



Tarım Bilimleri Dergisi
Tar. Bil. Der.

Dergi web sayfası:
www.agri.ankara.edu.tr/dergi

Journal of Agricultural Sciences

Journal homepage:
www.agri.ankara.edu.tr/journal

Application of Multivariate Statistical Analysis in the Assessment of Surface Water Quality in Seyfe Lake, Turkey

Sultan KIYMAZ^a, Ufuk KARADAVUT^b

^aAhi Evran University, Faculty of Agriculture, Department of Biosystems Engineering, Kırşehir, TURKEY

^bAhi Evran University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, Kırşehir, TURKEY

ARTICLE INFO

Research Article DOI: 10.1501/Tarimbil_0000001275

Corresponding Author: Sultan KIYMAZ, E-mail:skiyamaz@ahievran.edu.tr, Tel: +90 (386) 280 48 19

Received: 19 August 2013, Received in Revised Form: 21 January 2014, Accepted: 28 January 2014

ABSTRACT

Multivariate statistical methods are successfully used in many areas. Cluster Analysis and Principal Component Analysis methods are the most used of these methods. In this study, water quality data prepared for Seyfe Lake were evaluated. For this purpose, principal components and cluster analysis techniques were used. The 26 selected parameters were gathered at 11 of the points and evaluated. The cluster analysis was obtained from three different groups. These sampling points have different physicochemical characteristics and the pollution levels. According to the results of Principal Component Analysis, the six factors explained 92.72% of the total variance. The first factor 29.52%, the second factor 17.89%, the third factor 16.75%, the fourth 13.65%, the fifth 9.73%, and the sixth 5.17% of the cumulative variances explained respectively. Results reveal that sulfate, nitrate, total phosphorus, hardness, electrical conductivity, total dissolved solid, magnesium, potassium, sodium, calcium, biochemical oxygen demand, chemical oxygen demand, aluminum, iron, chromium, and lead were the most important parameters used to evaluate changes in water quality of the lake.

Keywords: Seyfe Lake; Multivariate data analysis; Quality of water; Pollution; Agricultural activities

Türkiye’de Seyfe Gölü Yüzey Su Kalitesinin Değerlendirilmesinde Çok Değişkenli İstatistiksel Analizlerin Uygulanması

ESER BİLGİSİ

Araştırma Makalesi

Sorumlu Yazar: Sultan KIYMAZ, E-posta:skiyamaz@ahievran.edu.tr, Tel: +90 (386) 280 48 19

Geliş Tarihi: 19 Ağustos 2013, Düzeltmelerin Gelişi: 21 Ocak 2014, Kabul: 28 Ocak 2014

ÖZET

Çok değişkenli istatistiksel yöntemler birçok alanda başarıyla kullanılmaktadır. Kümeleme ve temel bileşenler analiz yöntemleri en çok kullanılan yöntemlerdendir. Bu çalışmada, Seyfe Gölü’nden elde edilen su kalitesi verileri değerlendirildi. Bu amaç için temel bileşenler ve kümeleme analizi kullanıldı. On bir farklı noktadan alınan seçilmiş 26 parametre değerlendirilmiştir. Kümeleme analizi üç farklı gruptan elde edilmiştir. Bu örnekleme noktaları farklı

fiziko-kimyasal özelliklere ve kirlilik seviyelerine sahiptir. Temel bileşenler analizi sonuçlarına göre, toplam varyansın % 92.72'sini altı faktör açıklamıştır. Kümülatif varyansın % 29.52'sini birinci faktör, % 17.89'unu ikinci faktör, % 16.75'ini üçüncü faktör, % 13.65'ini dördüncü faktör, % 9.73'ünü beşinci faktör ve % 5.17'sini altıncı faktör sırasıyla açıklamıştır. Sonuçlar sülfat, nitrat, toplam fosfor, sertlik, elektriksel iletkenlik, toplam çözünmüş madde, magnezyum, potasyum, sodyum, kalsiyum, biyokimyasal oksijen, kimyasal oksijen, alüminyum, demir, krom ve kurşunun gölün su kalitesindeki değişimlerini değerlendirmek için kullanılan en önemli parametreler olduğunu ortaya koymuştur.

Anahtar Kelimeler: Seyfe gölü; Çok değişkenli veri analizi; Su kalitesi; Kirlilik; Tarımsal faaliyetler

1. Introduction

The lake water is an essential renewable source for human life, the environment, and industrial purposes. Continuous monitoring is required for the sustainability of water resources and registered changes (Giardino 2007). It is important for the early detection of pollution parameters. Surface water pollution is the easiest and the most affected waters of the pollutants. The main sources of water pollution are eutrophication and industrial pollution sources. As sources of pollutants include agricultural activities, and urban and industrial wastes include industrial applications. These pollutants reduce amount of oxygen in the water adversely and affect the biodiversity (Ouyang et al 2006). The determination of the impact amount and period is a major issue. Adverse changes can be identified in advance to see if this is possible; pollutants may possibly be minimized. Otherwise, water ecology may be polluted irrevocably.

In recent years, many studies were conducted on the topics and continue to be conducted. Water quality monitoring is important to determine the changes in chemical and physicochemical characteristics. In water quality monitoring, the number of variables is necessity to identify for pollution levels. Each data set gives valuable information about the topic being studied. Considerable improvement was achieved in determining the classification and modeling of water quality. The data obtained from monitoring is important to determine the classification and modeling of water quality (Boyacıoğlu 2006).

Multivariate data analysis methods used to decrease the number of characters are useful

techniques for chemical and physical measurements (Singh et al 2004). Multivariate data analysis methods as principal component analysis (PCA), factor analysis (FA) and cluster analysis (CA) will aid in understanding complex data more easily. Multivariate data analyses have been revealed to reduce data without losing the original information (Vega et al 1998; Singh et al 2004; Mendes 2011). Reducing the number of variables makes the assessment process more rapid. The reduction in variable number in the original data aids in the identification of the relationships between variables. The principle factor and principal component techniques are the most made use of multivariate data analysis methods in these fields. In terms of quality properties of water, the determination of quality properties data of water and features can aid in changing the issues.

Zhou et al (2007) explained these methods that are used to identify water pollution, water pollution parameters, and for the classification of stations by principal components analysis. Mendiguchía et al (2004) made of use the principal component analysis in order to determine the basic parameters of pollution. The evaluation of qualification of water resources in these methods successfully have been used by many researchers (Vega et al 1998; Lee et al 2001; Wunderlin et al 2001; Simeonova et al 2003; Simeonova et al 2004; Singh et al 2004; Shrestha & Kazama, 2007). According to these researchers, the principal component analysis of multivariate data analysis can be made use of successfully determine the quality of water.

Clustering analysis methods are one of the most common of multivariate statistical methods. Cluster

analyses are preferred to more clearly explain the similarities and differences among variables. Cluster analysis helps to determine the data and the properties and successful interpretations more clearly (Vega et al 1998). Cluster analysis can be used as a predetermined selection criterion (Shrestha & Kazama 2007). It can reveal high internal homogeneity or high external heterogeneity (Singh et al 2004). Cluster analysis produces results in the form of a dendrogram and clarifies the relationships.

Factor analysis was made use of decreasing the number of variables with a great number of internal relations. An analysis of the factors helps to explain the best way to obtain the least possible variables. This reduction of the data to obtain a new data set is computed by converting. The new variables that are arranged in order are important. Furthermore, the factor analysis includes a principal component analysis. Principal components analyses are mathematically calculated from the covariant and other cross-product matrices. The principal component technique explains the variance of great number of data and the interrelated variables. It produces a smaller number of non-dependent characters (Simeonova et al 2003). Principal component analyses explain to obtain the Eigenvalues and eigenvectors for the characters to better express the distribution of the measured parameters. Principle components are a linear combination of original characters and eigenvectors (Wunderlin et al 2001).

In the study it was aimed to evaluate with multivariate data analysis methods as PCA, CA, and FA on Seyfe lake water quality parameters and its variations.

2. Material and Methods

2.1. Research field

Seyfe Lake is located in Kırşehir Province, in the Middle Anatolia Region of Turkey (Figure 1), and average altitude ranges from 1110 to 1706 meters. The lake is located between 39°18’ north and 34°23’ east. Seyfe Lake has a 22 km² of surface area, a drainage basin area of 1430 km², and a maximum

depth of 1.65 meter. The water level of lake area ranges from 1560 hectares to 70000 hectares and the average area is 3400 hectares. It has length of 11 km and width of 6 km. The lake is principally fed by the Seyfe source and discharge is evaporated (DSİ 2004).

The climate is semi-arid and continental. Summers in the area are warm and droughty; winters are cold and precipitation. Average temperatures range between 9.5 °C and 11.4 °C. In the entire area, average annual precipitation ranges from 270 mm to 450 mm (Türkeş 2007). In addition, in 1994, Seyfe Lake was included in the Ramsar list. An area of 10700 hectares was declared “Ramsar”. It is also an essential habitat for the breeding of birds such as flamingos, white pelicans, spoonbills, little egrets, red-crested pochards, long legs, swordfish bills, spurred lapwings, Mediterranean gulls, black-headed gulls, and gull-billed tern broods. In this region, the economy is heavily dependent on agricultural and animal activities. Wheat, barley, sugar beets, and maize are the main crops grown near Seyfe Lake (DSİ 2004).

The lake water is mainly used for agriculture and drinking water purposes. Farming activities are present around the surface water stations and agricultural and animal husbandry activities are present near Seyfe Lake. Settlement areas (Malya, Eskidoğanlı, and Seyfe villages) are located near the sources that are fed the lake in basin. Therefore, Seyfe Lake is affected by point and non-point sources contamination. Non-point sources include fertilizers, waters of drainage, poor quality irrigation water, and livestock wastes. Point sources of pollution include domestic sewage and clothes washing machine runoff. Water quality varies continuously in the area due to these factors. Therefore, it is of great importance to determine changes in water quality.

2.2. Water sampling and analysis

In the field, representative sampling sites were identified. Samples were gathered at 30 days periodically between January and December 2009. Samples were taken in accordance with the irrigation period. Sampling stations are given in Figure 2.

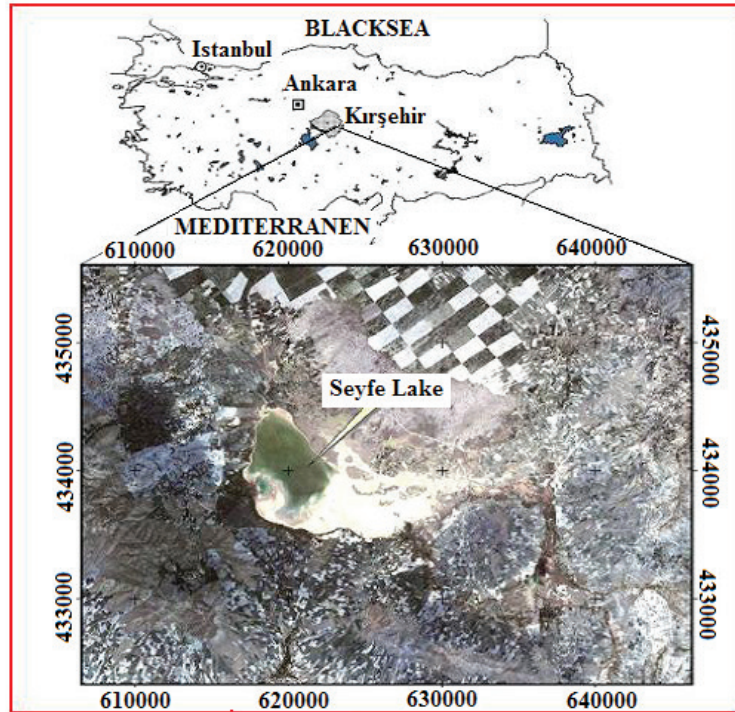


Figure 1- The location of the study area

Şekil 1- Çalışma alanının yeri

These are Malya (1), Eskidoğanlı (2), Seyfe (3), Badıllı (4-5), Yazıkınık (6-7), Budak (8-9), and Gümüşkümbet (10-11). Malya and Eskidoğanlı stations were situated in the Northern side of the lake. Seyfe, Badıllı, and Gümüşkümbet areas are situated in the Western side of the lake and Yazıkınık and Budak are situated in the Southern side of the lake. Monitoring network stations were chosen for determining the quality of lake water in study field. Data for these eleven water quality sampling station consist of numerous water quality parameters. The chosen parameters are given below: pH, water temperature (T), electrical conductivity (EC), dissolved oxygen (DO), total dissolved solids (TDS), ammonium nitrogen (NH₄-N), nitrate nitrogen (NO₃-N), nitrite nitrogen (NO₂-N), sulfate (SO₄), total phosphorus (PO₄-P), hardness (hard), total alkalinity (TA), biochemical oxygen demand (BOD), chemical oxygen demand

(COD), sodium (Na), potassium (K), calcium (Ca), magnesium (Mg), chloride (Cl), aluminum (Al), cadmium (Cd), lead (Pb), iron (Fe), copper (Cu), chromium (Cr), and zinc (Zn). Sampling and analysis for these parameters (except pH, T, EC, and DO) were determined in the Ankara Metropolitan Municipality Laboratory using standard methods for the Examination of Water and Wastewater (APHA 1992). T, EC, and DO were measured by a Wissenschaftlich-Technische-Werkstatlen 340i (WTW, Weilheim, Germany) field apparatus. The temperature and pH were measured with a pH meter WTW, electrical conductivity (TetraCon 325 WTW) and dissolved oxygen (oxygen meter, CelloX 325 WTW) were measured, respectively. Meters and probes were calibrated with standard solutions as suggested by the manufacturers. Most of parameters are expressed in milligrams per liter (mg L⁻¹), except pH, EC (µS cm⁻¹), temperature (°C).

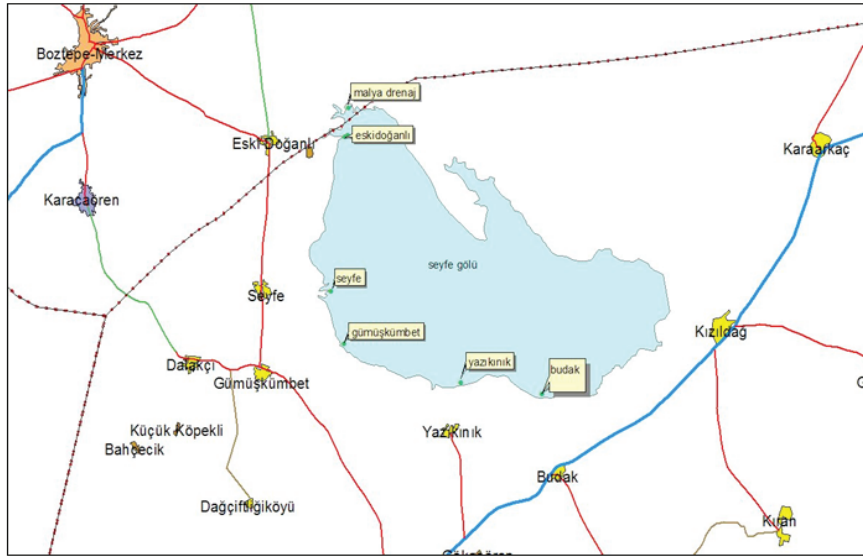


Figure 2- Location of the water sampling stations in the Seyfe Lake

Şekil 2- Seyfe gölündeki su örnekleme istasyonlarının yeri

2.3. Statistical analysis

We calculated descriptive statistics on all data sets as range, mean, standard deviations. Cluster analysis (CA) and principal component analysis (PCA) and factor analysis (FA) were applied to determine water quality. SPSS 16.0 version statistical program was used for descriptive statistics and multivariate data analysis. Eigenvalues measure the amount of the variation explained by each PC and will be largest for the first PC and smaller for the subsequent PCs. An eigenvalue greater than 1 indicates that PCs account for more variance than accounted by one of the original variables in standardized data (Athl 1996). Cluster analysis is very sensitive to outliers. Therefore outliers were identified using the methods outliers (DFBETA and DFBETAS methods). Outliers weren't evaluated.

3. Results and Discussion

The descriptive statistics concerning quality of water are presented in Table 1. The range, mean, and the standard deviation were shown for all parameters. When Table 1 is examined,

the 26 analyzed parameter values and standard deviation value range varies considerably. When analyzed with descriptive statistics related to water quality parameters EC, TDS, NO₃-N, Na, K, Mg, hardness, chloride and COD values seem to be very high standard deviation. These variables changed considerably in the lake water. It distressed in terms of irrigation water quality in the lake.

The cluster analysis provided sampled groups of similar stations and created a dendrogram (Figure 3). As results, three statistically significant clusters were obtained from 11 sampling stations of Seyfe Lake. Cluster I (sampling sites: Malya (1), Eskidoğanlı (2), Seyfe (3), Yazıkınık (6-7), and Budak (8-9)) were near agricultural areas, animal husbandry, and settlement areas, where the agricultural runoff to the lake and the domestic wastewater influenced the parameters. Cluster II (sampling sites: Badıllı (4-5)) corresponds to points in the lake, and Cluster III (sampling sites: Gümüşkumbet (10-11)) corresponds to remote locations that are relatively far from the lake.

Factor effects are given in Figure 4 according to eigenvalues. As shown in Figure, Cd, Cu, Cr and Pb are shown as stable in increasing direction. NO₃-N, NH₄-N, Fe, Al, Ca, EC, TDS, hard, BOI, KOI and Mg are shown in decreasing direction. Zn, K, SO₄, Na and Cl are shown as stable in decreasing direction. On other hand, NO₂-N, PO₄-P and TA are shown in decreasing direction.

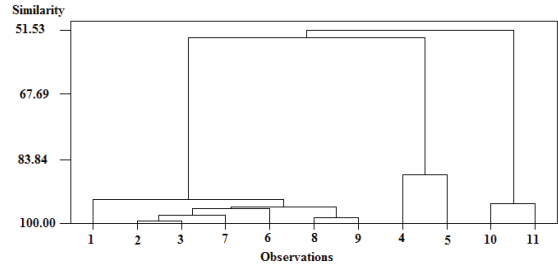


Figure 3- Dendrogram of sampling sites
Şekil 3- Örnekleme sitelerinin dendrogramı

Table 1- Descriptive statistics of the values of the samples and chemical analysis

Çizelge 1- Örneklere ait tanımlayıcı istatistikleri ve kimyasal analiz değerleri

Parameters	Range (Station No)		Mean	Standard deviation
	Minimum range	Maximum range		
T	14.68 (8)	21.98 (11)	19.22	2.82
pH	7.93 (6)	8.73 (4)	8.27	0.25
EC	63.0 (2)	7318,7 (11)	2371.82	2800.11
DO	8.19 (7)	11.08 (11)	9.08	0.87
TDS	434.67 (2)	52094.50 (11)	16702.26	19853.14
NH ₄ -N	0.42 (8)	4.86 (10)	1.64	1.22
NO ₂ -N	0.02 (5)	0.18 (8)	0.696	0.04
NO ₃ -N	9.31 (5)	276 (1)	53.53	76.94
PO ₄ -P	0.04 (7)	0.39 (1)	0.14	0.11
Na	1.80 (4)	4830 (5)	1314.51	1604.13
K	1(4)	250 (5)	50.74	77.29
Ca	74.08 (2)	1263.90 (11)	642.02	471.27
Mg	35.92 (2)	3901.56 (11)	977.14	1404.25
Hard	301.88 (2)	18077.87 (11)	5213.72	6608.30
TA	101.73 (5)	510.08 (11)	281.59	134.92
Cl	84.31 (2)	14663 (5)	3392.17	5307.15
SO ₄	45 (4)	4551 (5)	676.75	1301.88
COD	24 (2)	1382.67 (11)	206.21	395.53
BOD	10.67 (2)	88 (10)	36.85	28.51
Cd	0.001 (3)	0.003 (6, 7, 8)	0.002	0.0006
Cr	0.003 (2, 3, 4, 5)	0.035 (1)	0.003	0.0001
Cu	0.01 (2, 3, 4, 5)	0.006 (11)	0.0017	0.001
Zn	0.02 (1)	0.05 (11)	0.028	0.009
Al	0.016 (10)	0.40 (11)	0.12	0.09
Fe	0.003 (10)	0.30 (11)	0.084	0.08
Pb	0.008 (1, 3, 4, 5, 6, 7, 8, 9, 10, 11)	0.014 (2)	0.008	0.0019

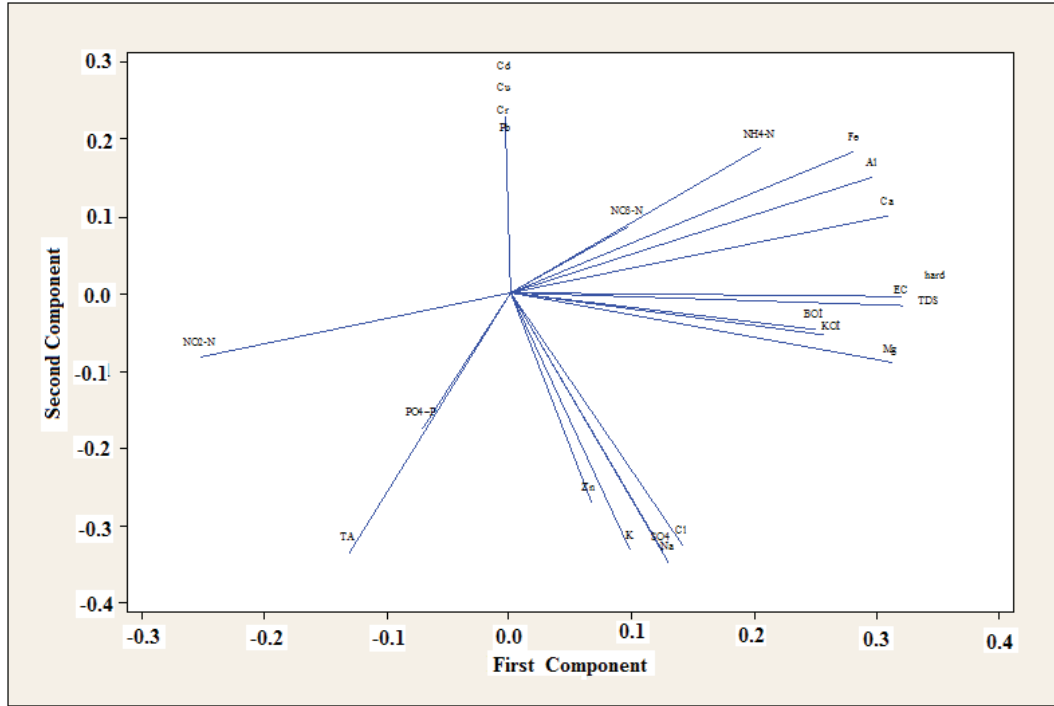


Figure 4- According to the eigenvalues of factors influence

Şekil 4- Özdeęerlere göre faktörlerin etki şekilleri

Malya and Eskidoęanlı are situated in the Northern side; Seyfe, Badıllı, and Gümüşkümbet are situated in the Western side; and Yazıkınık and Budak are situated in the Southern side of Seyfe Lake. Agricultural and domestic pollution (fertilizer, agricultural drainage water, irrigation water, livestock waste, and the polluted of settlement areas) affected all these stations. In particular, the main sources of Malya, Eskidoęanlı, and Seyfe were affected more than others.

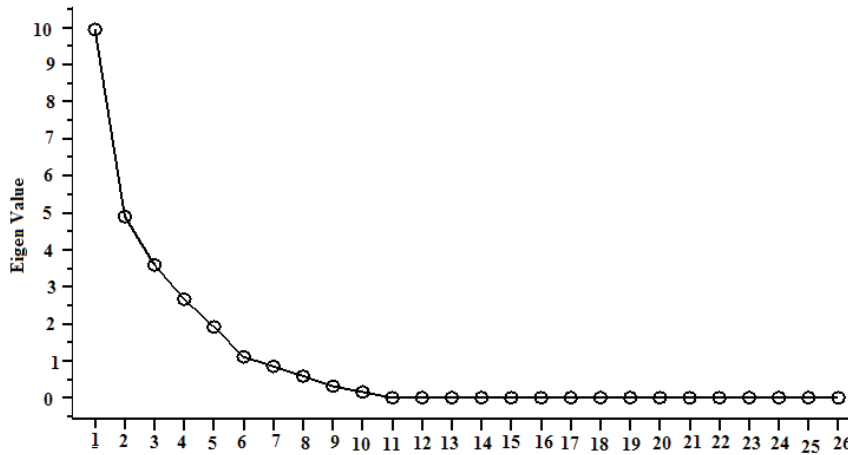
According to the results of the principal component analysis and factor analysis are shown in Table 2. The principal component analysis used Varimax rotation (Kuppusamy & Giridhar, 2006). Eigenvalue selection criteria were selected to explain the sources of variance as one and greater than one. The screen plots of Eigenvalues are shown in Figure 5. In Figure 5 are shown that six of Eigenvalues has over 1. Other Eigenvalues are below 1. Thus, a new

set of data is obtained. This may explain the variation of the data set with fewer variables. Total variance is very important in PCA analysis. The principal components are arranged according to size. The main component corresponds to each Eigenvalue. The variances of original variability and main components are equal. The total variance equals the sum of the Eigenvalues. The first component is the highest variance with linear component. The total variance is defined by the sum of the variance of the components. Eigenvalues belonging to the total variance explained before and after rotation are given in Table 2. Contributions to the total variance of groups with converted and unconverted Eigenvalues over 1 occurred at a rate of 92.72%. Eigenvalues indicate the degree of importance of one factor. Therewithal, the Eigenvalue that has the highest Eigenvalue makes the highest contribution to the variance (Shrestha & Kazama, 2007).

Table 2- The amount of variance explanation of components

Çizelge 2- Bileşenlerin varyansı açıklama miktarları

Component	Total variance explained before rotation			Total variance explained after rotation		
	Extraction sums of squared loading			Rotation sums of squared loading		
	Initial Eigen values			Eigen values		
	Total	% of variance	Cumulative (%)	Total	% of variance	Cumulative (%)
1	9.947	38.258	38.258	7.675	29.518	29.518
2	4.895	18.828	57.086	4.651	17.889	47.407
3	3.587	13.795	70.881	4.356	16.752	64.159
4	2.662	10.240	81.121	3.549	13.652	77.811
5	1.912	7.353	88.473	2.531	9.734	87.545
6	1.103	4.243	92.716	1.344	5.171	92.716

**Figure 5- Eigenvalues of principal components**

Şekil 5- Temel bileşenlerinin özdeğerleri

Six components obtained by the principal component analysis for the parameters are presented in Table 3. Factors are generally classified as “strong,” “moderate,” and “weak,” by Liu et al (2003). Also, he classified limits of their factors as >0.75 , $0.75-0.50$, and $0.50-0.30$, respectively. The first factor (F1) accounted 29.52% of the cumulative variance in the dataset. Hardness had the highest positive power. Mg, TDS, BOD, EC, and Ca had positive power. $\text{NH}_4\text{-N}$, DO, TA, COD, and T had a moderate positive loading. These results indicate that the lake water is contaminated. Hardness is evaluated as a source of physicochemical variability.

The region’s climate and geological characteristics play a very important role in shaping the hardness. Hardness, depending on the content of Ca and Mg, is known as a function of salt water. Total alkalinity and total hardness are known to be natural processes of the dissolution of soil components, in particular, the formation of calcium carbonate (Varol & Şen 2008). Yiğitbaşıoğlu (1993) reported in his study that Seyfe Lake and its surrounding region covers a large area, which is included the metamorphic series of the Paleozoic rocks.

Lake water is very important, especially in terms of local farmers because they use the lake water

Table 3- Results of the factor analysis

Çizelge 3- Faktör analizi sonuçları

Parameters	Component					
	1	2	3	4	5	6
EC	0.927*	0.268	0.187	-0.081	0.077	-0.103
TDS	0.930*	0.271	0.177	-0.078	0.071	-0.099
NH ₄ -N	0.733*	-0.358	-0.274	0.306	-0.126	0.356
NO ₂ -N	-0.342	-0.101	-0.321	0.266	-0.684	-0.190
NO ₃ -N	-0.146	-0.068	0.041	0.918*	-0.036	-0.195
PO ₄ -P	0.122	0.105	-0.035	0.822*	-0.044	0.119
Na	0.159	0.056	0.809*	0.522	-0.074	-0.034
K	0.027	-0.044	0.910*	0.354	0.022	0.033
Ca	0.786*	0.176	0.323	-0.140	0.160	-0.344
Mg	0.932*	0.350	0.041	-0.031	-0.006	0.004
Hard	0.944*	0.303	0.077	-0.038	0.015	-0.037
TA	0.597	0.247	-0.507	0.235	-0.340	0.141
Cl	0.125	-0.023	0.760*	-0.305	0.372	-0.335
SO ₄	0.034	0.040	0.942*	-0.016	0.020	0.028
COD	0.540	0.832*	-0.060	0.073	-0.014	-0.004
BOD	0.933*	0.202	0.133	-0.125	0.073	-0.171
Cd	0.306	0.063	0.068	-0.184	-0.856	0.096
Cr	-0.110	-0.098	0.124	0.935*	0.153	-0.079
Cu	0.439	0.842*	-0.130	0.230	-0.126	-0.004
Zn	0.141	0.691	0.261	-0.201	0.579	0.120
Al	0.236	0.958*	0.066	-0.077	0.009	-0.052
Fe	0.196	0.963*	0.013	-0.107	0.116	-0.057
Pb	-0.234	-0.030	-0.079	-0.160	0.125	0.839*
Ph	0.366	0.071	0.679	-0.354	0.310	-0.085
DO	0.664	0.477	-0.388	0.212	-0.022	0.136
T	0.528	0.163	0.211	0.225	0.713*	0.245

*, represent strong loadings

for irrigation. There is intense evapotranspiration in the region. Very high temperatures and a low amount of precipitation affect this region in July and August months. As a result, salinity and other adverse effects take a toll on the ecosystem of the lake, polluting it. The highest values of BOD and COD were observed at the Gümüşkümbet site (10-11). The maximum concentration of COD and BOD in the lake water may be because of the excess of the organic matter in the water. Organic matter in Lake can rise with the widespread use of fertilizers, chemicals, and domestic sewage. At the same time,

runoff from agricultural fields and organic deposits increased the organic matter in the lake.

Factor 2 (F2) stated 17.89% of the cumulative variance. Factor 2 contains a strong load of Fe, Al, Cu, and COD, whereas it contains a moderate load of Zn. The highest values of Fe, Al, Cu, Zn, and COD were observed at the Gümüşkümbet sampling site (11) due to local anthropogenic activities such as agricultural and domestic waste at this location. These variables were not at the level of pollution is reached. Therefore, the impact is less irrigation. However, Fe, Al, Cu, and Zn concentrations are

within the limits permitted by Turkish Standards (2004). F2 may be denoted as the 'heavy metal' parameters.

Factor 3 (F3) revealed 16.75% of the cumulative variance in the dataset and contain a strong load of SO_4 , K, Na, Cl and a moderate positive load of pH. This is on account of the human factor in the application of fertilizers and wastewater discharge into the lake from agriculture contributes to the pollution. Intensive agricultural activities have been reported around the lake (Kıymaz 2010).

Factor 4 (F4) accounted for 13.65% of the cumulative variance with a strong loading of Cr, nitrate, and total phosphorus. The concentration of $\text{NO}_3\text{-N}$ was obtained from 9.31 to 276 mg L^{-1} at the Badıllı (5) and Malya (1) sampling stations (Table 1). The highest and average values of $\text{NO}_3\text{-N}$ were above the permissible limit (20 mg L^{-1}) of the Turkish Norm values for irrigation water (Turkish Standards 2004). This factor includes point and non-point sources of contamination from household waste and agricultural fields. Ammonium sulfate fertilizers are widely used by farmers in agricultural areas and nitrates are more frequently used as the organic and inorganic fertilizers (Vega et al 1998). As a result, fertilizers can be transported through irrigation water and runoff into lakes. The amount of ammonium sulfate and nitrate in the lake has increased. Nitrate fertilizers in agricultural areas are the most common. In particular, pesticides and agricultural nitrate fertilizers containing nitrogen and phosphorus contribute to pollution. Sugar beet, cereals, barley, and maize are mainly grown in the lake basin. In addition, total phosphorus runoff from the agricultural areas causes pollution in the lake. The main source of nitrate is nitrogen. The sources of nitrogen are very diverse. These include geological resources, agricultural sources, and naturally decomposed organic matter (Madramootoo et al 1997).

Factor 5 (F5) explained 9.73% of the total variance. It contains a strong load on temperature and a moderate positive loading of Zn. However, it contains a strong negative load of Cd. This constituent is likely to be associated with

anthropogenic processes such as the application of micro-nutrients. Factor 6 (F6) explained only 5.17% of the cumulative variance. This factor is affected by Pb and it contains a weak load of $\text{NH}_4\text{-N}$. This factor is concerning to anthropogenic pollution from agricultural and domestic activities.

4. Conclusions

In the study, several multivariate data analysis methods were made use of assess changes in quality of the surface water of Lake Seyfe. Clustering methods allocated three clusters according to the water quality parameters provided from 11 sampling points. A sampling strategy that may reduce the number of sample points improved in the future will be able to make better designs for the areas studied. Factor analysis/principal component analysis on the status in water quality aided in the identification of the effective resources.

Water quality is constantly required to be monitored for changes especially degradations and pollutions. The main cause of degradation of the lake is the discharge of agricultural and household wastes. All stations are affected by intensive agricultural factors such as fertilizer, irrigation water, drainage water, and livestock waste. They are seriously negatively affected by wastes of residential areas close to the lake basin. Analysis of statistics resulted in six factors explained 92.72% of the cumulative variance in dataset. The first factor explained 29.52% of the cumulative variance, the second factor 17.89%, the third factor 16.75%, the fourth 13.65%, the fifth 9.73%, and the sixth 5.17% of the cumulative variances, respectively. Results reveal that sulfate, nitrate, total phosphorus, hardness, electrical conductivity, total dissolved solid, magnesium, potassium, sodium, calcium, biochemical oxygen demand, chemical oxygen demand, aluminum, iron, chromium, and lead were the most important parameters used to evaluate changes in water quality of the lake.

Water quality is considered the main factor for healthy of man and animals. Water quality is largely determined by natural processes and by anthropogenic

inputs. Ranging from source of pollution point such as from drainage pipes to non-point source pollution like pesticide and fertilizer runoff from farms, these pollutants and others have had adverse effects on the lakes. Adverse effects include reducing the water quality, contaminating soils, and damaging the lake ecosystems. This factors effected harmful repercussions on waterfowl, other wild species and livestock. One of the categories of chemical pollutants is organic contaminants. Also a result of synthetic fertilizers and pesticides are discovered by testing the tissue of plants and animals. This damage produces on the ecological imbalanced.

Multivariate statistical analysis manage numerous and complex data for decreasing the

variable number, and providing comprehensive data. In this study, multivariate statistical methods were expressed to be quite successful in determining the quality of water. Numerous variables were explained with a smaller number of variables. At the same time, these variables provide more useful information about the quality water of lake. This technique has been successfully applied for planning and monitoring in future.

Acknowledgements

This research was financially supported by the Ankara Water and Sewage Management Authority. We also thank to Mrs. Müfide Akgün, and Mr. Hasan Karakuş for their assistance in carrying out field.

Abbreviations and Symbols

T	water temperature, °C	K	potassium, mg L ⁻¹
EC	electrical conductivity, µS cm ⁻¹	Ca	calcium, mg L ⁻¹
DO	dissolved oxygen, mg L ⁻¹	Mg	magnesium, mg L ⁻¹
TDS	total dissolved solids, mg L ⁻¹	Cl	chloride, mg L ⁻¹
NH ₄ -N	ammonium nitrogen, mg L ⁻¹	Al	aluminum, mg L ⁻¹
NO ₃ -N	nitrate nitrogen, mg L ⁻¹	Cd	cadmium, mg L ⁻¹
NO ₂ -N	nitrite nitrogen, mg L ⁻¹	Pb	lead, mg L ⁻¹
SO ₄	sulphate, mg L ⁻¹	Fe	iron, mg L ⁻¹
PO ₄ -P	total phosphorus, mg L ⁻¹	Cu	copper, mg L ⁻¹
hard	hardness, mg L ⁻¹	Cr	chromium, mg L ⁻¹
TA	total alkalinity, mg L ⁻¹	Zn	zinc, mg L ⁻¹
BOD	biochemical oxygen demand, mg L ⁻¹	CA	cluster analysis
COD	chemical oxygen demand, mg L ⁻¹	PCA	principal component analysis
Na	sodium, mg L ⁻¹	FA	factor analysis

References

- Atlı H (1996). Uygulamalı Çok Değişkenli İstatistiksel Analiz. Akademi matbaası. Ankara
- AWWA-WPCF (1992). Standart methods for the examination of water and wastewater. 18 th Ed. Washington DC. American Public Health Association, American Water Works Association, and Water Pollution Control Federation, 137 pp.
- Boyacıoğlu H (2006). Surface water quality assessment using factor analysis. *Water SA* **32**(3): 389–393
- DSİ (2004). Kırşehir Seyfe Plain Hydrogeology Report. Kayseri, Turkey
- Giardino C, Brando V E, Dekker A G, Strombeck N & Candiani G (2007). Assessment of water quality in lake Gorda (Italy) using hyperion. *Remote Sensing of Environment* **109**: 183–195
- Kıymaz S (2010). Water resources management of Seyfe Lake wetlands and recommendation for resolve. *Journal of New World Sciences Academy* **5**(2): 174-186.
- Kuppasamy M R & Giridhar VV (2006). Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environmental International* **32**: 174–179

- Lee J Y, Cheon J Y, Lee K K, Lee S Y & Lee M H (2001). Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. *Journal of Environmental Quality* **30**: 1548–1563
- Liu CW, Lin K H & Kuo Y M (2003). Application of factor analysis in the assessment of groundwater quality in Blokfoot disease area in Taiwan. *Science of the Total Environment* **313**: 77–89
- Madramootoo C, Johnston W R & Willardson L S. (1997). Management of agricultural drainage water quality: Water Reports 13
- Mendiguchia C, Moreno C, Galindo-Riano M D & Garcia-Vargas M (2004). Using chemometric tools to assess anthropogenic effects in river water: A case study: Guadalquivir River (Spain). *Analytica Chimica Acta* **515**(1,5): 143–149
- Mendeş M (2011). Multivariate Multiple Regression Analysis Based on Principal Component Scores to Study Relationships between Some Pre- and Post-slaughter Traits of Broilers. *Tarım Bilimleri Dergisi* **17**(1): 77-83
- Ouyang Y, Nkedi-Kizza P, Wu Q T, Shinde D & Huang C H. Assessment of seasonal variations in surface water quality. *Water Research* **40**: 3800-3810
- 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**(4): 464–475
- Simeonov V, Simeonova P & Tsitouridou R (2004). Chemometric quality assessment of surface waters: two case studies. *Chemical and Engineering Ecology* **11**(6): 449–469
- Simeonova P, Simeonov V & Andreev G (2003). Environmetric analysis of the Struma River water quality. *Central European Journal of Chemistry* **2**: 121-126
- Singh K P, Malik A, Mohan D & Sinha S (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-a case study. *Water Research* **38**: 3980–3992
- Türkeş M (2007). Climate of southern part of the Middle Kızılırmak Sub-Region (Cappadocia District) and its vulnerability to desertification. *Ege University Journal of Geography* **14**: 75-99
- Turkish Standards 2004. Su kirliliği kontrol yönetmeliği (Water pollution control regulation). Turk Off Gaz. No: 25687
- Varol M & Şen B (2009). Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environment Monitoring Assessment* **159**: 543–553
- 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
- Wunderlin D A, Diaz M P, Ame M Y, Pesce S F, Hued A C & Bistoni M A (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). *Water Research* **35**: 2881–2894
- Yigitbaşıoğlu H (1993). Geomorphology of Seyfe Lake and its surroundings. Ankara University, Institute of Social Sciences. PhD Thesis (Unpublished), Turkey
- Zhou F, Liu Y & Guo H (2007). Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environment Monitoring Assessment* **132**: 1–13