Check for updates

http://www.e-journals.net

ISSN: 0973-4945; CODEN ECJHAO
E-Journal of Chemistry
2010, **7(2),** 593-599

Water Quality Assessment of Porsuk River, Turkey

SUHEYLA YEREL

Bilecik University,
Civil Engineering Department, Bilecik, Turkey
suheyla.yerel@bilecik.edu.tr

Received 7 August 2009; Revised 24 October 2009; Accepted 15 December 2009

Abstract: The surface water quality of Porsuk River in Turkey was evaluated by using the multivariate statistical techniques including principal component analysis, factor analysis and cluster analysis. When principal component analysis and factor analysis as applied to the surface water quality data obtain from the eleven different observation stations, three factors were determined, which were responsible from the 66.88% of total variance of the surface water quality in Porsuk River. Cluster analysis grouped eleven observation stations into two clusters under the similarity of surface water quality parameters. Based on the locations of the observation stations and variable concentrations at these stations, it was concluded that urban, industrial and agricultural discharge strongly affected east part of the region. Finally, this study shows that the usefulness of multivariate statistical techniques for analysis and interpretation of datasets and determination pollution factors for river water quality management.

Keywords: Multivariate statistical techniques, Principal component analysis, Factor analysis, Cluster analysis, Water quality, Porsuk River, Turkey

Introduction

The surface water quality is a matter of serious concern today. Rivers due to their role in carrying off the municipal and industrial wastewater and run off from agricultural land in their vast drainage basins are among the most vulnerable water bodies to pollution. The surface water quality in a region is largely determined both by the natural processes and the anthropogenic influences urban, industrial and agricultural activities and increasing exploitation of water resources[1]. Surface water quality in a region is largely determined in terms of its physical, chemical and biological parameters[2].

The particular problem in the case of water quality monitoring has a complexity associated with analyzing the large number of measured variables[2]. The application of different multivariate statistical techniques, such as cluster analysis, principal component

analysis and factor analysis, helps in the interpretation of complex data matrices for a better understanding of water quality and ecological status of the study region. 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[3-6].

Experimental

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. In addition, water in the Porsuk dam, which is utilized by the people who live in the city of Eskisehir, is taken from the Porsuk River that drains to the Kutahya plain[7].

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, watering animals as well as sports and leisure[8].

Dataset

Surface water quality dataset of eleven surface water quality observation stations comparing eleven water quality parameters monitoring monthly five years, were obtained from the General Directorate of State Hydraulic Works in Turkey. Coordinates of observation stations are given in Table 1. The selected surface water quality parameter for the determination of water quality characteristics are 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-}), total dissolved solids (TDS), phosphate (PO_4). The results were evaluated by using multivariate statistical analysis techniques.

Table 1. Observation stations coordinate.

Station No	UTM ZONE	UTM_X	UTM_Y
Station 1	36	247412	4362533
Station 2	36	245132	4374497
Station 3	36	246327	4379675
Station 4	36	369478	4395871
Station 5	36	394695	4393011
Station 6	36	244230	4372500
Station 7	36	244021	4369605
Station 8	36	244108	4369891
Station 9	36	244940	4373861
Station 10	36	289115	4405663
Station 11	36	266725	4391121

Principal component analysis and factor analysis

Principal component analysis was designed to transform the original variables into new, uncorrelated variables, called the principal components, which are linear combinations of the original variables. A principal component provides information on the most meaningful parameters, which describes a whole data set, affording data reduction with a minimum loss of the original information[9-10]. The principal component can be expressed as:

2962, 2010, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2010/438180 by Bilecik Seyh Edebali, Wiley Online Library on [26/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{13} + ... + a_{im}x_{mj} \quad (1)$$

Where, z is the component score, a is the component loading, x the measured value of variable, i is the component number, j the sample number and m the total number of variables.

Factor analysis follows principal component analysis. The main purpose of factor analysis is to reduce the contribution of less significant variables and to simplify even more of the data structure coming from principal component analysis. This purpose can be achieved by rotating the axis defined by principal component analysis according to well established rules, and constructing new variables, also called varifactors. As a result, a small number of factors will usually account for approximately the same amount of information as do the much larger set of original observations[10]. The factor analysis can be expressed as:

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + ... + a_{fm}f_{mi} + e_{fi} \quad (2)$$

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

Cluster analysis

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its objective is to sort cases into clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters[11]. In the case of cluster analysis, the similarities or dissimilarities are quantified through Euclidean distance measurements, the distance between two objects, i and j, is given as;

$$d_{ij}^2 = \sum_{k=1}^{m} \left(z_{ik} - z_{jk} \right)^2 \quad (3)$$

where d_{ij}^2 donates the Euclidean distance, z_{ik} and z_{jk} are the values of variable k for object i and j, respectively, and m is the number of variables[12]. Euclidean distance and the Ward method are used to obtain dendrogram.

Results and Discussion

Multivariate statistical techniques were applied to a surface water quality dataset collected from Porsuk River.

Application of factor analysis and principal component analysis to Porsuk River

The particular problem in the case of surface water quality monitoring is the complexity associated with analyzing the large number of measured parameters. Therefore, factor analysis used to obtain a smaller number of variables for the assessment of surface water quality. Table 2 shows the calculated factor loadings, eigenvalues, total variance and cumulative variance. From the results of the factor analysis, the first three eigenvalues were found to be higher than 1. The scree plot is shown in Figure 1.

The results of the factor analysis revealed that three factors accounted for 66.88% of the variance in datasets. Variables are grouped based on the factor loadings and the following factors are indicated:

Factor 1: Na^+, Cl^-, TDS, SO_4^{2-}, PO_4, BOD_5, COD

Factor 2: DO, pH

Factor 3: NO_3-N, T

2962, 2010, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2010/438180 by Bilecik Seyh Edebali, Wiley Online Library on [26/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Factor loading matrix, eigenvalues and variances.

Variables	Factor 1	Factor 2	Factor 3
Na	0.953		
Cl	0.957		
TDS	0.905		
SO4	0.840		
PO4	0.610		
BOD_5	0.532		
COD	0.453		
DO		0.864	
pH		0.834	
NO_3-N			0.912
T			-0.318
Eigenvalue	4.72	1.55	1.08
Total variance, %	42.93	14.10	9.85
Cumulative variance, %	42.93	57.03	66.88

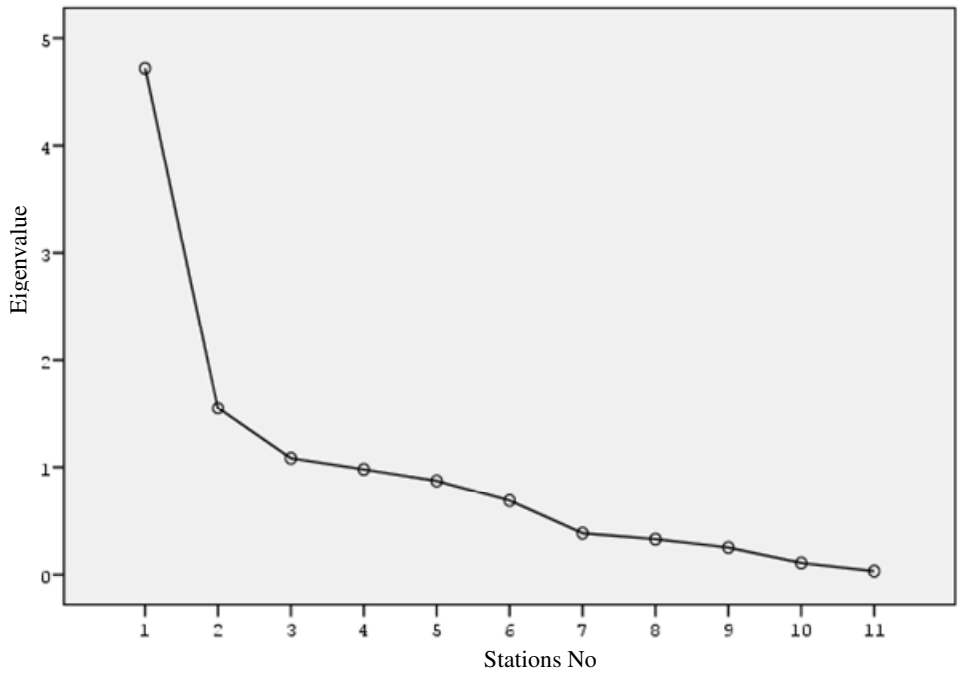

Figure 1. Scree plot of the eigenvalues.

The first factor including Na^+, Cl^-, TDS, SO_4^{2-}, PO4, BOD_5 is explained 42.93% of the variance. DO and pH were strongly correlated with Factor 2 and NO_3-N, T with Factor 3. Liu *et al.*[13] classified the factor loadings as "strong", "moderate" and "weak", corresponding to absolute loading values of >0.75, 0.75-0.50 and 0.50-0.30, respectively. The Factor 1 had a high positive. loading in Na^+, Cl^-, TDS, SO_4^{2-}, moderate positive loading in PO_4, BOD_5

2962, 2010, 2. Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2010/438180 by Bilecik Seyh Edebali, Wiley Online Library on [26/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and weak positive loading in COD. The second component (Factor 2) shows 14.10% of the total variance has strong positive loading of DO and pH. Thus, these parameters are due to anaerobic conditions in the river from the loading of organic matter and organic acids leading to an increase in pH. The third component (Factor 3), explaining 9.85% of total variance, has weak negative loading in T and strong positive loading in NO_3-N. This factor represents pollution from domestic waste and nutrient.

The principal component analysis was applied to dataset to confirm results of factor analysis. The principal component weights are presented in Table 3. This table shows that the first component (PCA1) was composed Na^+, Cl^-, TDS, SO_4^{2-}, PO_4, BOD_5 and COD. These constituents are common in urban, industrial and agricultural pollution in surface water. The second component (PCA2) includes DO and pH. Third component (PCA3) explains NO_3-N and T.

Table 3. Principal component analysis component weights.

Variables	PCA1	PCA2	PCA3
Na	-0.423		
Cl	-0.417		
TDS	-0.409		
SO_4	-0.360		
PO_4	-0.355		
BOD_5	-0.283		
COD	-0.234		
DO		-0.569	
pH		-0.614	
NO_3-N			0.875
T			-0.264

Determination of observation station similarity

Cluster analysis organizes sampling entities into discrete clusters, such that within-group similarity is maximized and among-group similarity is minimized according to some objective criteria[14]. In this study observation stations classification was performed by the use of cluster analysis. The dendrogram of the stations model resulting from the cluster analysis of measured dataset is presented in the Figure 2. Two major groups were formed by treating all the by clustering. Cluster I correspond to Stations 4 and 5. Cluster II correspond to Stations 2, 9, 6, 3, 8, 1, 7, 11 and 10. The classification to those clusters varies with the significance level. It is clearly shows that Cluster II was characterized by the biggest Euclidean distance to the Cluster I. Note also that the points 4 and 5 are localized east of the Porsuk River.

The dendrogram clarifies the abnormality of the observation stations 4 and 5, which make one group as Cluster I, which receive polluted effluents from non-point sources, *i.e.,* from agricultural, industrial and urban activities.

The data of the surface water quality parameters were to compare the aspects of the variation in surface water samples collected from eleven observation stations as shown in Figure 3. Among the mean concentrations, all parameters were found very high at observation stations 4 and 5, showing high pollution of these observation stations.

2962, 2010, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2010/438180 by Bilecik Seyh Edebali, Wiley Online Library on [26/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

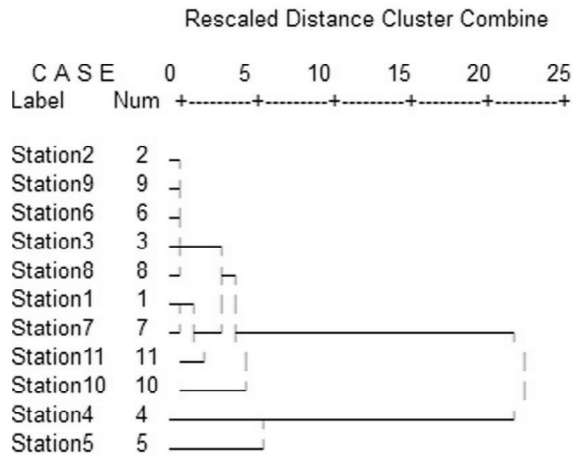

Figure 2. Dendrogram of the Ward method.

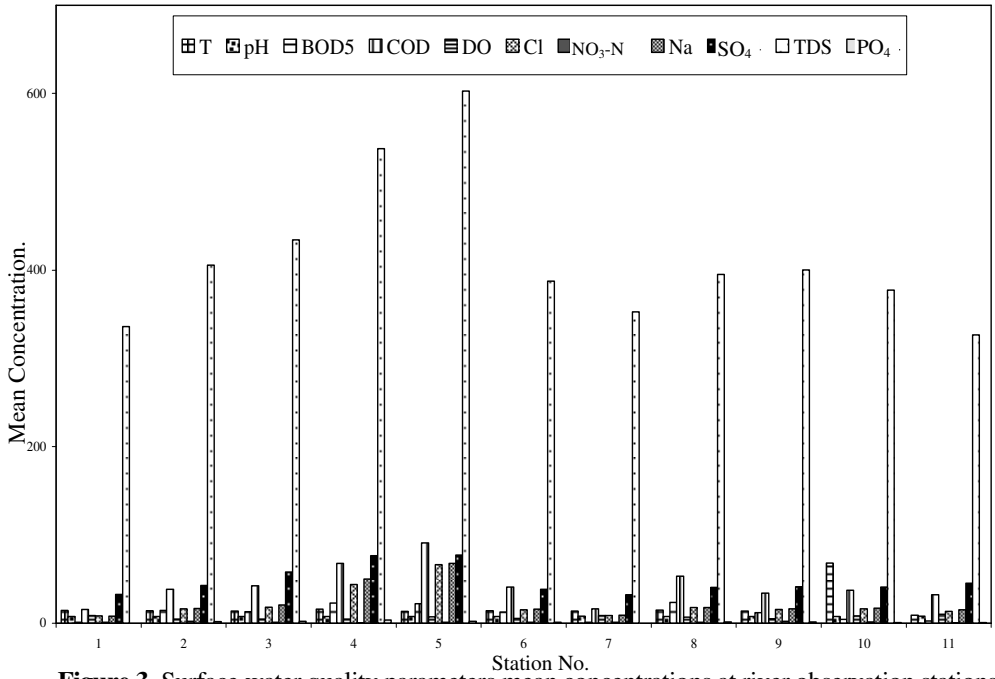

Figure 3. Surface water quality parameters mean concentrations at river observation stations.

Conclusion

This study presents the usefulness of multivariate statistical techniques of large and complex dataset in order to obtain better information concerning surface water quality. In this study, different techniques were applied to dataset obtain from Porsuk River.

Cluster analysis grouped eleven observation stations into two clusters of similar surface water quality characteristics. Based on obtained information, it is possible to design a future, optimal sampling strategy, which could reduce the number of observation stations. Principal component analysis and factor analysis helped in identify the factors responsible for surface

2962, 2010, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2010/438180 by Bilecik Seyh Edebali, Wiley Online Library on [26/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

water quality variations in three different factors. Based on the above results, it may be concluded that of the 66.88% of variances explained by the three factors, it is the anthropogenic factor that best explains the observed variances in the data (42.93%). The parameters loaded in the factor include Na^+, Cl^-, TDS, SO_4^{2-}, PO_4, BOD_5 and COD. These parameters were determined that urban, industrial and agricultural discharge strongly affected east part of the region. Thus, this study shows that usefulness of multivariate statistical techniques in surface water quality assessment and determination of pollution sources.

Acknowledgments

The author sincerely thanks the General Directorate of State Hydraulic Works in Turkey for their help in providing necessary data.

References

1. Singh K P, Malik A and Sinha S, *Analytica Chimica Acta,* 2005, **538**, 355-374.
2. Boyacioglu H, *Water SA,* 2006, **32**, 389-393.
3. Reghunant R, Murthy T R S and Raghavan B R, *Water Res.,* 2002, **36**, 2437-2442.
4. Simeonov V, Simeonov P and Tsitouridou R, *Chem Eng Ecol.,* 2004, **11**, 449-469.
5. Kazi T G, Arain M B, Jamali M K, Jalbani N, Afridi H I, Sarfraz R A, Big J A and Shah A Q, *Ecotoxicol Environl Saf.,* 2009, **72**, 301-309.
6. Yerel S, *Asian J Chem.*, 2009, **21**, 4054-4062.
7. Kaval N and Nalbantcillar M T, *CLEAN-Soil, Air, Water,* 2007, **35(6)**, 585.
8. Kutlu M, Aydogan G, Susuz F and Ozata A, *Environ Toxicol Pharmacol.,* 2004, **17**, 111-116.
9. Helena B, Pardo R, Vega M, Barrado E, Fernandez J M and Fernandez L, *Water Res*., 2000, **34**, 807.
10. Shrestha S and Kazama F, *Environmental Modelling & Software,* 2007, **22**, 464.
11. Johnson R A and Wichern D W, Applied Multivariate Statistical Analysis; Pearson Education International, London, 2002, 767.
12. Everitt B S, Cluster Analysis; John Wiley & Sons Inc, New York, 1993, 170.
13. Liu C W, Lin K H and Kuo Y M, *Science of the Total Environment,* 2003, **313**, 77-89.
14. McGarial K, Cushman S and Stafford S, Multivariate Statistics for Wildlife & Ecology Research; Springer, New York, 2000, 283.

2962, 2010, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2010/438180 by Bilecik Seyh Edebali, Wiley Online Library on [26/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License