Application of Cluster Analysis and Multivariate Statistical Techniques Associated with Water Quality Index to Evaluate Water Quality of Tigris River in Iraq

Mohammed A. Ibrahim
Al-Nahrain University
Ziad Tark Abd Ali
University of Baghdad
Haitham A. Hussein
Al-Nahrain University
Baghdad \Iraq

Abstract

The study comprised investigation of the seasonal and spatial variations in water quality of the Tigris River in Iraq for the period (2005-2012). Cluster analysis (CA) and principal component analysis (PCA) were applied to analyze the similarities among sampling sites to identify the source apportionment of pollution parameters in river. Three principal components were extracted explaining 88.64 % of the total variance and indicating that the main components affecting water quality are soil leaching and runoff associated with dissolved ions, domestic wastewater discharge and industrial effluents from point source. In addition to the multivariate statistical techniques, water quality index (WQI) was used for the assessment of water quality through an index number. The WQI of the river for the studied period was 79.7 units classified as fair quality, which requires treatment for drinking use and no treatment for irrigation purposes.

Keywords: **Cluster Analysis, Principal Components, Water Quality Index, Tigris River.**

1-Introduction

Quality of water is defined in terms of physical, chemical and biological parameters. Increasing levels of water pollution necessitate development of water quality indices that quantify and develop the quality of a given water body. The water quality index (WQI) is a single numeric score that describes the water quality condition at a particular time and location [1]. After the initial concept of the index method proposed by Horton in 1965 [2], many studies on water quality index were carried such as [3, 4, 5, 6, 7, 8, 9]. Cluster analysis (CA) and principal component analysis (PCA) are main techniques of multivariate analysis approach, which provides useful and accurate approach to understand the water quality of the studied area and identifying the pollution source. The CA analysis helps in grouping objects (cases) into

classes (clusters) on the basis of similarities within a class and dissimilarities between different classes [10]. Hence, the PCA is one of the most powerful and common techniques used for reducing the dimensionality of the large set of data without loss of information [11]. Objective of this study is to determine the seasonal and spatial variations of water qualit
y for the Tigris River in Iraq .

The Cluster Analysis and principal component analysis associated with water quality index are used to determine the similarities or dissimilarities between sampling sites, the influence of possible sources on the physicochemical parameters, source apportioning for the estimation of the contribution of possible sources on the concentration of determined parameters, and a global water quality status for the studied water region.

2Materials and Methods

2.1 Monitoring Area

A total of thirteen sampling stations for water quality monitoring were selected along the Tigris River in Iraq **Table1**. It originates from Turkey entering Iraqi border from north at Phishkhabour and discharge in south at Shat Al-Arab

Sampling Location	Governorate	Location No.	Distance (km)	Latitude	Longitude
Phishkhabour	Dohuk	S1	0	37° 5' 45.4" N	42° 2' 13.76" E
Al-Mosul Dam	Neynavah	S2	66.15	36° 37' 40" N	42° 49' 11.9" E
Al-Mosul	Neynavah	S3	42.4	36° 20.80' N	43° 8.417' E
Al-Sherqat	Salah Al-Din	S4	84.35	35° 35.23' N	43° 15.555' E
Tikrit	Salah Al-Din	S5	118.32	34° 36.33' N	43° 41.761' E
Theraa Dijla	Salah Al-Din	S6	114.58	33° 45' 40" N	44° 2' 14.46" E
A-Tarmiyah	Salah Al-Din	S7	10.18	33° 25.73' N	44° 23.987' E
Al-Muthana Bridge	Baghdad	S8	27.07	33° 25.28' N	44° 20.83' E
Al-Shuhadaa Bridge	Baghdad	S9	10.65	32° 20' 20 " N	44° 23' 18.1" E
Al-Azzizeyah	Waset	S10	79.69	32° 54.09' N	45° 4.031' E
Al-Kut	Waset	S11	78.52	32° 31.63' N	45° 46.99' E
Al-Umara	Misan	S12	148.28	31° 51.34' N	47° 8.618' E
Al-Qurnah	Al-Basrah	S13	98.92	31° 0.683' N	47° 26.301' E

Table 1 Sampling Site Locations

2.2 Analytical Procedures

Based on extensive study, 11 parameters were analyzed using data retrieved from (ministry of Water Resources-Environmental Studies Center). The selected parameters were monitored for twelve months through eight years from 2005 to 2012. The studied parameters are biochemical oxygen demand (BOD), fecal coliforms (Ecoli), chemical oxygen demand (COD), total dissolved solids (TDS), total hardness (TH), calcium (Ca), chlorides (CL), sulfates (SO4), electrical conductivity (Ec), temperature, and pH.

The chosen parameters have verified weight factors to calculate WQI in many literature [11, 12, 13]. In addition these parameters are potential contaminants to surface waters.

C2.3 Cluster Analysis (CA)

CA could group the objects (cases) into classes (clusters) on the basis of similarities within the class and dissimilarities between different classes. The classification of this method represents a convenient method for organizing large data set so that it can be understood more easily and information retrieved more efficiently [14]. In hierarchcal cluster analysis, the distance between samples is used as a measure of similarity. The Ward's method, using Euclidean distances as a measure of similarity, posses a grouping mechanism small space distorting effect uses more information about cluster content than other methods, and has been proved to be extremely powerful [15].

2.4 Principal Component Analysis (PCA)

Principal components (PCs) are uncorrelated variables, obtained by multiplying the original correlated variables with eigenvectors. Thus the PCs are weighted linear combinations of the original variables [16]. PCA attempts to extract a lower dimensional linear structure for the data set i.e. reducing number of variables by reducing the contribution of less significant variables [17].

PCA was performed to predict the significance of the studied parameters in monitoring stations. The analysis was based on the collected data of the mean annual values of the 11 water quality parameters. The PCs were subjected to varimax rotation generating varifactors.

2.5 Water Quality Index (WQI)

The concept of WQI is based on the comparison of the water quality parameters with respective regulatory standards and gives a single value to the water quality of the source, which translates the list of constituents and their concentrations present in the sample [18]. The significance of WQI can be appreciated as the water resources play a crucial role in the overall environment. The WQI was calculated by transforming the parameters having different units of measurements into a unit less sub-index values. This may be done by scaling each single parameter into (0-100) score [11], using weighting factors that reflect the importance of each parameter as indicator of the water quality. These sub-indices are then averaged a water quality value [12].

The equation for calculating WQI is [9]:

$$WQI = \frac{\sum_{i=1}^{n} C_i * P_i}{\sum P_i} \qquad (1)$$

Where:

n is the total number of parameters.

C_i is sub-index of parameter (i), a number from (0 to 100) having 100 as highest water quality.

P_i is the relative weight assigned to each parameter.

Pi value ranges from (1 to 4) with 4 assigned to a parameter that has the most important impact on the environment and value of 1 assigned to the parameter that has a smaller impact [12]. P_i values for each parameter are presented in **Table2.** Rating curves were selected as the most efficient method of transforming information on the individual determined concentrations to the same scale. The rating sub-index curves were based on the ratings proposed by [19,12, 13] as shown in **Figure1.** ..

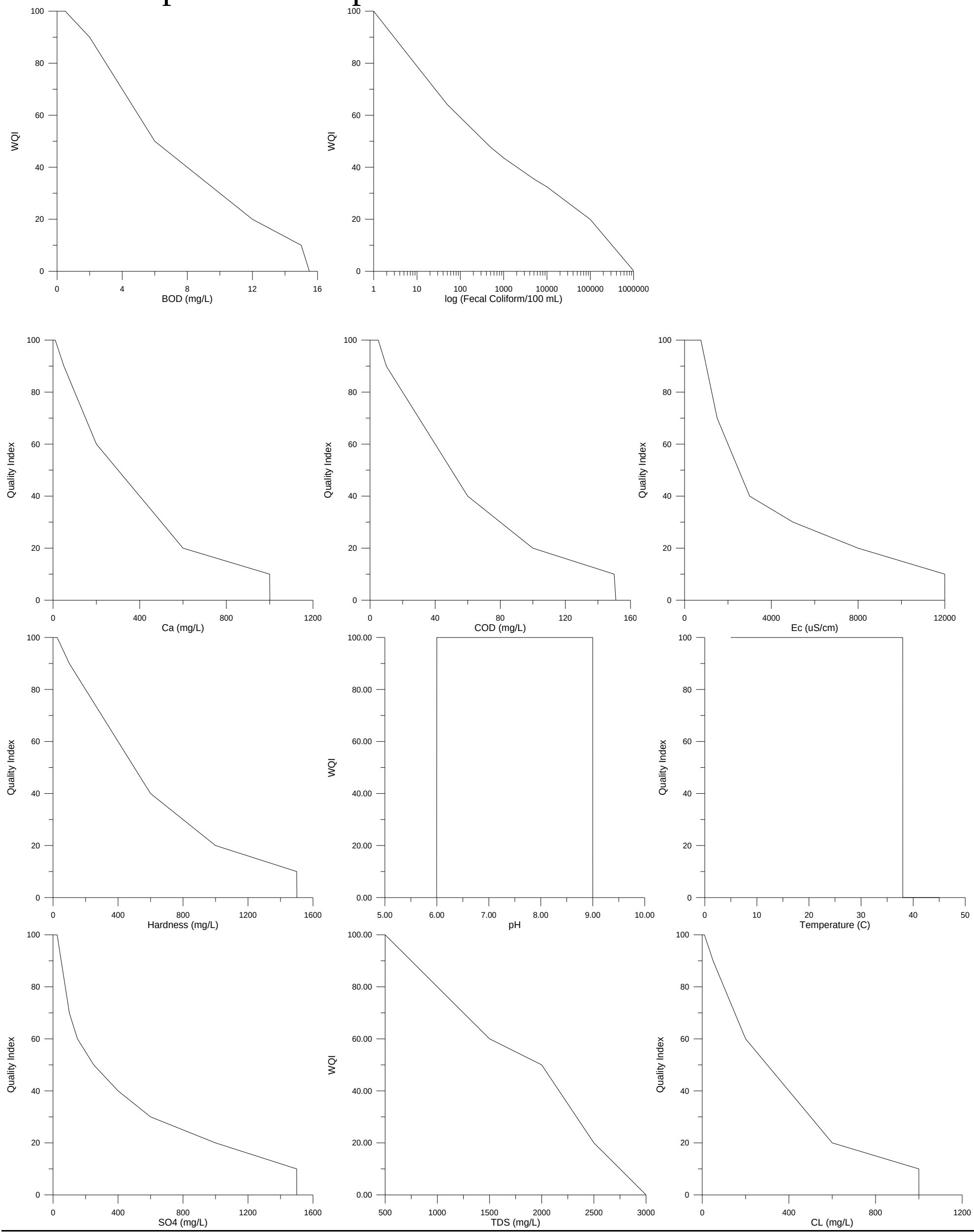


Figure 1 the Assigned Rating Curves for the Studied Parameters

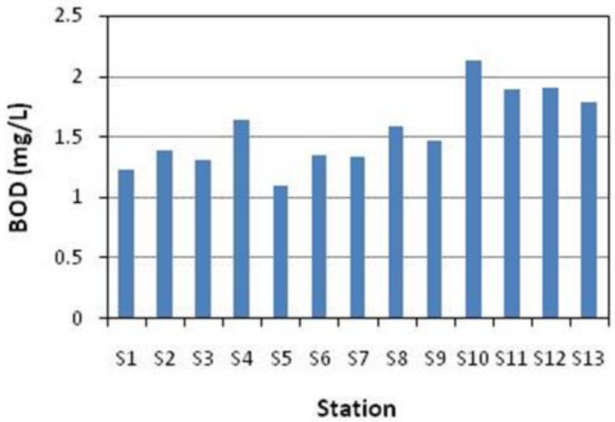
Parameter	BOD	Ecoli /100ml	COD	TDS	TH	Ca	CL	SO4	EC	Temp.	pH
Units	(mg/L)	CFU/100 ml	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	(mg/L)	µS/cm	C°	pH unit
Relative Weight (Pi)	4	3	4	3	2	1	1	2	2	1	1

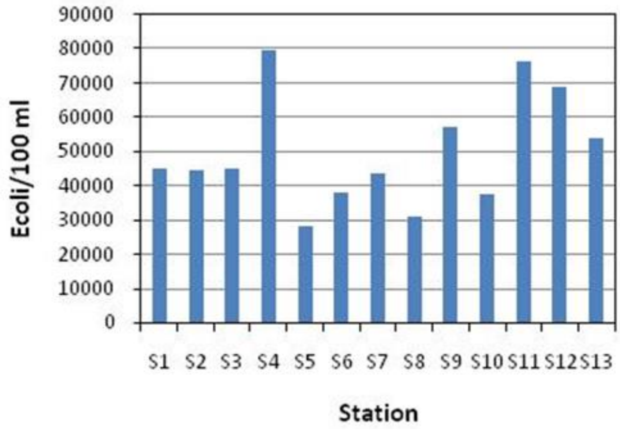
Table 2 Parameters Used in WQI Calculation and their Relative Weights [12]

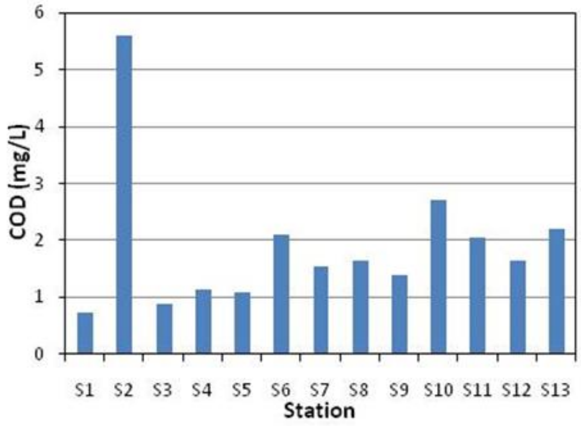
The WQI is ranked here by relating it to one of the following categories [9]; excellent for WQI in the range (95-100), good for range (80-94), fair for (65-79), marginal if the range (45-64), and poor when WQI value (0-44).

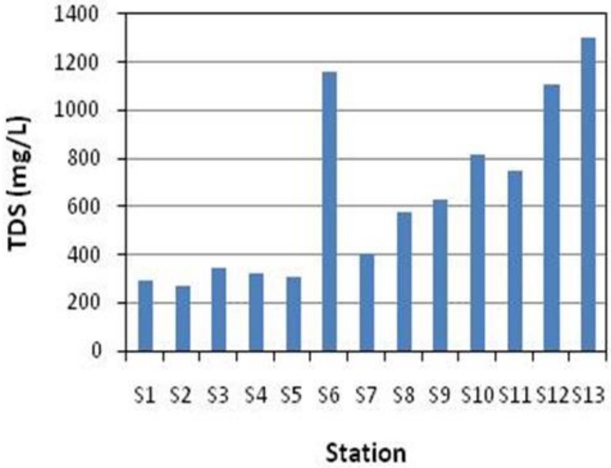
3. Results and Discussion

The total average results of the water quality variables are presented in **Figures 2, 3**









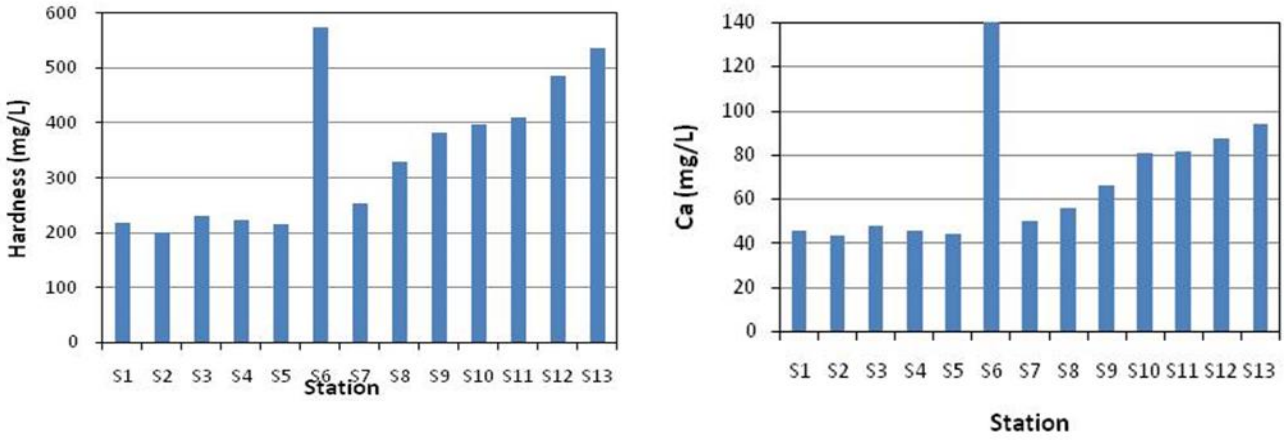
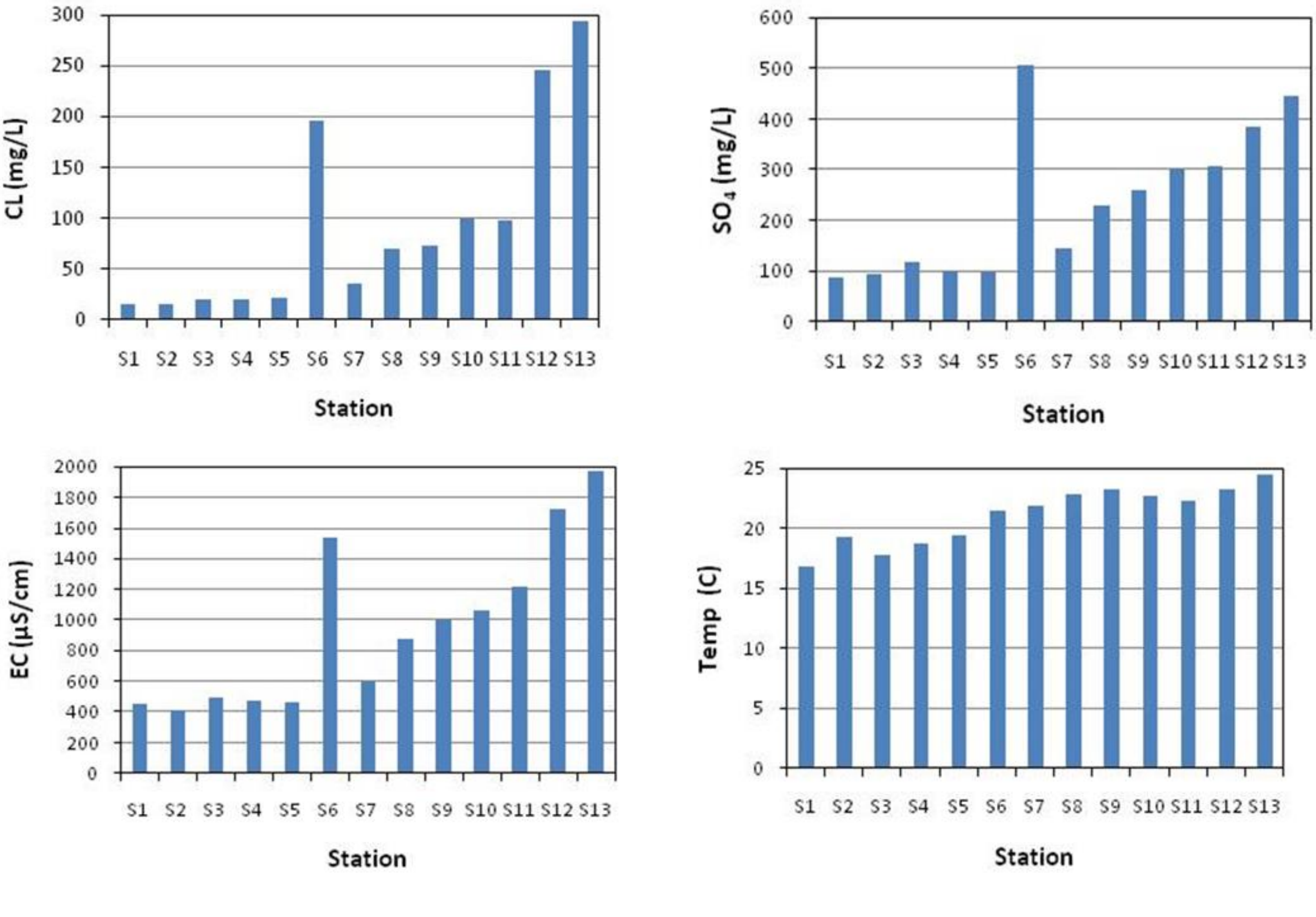


Figure 2 Spatial variations of BOD, Ecoli, COD, TDS, Hardness, and Ca at the Tigris river in Iraq



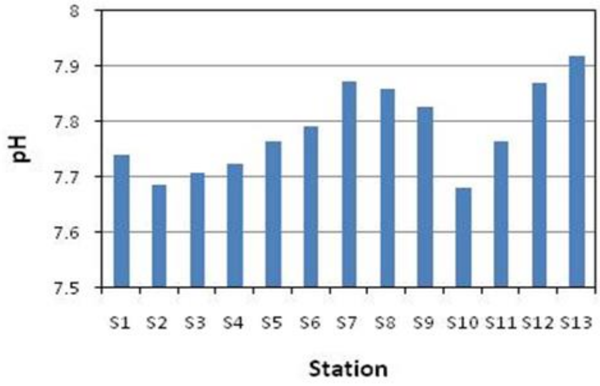


Figure 3 Spatial variations of CL, SO4, EC, Temp, and pH at the Tigris river in Iraq

3.1 Source Identification Using Principal Component Analysis (PCA)

The correlation matrix consisting of 11 water quality parameters is shown

in **Table3** A strong positive correlation between hardness, TDS, Ca, CL, SO_4, can be observed. The Ecoli is positively correlated with COD

The rotation of principal components was executed by the Varimax method with Kaiser Normalization. Eigen value one-criterion was adopted [13]. The results of PCA indicates three main controlling components with (Eigen value >1). Following the above criterion those components higher than 0.6 may be taken into consideration for the interpretation of the PCA [13] as shown in **Table4**. The scree-plot of the PCA is shown in **Figure4.** Three PCs are extracted, together explaining 88.64% of the variance information contained in the original data

Table (3) Correlation Matrix of Tigris River Water Parameters

	BOD	Ecoli	COD	TDS	TH	Ca	CL	SO_4	EC	Temp.	pH
BOD	1.000										
Ecoli	-0.761	1.000									
COD	-0.335	0.421	1.000								
TDS	-0.319	-0.693	-0.504	1.000							
TH	0.070	-0.362	-0.088	0.826	1.000						
Ca	-0.154	-0.105	-0.374	0.509	0.800	1.000					
CL	0.0725	-0.471	-0.47	0.632	0.720	0.762	1.000				
SO4	-0.169	-0.548	-0.031	0.969	0.936	0.660	0.640	1.000			
EC	0.190	-0.592	-0.241	0.880	0.918	0.675	0.760	0.904	1.000		
Temp.	0.349	-0.721	0.089	0.606	0.133	-0.253	0.156	0.436	0.312	1.000	
pH	0.355	-0.650	-0.204	0.406	0.038	-0.223	0.275	0.261	0.183	0.862	1.000

Table (4) PC Loadings (Varimax Normalized) of Tigris River Water Parameters

Parameter	PC 1	PC 2	PC 3
BOD	-0.025	0.646	0.461
Ecoli	-0.360	-0.820	-0.377
COD	-0.118	-0.069	-0.944
TDS	0.848	0.510	-0.081
TH	0.982	0.022	-0.010
Ca	0.846	-0.369	0.303
CL	0.783	0.087	0.406
SO_4	0.933	0.298	-0.109
EC	0.923	0.223	0.120
Temp.	0.157	0.926	-0.235
pH	0.050	0.869	0.056
Eigen value	5.67	2.64	1.44
% Total Variance	51.55	24.04	13.05
% Cumulative Variance	51.55	75.59	88.64

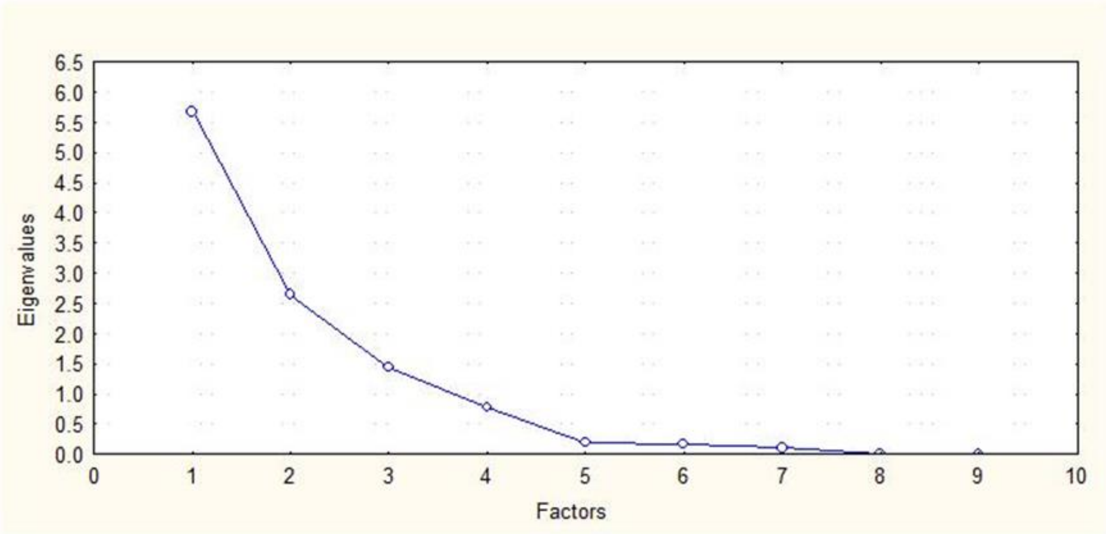


Figure (4) Scree-Plot for the PCA of Tigris River Water Parameters

PC1 explains 51.55% of the variance and is mainly contributed by TDS, hardness, Ca, CL, SO_4, and EC. These PCs represents soil leaching processes and contribution of dissolved ions in the river water quality

PC2 explains 24.04% of the variance and is mainly participated by BOD, Ecoli, Temperature and pH. and This PC can be interpreted as representing the domestic wastewater discharge activities.

PC3 explains 13.05% of the variance and consist of strong loading of COD. This PC can be interpreted as representing influences from point source, such as industrial effluents.

3.2 Spatial Similarity and Site Grouping Using Cluster Analysis (CA)

CA was applied to detect similarity groups between the sampling sites. The dataset was treated by the Ward's method of linkage with Euclidean distance as a measure of similarity.

The first cluster contained two sampling sites (S6, S10) corresponds to Theraa Dijla, Azziziyah (with a maximum distance of 0.1). The second cluster contained the sites (S2, S11) corresponds to Mosul Dam, Al-Kut Dam (with a maximum distance of 0.24). The third cluster contained the sites (S4, S9) corresponds to Al-Sharqat, Al-Shuhadaa Bridge (with a maximum distance of 0.3). The fourth cluster contained (S3, S7) corresponds to Al-Mosu, Al-Tarmiyah (with a maximum distance of 0.43). The lower distance in cluster gives a higher similarity between the groups in the same cluster as shown in **Figure5.**

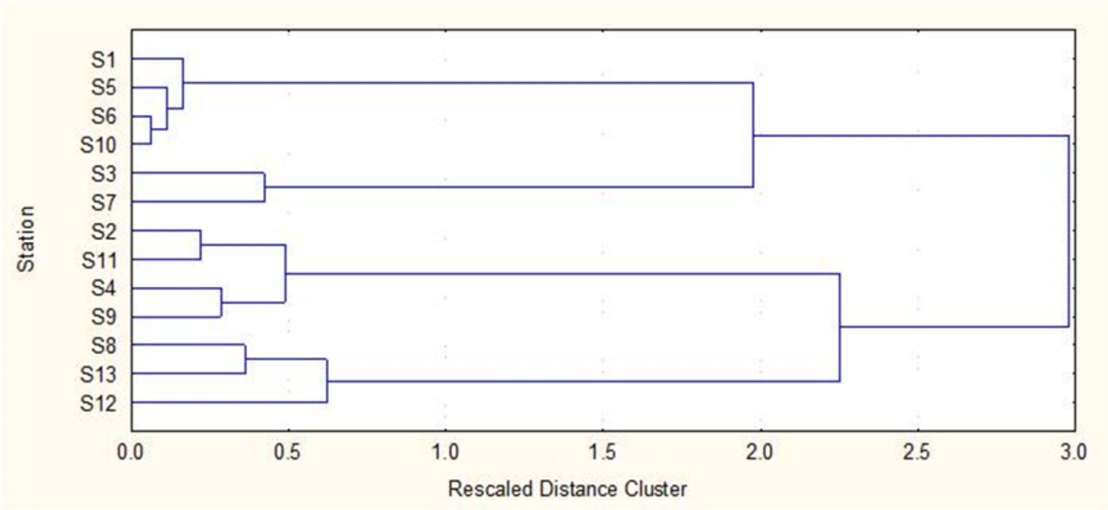


Figure (5) The Dendogram Showing the Clustering of Tigris River Water Quality Parameters

3.3 Water Quality Index (WQI)

The spatial WQI enabled classify the river system according to the mean of eight years period of study (2005-2012) as:

Stations (S1, S2, S3, S4, S5, S7, S8) classified as good (WQI = 80.06 – 86.1 unit) and stations (S6, S9, S10, S11, S12, S13) classified as fair (WQI = 70.2 – 79.2 unit) as shown in **Figure6.**

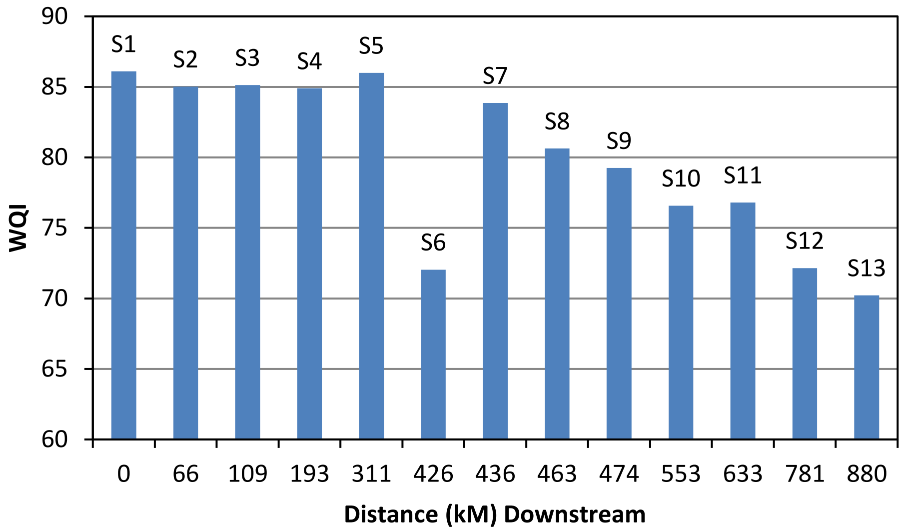


Figure (6) Spatial Variation of Mean WQI scores for Monitoring Stations in the Tigris River from 2005 to 2012

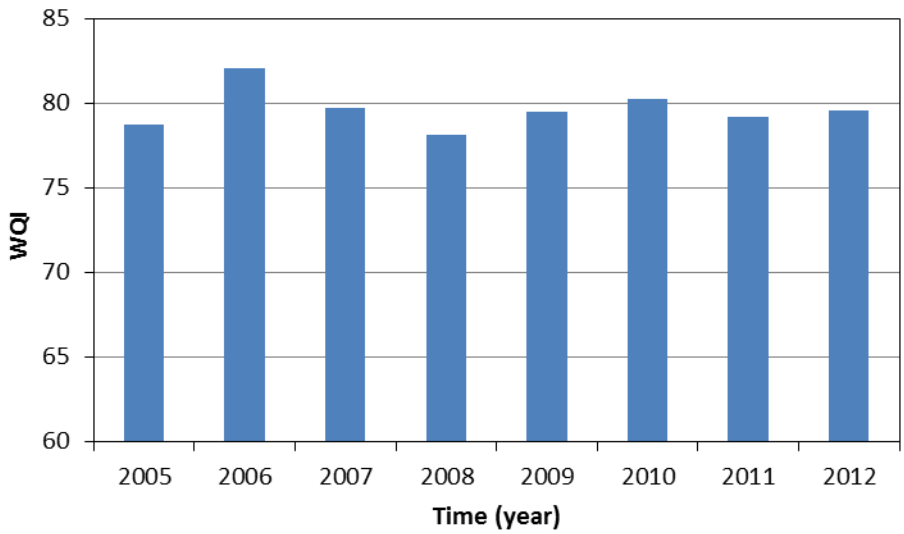


Figure (7) Temporal Variation of the Mean WQI scores for the 13 Monitoring Stations in the Tigris River

4. Conclusion

The study investigated how multivariate statistical techniques associated with quality index are effective in deriving the information from complex water quality dataset. PCA identified three major components responsible for water quality variations through the river; the first is related to soil leaching associated with dissolved ions, the second is associated with domestic wastewater discharge, while the third is related to industrial effluents from point source.

CA permitted the determination of spatial similarities between sampling

sites, four clusters were extracted according to cluster analysis.
WQI was used to evaluate spatial and seasonal changes in the water quality. It was found that mean WQI score is 79.7 units for the studied period, classified as fair quality, which needs treatment for drinking water use and no treatment for irrigation purposes

References

[1] Kaurish, F.W., and Younos,(2007), " Developing a Standarized Water Quality Index for Evaluating Surface Water Quality", Journal of the American Water Resources association (JAWRA), 43(2), 533-545.

[2] Horton, K, (1965), "An Index Number System for Rating Water Quality", Journal of Water Pollution Control Federation, 37(3), 300-306.

[3] Harkins, R.D, (1974), "An Objective Water Quality Index", Journal of Water Environment Federation, 46(3), 588-591.

[4] Bhargava, D.S., 1983, "Use of a Water Quality Index for River Classificationand Zoning of Ganga River", Journal of Environmental Pollution, 6, 51-67.

[5] Dinius, S.H., (1987), "Design of an Index of Water Quality", Water Resources Bulletin, 23(5), 1987.

[6] Dojlido , J., Raniszewki, J., and Woyciechowsk, J., (1994), " Water Quality Index Applied to Rivers Vistula River Basin in Poland", Environmental Monitoring and Assessment", 33, 33-42.

[7] Said, A., Stevens, D.K., and Sehlke, G., (2004), " An Innovative Index for Evaluating Water Quality in Streams", Environmental Management, 34(3), 406-414.

[8] Fullazzaky, M.A., (2009), "Water Quality Evaluation System to Assess the Branatas River Water", Water Resources Management", 23, 3019-3033.

[9] Sharma, D., and Kansal, A.,(2011), " Water Quality Analysis of River Yamuna Using Water Quality Index in the National Capital Territory, India", Applied World Science,1, 147-157.

[10] Rango, G., Deluca, M., and Ioele, G., (2007), "An Application of Cluster Analysis and Multivariate Classification Methods to Spring Water Monitoring Data", Microchemical Journal, 87, 119-127.

[11] Psce, S.F, and Wunderlin, D.A., (2000), "Use of Water Quality Indices to Verify the Impact of Cordoba City (Argentina) on Suquia River", Water Research, 34(11), 2915-2926.

[12] Kannel, P. Lee., S., and Kannel, S., (2007), "Application of Water Quality Indices and Dissolved Oxygen as Indicator for River Water Classification and Urban Impact Assessment", Environmental Monitoring and Assessment, 132, 93-110.

[13] Diaz, J. and Lopez, E., (2007), "Water Quality in the Rio River, Mexico: An Overview of the Last Twentieth Century", Water Resources Management, 21, 1797-1812.

[14] Everitt, B.S., Landau, S., Leese, M., and Stahl, D., (2011), Cluster Analysis, 5th ed. John Wiley.

[15] Kaufman, L., and Rousseeuw, P.J.,(1994), Finding Groups in Data: An Introduction to Cluster Analysis" , John Wiley.

[16] Basilevsky, A., (1994), Statistical Factor Analysis and Related Methods: Theory and Applications, John Wiley.

[17] Singh, K.P., Malik, A., and Sinha, S., (2005), "Water Quality assessment and Apportionment of Pollution Sources of Gomti River (India) Using Multivariate Statistical Techniques- a Case Study", Analytica Chemica Acta, 538, 355-374.

[18] Boyacioglu, H., (2010), "Utilization of Water Quality Index Method as a Classification Tool", Environmental Monitoring assessment", 167, 115-124.

[19] Lio, S., Lo, S., and Wang, S., (2004), " A Generalized Index for Taiwan", Environmental Monitoring Assessment, 96, 35-52.

تطبيقـــات التحليـــل العنقـــودي وتقنيـــات المتغيـــرات الأحصـــائية المتعـــددة مقترنة بمؤشر نوعية المياه لتقييم نوعية مياه نهر دجلة في العراق

محمد عبد الخالق ابراهيم
كلية الهندسة/جامعة النهرين
زياد طارق عبد علي
كلية الهندسة/جامعة بغداد
هيثم علاء حسين
كلية الهندسة/جامعة النهرين
بغداد / العراق

الخلاصة

الدراسة تشمل التحريات الموسمية والموقعية لنوعية مياه نهر دجلة في العراق للفترة (2005-2012). تم تطبيق التحليل العنقودي وتحليل المركب الأساسي وذلك لتحليل مدى التشابه بين مواقع اخذ العينات من اجل تحديد مصدر التلوث. بينت النتائج التحليلية للمركب الأساسي وجود ثلاثة مركبات تسبب حدوث التغيير في نوعية المياه، المركب الاول ناتج من السيح وغسل التربة مصحوبة بالأيونات الذائبة، الثاني ناتج من التلوث بمياه الصرف الصحي، والثالث ناتج من التلوث الصناعي للمنشأت الصناعية.

بالأضافة للطرق الاحصائية تم استخدام مؤشر (دليل) نوعية المياه. واظهرت نتائج التحليل لفترة الدراسة بأن مؤشر نوعية مياه دجلة 79.7 وحدة ويصنف النهر من خلالها بنوعية مقبولة نسبيا، وتبين القيمة حاجة مياه النهر الى المعالجة بالطرق التقليدية لأغراض الشرب ولاتحتاج الى معالجة لأغراض الري.