Application of Factor Analysis in the Assessment of Water Quality in Sakarya River (Turkey)

Suheyla Yerel
Bozuyuk Vocational School
Bilecik University, Bilecik, Turkey
suheyla.yerel@bilecik.edu.tr

Nurgul Ozbay
Engineering Faculty
Bilecik University, Bilecik, Turkey
nurgul.ozbay@bilecik.edu.tr

Huseyin Ankara
Department of Mining Engineering
Eskisehir Osmangazi University
Eskisehir, Turkey
hankara@ogu.edu.tr

Abstract: Factor analysis is applied to the dataset on surface water quality of the Sakarya River (Turkey), generated one year monitoring at five monitoring stations for eight parameters. This study presents of factor analysis technique for evaluation of large complex dataset with a view to get better information about the surface water quality and design the monitoring stations for effective management of water sources. Three factors were determined, which were responsible from the 88.886% of total variance of the surface water quality in the Sakarya River (Turkey). The first factor explained 43.639% of the total variance. The second factor explained 27.914% and the third factor 17.332% explained of the variance, respectively. This study showed that, factor analysis help decision makers to judge effectiveness of surface water quality programs.

1. Introduction

One of the major concerns in hydrological studies understands the factor and process that control and affect water quality. Water quality "reflects the composition of water as affected by natural processes and by humans' cultural activities, expressed in terms of measurable quantities and related to intended water use" (Novanty and Chesters, 1981). Surface waters contain many chemical species in the dissolved state that play an important role in the survival aquatic ecosystems (Santos-Roman et al., 2003).

Water quality monitoring has one of the highest priorities in environmental protection policy (Simeonov et al., 2002). The main objective is to control and minimize the incidence of pollutant-oriented problems, and to appropriate quality to serve various purposes such as drinking water, irrigation water, etc (Boyacioglu, 2006).

The quality of water is identified in terms of its physical, chemical and biological parameters (Sargaonkar and Deshpande, 2003). The particular problem in the case of water quality monitoring is the complexity associated with analyzing the large number of measured variables (Saffran, 2001). The data sets contain rich information about the behavior of the water resources. The classification, modeling and interpretation of monitoring data are the most important steps in the assessment of water quality (Boyacioglu, 2006).

In this study evaluated the water quality parameters using factor analysis. This analysis was applied to a surface water quality dataset obtain by Sakarya river in Turkey.

2. Materials and Methods

2.1. Dataset

Surface water quality datasets of five surface water quality observation stations comparing eight parameters monitoring monthly a years, were obtained from DSI. Observation stations are seen fig. 1.

The selected surface water quality parameter for the determination of water quality characteristics are; Total dissolved solids (TDS), Sulphate (SO_4), Chloride (Cl), Biochemical oxygen demand (BOD_5), Nitrate nitrogen (NO_3-N), Chemical oxygen demand (COD), Dissolved oxygen (DO) and Sodium (Na). The results were evaluated by using factor analysis.

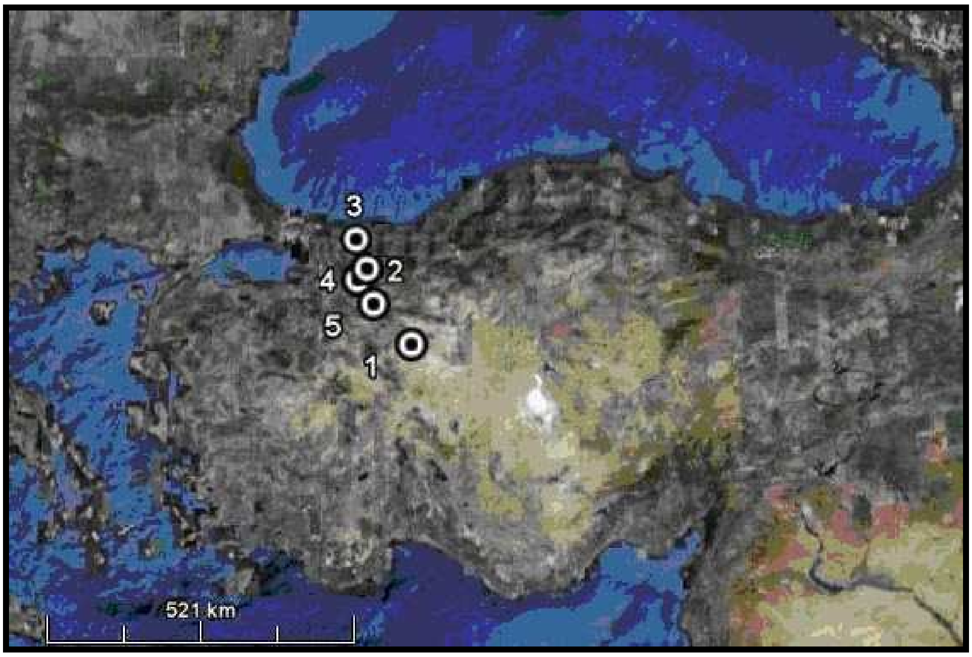

Figure 1. Sakarya river monitoring stations

2.2. Factor Analysis

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 (Shrestha and Kazama, 2007). 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.

3. Results and Discussion

Surface water quality parameters were grouped using factor analysis in this study. The eigenvalues for different factors, eigenvalues and total variance are given in Tab. 1. The table shows that, the three eigenvalues were higher than 1. 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.

Parameters	Factor 1	Factor 2	Factor 3
TDS	0.952	-	-
SO_4	0.946	-	-
Cl	0.887	-	-
BOD_5	0.857	-	-
NO_3-N	-	0.943	-
COD	-	0.777	-
DO	-	-	0.843
Na	-	-	0.708
Eigenvalue	3.491	2.233	1.387
Total variance (%)	43.639	27.914	17.332

Table 1. Rotated factor loadings matrix

The factor analysis generated three significant factors, which explained approximately 88.886% of the variance in monitoring stations dataset. Parameters were grouped based on the factor loadings and the following factors were given:

Factor 1: TDS, SO_4, Cl and BOD_5
Factor 2: NO_3-N and COD
Factor 3: DO and Na

The first factor (Factor 1) is explained 43.639% of the total variance. The second factor (Factor 2) is positively loaded with parameters NO_3-N and COD. This factor accounts for 27.914% of the total variance. Factor 3 explained 17.332% of the total variance and related to the parameters DO and Na.

The data of the Factor 1 were calculation into mean value to compare the aspects of the variation in surface water quality data obtained from five different monitoring stations as presented by fig. 2. Among the mean value, all parameters were found to be high at Station 1 showing high pollution of these sites.

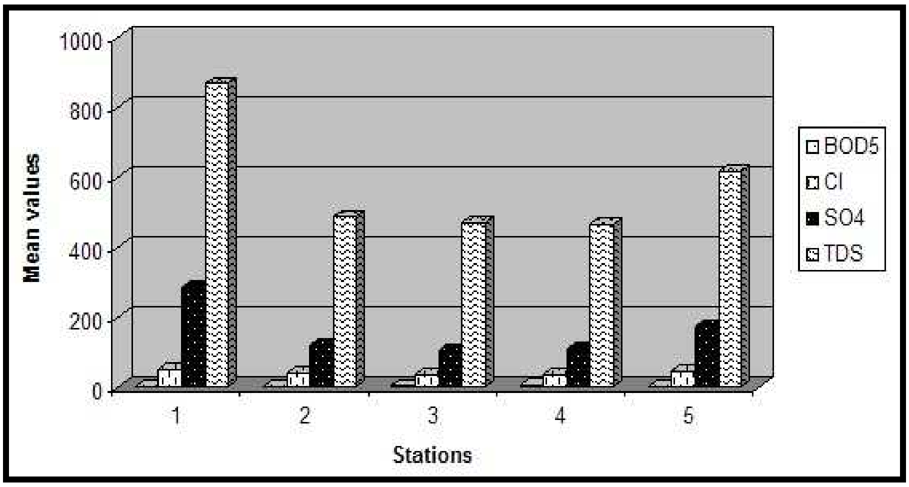

Figure 2. TDS, SO_4, Cl and BOD_5 mean value at monitoring stations

4. Conclusion

In this study, factor analysis was applied to surface water quality dataset. The analysis was used to classify surface water quality parameters. Based on the above results, monitoring stations were explained by the three factors, it is the Factor 1 (TDS, SO_4, Cl and BOD_5) that best observed variances in the data. These results represent that pollution caused differences in terms of water quality in the Station 1 of the river. Finally, it was determined that the factor analysis usefulness for interpretation of water quality dataset.

5. References

1. Boyacioglu, H. (2006). Surface water quality assessment using factor analysis. *Water SA*, 32, 389-394.

2. Novanty, V., & Chesters, G. (1981). *Handbook of non-point pollution, Sources and Management*. Van Nostrand Reinhold Company, New York.

3. Saffran, K. (2001). *Canadian water quality guidelines for the protection of aquatic life,* CCME water quality index 1,0 User's manual, ISBN 1-896997-34-1.

4. Santos-Roman, D.M., Warner, G.S., & Scatena, F. (2003). Multivariate analysis of water quality and physical characteristics of selected watersheds in Puerto Rico. *Journal of the American Water Resources Association*, 829-839.

5. Sargaonkar, A., & Deshpande, V. (2003). Development of an overall index of pollution surface water based on a general classification scheme in Indian context. *Environ. Monit. Assess*., 89, 43-67.

6. Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques. A case study of Fuji river basin, Japan. *Environmental Modelling and Software*, 22, 464-475.

7. Simeonov, V., Einax, J.W., Stanimirova, I., & Kraft J. (2002). Environmetric modeling and interpretation of river water monitoring data. *Anal. Bional Chem*., 374, 898-905.