Asian Journal of Chemistry Vol. 22, No. 4 (2010), 2928-2934

Determination of Surface Water Quality Parameters of Sakarya River, Turkey

SUHEYLA YEREL
Bozuyuk Vocational School, Bilecik University, Bozuyuk, Bilecik, Turkey
Fax: (90)(228)3141195; Tel: (90)(228)3141195; E-mail: syerel@anadolu.edu.tr

This study aims to investigate the changes of the surface water quality parameters on the Sakarya river in Turkey. For this purpose water quality samples were collected from 5 monitoring stations in the river. Variations in the surface water quality parameters of the monitoring stations have been examined by environmetric methods including ordinal logistic regression technique and cluster analysis. Ordinal logistic regression results showed that TDS, SO_4 and NO_3-N were identified as the water quality parameters that displayed essential variations. The cluster analysis of the monitoring stations determined two clusters. It was concluded that agricultural pollution affected the southeast part of the river. These methods are believed to help to water quality issues.

Key Words: Water quality, Environmetric methods, Ordinal logistic regression technique, Cluster analysis, Sakarya river, Turkey.

INTRODUCTION

One of the major concerns in hydrological studies understands the factors and processes that control and affect water quality. Water quality "reflects the composition of water as affected by natural process and by humans' cultural activities, expressed in terms of measurable quantities and related to intended water use"[1]. Surface waters contain many chemical species in the dissolved state that play an important role in the survival of aquatic ecosystems[2].

Turkey, it is believed, will be a water scarce country in the future because of over population, drought and environmental pollution. Therefore, water quality observations have started to gain great importance nowadays. For this purpose, water quality data are collected regularly from the surface water quality monitoring stations installed river to determine the changes in the pollution sources and therefore pollution levels of the rivers[3].

The purpose of this study is to identify the surface water quality for the Sakarya river (Turkey) by environmetric methods. One of the environmetric methods used for the assessment of water quality is the ordinal logistic regression technique. The technique related to probability of a contamination to exceed a threshold concentration to a set of possible influence variables. Ordinal logistic regression technique has been used extensively in epidemiological studies and recently, it is becoming a common technique in environmental research[4]. Some studies conducted in hydrological sciences by using ordinal logistic regression technique are as follows: Wear *et al*.[5] modeled the effect of urbanization on water quality in the Southern Appalachian Mountains using the ordinal logistic regression technique. Twarakavi and Kaluarachchi[6] showed the aquifer vulnerability to heavy metal contamination is estimated using the method of ordinal logistic regression and univariate analysis. Gunter *et al*.[7] studied the probability analysis for predicting local water quality regulations in the southern United States. The analysis uses an ordinal logistic regression model to estimate the influence of spatial and demographic factors on the probability that most of countries adopt local regulations to improve water quality.

Cluster analysis is another environmetric method. The analysis has been used hydrological processing for many years. The intension underlying the use of cluster analysis is to achieve great efficiency of datasets clusters from the original data. The results of cluster analysis help in interpreting the data and indicate patterns[8].

The aim of this study is to identify the most important variables causing difference in the surface water quality for the Sakarya river by using environmetric methods. The results of these methods may present the pollution sources in the river.

EXPERIMENTAL

The studied area of Sakarya river is passed through Sakarya (Turkey) province which is located in north-west part of Turkey. There are many industrial plants in Sakarya province. Although some plants have their own waste disposal units, some of them discharge their waste into the small streams which feeds and pollute the Sakarya river[9].

The topography of the area has the characteristics of a high plain. The plain has been fragmented in the direction of Sakarya river by the large and deep valleys of the stream network that emerges in the basin. In the areas where magmatic and metamorphic rocks are dominant, the topography is sharp due to faulting. A flatter topography dominates the areas where neogene series crop out. The maximum and minimum elevations in the study area are 1,800 and 800 m, respectively and the average elevation is about 1,100 m. A continental climate with cold winters and warm summers is dominant[10].

Dataset: Surface water quality dataset were collected from 5 surface water quality stations for the length of 5 years. These dataset obtained from DSI. The selected surface water quality parameters, their units are summarized in Table-1. Water quality data descriptions were evaluated by using the environmetric methods. All statistical computation was made using the MINITAB statistical software.

Environmetric methods: In this study, surface water quality dataset were performed environmetric methods including ordinal logistic regression technique and cluster analysis.

Ordinal logistic regression technique: Ordinal logistic regression technique can be used to determined relation of dependent and independent variables. It is useful when the dependent variable is observation stations. The ordinal logistic

TABLE-1
SURFACE WATER QUALITY PARAMETERS ASSOCIATED
WITH THEIR ABBREVIATIONS AND UNITS

No.	Parameter	Abbreviations	Units
1	Temperature	T	ºC
2	pH	pH	pH unit
3	Chloride	Cl	mg/L
4	Sulphate	SO_4	mg/L
5	Sodium	Na	mg/L
6	Nitrate nitrogen	NO_3-N	mg/L
7	Total dissolved solids	TDS	mg/L
8	Dissolved oxygen	DO	mg/L
9	Chemical oxygen demand	COD	mg/L
10	Biochemical oxygen demand	BOD_5	mg/L

regression is the production of data matrix. While each row data represents an individual case expressed using a water quality unit, column data show the independent (surface water quality parameters) and dependent variables (observation stations) in the data matrix.

Ordinal logistic regression involves fitting the dependent variable by using eqn. 1.

$$Y = \ln\left[\frac{p}{(1-p)}\right] = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \quad (1)$$

where, p/(1-p) is the so called odds ratio, β_0 is the intercept and β_1, β_2 … β_n are the coefficients that measure the contribution of independent factors to the variations in Y[11].

The ordinal logistic regression model is fitted to the observed responses using the maximum likelihood approach. Usually, the method of maximum likelihood produces unknown parameter values that best match the predicted and observed probability values[6].

Cluster analysis: Cluster analysis is a statistical technique, whose primary purpose is to classify the objects of the system into categories or clusters based on their similarities and the objective is to find an optimal grouping for which the observations or objects within each cluster are similar, but the clusters are dissimilar to each other. Hierarchical clustering is the most common approach in which clusters are formed sequentially. The most similar objects are first grouped and these initial groups are merged according to their similarities. Eventually as the similarity decreases all subgroups are merged into a single cluster. Cluster analysis was applied to surface water quality data using a single linkage method. In the single linkage method, the distances or similarities between two clusters A and B are defined as the minimum distance between a point in A and a point in B:

$$D(A, B) = \min\{d(x_i + x_j), \text{ for } x_i \text{ in A and } x_j \text{ in B}\} \quad (2)$$

where $d(x_i + x_j)$ is the Euclidean distance in eqn. 2. At each step the distance is found for every pair of clusters and the two clusters with smallest distance are merged. After two clusters are merged the procedure is repeated for the next step: the distances between all pairs of clusters are calculated again and the pair with minimum distance is merged into a single cluster. The result of a hierarchical clustering procedure can be displayed graphically using a dendrogram which shows all the steps in the hierarchical procedure[12-14].

RESULTS AND DISCUSSION

In this study, surface water quality parameters were examined by environmetric methods including ordinal logistic regression technique and cluster analysis. Ordinal logistic regression technique was applied to the Sakarya river to relate the independent variables (water quality parameters) and dependent variable (monitoring stations). Ordinal logistic regression technique was used to determine whether the surface water quality parameters vary between monitoring stations or not.

The odds ratio results achieved by within the adjusted 95 % confidence interval using ordinal logistic regression technique were assessed and presented in Table-2. Table-2 shows that the regression coefficient of TDS is 1.50, which is the highest among all parameters with an odds ratio of 4.5. The coefficient of SO_4 is -0.29, with an odds ratio of 0.75 and the regression coefficient of NO_3-N is -0.45, with an odds ratio of 0.63. The ordinal logistic regression coefficients of the other parameters are close to zero, with odds ratio approaching 1.00. By considering the odds ratio values, TDS, SO_4 and NO_3-N have the most different values for 1.00. But, the odds ratio values for temperature, pH, Cl, Na, DO, COD and BOD_5, have close values for 1.00. The odds ratio values different for 1.00 indicated that significant variations exist among the monitoring stations related to water quality parameters. According to these results, the analyses generated for two different groups indicated; group I: TDS, SO_4 and NO_3-N and group II: temperature, pH, Cl, Na, DO, COD and BOD_5.

TABLE-2
THE ORDINAL LOGISTIC REGRESSION TECHNIQUE
RESULTS OF SAKARYA RIVER IN TURKEY

No	Parameters	Coefficient	SE Coefficient	Odds ratio
1	Temperature	-0.04	0.06	0.96
2	pH	-0.00	0.01	1.00
3	Cl	-0.01	0.04	0.99
4	SO_4	-0.29	0.21	0.75
5	Na	-0.12	0.03	0.93
6	NO_3-N	0.45	0.47	0.63
7	TDS	1.50	0.86	4.50
8	DO	0.04	0.01	1.04
9	COD	-0.02	0.01	0.98
10	BOD_5	0.17	0.23	1.18

Surface and subsurface drainage effluent contains substances that are potential pollutions. High NO_3-N concentrations in subsurface drainage can originate from a number of sources such as geologic, natural organic matter decomposition and deep percolation of NO_3-N resulting from fertilizer applications[15]. As has been mentioned above group I parameters are mainly found in agricultural drainage water. Thus, group I can be called as "agricultural pollution group".

Cluster analysis was clustered monitoring stations by using water quality dataset and graphed dendrogram. The number of cluster for stations was selected five. Then, for each cluster pair belonging to monitoring stations, Euclidean distances were calculated and distance matrix determined. Using the distance matrices, similarity matrix were formed and presented in Table-3. Regarding table, it was seen that the 4th cluster joins with the 2nd cluster and formed new cluster called 2, the 3rd cluster joins with similarity values of 67.38 % with 2nd cluster and formed new cluster called 2 and the steps repeated until all clusters were joined one cluster. After, dendrogram were drawn using cluster joined *versus* the similarity values (Fig. 1). It was observed from dendrogram that the monitoring stations were formed into two clusters. Cluster I is formed by stations 2, 4, 3 and 5. Cluster II is composed station 1. Cluster I joins with cluster II with similarities of 43.64 %.

TABLE-3
SIMILARITY VALUES, CLUSTER JOINED AND NEW CLUSTER FOR MONITORING STATIONS

Step	Number of clusters	Similarity values (%)	Clusters joined	New cluster	Number of observation in new cluster
1	4	84.20	2-4	2	2
2	3	67.38	2-3	2	3
3	2	63.31	2-5	2	4
4	1	43.64	1-2	1	5

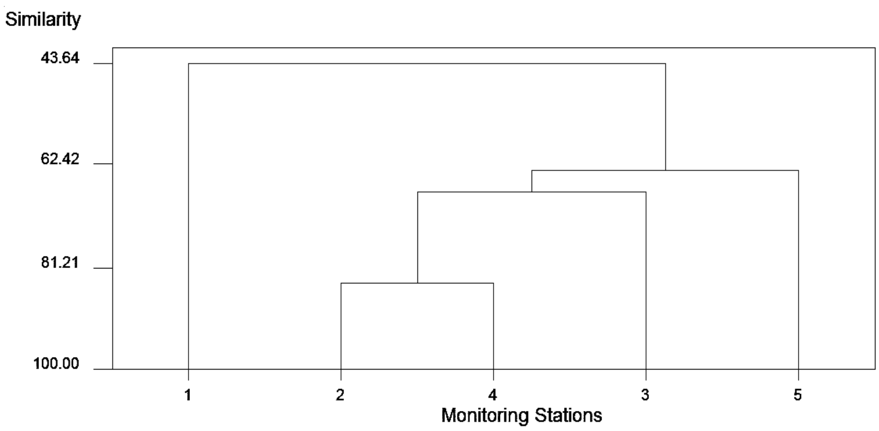

Fig. 1. Dendrogram for monitoring stations

By the representation of the mean values *versus* monitoring stations in the river may be determined (Fig. 2). It was observed that the mean values of TDS, SO_4 and NO_3-N at station 1 were relatively high than the other monitoring stations. Thus, it was determined that the southeast part of Sakarya river is affected by agricultural pollution.

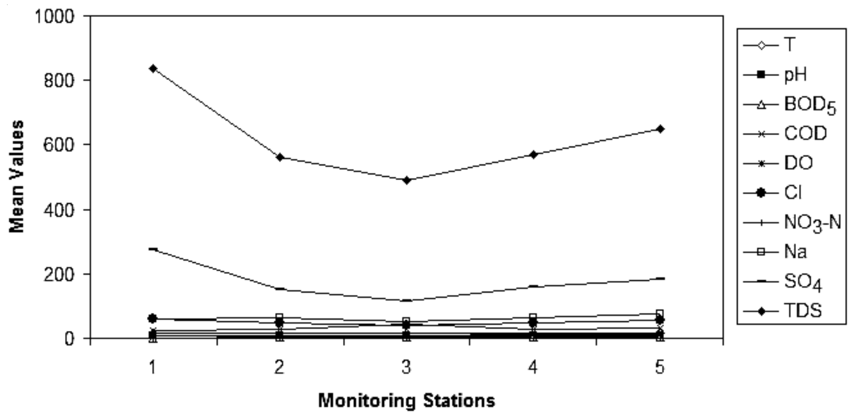

Fig. 2. Mean values concentrations at river monitoring stations

Conclusion

In this study, environmetric methods including ordinal logistic regression technique and cluster analysis were applied to surface water quality dataset obtained from Sakarya river in Turkey. Ordinal logistic regression technique was applied to determine the variations among the monitoring stations and TDS, SO_4 and NO_3-N were identified as the parameters that displayed essential variations. Afterwards, cluster analysis grouped the monitoring stations into two clusters of similarity of the surface water quality characteristics. These results represent that agricultural pollution affected the southeast part of the river. Finally, this study shows that the usefulness of environmetric methods including ordinal logistic regression technique and cluster analysis for the interpretation of dataset, identification of pollution sources and understanding of variations in surface water quality effective river water management.

ACKNOWLEDGEMENT

The author sincerely thanks the Devlet Su Isleri Genel Mudurlugu (DSI) in Turkey for their help in providing the necessary data.

REFERENCES

1. V. Novotny and G. Chesters, Handbook of Non-Point Pollution, Sources and Management, Van Nostrand Reinhold Company, New York (1981).
2. D.M. Santos-Roman, G.S. Warner and F. Scatena, *J. Am. Water Res. Assoc.*, **39**, 829 (2003).
3. C.F. Iscen, A. Altin, B. Senoglu and H.S. Yavuz, *Environ. Monit. Assess.*, **151**, 259 (2008).

4. D.W. Hosmer and S. Lemeshow, Applied Logistic Regression, John Wiley and Sons, New York (1989).
5. D. Wear, M. Turnet and R. Naiman, *Ecol. Appl.*, **8**, 619 (1998).
6. N.K.C. Twarakavi and J.J. Kaluarachchi, *Ground Water*, **43**, 200 (2005).
7. J.T. Gunter, D.G. Hodges and J.L. Regens, *Forest Policy Econom.*, **7**, 868 (2005).
8. K.P. Singh, A. Malik and S. Sinha, *Anal. Chim. Acta*, **538**, 355 (2005).
9. M.S. Dundar and H. Altundag, *Environ. Monit. Assess.*, **128**, 177 (2007).
10. G. Gunay, *Environ. Geol.*, **51**, 229 (2006).
11. S. Kalayci, SPSS Uygulamali Cok Degiskenli Istatistik Teknikleri, Asil Press, Turkey (2005).
12. A.F.M. Alkarkhi and A.M. Ahmad Easa, Environmentalist, doi 10.1007/s10669-008-9190-4 (2008).
13. R.A. Johnson and D.W. Wichern Applied Multivariate Statistical Analysis, Pearson Education International, Englewood Cliffs, New Jersey (2002).
14. C.R. Alvin, Methods of Multivariate Analysis, Wiley Inc., USA (2002).
15. FAO, Management and Agricultural Drainage Water Quality, In: CA Madramootoo, WR Johnson and LS Willardson (eds), ISBN: 92-5-104058-3 (1997).

(*Received*: 22 May 2009; *Accepted*: 23 December 2009) AJC-8220