Asian Journal of Chemistry Vol. 22, No. 5 (2010), 3959-3965

Determination of Water Quality Parameters of Yesilirmak River, Turkey

SUHEYLA YEREL*, HUSEYIN ANKARA† and NURGUL OZBAY‡
Bozuyuk Vocational School, Bilecik University, Bozuyuk, Bilecik, Turkey
E-mail: syerel@anadolu.edu.tr; syrel@gmail.com

In this study, the surface water quality of the Yesilirmak river in Turkey is assessed by using multivariate statistical techniques. These techniques were applied to the water quality parameters obtained from the six different surface water quality monitoring stations. Factor analysis represents that agricultural, domestic waste and organic pollution caused differences in terms of water quality. Hierarchical cluster analysis revealed two different clusters of similarities between the monitoring stations, reflecting different chemical properties and pollution levels in the studied river. These results present that agricultural, domestic waste and organic pollution caused differences in terms of river quality in the north and northwest part of the Yesilirmak river, Turkey. Thus, this paper represents the usefulness of multivariate statistical techniques, including factor analysis and hierarchical cluster analysis, for helping determination makers in hydrochemical studies.

Key Words: Factor analysis, Hierarchical cluster analysis, Water quality, Pollution, Yesilirmak river.

INTRODUCTION

The surface water quality is truly a sensitive issue today because of its effects on human health and aquatic ecosystems. Rivers are highly vulnerable to pollution attributing to their role in carrying off the municipal and industrial wastewater and run-off from agriculture in their vast drainage basins. Anthropogenic influences, as well as natural processes, deteriorate surface water and impair their use for drinking, industrial, agricultural, recreation or other purposes[1-3].

A particular problem in the case of water quality monitoring is the complexity associated with analyzing the large number of measured variables. The data sets contain rich information about the behaviour of the water resources. Classification, modeling and interpretation of monitored data are the most important steps in the assessment of water quality[4,5].

Multivariate statistical techniques including factor analysis and hierarchical cluster analysis are known as a suitable tool for obtaining consequentially reduced

†Department of Mining Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey.
‡Engineering Faculty, Bilecik University, Bilecik, Turkey.

data and interpreting various parameters[6] and, since multivariate statistical techniques can reflect spatial variations, used for analyzing surface water quality data. These techniques allow the identification of the possible sources that influence water systems and offer a valuable tool for reliable management of water resources as well as rapid solution for pollution problems[7-9].

In this study multivariate statistical techniques including factor analysis and hierarchical cluster analysis were used to determine the sources of the surface water quality inputs and to group monitoring stations using water quality data set collected from the Yesilirmak river in Turkey.

EXPERIMENTAL

Description of study area: The Yesilirmak river basin is one of the 26 major basins in Turkey. The water of the river is mostly used for irrigation in addition to drinking, swimming, fishing, wildlife habitat, *etc*. The river originates north of Sivas, flows approximately 519 km and reaches the Black Sea. It drains an area of 38,730 km^2, which is about 5 % of Turkey's surface area[10,11]. Fig. 1 shows the location of the monitoring stations (S1, S2, S3, S4, S5 and S6) in Yesilirmak river, Turkey.

Fig. 1. Location of monitoring stations

Data set: Surface water quality data set have been collected monthly by analysis from the six monitoring stations in the Yesilirmak river, Turkey. The selected surface water quality parameter for the determination of water quality characteristics are chloride (Cl^-), sodium (Na^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), sulphate (SO_4^{2-}), electrical conductivity (EC), total dissolved solids (TDS), dissolved oxygen (DO), biochemical oxygen demand (BOD_5) and chemical oxygen demand (COD). The parameters were evaluated by using multivariate statistical analysis techniques. All multivariate statistical computations were made using SPSS and Minitab statistical software.

Multivariate statistical analysis: Modern chemometrics is a branch of chemistry (often related to analytical chemistry) that deals with the application of mathematical and statistical methods in order to evaluate, classify, model and interpret chemical and analytical processes and experiments and to extract a maximum of chemical and analytical information from experimental data. When the methods of chemometrics are applied to data sets obtained by monitoring various environmental compartments (surface water, atmosphere, soil *etc*.), the term "environmetrics" is used to stress the information ability to the methods to gain specific information from samples of the total environment[12-14].

The most important methods of multivariate statistics employed in environmetrics are factor analysis and hierarchical cluster analysis. In this study, surface water quality data set were subjected to multivariate statistical analysis techniques including factor analysis and hierarchical cluster analysis.

Factor analysis: Factor analysis is a multivariate statistical technique that attempts to extract a lower dimensional linear structure from the data. The main purpose of factor analysis is to reduce the contribution of less significant variables and to simplify even more of the data structure. As a result, a small number of factors will usually account for approximately the same amount of information as the much larger set of original observations[15]. The factor analysis can be expressed as:

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \ldots + a_{fm}f_{mi} + e_{fi} \qquad i = 1, 2, \ldots, p \qquad (1)$$

where z is the measured variable; a is the factor loading; f represents the factor score, e is the residual term accounting for errors or other source of variation; i is the sample number and m represents the total number of factors.

Hierarchical cluster analysis: Cluster analysis is an unsupervised pattern recognition method that divides a large group of cases into smaller groups of clusters of relatively similar cases that are dissimilar to other groups. Hierarchical cluster analysis, the most common approach, starts with each case in a separate cluster and joins the clusters together step by step until only one cluster remains[16,17]. The Euclidean distance usually given the similarity between two samples and a distance can be represented by the difference between transformed values of the samples[18,19]. In this study, hierarchical cluster analysis was performed using Ward's method with squared Euclidean distance as a measure of similarity. Ward's method uses to calculate the distances between clusters to minimize the sum of square of any two possible clusters at each step. Spatial variations in surface water quality were determined from hierarchical cluster analysis using the linkage distance.

RESULTS AND DISCUSSION

In this study, surface water quality data set belonging to Yesilirmak river (Turkey) were classified by using factor analysis. The correlations matrix of surface water quality parameters was generated and factor extracted by the centroid method, rotated by varimax. Factor analysis scree plot was presented in Fig. 2. When this plot is

examined, it is observed that the number of factor is three. The factors were explained 69.612 % of the variance in data set. Results of factor analysis including factor loading, total and cumulative variance values are given in Table-1.

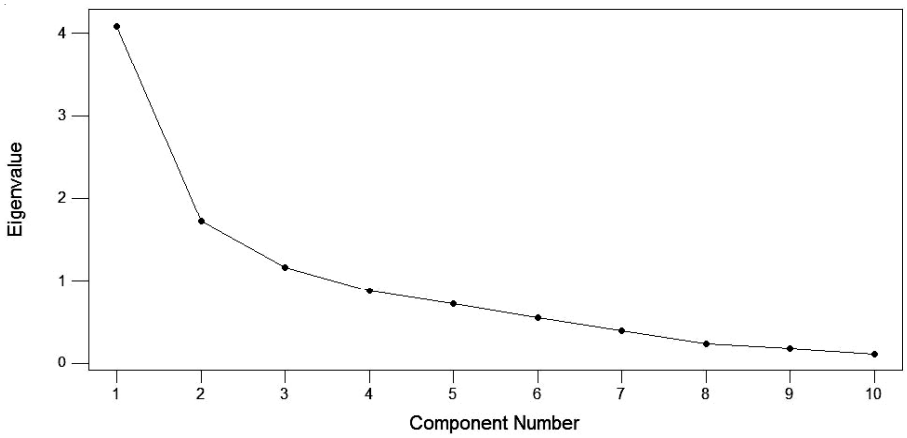

Fig. 2. Scree plot of the eigenvalues

TABLE-1
FACTOR LOADING, TOTAL AND CUMULATIVE VARIANCES

Variables	Factor 1	Factor 2	Factor 3
Cl^-	0.844	–	–
EC	0.831	–	–
Mg^{2+}	0.763	–	–
Ca^{2+}	0.638	–	–
BOD_5	0.637	–	–
TDS	0.618	–	–
DO	–	0.817	–
COD	–	0.813	–
Na^+	–	–	0.931
SO_4^{2-}	–	–	0.883
Total variance (%)	32.789	19.178	17.646
Cumulative variance (%)	32.789	51.967	69.612

When the results of factor analysis were examined, parameters were grouped based on the factor loading and the following factors were indicated: **Factor 1**: Chloride (Cl^-), electrical conductivity (EC), magnesium (Mg^{2+}), calcium (Ca^{2+}), biochemical oxygen demand (BOD_5), total dissolved solids (TDS). **Factor 2**: Dissolved oxygen (DO), chemical oxygen demand (COD). **Factor 3**: Sodium (Na^+), sulphate (SO_4^{2-}).

In view of this analysis, it was determined that Factor 1, namely organic pollution factor, was composed Cl^-, EC, Mg^{2+}, Ca^{2+}, BOD_5 and TDS, which were 0.844, 0.831, 0.763, 0.638, 0.637 and 0.618, respectively. This factor was explained 32.789 %

of the variance. BOD_5 is local anthropogenic pollution and also addition of local domestic waste of the monitoring stations[20]. The discharges of the surface water from many factors and especially from municipal, fertilizers and factories waste contribute to the pollution of the river[21]. Thus, the factor was mainly found agricultural pollution and domestic waste. Factor 2 was formed, DO and COD, which were 0.817 and 0.813, respectively. High positive loadings indicated strong linear correlation between the factor and parameters[22]. The factor can donate by oxygen content. Factor 3 was explaining Na^+ and SO_4^{2-}, which were 0.931 and 0.883, respectively. This factor was represented organic pollution.

Hierarchical cluster analysis was used to detect the similarity groups between the monitoring stations[15]. This analysis was applied on surface water quality data set, to detect spatial similarity for clustering of monitoring stations. The dendrogram resulting from the hierarchical cluster analysis of measured data set is presented in the Fig. 3.

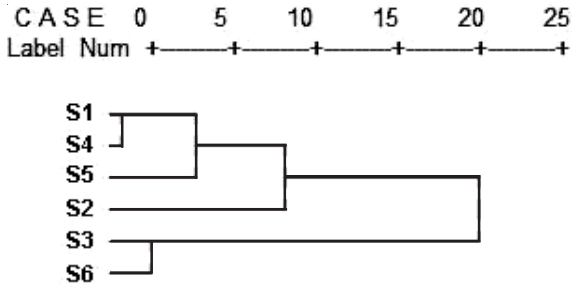

Fig. 3. Dendrogram of the Ward method

The clustering procedure generated two groups of stations in a convincing way, as the sites in these groups have similar characteristic features and natural background source types. Cluster I (S1, S2, S4, S5) and cluster 2 (S3, S6) correspond to a relatively high pollution and low pollution regions, respectively. The mean values of surface water quality parameters in the Yesilirmak river are presented in Fig. 4. A close examination of surface water quality of mean values graph reveals that the highest concentrations were observed in the S1, S2, S4 and S5. These analyses results present that agricultural, domestic waste and organic pollution caused differences in terms of river quality in the north and northwest part of the river.

As has been discussed above for rapid assessment of water quality, only one site in each cluster may be serving as good in spatial assessment of the surface water quality as the whole network. It is evident that hierarchical cluster analysis technique is useful in the offering reliable classification of surface waters in the whole region and will make possible assessment in an optimal manner. Thus, the number of observation stations and cost in the monitoring network will be reduced without loosing any significance of the outcome. The same aspects are also reported by other researchers[9].

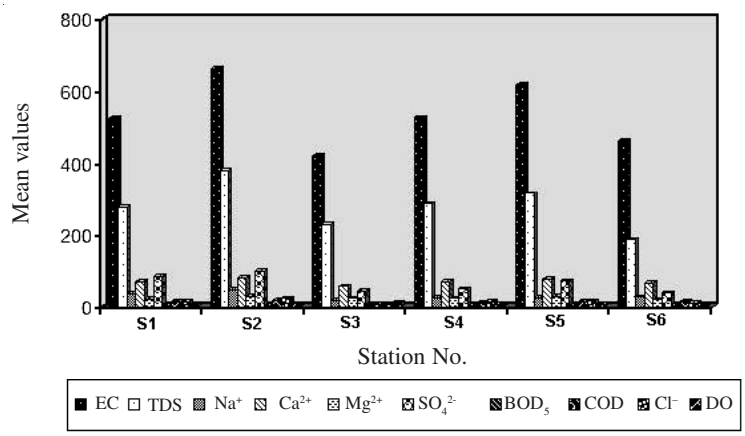

Fig. 4. Water quality parameters mean values of monitoring stations

Conclusion

In this study, different multivariate statistical techniques including factor analysis and hierarchical cluster analysis were applied to data set obtain from Yesilirmak river. The results were useful for surface water quality management.

Factor analyses helped in identify the factors responsible for surface water quality variations in three different factors. Based on the above results, it was resulted that of the factor analyses explained by the three factors.

Hierarchical cluster analysis grouped six monitoring stations into two clusters of similar surface water quality characteristics. Based on obtained information, it was possible to design a future, optimal sampling strategy, which could reduce the number of monitoring stations. These results represent that agricultural, domestic waste and organic pollution caused differences in terms of water quality in the north-northwest part of the area.

Thus, this paper show the usefulness of multivariate statistical techniques such as factor analysis and hierarchical cluster analysis in surface water quality assessment, determination of pollution sources with a view to get better information about the surface water quality.

ACKNOWLEDGEMENT

The authors would like to thank the General Directorate of State Hydraulic Works in Turkey for their help in providing necessary data.

REFERENCES

1. S.R. Carpenter, N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharply and V.H. Smith, *Ecol. Appl.*, **8**, 559 (1998).
2. H.P. Jarvie, B.A. Whitton and C. Neal, *Sci. Total Environ.*, **210-211**, 79 (1998).
3. Y. Zhang, F. Guo, W. Meng and W. Xi-Qin, *Environ. Monit. Assess.*, **152**, 105 (2009).
4. H. Boyacioglu, *Water SA*, **32**, 389 (2007).

5. A. Altin, Z. Filiz and C.F. Iscen, *Environ. Monit. Assess.*, **158**, 51 (2009).
6. S. Han, E. Kim and S. Kim, *KSCE J. Civil Eng*., **13**, 97 (2009).
7. R. Reghunath, T.R.S. Murthy and B.R. Raghavan, *Water Res*., **36**, 2437 (2002).
8. V. Simeonov, P. Simeonov and R. Tsitouridou, *Chem. Eng. Ecol*., **11**, 449 (2004).
9. T.G. Kazi, M.B. Arain, M.K. Jamali, N. Jalbani, H.I. Afridi, R.A. Sarfraz, J.A. Big and A.Q. Shah, *Ecotoxicol. Environ. Safety*, **72**, 301 (2009).
10. N. Munsuz and I. Unver, Waters of Turkey, Publication number: 882, University of Ankara, Agricultural Faculty Publications, Ankara (1983) (in Turkish).
11. A. Kurunc, K. Yurekli and O. Cevik, *Environ. Model Software*, **20**, 1195 (2005).
12. V. Simeonov, E. Adam, S. Tsakovski, P. Mandjukov and J. Stratis, *Toxicol. Environ. Chem*., **50**, 57 (1995).
13. T. Spanos, V. Simeonov, J. Stratis and X. Xatzixristou, *Mikrochim. Acta*, **141**, 35 (2003).
14. P. Simeonova, V. Lovchinov, D. Dimitrov, I. Radulov, *Environ. Monit. Assess.*, doi:10.1007/s10661-009-0888-7 (2009).
15. S. Shrestha and F. Kazama, *Environ. Modelling Software*, **22**, 464 (2007).
16. J. Lattin, D. Carroll and P. Green, Analyzing Multivariate Data, New York: Duxbury (2003).
17. J. McKenna, *Environ. Modelling Software*, **18**, 205 (2003).
18. M. Otto, Multivariate Methods, Weinheim: Wiley-VCH (1998).
19. F. Zhou, Y. Liu and H. Guo, *Environ. Monit. Assess.*, **132**, 1 (2007).
20. A.M. Silva, E.L.B. Novelli, M.L. Fascinelli and J.A. Almeida, *Environ. Pollut*., **105**, 243 (1999).
21. S. Yerel, *Asian J. Chem*., **21**, 4054 (2009).
22. H. Boyacioglu and H. Boyacioglu, *Environ. Geol*., **54**, 271 (2008).

(*Received*: 31 August 2009; *Accepted*: 23 January 2010) AJC-8359