and 2 is shown in Figure 3. Among the three groups, group 1 is clearly separated from groups 2 and 3, whereas little overlap is seen between subgroups 3-I and 3-II.

4.3. Hydrogeochemical processes

4.3.1. Saturation state

In order to investigate thermodynamic controls on the water composition, equilibrium speciation calculation was made using PHREEQC [16]. These calculations provided saturation indices (SI) of minerals that might be reacting in the system. The SI of a particular mineral can be defined as Eq(2)

$$SI = \log (IAP / K_T) \tag{2}$$

where IAP is the ion activity product of the mineral-water reaction and K_T is the thermodynamic equilibrium constant adjusted to the temperature of the given sample. The SI values of the three groups were listed in Table 5. All groups are super-saturation with Ca-montmorillonite, dolomite, gibbsite, K-mica, kaolinite and quartz, and all these groups are under-saturation with albite, anhydrite, anorthite, $CO_{2(g)}$, gypsum, halite, melanterite and smithsonite.

4.3.2. Variation of saturation indices between the three groups

The ANOVA results presented in Table 6 indicate that the saturation indices of each of the mineral phases are significantly except chalcedony and quartz ($p > 0.05$) [17].

The saturation indices of Ca-montmorillonite, illite, gibbsite, K-mica and kaolinite are increases from group 1 to subgroup 3-II, and these mineral phases are super-saturation in all groups. The SI of albite, anhydrite, anorthite, gypsum and chalcedony are increases in the groundwater from the group 2 to subgroup 3-II, and the SI of these minerals are under-

4

saturation. The variation of SI of aragonite, calcite, chlorite and dolomite between the three groups are significant and the values of SI of these minerals are deceases from group 1 to subgroup 3-II. The SI of anorthite, halite, melanterite and siderite are under-saturate in the three groups and their variations are significant. Most silicate and carbonate minerals are super-saturation in the three groups but all evaporite minerals are under-saturation in these groups.

5. CONCLUSION

In this study, selected statistical methods (cluster analysis (CA), discriminant analysis (DA) and analysis of variance (ANOVA)) were used to determine the spatial variations of hydrochemical elements, controlling factor and to identify the origin of these elements in surface and groundwater of Soummam basin, Algeria. For this purpose, 20 sample wells and 4 points of surface waters were collected. Hierarchical cluster analysis grouped 24 sampling waters into three groups (group 1 (surface water), groups 2 and 3 (groundwater)) based on the similarity of water quality characteristics. This grouping was supported by the step-wise discriminant analysis (DA) as 98.6% of the water samples were grouped. The ANOVA results indicate that the saturation indices of each of the mineral phases are significantly except chalcedony and quartz ($p > 0.05$). Therefore, the spatial similarities and differences may allow optimization of a monitoring program in the future by decreasing the number of sampling stations, the number of parameters monitored, and thus, the subsequent costs.

REFERENCES

[1] Fattal B., Gutman-Bass, N., Agursky, T. and Shuval, H.I., Evaluation of health risk associated with drinking water quality in agricultural communities. *Water Sci Technol*, 20, (1988), 409-415.

[2] Medema, G.J., Payment, P., Dufour, A., Robertson, W., Waite, M. and Hunter, P., *Safe drinking water: an ongoing challenge. In: Dufour A et al (eds) Assessing microbial safety of drinking water. Improving approaches and methods*. World Health Organization, Geneva, (2003).

[3] Mendiguchia, C., Moreno, C., Galindo-Riaño, M.D. and Garcia-Vargas, M., Using chemometric tools to assess anthropogenic effects in river water. A case study Q Guadalquivir River (Spain). *Anal Chim Acta*, 515, (2004), 143–149.

[4] Singh, K.P., Malik, A. and Sinha, S., Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Anal Chim Acta*, 538, (2005), 355–374.

[5] Zhou, F., Liu, Y. and Guo, H., Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environ Monit Assess*, 132, (2007), 1–13.

[6] Appelo, C.A. and Postma, D., *Geochemistry, groundwater and pollution*. Balkema, Rotterdam, (1993).

[7] Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. and Kaufman, L., *Chemometrices*: A Textbook. Elsevier, Amsterdam, (1988).

[8] Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and Fernandez, L., Temporal evolution of groundwater composition in an alluvial (Pisuerga river, Spain) by principal component analysis. *Water Research*, 34, (2000), 807–816.

[9] Simeonov, V., Simeonova, P. and Tsitouridou, R., Chemometric quality assessment of surface waters two case studies. *Chemical and Engineering Ecology*, 11 (6), (2004), 449–469.

[10] Singh, K.P., Malik, A., Mohan, D. and Sinha, S., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. *Water Research*, 38, (2004), 3980–3992.

[11] Belkhiri, L., Boudoukha, A., Mouni, L. and Baouz, T., Application of multivariate statistical methods and inverse geochemical modeling for characterization of groundwater - A case study: Ain Azel plain (Algeria). *Geoderma*, 159, (2010), 390-398.

[12] APHA, *Standard methods for the examination of water and wastewater*, 19th edn. DC: American public Health Association, Washington, (1995).

[13] Liu, C.W., Lin, K.H. and Kuo, Y.M., Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *The Science of the Total Environment*, 313, (2003), 77-89.

[14] Lattin, J., Carroll, D., Green, P., *Analyzing multivariate data*. New York: Duxbury, (2003).

[15] WHO, *Guidelines for Drinking-Water Quality*, third ed., vol.1 -Recommendations. Word Health Organization, Geneva, (2006).

[16] Parkhurst, D.L. and Appelo, C.A.J., *User's Guide to PHREEQC (Version 2)-A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations*. United States Geological Survey, Water Resources Investigations Report 99-4259, Washington, DC, (1999).

[17] Alkarkhi, F.M.A., Ismail, N. and Easa, A.M., Assessment of arsenic and heavy metal contents in cockles (Anadara granosa) using multivariate statistical techniques. *Journal of Hazardous Materials*, 150, (2008), 783–789.

Table 1. Mean values of the parameters in the three principal water groups

	Cluster 1	Cluster 2	Cluster 3-I	Cluster 3-II
T	24	19	19	20
pH	8.03	7.76	7.45	7.16
EC	3610	1395	1595	2183
Ca	428.75	124.2	123.5	166.22
Mg	468.75	79.72	251	262.22
Na	517.5	33.83	92.73	157.24
K	2.78	3.39	7.02	6.49
Cl	483.68	142.53	48.99	43.58
SO_4	393	160.2	387.83	575.78
HCO_3	308.18	275.05	319.64	347.29
NO_3	44.67	47.59	25.23	65.24
Fe	2.798	0.015	0.013	0.024
Zn	0.165	0.174	1.144	0.914
SiO_2	13.28	12.04	11.62	13.48
Al	0.02	0.02	0.017	0.038

All values are in mg/l except pH, T (°C) and EC (µSiemens/cm)

Table 2. Wilks' lambda and chi-square test of DA of water quality

Functions	*R*	Wilks' lambda	Chi-square	*p*-level
1	0.98	0.0015	109.989	0.000000
2	0.96	0.0510	50.596	0.000005
3	0.62	0.6134	8.307	0.216441

Table 3. Classification functions coefficients for DA of water quality

Discriminant function	1	2	3
Percent separation	73.2	25.3	1.4
Cumulative separation	73.2	98.6	100.0
Parameter	Factor structure coefficients (correlations between parameters in the model and the discriminant functions)		
Na	0.48	-0.40	-0.09
SO_4	-0.07	-0.50	-0.33
NO_3	-0.01	-0.05	-0.55
Cl	0.26	0.03	-0.05
Mg	0.23	-0.40	0.48
Fe	0.20	-0.06	0.06
EC	0.15	-0.17	-0.20
SiO_2	0.02	-0.09	-0.38

Table 4. Classification matrix for discriminant analysis of the water quality

Prior classification	% correct	Assigned by DA			
		Cluster 1	Cluster 2	Cluster 3-I	Cluster 3-II
Cluster 1	100	4	0	0	0
Cluster 2	100	0	5	0	0
Cluster 3-I	100	0	0	6	0
Cluster 3-II	100	0	0	0	9
Total	100	4	5	6	9

Table 5. Mean values of the saturation indices (SI) and analysis of variance for SI in the three principal water groups

| Phase	Cluster 1	Cluster 2	Cluster 3-I	Cluster 3-II	Sum of Squares	df	Mean Square	F-ratio	p-level
Albite	-0.93	-1.94	-1.59	-0.92	4.42	23	1.47	10.39	0.000248
Anhydrite	-1.15	-1.62	-1.44	-1.19	0.81	23	0.27	12.17	0.000094
Anorthite	-2.56	-2.88	-3.11	-2.28	2.85	23	0.95	3.21	0.045232
Aragonite	1.12	0.45	0.06	-0.07	4.37	23	1.46	14.85	0.000026
Ca-Montmorillonite	0.68	1.66	2.1	3.59	27.89	23	9.3	11.97	0.000104
Calcite	1.27	0.6	0.21	0.08	4.35	23	1.45	14.81	0.000026
Chalcedony	-0.09	-0.09	-0.1	-0.03	0.02	23	0.01	1.47	0.253728
Chlorite	6.7	0.56	0.09	-1.45	187.71	23	62.57	9.47	0.000423
$CO_{2(g)}$	-2.73	-2.41	-2.08	-1.73	3.27	23	1.09	7	0.002100
Dolomite	2.9	1.28	1.03	0.66	14.35	23	4.78	15.77	0.000017
Gibbsite	0.3	0.8	1.05	1.63	5.58	23	1.86	12.92	0.000064
Gypsum	-0.92	-1.38	-1.2	-0.95	0.78	23	0.26	11.85	0.000111
Halite	-5.34	-7.09	-7	-6.86	8.95	23	2.98	46.12	0.000000
Illite	0.45	1.02	1.55	2.8	19.64	23	6.55	12.87	0.000066
K-feldspar	-0.85	-0.74	-0.48	0.04	3.15	23	1.05	4.11	0.019976
K-mica	5.34	6.44	7.2	8.88	41.5	23	13.83	15.77	0.000017
Kaolinite	2.1	3.12	3.61	4.89	24.65	23	8.22	13.17	0.000057
Melanterite	-6.04	-8.41	-7.78	-7.43	13.14	23	4.38	12.37	0.000085
Quartz	0.35	0.36	0.35	0.41	0.02	23	0.01	1.53	0.236753
Siderite	0.91	-1.72	-1.66	-1.68	22.45	23	7.48	17.52	0.000008
Smithsonite	-1.56	-1.48	-0.82	-1.14	1.8	23	0.6	5.98	0.004405

df: degrees of Freedom
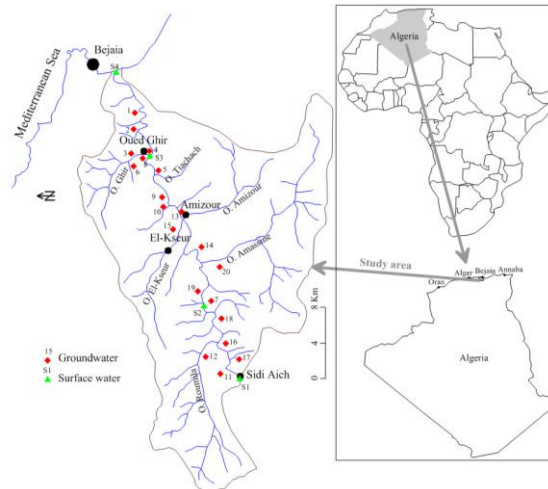
Figure 1. Map showing the water sampling locations of the study area
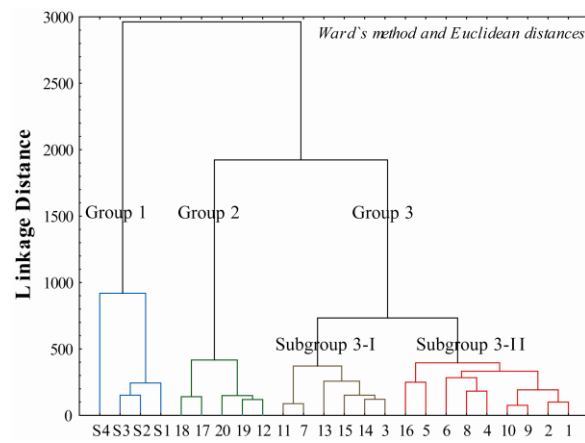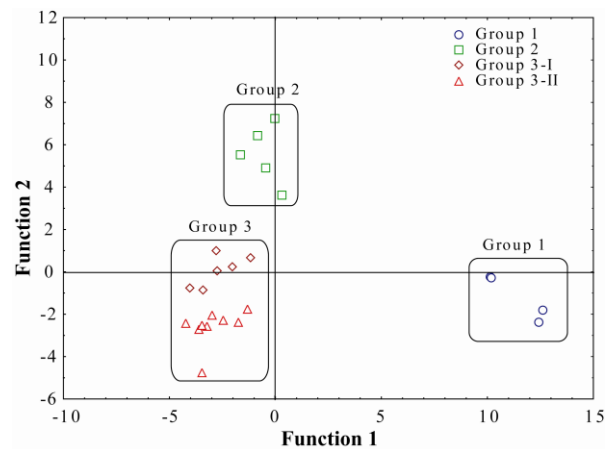

Figure 2. Dendogram of Q-mode cluster analysis

Figure 3. Bivariate plot of discriminant function 1 and 2