Journal of Applied Biological Sciences 7 (3): 79-83, 2013
ISSN: 1307-1130, E-ISSN: 2146-0108, www.nobel.gen.tr

Use of Statistical Methods in Water Quality Assessment: A Case Study of Balkan Arboretum Area in Trakya University (Edirne, Turkey)

Cem TOKATLI

Trakya University, İpsala Vocational School, Department of Laboratory Technology

*Corresponding author:
E-mail: tokatlicem@gmail.com

Received: November 30, 2013
Accepted: December 29, 2013

Abstract

The aim of this study was to evaluate the quality of freshwater resources of Balkan Arboretum Area located in Trakya University (Güllapoğlu Stream, Artificial Pond and Güllapoğlu Well) by a statistical approach using Pearson Correlation Index (PCI), Factor Analysis (FA) and Cluster Analysis (CA). For this purpose, total of 9 water quality parameters (dissolved oxygen, temperature, conductivity, pH, nitrate, nitrite, phosphate, sulphate and cyanide) were measured in water samples, which were collected monthly from 5 stations between the dates of June 2013 – October 2013. According to results of PCI, significant relations were recorded between temperature – nitrate and nitrate – sulphate parameters ($p<0.01$). According to results of FA, 2 factors named as "Agricultural Factor" and "Climatic Factor" explained 56.66% of the total variance. According to CA, 3 statistically significant clusters were formed, which were corresponded to Lentic and Lotic 1; Lotic 2 and Lotic 3; and Well stations.

Keywords: Balkan Arboretum Area, Freshwater Quality, Factor Analysis, Cluster Analysis, Pearson Correlation Index

INTRODUCTION

Statistical methods are being widely used in large numbers of countries in water quality assessment studies. Especially multivariate statistical techniques help the interpretation of complex data matrices to better understand the ecological status of the investigated ecosystems. Factor and Cluster Analysis are the most convenient multivariate statistical methods that are used commonly all over the world in order to evaluate the water quality of many different aquatic ecosystems [1, 2, 3, 4, 5].

Balkan Arboretum Area, which is located in the north east part of the Balkan Campus of Trakya University, contains a few important freshwater resources including lotic, lentic and groundwater ecosystems (Güllapoğlu Stream, Artificial Pond and Güllapoğlu Well).

In the present study, some mono (Pearson Correlation Index) and multi (Factor and Cluster Analysis) statistical methods were applied to lymnological data detected from freshwater resources of Balkan Arboretum Area in order to evaluate the water quality and observe the stational – habitatal differences of the region in a statistical approach.

MATERIALS AND METHODS

Study Area and Collection of Samples

Balkan Campus is the biggest campus of the Trakya University and has an area of 2.215.744 square meter including 256.835 square meter closed area [6]. Balkan Arboretum Area, where constitutes the study area of the present investigation, is located in the Balkan Campus of Trakya University and contains an important freshwater potential.

Water samples were collected monthly from five selected stations between the dates of June 2013 – October 2013 from Güllapoğlu Stream (three stations), Artificial Pond (one station) and Güllapoğlu Well (one station) that are located in the Balkan Arboretum Area. Map of Trakya University and selected stations were given in Figure 1.

Physical and Chemical Analysis

Dissolved oxygen, temperature, conductivity and pH parameters were determined by using "Hach Lange HQ40D Multiparameter" device during the field studies and nitrate, nitrite, phosphate, sulphate and cyanide parameters were determined by using "Hach Lange DR890 Colorimeter" device during the laboratory studies.

Figure 1. Study area and selected stations

Statistical Analysis

Pearson Correlation Index (PCI) was applied to the results in order to determine the relations between the psychochemical parameters by using the SPSS 17 package program. Factor Analysis (FA) was applied to the results in order to determine the effective varifactors on freshwater resources of Balkan Arboretum Area according to correlated variables by using the SPSS 17 package program. Cluster Analysis (CA) was applied to the results in order to classify the stations according to ecological status by using the Past package program.

RESULTS AND DISCUSSION

Pearson Correlation Index

The relations between the levels of physical and chemical parameters in freshwater resources of Balkan Arboretum Area were determined by using mensal data, (n = 25 for all parameters) and all detected relations are given in Table 1. It was found that, the relations between dissolved oxygen – temperature and conductivity (-), temperature – conductivity (-) and phosphate (+), nitrate – sulphate and cyanide (+), sulphate – cyanide (+) levels were directly proportional at the 0.05 significance level; and temperature – nitrate (+), nitrate – sulphate (+) levels were directly proportional at the 0.01 significance level.

Factor Analysis (FA)

Principal Component Analysis (PCA) is a powerful pattern recognition tool. It attempts to explain the variance of a large dataset of inter correlated variables with a smaller set of independent variables. Factor Analysis (FA) reduces the contribution of less significant variables and makes new group of variables detected from PCA. New group of variables are extracted through rotating the axis defined by PCA. A varifactor can include unobservable, hypothetical, latent variables, while a principle component is a linear combination of observable variables [4, 5, 7, 8, 9].

Table 1. Pearson Correlation Index coefficients

	DO	Temp	Cond	pH	NO_3	NO_2	PO_4	SO_4	CN
DO	1								
Temp	**-.502***	1							
Cond	**-.455***	**.403***	1						
pH	**.407***	-.055	.086	1					
NO_3	.223	**-.516****	-.278	-.129	1				
NO_2	.120	-.392	-.275	-.234	.293	1			
PO_4	.063	**-.408***	.089	.186	.050	.313	1		
SO_4	.081	-.288	.092	-.161	**.547****	**.412***	.320	1	
CN	.217	-.143	-.065	.242	.267	**.414***	.228	**.496***	1

DO: Dissolved oxygen; Temp: Temperature; Cond: Conductivity
*: Correlation is significant at the 0,05 level (p<0.05); **: Correlation is significant at the 0,01 level (p<0.01)

FA was used to determine the effective varifactors on freshwater resources of Balkan Arboretum Area by using correlated variables. Uncorrelated variables (pH) were removed from the data set in order to increase the reliability of FA. A total of eight variables were used to detect the varifactors (n = 25 for all parameters). Result of KMO (Kaiser-Meyer-Olkin) test that presents the measure of sampling adequacy was 0.59 and this value means that, the sampling adequacy was in a good level for the present application (>0.5) [10].

Eigenvalues higher than one were taken as criterion for evaluate the principal components that required to explain the sources of variance in the data. According to rotated cumulative percentage variance, two factors explained 56% of the total variance and the scree plot of FA is given in Figure 2.

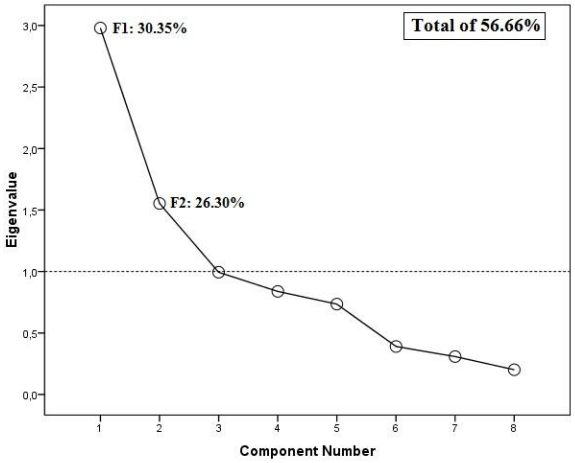

Figure 2. Scree plot of FA

The parameter loadings higher than 0.5 calculated after rotation for two components are given in Figure 3. Also component plot in rotated space, which shows the related variables of two factors, is given in Figure 4. Liu et al. [10] classified the factor loadings according to loading values as "strong (>0.75)", "moderate (0.75 – 0.50)" and "weak (0.50 – 0.30)".

First factor (F1), named as "Agricultural Factor" explained 30.35% of total variance and it was related to the variables of sulphate, cyanide, nitrite, phosphate and nitrate parameters. All parameters were strong positively loaded with this factor (Figure 3, 4).

Second factor (F2), named as "Climatic Factor" explained 26.30% of total variance and it was related to the variables of temperature, dissolved oxygen and conductivity parameters. Dissolved oxygen parameter was strong positively and temperature and conductivity parameters were strong negatively loaded with this factor (Figure 3, 4).

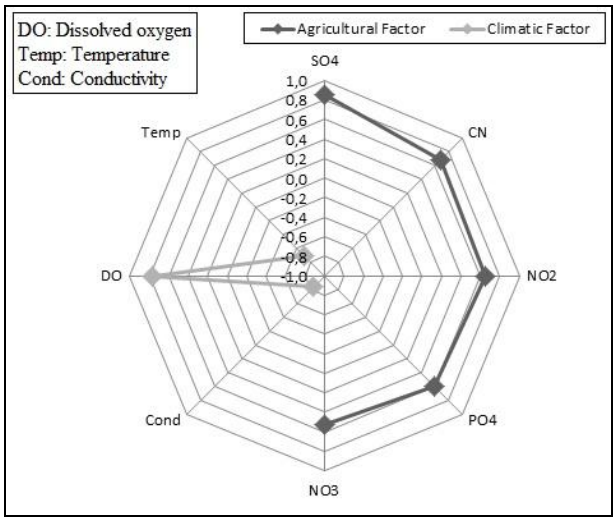

Figure 3. Rotated component matrix

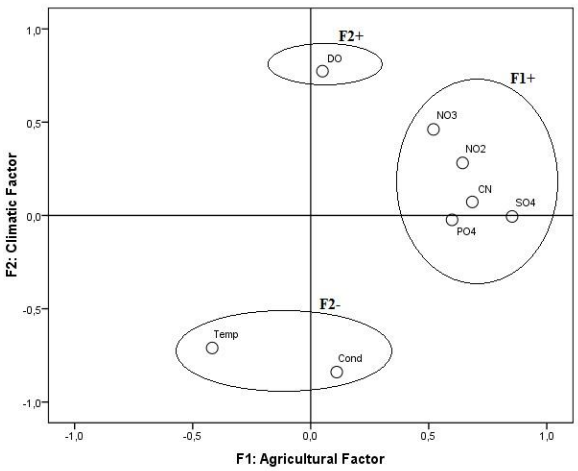

Figure 4. Rotated component matrix

Cluster Analysis (CA)

Cluster Analysis (CA) that classifies the objects is an important group of multivariate statistical techniques. Assembling objects based on the characteristics they possess is the primary purpose of Cluster Analysis (CA). One of the most common approaches in CA is the hierarchical agglomerative clustering, which provides intuitive similarity relationships between any one sample and the entire data set and is typically illustrated by a dendrogram in order to provide visual summaries of the clustering processes [1, 4, 5, 11].

CA was applied to detected data to classify the stations according to physicochemical status. The diagram of CA calculated by using temperature, conductivity, dissolved oxygen, pH, nitrate, nitrite, phosphate, sulphate and cyanide levels of freshwater resources of Balkan Arboretum Area was given in Figure 5.

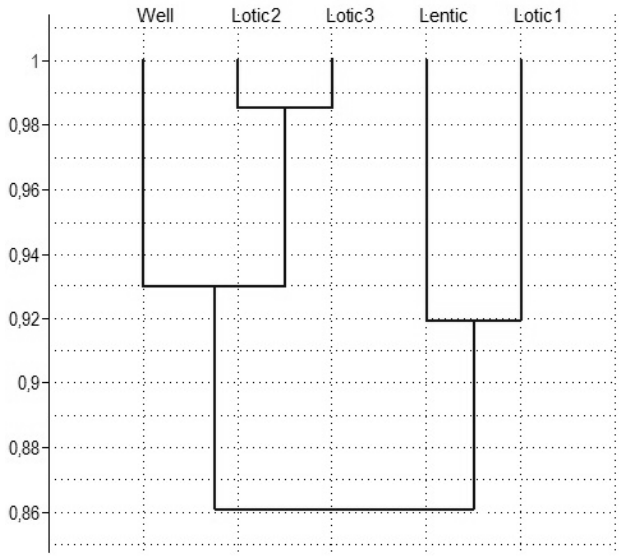

Figure 5. Diagram of CA

According to results of CA, three statistically significant clusters were formed: Cluster 1 corresponded to “Lentic” and “Lotic 1” stations; Cluster 2 corresponded to “Lotic 2” and “Lotic 3” stations; Cluster 3 corresponded to “Well” station. Maximum similarity was observed between Lotic 2 and Lotic 3 stations (98%) and minimum similarity was observed between Lentic and Well stations (79%) (Table 2).

Table 2. Pearson Correlation Index coefficients

	Lotic1	Lotic2	Lotic3	Letic	Well
Lotic1	1.000				
Lotic2	0.908	1.000			
Lotic3	0.911	**0.985**	1.000		
Letic	0.919	0.846	0.857	1.000	
Well	0.846	0.932	0.928	**0.798**	1.000

DISCUSSION

Dissolved oxygen is an important water quality parameter. The amount of dissolved oxygen in water depends on many environmental factors like temperature, density of dissolved salts and biological processes. As it is known, dissolved oxygen in water is closely related to the temperature and conductivity parameters and solubility of oxygen in water decreases with increasing temperature and conductivity [12, 13]. As similar to the literature knowledge, significant negative correlations were recorded between dissolved oxygen – temperature and conductivity parameters ($p<0.05$). Also temperature and conductivity parameter were strong negatively and dissolved oxygen parameter was strong positively loaded with second factor (F2), which named as "Climatic Factor" and explained 26.3% of total variance.

Agricultural activities and irrigation practices could significantly raise the concentrations of phosphate and nitrogenous compounds in close aquatic ecosystems to the agricultural lands [12, 13]. According to results of FA, “Agricultural Factor”, which was explained 30.35% of total variance and strong positively related to the variables of sulphate, cyanide, nitrite, phosphate and nitrate parameters, was identified as the most effective component for the freshwater resources of Balkan Arboretum Area.

Two of the most widely used multivariate statistical techniques are Factor and Cluster Analysis. They provide valuable and easy explaining data and helps in the interpretation of complex data matrices for a better understanding of water quality and ecological status of the aquatic system. They are being used in large numbers of countries in order to assess many different freshwater habitats [14, 15, 16].

In a study performed in Xiangjiang watershed in China, FA and CA were used to evaluate the quality of lotic and lentic habitats of a surfacewater ecosystem as similar to the present study. FA reduced the data sets in 4 latent factors for 3 different sites accounting for 71.62%, 71.77% and 72.01% of the total variance and CA grouped 34 sampling sites into 3 clusters according to the similarities of water quality characteristics [17].

In a study performed in Uluabat Lake in Turkey, FA and CA were used to assess the water quality of a lentic ecosystem. According to FA, 3 factors explained 77.35% of total variance and CA grouped 12 sampling sites into 2 clusters. Phosphate and sulphate parameters were located in the same factor (Microbial Factor) with positive factor loadings as detected in the present study [18].

In another study performed in Turkey, groundwater quality of Türkmen Mountain was evaluated by using FA. According to rotated cumulative percentage variance, 4 factors explain 79.25% of the total variance. As similar to the present study “Agricultural Factor” that explained 14.06% of total variance was an effective component on groundwater quality of Türkmen Mountain and nitrite – phosphate parameters were positively loaded with this factor. Also dissolved oxygen and conductivity parameters had opposite factor loadings as in freshwater resources of Balkan Arboretum Area [5].

CONCLUSION

Multivariate statistical techniques were used to evaluate the freshwater resources quality of the Balkan Arboretum Area by using a large number of physico – chemical data that were difficult to evaluate without using any multivariate statistical method. Factor Analysis helped to identify the effective factors on the system and 2 effective factors were determined that were explained 56.66% of the total variance. Also Cluster Analysis helped to group the sampling stations into three statistically significant clusters of similar water quality characteristics.

In conclusion of the present study, multistatistical techniques are necessary for a sophisticated environmental evaluation especially in water quality assessment studies because of obtained large numbers of different parameters and difficulty of the interpretations of all parameters. Results of this study reveal the benefits of statistical approaches in freshwater evaluation studies.

Acknowledgment

The author would like to acknowledge the support provided by Trakya University.

REFERENCES

[1] Shrestha S., Kazama F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin; Japan. Environmental Modelling & Software, 22, 464–475.

[2] Akin, B. S., Atıcı, T., Katircioglu, H., Keskin, F., 2010. Investigation of water quality on Gökçeekaya dam lake using multivariate statistical analysis, in Eskişehir, Turkey. Environ Earth Sci, DOI 10.1007/s12665-010-0798-6.

[3] Najar, IA., Khan, AB. 2012. Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis. Environ Earth Sciences, 66:2367–2378.

[4] Tokatlı, C., Çiçek, A., Emiroğlu, Ö., Arslan, N., Köse, E., Dayıoğlu, H., 2013. Statistical Approaches to Evaluate the Aquatic Ecosystem Qualities of a Significant Mining Area: Emet Stream Basin (Turkey). Environmental Earth Sciences, DOI 10.1007/s12665-013-2624-4.

[5] Tokatlı, C., Çiçek, A., Köse, E., 2013. Groundwater Quality of Türkmen Mountain (Turkey). Polish Journal of Environmental Studies, 22 (4), 1197-1208.

[6] http://www-en.trakya.edu.tr/

[7] Vega, M., Pardo, R., Barrato, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research, 32, 3581–3592.

[8] Wunderlin DA., Diaz MP., Ame MV., Pesce SF., Hued AC., Bistoni MA., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquira river basin (Cordoba — Argentina). Water Research, 35, 1894–2881.

[9] Simeonov V., Stratis JA., Samara C., Zachariadis G., Voutsa D., Anthemidis A., 2003. Assessment of the surfacewater quality in Northern Greece. Water Research, 37, 4119–4124.

[10] Liu, CW., Lin, KH., Kuo, YM., 2003. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science of the Total Environment, 313, 77–89.

[11] Tabachnick, BG., Fidell LS., 1996. Using multivariate statistics (3rd ed.). New York: Harper Collins College Publishers.

[12] Wetzel, RG., 2001. Limnology: Lake and River Ecosystems. Elsevier Academic Press, 1006 pages.

[13] Manahan, SE., 2011. Water Chemistry: Green Science and Technology of Nature's Most Renewable Resource. Taylor & Francis Group, CRC Press, 398 pages.

[14] Kazi, TG., Arain, MB., Jamali, MK., Jalbani, N., Afridi, HI., Sarfraz, RA., Baig, JA., Shah, AQ., 2009. Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. Ecotoxicol Environ Saf 72:301–309.

[15] Çiçek, A., Tokatlı, C., Köse, E., 2013. An Application on Using Multivariate Statistical Techniques to Evaluate the Sediment Quality. Journal of Research in Ecology, 2 (1): 075-082.

[16] Çiçek, A., Bakış, R., Uğurluoğlu, A., Köse, E., Tokatlı, C., 2013. The Effects of Large Borate Deposits on Groundwater Quality of Seydisuyu Basin (Turkey). Polish Journal of Environmental Studies, 22 (4): 1031-1037.

[17] Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q., Wang, Y., Ye, F., 2009. Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. Environ Monit Assess, 152:123–131.

[18] Iscen, CF., Emiroglu, Ö., Ilhan, S., Arslan, N., Yılmaz, V., Ahiska, S., 2007. Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. Environ Monit Assess DOI 10.1007/s10661-007-9989-3.