Asian Journal of Chemistry Vol. 21, No. 9 (2009), 7234-7240

Investigation of Water Quality Characteristics by Using Factor and Multidimensional Scaling Analyses in Porsuk 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

Surface water quality of the Porsuk river in Turkey is evaluated by using the factor analysis (FA) and multidimensional scaling (MDS) analysis in this study. These analyses were applied to the physical and inorganic chemical parameters obtained from the nine different water quality monitoring stations. Three factors were determined, which were responsible from the 67.35 % of total variance of the surface water quality in the river, when factor analysis was applied to the surface water quality dataset obtained from the monitoring stations. Multidimensional scaling analysis generated three groups and determined that Cl^-, Na^+, SO_4^{2-} and COD are the most important parameters causing differences in the surface water quality. Thus, this study shows the usefulness of statistical analysis, including factor analysis and multidimensional scaling analyses, for helping decision makers in water quality management.

Key Words: Factor analysis, Multidimensional scaling analysis, Porsuk river, Water quality.

INTRODUCTION

The quality of surface waters has recently became as significant as their quantity in water resources management since the former directly affects the amount of water that can be used for various purposes such as drinking, agricultural, recreational and industrial uses, *etc*. Water quality assessment encompasses monitoring, data evaluation, reporting and dissemination of the aquatic environment. Major objectives of water quality assessment are describing water quality at regional or national scales, investigating spatial trends and determining if the water quality meets previously defined objectives for designated uses *etc.*[1]. The quality of water is identified in terms of its physical and chemical parameters. A balanced ecosystem in one in which living things and the environment interact beneficially with one another and such an ecosystem can not be achieved by polluted surface waters. Water quality obviously plays a critical role in this relationship[2] as it is the key to the maintenance of a well-balanced environment.

Various statistical techniques, such as factor analysis (FA) and multidimensional scaling (MDS) analysis, help in the interpretation of complex datasets, such as those created by long-term water quality monitoring programs, in order to allow a

better understanding of the spatial variations in river water quality and in the identification of discriminant parameters that are of use in optimizing monitoring stations[3,4]. These techniques permit identification of the possible factors/sources that are responsible for the variations in water quality, influencing the water system and apportionment of the sources. Thus, these techniques are valuable tools for developing appropriate strategies for effective management of the water resources[5].

The aim of this study is to apply the statistical techniques including factor analysis and multidimensional scaling analysis in the evaluation of river water quality parameters and to extract those parameters most relevant in assessing variations in Porsuk river water quality.

EXPERIMENTAL

Study area: The Kutahya region in western Turkey comprises mountainous areas hosting formations with vast amounts of metallic mineralization and plains, covering an area of 2540 km^2 with an elevation of 930 m, where wet agriculture is performed. The drinking and domestic water needs of the rural area are met by groundwater while that of the city of Kutahya is supplied by spring waters. Additionally, water in the Porsuk dam, which is utilized by the citizens of Eskisehir city, is taken from the Porsuk River that drains to the Kutahya plain[6].

The Porsuk river covers a significant urban, agricultural and industrial area of the region. The water of the Porsuk river basin is used for a public and industrial water supply, irrigation and watering animals as well as sports and leisure[7].

Dataset: Surface water quality datasets of nine surface water quality monitoring of Porsuk river stations, comparing 10 water quality parameters monitored monthly for four years, were obtained from Devlet Su Isleri Genel Mudurlugu (DSI). The selected water quality parameter for the determination of water quality characteristics were; temperature (T), pH, biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), dissolved oxygen (DO), chloride (Cl^-), nitrate nitrogen (NO_3-N), sodium (Na^+), sulphate (SO_4^{2-}) and phosphate (PO_4^{3-}). These parameters were evaluated by using factor analysis and multidimensional scaling analyses. All statistical computation was implemented by using SPSS statistical software.

Statistical analysis: In this study, water quality dataset were performed statistical techniques including factor analysis (FA) and multidimensional scaling (MDS) analysis.

Factor analysis (FA): Factor analysis is a 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[3]. The factor analysis can be expressed as:

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

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.

Multidimensional scaling (MDS) analysis: Multidimensional scaling analysis is an appropriate exploratory technique for treating problems with such a need for exploration[8]. The input of the procedure is the proximity matrix of the objects under investigation. It contains the values of a quantitative measure of the pair wise dissimilarities between the objects. The output, however, is a spatial configuration of points in some preferably low-dimensional space. This is the so called final configuration of the MDS analysis solution. The goal of the procedure is to represent the observed structure of the proximity matrix in terms of the interpoint distance in the final configuration as possible. While aiming to reproduce the original rank order of the proximities as close as possible, non-metric MDS analysis iteratively rearranges the points at a fixed dimension of the space until a suitable goodness of fit measure is optimized. The stress function is commonly minimized in this analysis. The number of the dimensions of a correct MDS analysis solution should be sufficient enough to reveal the hidden structure underlying the proximity data. The final step in the procedure is the interpretation of the meaningful dimensions of the MDS analysis solution[9].

RESULTS AND DISCUSSION

Surface water quality parameters were grouped using factor analysis in this study. The eigen values for different components, percentage variance, cumulative percentage variance and component loadings are given in Table-1. Table-1 shows that, the first three eigen values were higher than 1 whereas other eigen values were found to be less than 1. The first three components were extracted and the other components have been eliminated according to these results. In other words, majority of the total variance of the dataset has been investigated by the first three factors Varimax rotation was then used to obtain readily interpretable factor loadings[10]. The surface water quality parameters, loading for three factors from the factor analysis of the dataset, are given in Table-2. Three factors were explaining 67.35 % of the variance in the dataset.

Parameters were grouped based on the factor loadings and the following factors were given: Factor-1: Cl^-, Na^+, SO_4^{2-}, COD; Factor-2: DO, pH, PO_4^{3-}; Factor-3: T, NO_3-N, BOD_5.

Liu *et al.*[11] presented the factor loadings as strong, moderate and weak corresponding to loading values of > 0.75, 0.75-0.50 and 0.50-0.30, respectively. Factor 1 had strong positive loadings in Cl^-, Na^+ and SO_4^{2-}, which were 0.934, 0.913 and 0.780, respectively and a moderate positive loading in COD which were 0.529. The first factor can be donated as agricultural pollution.

The second factor (Factor 2), explaining 23.55 % of total variance, had strong negative loadings on DO and pH and a moderate positive loading on PO_4^{3-}. The

TABLE-1
TOTAL VARIANCE OF PORSUK RIVER

Component	Initial eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative (%)	Total	% of variance	Cumulative (%)	Total	% of variance	Cumulative (%)
1	3.73	37.34	37.34	3.73	37.34	37.34	3.19	31.90	31.90
2	1.87	18.74	56.08	1.87	18.74	56.08	2.36	23.55	55.45
3	1.13	11.27	67.35	1.13	11.27	67.35	1.19	11.90	67.35
4	0.93	9.29	76.64	–	–	–	–	–	–
5	0.89	8.86	85.49	–	–	–	–	–	–
6	0.54	5.44	90.94	–	–	–	–	–	–
7	0.40	4.03	94.97	–	–	–	–	–	–
8	0.35	3.50	98.47	–	–	–	–	–	–
9	0.11	1.12	99.58	–	–	–	–	–	–
10	0.04	0.42	100.00	–	–	–	–	–	–

TABLE-2
FACTOR LOADING (VARIMAX ROTATION) MATRIX

Variables	Factor		
	1	2	3
Cl^-	0.934		
Na^+	0.913		
SO_4^{2-}	0.780		
COD	0.529		
DO		-0.913	
pH		-0.834	
PO_4^{3-}		0.615	
T			0.608
NO_3-N			0.594
BOD_5			0.544

strong negative loadings on DO and pH was due to anaerobic conditions in the river from the loading of high dissolved organic matter; which results in formation of ammonia and organic acids leading to a decrease in pH[3]. Factor 3, on the other hand, had moderate positive loadings in T, NO_3-N and BOD_5, which were 0.608, 0.594 and 0.544, respectively. This factor indicates influence of industrial activities.

The surface water quality parameters were analyzed by using MDS analysis in this study to determinate similarities or differences between monitoring stations. Standardized residual sum of squares, used to assess how well a particular configuration reproduces the observed distance matrix, was found to be 0.005 even for MDS analysis. Standardized residual sum of squares values close to zero shows

that the fit is almost perfect and therefore, results of the MDS analysis was reasonable and reliable[12]. The coordinate values of the monitoring stations can be seen in Table-3. This coordinate values were used to generate a plot given in Fig. 1.

TABLE-3
MULTIDIMENSIONAL SCALING ANALYSIS RESULTS

Stations No	Dimensions	
	1	2
S1	0.629	0.352
S2	0.169	0.413
S3	-1.653	1.668
S4	-2.333	-1.385
S5	0.469	0.182
S6	1.482	-0.259
S7	0.461	-0.605
S8	0.541	0.110
S9	0.574	-1.020

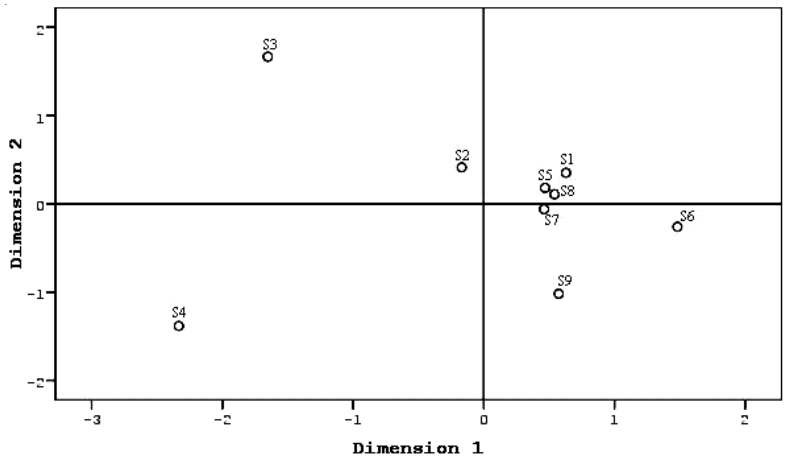

Fig. 1. Dimension 1 and dimension 2 plot

The similarities or differences among nine monitoring stations were found according to dimension-1 and dimension-2 *via* MDS analysis. Fig. 1 ideally illustrates that three groups of objects has similarities. Monitoring stations summed up under each group can be seen in Fig. 1. Based on the results of MDS analysis, it was concluded that: Group-1: S3; Group-2: S4; Group-3: S1, S2, S5, S6, S7, S8, S9.

Group 1 and group 2 are consisted of sites located at the northeast whereas the sites in group 3 were located at the northwest part of the river. The common feature of these sites was relatively high levels of Cl^-, Na^+, SO_4^{2-} and COD values compared to the other monitoring stations (Figs. 2 and 3). Thus, northeast part of the river was affected by agricultural pollution.

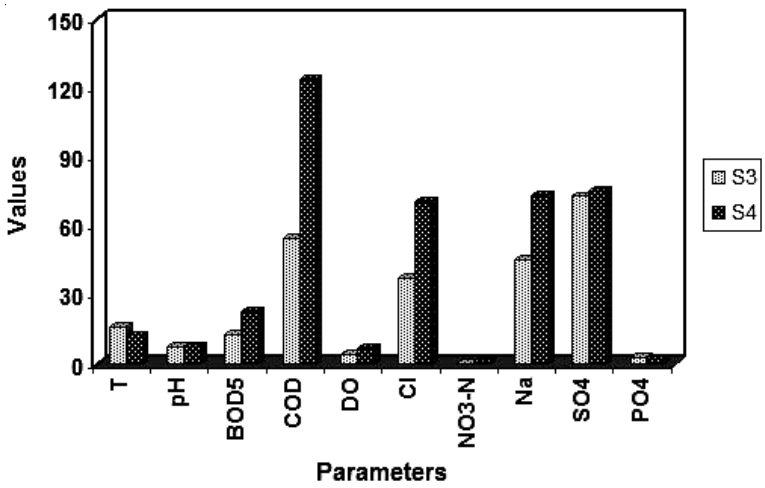

Fig. 2. Comparison of S3 and S4

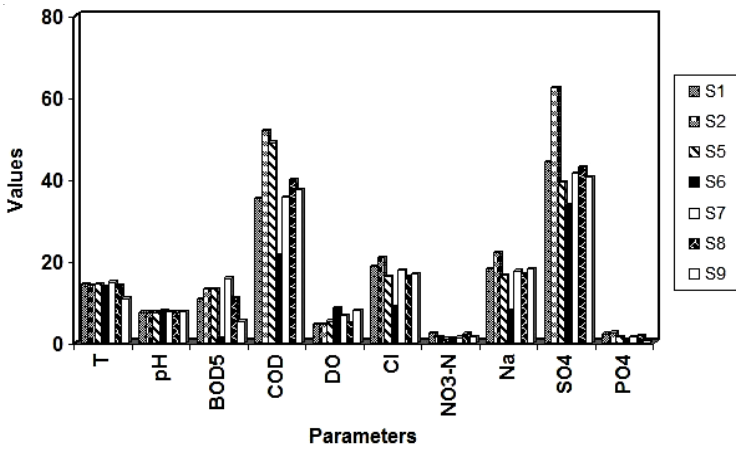

Fig. 3. Comparison of S1, S2, S5, S6, S7, S8 and S9

Conclusion

Statistical analysis methods, including factor analysis and multidimensional scaling (MDS) analysis, can be used successfully to acquire the relevant information from surface water quality dataset. Water quality parameters, which were obtained from nine monitoring stations at Porsuk river in Turkey, were evaluated in this study using factor analysis and multidimensional scaling analysis. Three factors explaining the 67.35 % of the total variance in the water quality dataset were determined. It may be concluded that the monitoring stations were explained by these three factors and Factor 1 (Cl^-, Na^+, SO_4^{2-} and COD) is that most effective among observed variances in the data. These results of factor analysis show that agricultural discharge and industrial activities were main sources of the contamination in the river.

Multidimensional scaling analysis was used to classify monitoring stations with similarities in the river. Group 1 and group 2 are consisted of sites located at the northeast part of the study area. The common feature of these sites was relatively high levels of Cl^-, Na^+, SO_4^{2-} and COD values compared to the other monitoring stations. Thus, northeast part of the river was affected by agricultural pollution.

This study shows the usefulness of factory and multidimensional scaling analysis in water quality assessment and determination of pollution sources with a view to get better information about the water quality and more effective management of hydrological issues.

ACKNOWLEDGEMENT

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

REFERENCES

1. H. Boyacioglu and H. Boyacioglu, *Bull. Environ. Contam. Toxicol.*, **80**, 469 (2008).
2. F.W. Ntengwe, *Phys. Chem. Earth*, **31**, 832 (2006).
3. S. Shrestha and F. Kazama, *Environ. Model. Software*, **22**, 464 (2007).
4. V. Simeonov, J.A. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis, M. Sojoniou and Th. Kouimtzis, *Water Res.*, **37**, 4119 (2003).
5. K.P. Singh, A. Malik and S. Sinha, *Anal. Chim. Acta*, **538**, 355 (2005).
6. N. Kaval and M.T. Nalbantcillar, *Clean*, **35**, 585 (2007).
7. M. Kutlu, G. Aydogan, F. Susuz and A. Ozata, *Environ. Toxicol. Pharmacol.*, **17**, 111 (2004).
8. I. Borg and P. Groenen, Modern Multidimensional Scaling Theory and Applications, New York, Springer, p. 471 (1997).
9. S. Evtimov and M. Ivanov, *Theor. Appl. Climatol.*, **90**, 161 (2007).
10. R.A. Johnson and D.W. Wichern, Applied Multivariate Statistical Analysis, Pearson Education International, edn. 5 (2002).
11. C.W. Liu, K.H. Lin and Y.M. Kuo, *Sci. Total Environ.*, **313**, 77 (2003).
12. C.F. Iscen, A. Altin, B. Senoglu and H.S. Yavuz, *Environ. Monit. Assess.*, **144**, 269 (2008).

(*Received*: 17 February 2009; *Accepted*: 19 August 2009) AJC-7757