Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 3, No. 1, p. 95-102, 2013
http://www.innspub.net

RESEARCH PAPER OPEN ACCESS

Evaluation of groundwater pollution using multivariate statistical analysis: A case study from Burimi area, Kosovo

Sabri Avdullahi*, Islam Fejza, Ahmet Tmava

Faculty of Geosciences and Technology, Str. Parku industrial No NN, 40000 Mitrovica, Republic of Kosovo

Article published on January 21, 2013

Key words: Groundwater quality, Cluster analyses, Principe component, factor analysis, Burimi .

Abstract

Groundwater pollution can be described as degrading of water quality for any usage. Sources of pollution are grouped into two as natural pollution and man-made pollution. The source of ground-water pollution is many and varied because in addition to natural processes practically every type of facility or structure installed by man, and each of his activities may eventually contribute to ground-water quality problems. The aim of this work is to analyze hydro chemical data of groundwater from the 22 sampling points in this aquifer. The descriptive statistical analysis was done besides the principal component, Pearson correlation, and regression analysis. The principal component analysis identified three factors that are responsible for the data structure explaining 82.89% of the total variance of the data set. Factor 1 to 3 explains variance of 54.02%, 16.81% and 12.06% respectively. This study indicates the necessary and usefulness of multivariate statistical techniques to get better information about the water quality and to prevent the pollution caused by households and industries in the future.

***Corresponding Author:** Sabri Avdullahi ✉ sabri.avdullahi@uni-pr.edu

Introduction

Groundwater reserves in Kosovo are not researched enough. These reserves are important for drinking water supply needs, industry, agriculture, etc. According to the Appelo & Postma (1993) the chemical composition of groundwater is a measure of its suitability as a source of water for human consumption and for other purposes, and also influences ecosystem health and function. Groundwater quality is a very sensitive issue, which transcends national boundaries. It is influenced by many factors, including atmospheric chemistry, the underlying geology, the vegetation and anthropogenic agents. Previous studies (Ackah *et al.,* 2011, Sayyed and Wagh, 2011) have revealed that the groundwater quality is one of the most important aspects in water resource studies. It is, thus, important to detect changes and early warnings of change both in natural systems and resulting from pollution.

Generally, the quality of water is controlled by many factors that include composition of recharge water, geological structure and mineralogy of the watersheds and aquifers as well as the residence time and reactions that take place within the aquifer and anthropogenic factors (Drever, 1988, Fetter, 1994, Appelo and Postma, 2005).

The main factors influencing the transport of the pollutants in the ground are: the underground water level, the quantity of pollutants, their type, and soil bedding (Bedient *et al.*, 1999). The ground water quality depends not only on natural factors such as the lithology of the aquifer, the quality of recharge water and the type of interaction between water and aquifer, but also on human activities, which can alter these groundwater systems either by polluting them or by changing the hydrological cycle (Farooq *et al.,* 2010). Groundwater has been associated with water quality problems and the practice of discharging untreated domestic and industrial waste into the water course has emerged to an alarming level.

Water is polluted not only by industries but also by households (Loganathan *et al.,* 2011, Ravichandran and Jayaprakash, 2011, Dhiviyaa *et al.,* 2011). Both industries and household wastewater contain chemicals and biological matter that impose high demands on the oxygen present in water.

According to the Ayoko *et al.* (2007), water quality depends on a variety of physical-chemical parameters and meaningful prediction, ranking analysis or pattern recognition of the quality of water requires multivariate projection methods for simultaneous and systematic interpretation. In recent years, with increasing number of chemical and physical variables of groundwater, a wide range of statistical methods is now applied for proper analysis and interpretation of data (Ashley and Lloyd, 1978, Usunoff & Guzman, 1989, Suk and Lee, 1999 and Sanchez-Martos *et al.,* 2001). Multivariate statistical analysis comprises a number of statistical methods or a set of algorithms that may be applied to several fields of empirical investigation. In the present study statistical software SPSS software version 19 is used to carry out the statistical analysis. Besides, Pearson`s correlation coefficient and Principal Component Analysis (PCA), Regression analysis was also performed.

Cluster analysis is the name given to an assortment of techniques designed to perform classification by assigning observation to group so each is more or less homogeneous and distinct from other groups (Hussain *et al.,* 2008). The most commonly used measure of correlation is Pearson's r-value. According to the Helsel and Hirsch (2002), is also called the linear correlation coefficient because are measuring the linear association between two variables. The data were statistically computed using the correlation coefficient in order to indicate the sufficiency of one variable to predict the other (Davis, 1986). A low distance shows the two objects are similar or 'close together' whereas a large distance indicates dissimilarity.

Principal component analysis (PCA) has been the most frequently employed factor analytic approach. PCA is theoretically the optimum transform for a given data in the least square terms (Usunoff and Guzman, 1989, Tabachnick and Fidell, 2006). This technique aims to transform the observed variables to a set of variables, which are interrelated and arranged in decreasing order of importance. The weights of the original variables in each factor are called loadings; each factor is associated with a particular variable. Commonality is a measure of how well; the variance of the variable is described by a particular set of factors (Grande *et al.,* 2003). The numbers of factors, called principal components (PC), were defined according to the criterion that the only factors that account for variance greater than 1 should be included. Simple linear regression analysis was performed to evaluate the statistically significant variables of the system. The variables shown significance in the correlation analysis are subject to regression analysis and predictive model for the same is prepared.

Material and methods

Study area

The study area is located in the western part of Kosova and belongs to the Drini i Bardhe basin (Sabri *et al.* 2007). The quantities of rainfalls are changeable depending from the seasons of the year. The average rainfall of the area is around 855 mm year^{-1}, from which 317mm flows, while 538mm evaporates (Sabri *et al* 2008). The area falls under continental zone and is characterized by hot summer and cold winter. During summer, the temperature shoots up to 32^{o}C and winter sometime temperature falls to -20^{o}C. The terms of geology characteristic, several investigations have discussed various aspects concerning the geology of Burimi region (Fig. 1). The northern part of the aquifer is comprised from two main formations. The first formation composed of age middle Jurassic, which was built by volcanogenic sedimentary formations (sandy clay, conglomerates, etc.). The second formation composed of age Upper Triassic and is consisted mainly from limestone and dolomite. Geologically, the area is underlain by alluvial deposits of Holocene. Alluvium comprises from the sand, gravel, clay and lignite.

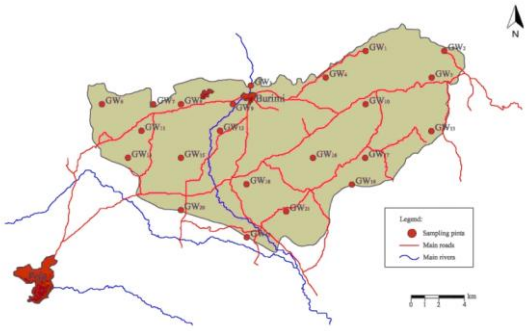

Fig. 1. Showing the study area.

The hydro chemical data from 22 locations in the area was used in the present study. The summary of the results is given in the Table 1. All the samples, collected in tight capped high-quality polyethylene bottles, were immediately transported to the laboratory under low temperature conditions in the icebox and stored in the laboratory. All parameters were determined in the laboratory following the standard protocols (Apha, 1985). The ground water samples were analyzed for parameters, which include pH, electrical conductivity (EC), total dissolved solids (TDS), bicarbonates (HCO_3^-), chloride (Cl^-), sulfate (SO_4^{2-}), calcium (Ca^{2+}), magnesium (Mg^{2+}) and total hardness (TH). Electrical conductivity was measured at 25°C with a conductivity meter. The pH was measured using a pH-meter.

Results and discussion

Cluster analysis

Cluster analysis is the method used for finding different classes and groups within the obtained data. 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). The cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess (Danielsson, 1999). The levels of similarity at which observations are merged are used to construct a dendrogram (Chen, 2007). The Euclidean distance usually gives the similarity between two samples, and the distance can be presented by the difference in analytical values from the samples (Otto, 1998).

Table 1. Descriptive statistics data of water analysis (lions in mg L^{-1}).

Variables	Mean	Median	Std. Dev.	Range	Minimum	Maximum
pH	7.41	7.47	0.40	1.51	6.58	8.09
EC	685.41	606.0	368.22	1251.0	130	1381
HCO_3^-	321.93	294.3	135.32	747.3	24.4	771.7
Cl^-	30.36	19.25	23.74	131.5	2.5	134
SO_4^{2-}	36.65	11.28	23.74	212.0	0.05	212
Ca^{2+}	81.92	67.15	37.81	226.93	14.03	240.96
Mg^{2+}	33.60	20.04	39.29	163.44	4.25	167.69
TH	312.91	265.11	163.01	620.41	80.01	700.42
TDS	360.09	322.50	194.38	673.0	137	810

Table 2. Cluster groups and their members.

Group	Members (location/sample)
A	3, 7, 11
B	8, 9, 15, 10, 14
C	20, 22, 17, 18, 4, 16
D	2, 13, 12, 6
E	1, 5, 21, 19

Table 3. Pearson correlation.

	pH	EC	HCO_3^-	Cl^-	SO_4^{2-}	Ca^{2+}	Mg^{2+}	TH	TDS
pH	1.000								
EC	-.275	1.000							
HCO_3^-	.014	.645	1.000						
Cl^-	-.017	.501	.631	1.000					
SO_4^-	-.410	.574	.161	-.006	1.000				
Ca^{2+}	-.179	.585	.086	.258	.147	1.000			
Mg^{2+}	-.067	.721	.651	.430	.375	.039	1.000		
TH	-.255	.898	.517	.368	.543	.718	.613	1.000	
TDS	-.175	.971	.688	.557	.500	.490	.802	.827	1.000

Table 4. Showing result of principal component analysis.

	Factor 1	Factor 2	Factor 3
pH	-.270	.651	.240
EC	.980	-.088	.034
HCO_3^-	.720	.520	-.140
Cl^-	.594	.502	.210
SO_4^{2-}	.552	-.532	-.464
Ca^{2+}	.537	-.372	.752
Mg^{2+}	.779	.288	-.381
TH	.907	-.226	.189
TDS	.968	.066	-.015
% of variance	54.025	16.814	12.060
Eigen Values	4.862	1.513	1.085
Cumulative	54.025	70.025	82.899

Table 5. Regression equations for various water quality parameters.

Variables	β_0 β_1
HCO_3^-	= 0.300 + 116.634EC
Cl^-	= 0.048 - 2.711EC
SO_4^{2-}	= 0.038 - 26.453EC
Ca^{2+}	= 0.085 + 23.428EC
Mg^{2+}	= 0.771 - 19.138EC
TH	= 0.398 + 40.356
TDS	= 0.513 + 8.738

The results of the cluster analysis are presented in figure 2. The data set were classified in six groups named as A, B, C, D, and E. C contain six samples; B contains five samples; D and E contain four samples, and A contains only three. Clusters of samples are listed in Table 2, which indicate that each cluster has a water quality of its own, which is different from the other clusters. Group A consist of the samples from location no 3, 7, and 11. The values of EC, HCO_3^- and TH are in a narrow range. The water of the group is low in SO_4^{2-}. The group B samples from location no 8, 9, 10, 15 and 14, has EC, TH, and TDS in the close range. The water type of the area is dominated by pH, while the TDS is the lowest. The group C contains of the samples from location no 20, 22, 17, 18, 4 and 16 is dominated by TH and SO_4^{2-} and has low Cl^-concentrations. In the group D sample from location 2, 13, 12 and 6, the values of EC, HCO_3^- and Ca^{2+} is dominant in this group. In the group E samples from location no 1, 5, 21 and 19 the values of EC and Mg^{2+} are high.

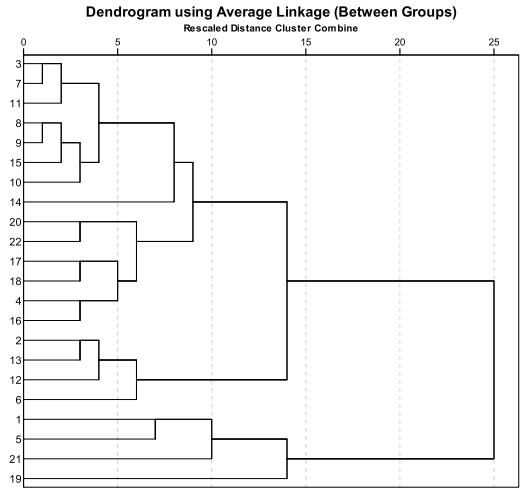

Fig. 2. Dendrogram using average linkage.

Pearson correlation coefficients

The close inspection of the correlation matrix was useful because it can point out associations between variables that can show the overall coherence of the data set and indicate the participation of the individual chemical parameters in several influence factors, a fact which commonly occurred in hydrochemistry (Helena *et al.*, 2000). The Pearson correlation coefficient matrix is given in the Table 3. The variables having coefficient value (r) >0.5 are considered significant. Inspection of the table reveals that EC is positively related with HCO_3^-, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, TH and TDS. The same matrix gives the maximum variance as shown in the principal component analysis-factor 1. This further substantiates the significance of the analysis. HCO_3^- shows the correlation with Cl^-, Mg^{2+}, TH and TDS. Cl^- is related to TDS. SO_4^{2-} in the groundwater of the area shows correlation only with TH. Ca^{2+} is positively related with TH. Mg^{2+} is positively related with TH and TDS. The variation in relationship indicates the complexity of the quality of groundwater.

Principal component analysis

PCA reflects 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. In all the Principal component analysis generated tree significant factors (Table 4). Factor analysis is a multivariate analytical technique, which derives a subset of uncorrelated variables called factors that explain the variance

observed in the original data set (Anazawa and Ohmori, 2005, Brown, 1998).

The analysis generated three factors which together account for 82.89 % of variance. The factors are given in descending order depending on the variance. Factor 1 exhibit 54% of the total variance of 82.89% with high loadings of Mg^{2+}, HCO_3^-, pH, and EC, and moderate to loadings of Ca^{2+}, SO_4^{2-} and Cl^-. This factor reveals that the EC and TDS in the study area are mainly due to Mg^{2+} and HCO_3^-, though Ca^{2+}, SO_4^{2-} and Cl^- also play a substantial role to determining e EC and TDS. This factor accounts for temporary hardness of the water.

Factor 2 exhibits 16 % of the total variance with positive loading on pH, HCO_3^- and Cl^- and negative loading on SO_4^{2-}. This water is bicarbonate and chloride dominated, and it also has low concentrations of sulphate. Main source of HCO_3^- ions in the water in the alluvium of this region is due to dissolution limestone, dolomite and from anthropogenic activities.

Factor 3 exhibits 12 % of the total variance with positive loading on Ca. Ca^{2+} is a significant variable in this factor, which happens to be one of the major ions in the hydrosphere and the most abundant divalent cation in the biosphere. It is an essential element for both plants and animals. The variable Ca^{2+} contribute most strongly to the third factor and probably represents the presence of carbonate mineral in the aquifer. The clustering of the present variables further explains the dissolution of soils and mineral in the sediments containing groundwater.

Simple Linear Regression Analysis

Linear regression is one of the modelling techniques to investigate the relationship between a dependent variable and several independent variables. Linear regression analysis is an important tool for the statistical analysis of the water resource`s data. It is used to describe the covariation between some variable of interest and one or more other variables. Regression analysis is performed to estimate or predict values of one variable based on knowledge of another variable, for which more data are available. Values of r^2 close to 1 are often incorrectly deemed an indicator of a good model. According to the Helsel and Hirsh (2002) the r^2 near 1 can result from a poor regression model; lower r^2 models may sometime be preferable.

A positive correlation between EC and HCO_3^-, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, TH and TDS, is used to carry out the regression analysis (Table 5).

The model for simple linear regression is

$Y = \beta_0 + \beta_1 X$

Where:

Y is the dependent variable

X is the independent variable

β_0 is the intercept which is the coefficient of regression

β_1 is the slope

The results of the analysis show that by measuring the EC value the other variable that is HCO_3^-, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, TH and TDS can be calculated. TDS and EC show a perfect relation that is clear from high r^2 value.

Conclusion

Multivariate statistical techniques, including cluster and principal component analysis can successfully be used to derive information from the data set about the possible influences of the environment on groundwater quality and also identify natural groupings in the set of data. The Principal component analysis of the hydro chemical data reduces the original data matrix into three components that explains 82.89 % of the total variance. The regression analysis confirms the positive relation of EC and HCO_3^-, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+} and Total Hardness and Total Dissolve Solid in the study area. The water samples are mainly of caladium-bicarbonate type, pointing to the hardness of groundwater. Contamination of the ground water in the aquifer in the area of research is due to anthropogenic factors.

References

Ackah M, Agyemang O, Anim AK, *et al*. 2011. Assessment of groundwater quality for drinking and irrigation: the case study of Teiman-Oyarifa Community, Ga East Municipality, Ghana. Proceedings of the International Academy of Ecology and Environmental Sciences **1(3-4),** 186-194

Alberto WD, Del PDM, Valeria AM, Fabiana PS, Cecilia HA, De Los ABM. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquıa River Basin (Cordoba-Argentina). Water Res **35,** 2881-2894.

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

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

APHA, AWWA, WEF. 1985. Standard methods for the examination of water and wastewater (18th Ed.). American Water Work s Association, Washington DC.

Appelo CAJ, Postma D. 2005. Ion exchange. In: Geochemistry, Groundwater and Pollution. 2nd edn, Balkema, 241-309.

Ashley RP, Lloyd JW. 1978. An example of the uses of Facto analysis and cluster analysis in groundwater chemistry interpretation. Journal of Hydrology **39,** 355-364.

Avdullahi S, Fejza I, Syla A. 2008. Water resources in Kosova. Journal of International Environmental Application & Science **3 (6)**, 51-56,

Avdullahi S, Fejza I, Tmava A, Syla A. 2007 Water resources in Drin i Bardh River Basin, Kosova. International Journal of Natural and Engineering Sciences **2 (3),** 105-109.

Ayoko GA, Singh K, Balerea S, Kokot S. 2007. Exploratory multivariate modelling and prediction of the physico-chemical properties of surface water and groundwater. Journal of Hydrology **336**, 115-124

Bedient PB, Rifai HS, Newell CJ. 1999. Ground Water Contamination Transport and Remediation, 2nd edition. Upper Saddle River, New Jersey: Prentice Hall.

Brown CE. 1998. Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer, New York.

Chen K, Jiao JJ, Huang J. Huang R. 2007, Multivariate statistical evaluation of trace elements in groundwater in a coastal area in Shenzhen, China, Environmental Pollution **147 (3),** 771-780.

Danielsson A, Cato I, Carman R, Rahm L. 1999.Spatial clustering of metals in the sediments of the Skagerrak/Kattegat. Appl. Geochem. **14**, 689-706.

Davis JC. 1986. Statistics and data analysis in geology (2nd. Ed), John Willey and Sons, New York.

Dhiviyaa TS, Venkatesa T, Punithavathi L, Karunanithi S, Bhaskaran A. 2011 Groundwater pollution in the Palar Riverbed near Vellore, Tamil Nadu India. Indian J.Sci.Technol **4(1),** 19-21.

Drever JI. 1988. The Geochemistry of Natural Waters. 2nd ed. Prentice Hall, Englewood Cliffs, NY, 437pp.

Farnham IM, Stetzenbach KJ, Singh AK, Johannesson KH. 2000. Deciphering groundwater flow systems in Oasis Valley, Nevada, Using trace element chemistry, multivariate statistics

and Geographical Information System. Math. Geol. **32,** 943-968.

Farooq MA, Malik MA, Hussain A, Abbasi HN. 2010. Multivariate Statistical Approach for the Assessment of Salinity in the Periphery of Karachi, Pakistan. World Applied Sciences Journal **11 (4),** 379-387.

Fetter CW. 1994. Applied Hydrogeology. Macmillan College Publishing Company, New York.

Grande JA, Borrego J, Torre ML, Sainz A. 2003. Application of cluster analysis to the geochemistry zonation of the estuary waters in the tinto and odiel rivers (Huelva, Spain). Environmental Geochemistry and Health **25**, 233-246.

Helena B, Pardo R, Vega M, Barrado E, Fernandez J, Fernandez L. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis, Water Res **34(3),** 807–816.

Helsel DR, Hirsch RM. 2002. Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chapt. A3, 510 p

Hussain AI, Anwar F, Sherazi STH, Przybylski R. 2008. Chemical composition. Antioxidant and antimicrobial activities of basil (Ocimum basilicum) essential oils depends on seasonal variations. Food Chemistry **108**, 986-995

Ravichandran K, Jayaprakash M. 2011. Seasonal variation on physico-chemical parameters and trace metals in groundwater of an industrial area of north Chennai, India. Indian J.Sci.Technol **4 (6)**, 646-649.

Loganathan BG, Sajwan KS, Sinclair E, Senthil K, Kannan K. 2007. Perfluoroalkyl sulfonates and perfluorocarboxylates in two wastewater treatment facilities in Kentucky and Georgia. Water Research **41 (20),** 4611-4620.

Meng SX, Maynard JB. 2001. Use of statistical analysis to formulate conceptual models of geochemical behavior: water chemical data from Butucatu aquifer in Sao Paulo State, Brazil. J. Hydrol. **250**, 78- 97

Otto M, 1998. Multivariate methods. In: Kellner, R., Mermet, J. M., Otto, M., Widmer, H.M.(Eds), Analytical Chemistry. Wiley-VCH, Weinheim, Germany

Sánchez-Martos F, Jimenezespinosa R, Pulido-Bosch A. 2001. Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, southeastern Spain. Hydrological Sciences **46 (2),** 227-242.

Sayyed MRG, Wagh GS. 2011. An assessment of groundwater quality for agricultural use: a case study from solid waste disposal site SE of Pune, India. Proceedings of the International Academy of Ecology and Environmental Sciences, **1(3-4)**, 195-201.

Suk H, Lee KK. 1999 Characterization of a groundwater hydrochemical system through multivariate analysis clustering into groundwater zones. Ground Water **37,** 358-366.

Tabachnick BG, Fidell L. 2006 Using Multivariate Statistics (5th Ed.). Allyn & Bacon, NY.

Usunnoff EJ, Guzman-Gusman A. 1989. Multivariate analysis in hydrochemistry. An example of the use of factor analysis and correspondence analysis. Groundwater **27(1),** 27-33.

Williams RE. 1982. Statistical identification of hydraulic connections between the surface of a mountain and internal mineralized sources. Groundwater **20**, 466-478.