Journal of Environmental Protection, **2012, 3, 1629-1633**
http://dx.doi.org/10.4236/jep.2012.312180 Published Online December 2012 (http://www.SciRP.org/journal/jep)

Assessment of Water Quality of Euphrates River Using Cluster Analysis

Emad A. Mohammad Salah[1], Ahmed M. Turki[2], Eethar M. Al-Othman[2]

[1]Department of Applied Geology, University of Anbar, Ramadi, Iraq; [2]Department of Biology, University of Anbar, Ramadi, Iraq.
Email: ealheety@yahoo.com

Received September 15th, 2012; revised October 9th, 2012; accepted November 13th, 2012

ABSTRACT

Multivariate statistical method including cluster analysis (CA) was used to assess temporal and spatial variations in the water quality of Euphrates River, Iraq, for a period 2008-2009 using 16 parameters at 11 sampling sites. Hierarchical CA grouped the 8 months into three periods (I, II and III) and classified the 11 sampling sites into two groups (I and II) based on similarities of water quality characteristics. The temporal pattern shows that April has higher pollution level relative to the other months. Spatially, sampling site 7 (S7) has lower pollution level while the other sampling sites have higher pollution level. Thus, this study shows usefulness of cluster analysis method for analyzing and interpreting of surface water dataset to assess the temporal and spatial variations in the water quality parameters and the optimization of regional water quality sampling network.

Keywords: Cluster Analysis; Surface Water; Euphrates River; Iraq

1. Introduction

Globally, pollution of rivers and streams has become one of the most crucial environmental problems of the 20th century [1]. It is important to control water pollution, monitor water quality [2,3]. The application of different multivariate statistical techniques, such as cluster analysis (CA), principle component analysis (PCA) and factor analysis (FA) help to identify important components or factors accounting for most of the variances of a system [4,5] and interpretation of the complex databases offers a better understanding of the temporal and spatial variations in the identification of discriminate parameters that are of use in optimizing monitoring network [5,6,7]. Multivariate statistical techniques have been applied in water quality assessment and sources apportionment of water bodies over the last decade [3,9-20].

The aim of this study is to identify water quality parameters responsible for temporal and spatial variations in Euphrates river water using cluster analysis.

2. Materials and Methods

2.1. Study Area

The study area is located in Al-Anbar governorate between latitudes 33˚24'N - 33˚39'N and longitudes 42˚47'E - 43˚16'E, **Figure 1**. The area includes the largest urban centers in Al-Anbar governorate (Ramadi and Heet cites).

2.2. Sampling, Measuring and Analysis

Eleven sampling stations were chosen. Coordinates of sampling were listed in **Table 1**. The sampling process was carried out during 2008-2009. The number of samples are 16 for each sampling station, two samples per month.

Measuring and analysis was done upon 16 physical, chemical, microbiological parameters. These parameters were sampled monthly, **Table 2**. Electrical conductivity (EC), total dissolved solids (TDS) and dissolved oxygen were measured at the time of sampling in the field using portable EC meter, WTW model, and portable HANNA dissolved oxygen meter, H19142 model. Total suspended solids (TSS), turbidity, total hardness, biological oxygen demand(BOD), K^+, Na^+, Ca^{2+}, Cl^-, SO_4^{2-}, PO_4^{3-}, HCO_3^-, NO_3^-, and total coliform (*T. coli*) were determined according to APHA [21].

2.3. Cluster Analysis

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its object is to sort cases, data, or objects (events, people , things, etc.) into groups or clusters. The resulting clusters of objects should exhibit high internal (within-clusters) homogeneity and high external (between-clusters) heterogeneity [22]. Hierarchical CA, the most common approach, starts with each case in a separate cluster and joins clusters together step by step until only one cluster remains [23,24]. The

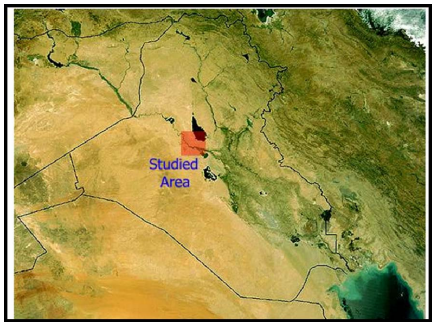

Figure 1. Study area location map.

Table 1. Sampling sites data.

Site Number	Site Name	Coordinates	
		Latitude (N)	Longitude (E)
1	$Heet_1$	33˚38'49"	42˚49'20"
2	$Heet_2$	33˚38'45"	42˚49'28"
3	$Heet_3$	33˚38'41"	42˚49'36"
4	$Heet_4$	33˚38'26"	42˚49'26"
5	$Heet_5$	33˚38'18"	42˚50'22"
6	$Heet_6$	33˚38'15"	42˚50'33"
7	$Heet_7$	33˚38'16"	42˚50'35"
8	Mohammady	33˚34'52"	42˚53'20"
9	$Ramadi_1$	33˚27'26"	43˚19'50"
10	$Ramadi_2$	33˚27'18"	43˚19'47"
11	$Ramad_3$	33˚26'16"	43˚16'52"

Table 2. Euphrates river water quality parameters.

Parameter	Symbol	Units
Turbidity	Turb.	NTU
Electrical Conductivity	EC	µm/cm
Total Dissolved Solids	TDS	mg/l
Total Suspended Solids	TSS	mg/l
Total Hardness	TH	mg/l
Dissolved Oxygen	DO	mg/l
Biological Demand Oxygen	BOD	mg/l
Potassium	K	mg/l
Sodium	Na	mg/l
Calcium	Ca	mg/l
Choride	Cl^-	mg/l
Sulfate	SO_4^{2-}	mg/l
Phosphate	PO_4^{3-}	mg/l
Bicarbonate	HCO_3^-	mg/l
Nitrate	NO_3^-	mg/l
Total Coliform	*T. coli*	MNP/100 ml

Euclidean distance usually gives the similarity between two samples, and a distance can be represented by the difference between analytical values from the samples [25]. The squared Euclidean distance (D^2) between location I and location II is calculated from normalized values as Follows :

$$D^2 = \left(Z_{DO1} - Z_{DO2}\right)^2 + \left(Z_{BOD1} - Z_{BOD2}\right)^2 + \cdots \quad (1)$$

where Z_{DO1} and Z_{DO2} are the normalized values of DO at locations 1 and 2. Similarly, Z_{BOD1} and Z_{BOD2} are similar values of BOD.

The results of the application of the clustering technique are best described using a dendogram or binary tree. The dendogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data [5,26]. In this study, hierarchical CA was performed on the normalized dataset using Ward's method with squared Euclidean distances as a measure of similarity. The Ward's method uses an analysis of variance (ANOVA) to evaluate the distance between clusters to minimize the sum of squares of any two clusters at each step. Both temporal and spatial variations of water quality were determined from hierarchical CA using linkage distance. Cluster analysis requires variables to conform to normal distribution. The normality of the data distribution was analyzed by one sample Kolmogorov-Smirnov test. The cluster analysis should be data standardization (mean = 0; variance = 1). The standardization tends to increase the influence of variables whose variance is small and reduce the influence of those whose variance is large [27]. This will also minimize the effects of scale of measurement of data. All the mathematical and statistical calculations were done by statistica 7 software.

3. Results and Discussion

3.1. Temporal Similarity and Period Grouping

An initial exploratory approach involved the use of hierarchical cluster analysis on standardized log—transformed data sorted by season. Temporal CA generated a dendogram as shown in **Figure 2** grouping 8 months into three clusters. Cluster I comprised April and the cluster II included May and June, while the cluster III consisted from the rest of months (November, January, February and March). The cluster III, approximately corresponding to the wet season in Iraq. **Figure 2** shows that the temporal patterns to water quality were not purely consistent with the dry/wet seasons. Among the monitoring months, April has the highest pollution level and the other months (November, December, January, February, March, May and June) have the lowest pollution level. The temporal variation in physical, chemical and microbiological

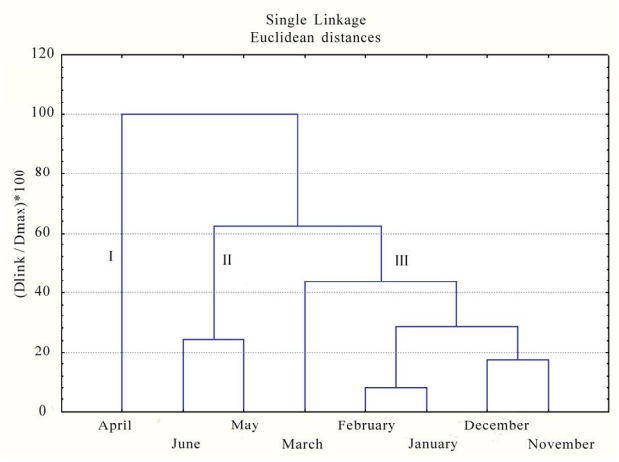

Figure 2. Dendogram of temporal clustering of sampling periods.

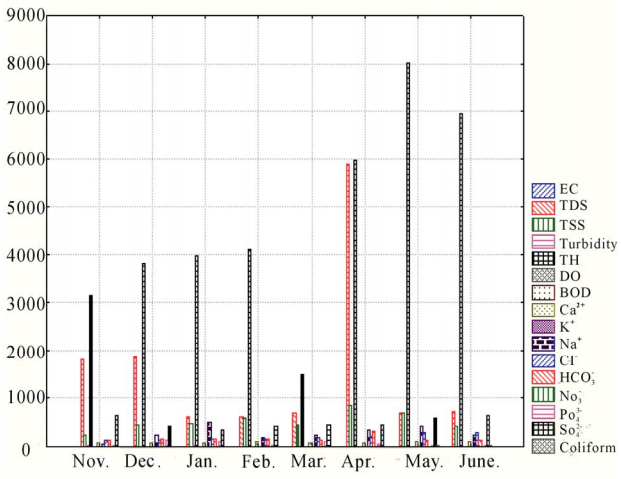

Figure 3. Temporal variation of water quality parameters at Euphrates river.

parameters level (**Figure 3**) demonstrated that April has highest level of pollution. The high pollution level in April is attributed to that the highest level of total dissolved solids (TDS) was reported in this month. High concentration of TDS was reported in April in the study area [28].

3.2. Spatial Similarity and Site Grouping

In this study, sampling sites classification was performed by the use of cluster analysis (z-transformation of the input data, Euclidean distance as similarity measure and Ward's method of linkage) based on the standardized mean of 16 measured parameters. With regard to dendogram, the sampling sites were grouped into two statistically significant clusters, **Figure 4**. Grouped sites under each cluster can be seen in **Figure 4**. Cluster I included sampling site 7 (S7). Cluster II comprised the sampling sites 1 - 6, and 8 - 11. Among the sampling sites, site 7 (S7) has lowest pollution while the other sites (1 - 6 and 8 - 11) have the highest pollution level. This result in good agreement with the variation in water quality parameters measured in the sampling sites as shown in **Figure 5**. Among the mean concentrations, most parameters were found high at sampling sites (1 - 6 and 8 - 11) and less in Site 7 (S7).

The results showed that the CA technique is useful in classification of river water in the study region and the number of sampling sites and associated monitoring costs can be reduced without missing much information. This result was in accordance with results of many studies carried out in other rivers [7,11-13,17,29,30].

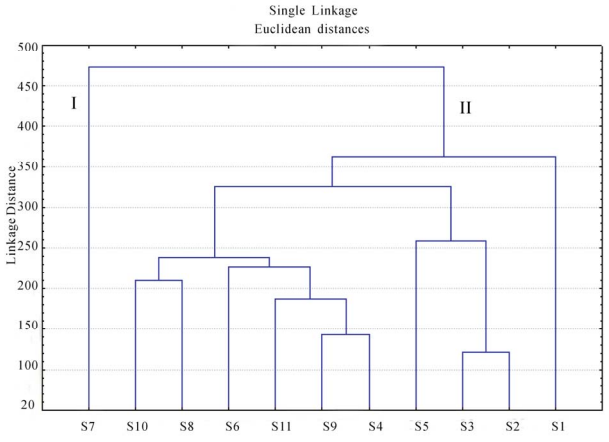

Figure 4. Dendogram of spatial clustering of sampling sites.

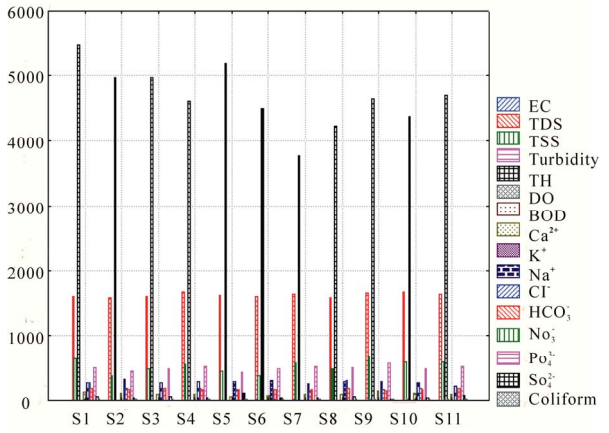

Figure 5. Spatial variation of water quality parameters at Euphrates river.

4. Conclusion

In this study, cluster analysis was applied to dataset for Euphrates River, Iraq. The results of this study show the importance and usefulness of cluster analysis of large and complex databases to obtain better information concerning the surface water quality. Hierarchical CA grouped the 8 months into three clusters and classified 11

sampling sites into two clusters based on the similarity sites of water quality parameters. The temporal pattern shows that April has high pollution level comparison with the rest of months. The spatial pattern shows that the sampling site 7 (S7) has lowest level of pollution while the other sampling sites have highest pollution level. Based on the information obtained, it is possible to design an optimal future sampling strategy which could reduce sampling frequency, number of sampling sites and associated sampling costs.

REFERENCES

[1] D. Otieno, "Determination of some Physico—Chemical Parameters of the Nairobi River, Kenya," *Journal of Applied Science Environmental Management*, Vol. 12, No. 1, 2008, pp. 57-62.

[2] W. Dixon and B. Chiswell, "Review of Aquatic Monitoring Program Design," *Water Research*, Vol. 30, No. 9, 1996, pp. 1935-1948. doi:10.1016/0043-1354(96)00087-5

[3] K. Singh, A. Malik, D. Mohan and S. Sinha, "Multivariate Statistical Techniques for the Evaluation of Spatial and temporal Variations in Water Quality of Gomti River (India)—A Case Study," *Water Research*, Vol. 38, No. 18, 2004, pp. 3980-3992. doi:10.1016/j.watres.2004.06.011

[4] Y. Ouyang, P. Nkedi-Kizza, Q. Wu and C. Huang, "Assessment of Seasonal Variations in Surface Water Quality," *Water Research*, Vol. 40, No. 20, 2006, pp. 3800-3810. doi:10.1016/j.watres.2006.08.030

[5] S. Shrestha and F. Kazama, "Assessment of Surface Water Quality using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan," *Environmental Modeling & Software*, Vol. 22, No. 4, 2007, pp. 464-475. doi:10.1016/j.envsoft.2006.02.001

[6] P. Simeonova, V. Simeonova and G. Andreev, "Water Quality Study of the Struma River Basin, Bulgaria (1989-1998)," *Central European Journal of Chemistry*, Vol. 1, No. 2, 2003, pp. 136-212.

[7] E. Fataei and S. Shiralipoor, "Evaluation of Surface Water Quality Using Cluster Analysis: A Case Study," *World Journal of Fish and Marine Sciences*, Vol. 3, 2011, pp. 366-370.

[8] D. Wunderlin, M. Diaz, M. Ame, S. Pesce, A. Hued and M. Bistoni, "Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality, A Case Study: Suquia River Basin (Cordoba—Argentina)," *Water Research*, Vol. 35, No. 12, 2001, pp. 2881-2894. doi:10.1016/S0043-1354(00)00592-3

[9] J. Grande, J. Borrego, J. Morales and M. De La Torre, "A Description of Show Metal Pollution Occurs in the Tinto—Odiel Rias (Huelva-Spain) through the Application of Cluster Analysis," *Marine Pollution Bulletin*, Vol. 46, No. 4, 2003, pp. 475-480. doi:10.1016/S0025-326X(02)00452-6

[10] C. Iscen, Ö. Emiroglu, S. Ilhan, N. Arslan, V. Yilmaz and S. Ahiska, "Application of Multivariate Statistical Techniques in the Assessment of Surface Water Quality in Uluabat Lake, Turkey," *Environmental Monitoring and Assessment*, Vol. 144, No. 1-3, 2008, pp. 269-276. doi:10.1007/s10661-007-9989-3

[11] A. Alkarkhi, A. Ahmad and A. Easa, "Assessment of Surface Water Quality of Selected Estuaries of Malaysia: Multivariate Statistical Techniques," *Environmentalist*, Vol. 29, No. 3, 2009, pp. 255-262. doi:10.1007/s10669-008-9190-4

[12] I. Gupta, S. Dhage and R. Kumar, "Study of Variations in Water Quality of Mumbai Coast through Multivariate Analysis Techniques," *Indian Journal of Marine Sciences*, Vol. 38, No. 2, 2009, pp. 170-177.

[13] Q. Zhang, Z. Li, G. Zeng, J. Li, Y. Fang, Q. Yuan, Y. Wang and F. Ye, "Assessment of Surface Water Quality Using Multivariate Statistical Techniques in Red Soil Hilly Region: A Case Study of Xiangjiang Watershed, China," *Environmental Monitoring and Assessment*, Vol. 152, No. 1-4, 2009, pp. 123-131. doi:10.1007/s10661-008-031-y

[14] S. Thareja and P. Trivedi, "Assessment of Water Quality of Bennithora River in Karnataka through Multivariate Analysis," *Nature and Science*, Vol. 8, No. 6, 2010, pp. 51-56.

[15] A. Ato, O. Samuel, Y. Oscar, P. Moi and B. Akoto, "Mining and Heavy Metal Pollution: Assessment of Aquatic Environments in Tarkwa (Ghana) Using Multivariate Statistical Analysis," *Journal of Environmental Statistics*, Vol. 1, No. 4, 2010, pp. 1-13.

[16] S. Shivani, S. Anulool, M. Negi and P. Tandon, "Evaluation of Effect of Drains on Water Quality of River Gomti in Lucknow City Using Multivariate Statistical Techniques," *International Journal of Environmental Sciences*, Vol. 2, No. 1, 2011, pp. 1-7.

[17] E. Fataei, "Assessment of Surface Water Quality Using Principle Component Analysis and Factor Analysis," *World Journal of Fish and Marine Sciences*, Vol. 3, No. 5, 2011, pp. 159-166.

[18] S. Yerel and H. Ankara, "Application of Multivariate Statistical Techniques in the Assessment of Water Quality in Sakarya River, Turkey," *Journal Geological Society of India*, Vol. 79, 1, 2012, pp. 89-93. doi:10.1007/s12594-012-0019-x

[19] A. Mustapha, "Identification of Anthropogenic Influences on Water Quality of Jakara River, Northwestern Nigeria," *Journal of Applied Sciences in Environmental Sanitation*, Vol. 7, No. 1, 2012, pp. 11-20.

[20] E. Singovszka and M. Balintova, "Application Factor Analysis for the Evaluation Surface Water and Sediment Quality," *Chemical Engineering Transactions*, Vol. 26, 2012, pp.183-188. doi:10.3303/CET1226031

[21] American Public Health Association, "Standard Methods for the Examination of Water & Wastewater," 20th Edition, Washington DC, 1998.

[22] K. McGarial, S. Cushman and S. Stafford, "Multivariate Statistics for Wildlife and Ecology Research," Springer, New York, 2000. doi:10.1007/978-1-4612-1288-1

[23] J. Lattin, D. Carroll and P. Green, "Analyzing Multivariate Data," Duxbury, New York, 2003.

[24] J. McKenna, "An Enhanced Cluster Analysis Program with Bootstrap Significant Testing for Ecological Community Analysis," *Environmental Modeling and Software*, Vol. 18, No. 3, 2003, pp. 205-220. doi:10.1016/S1364-8152(02)00094-4

[25] M. Otto, "Multivariate Methods," In: R. Kellner, J. M. Mermet, M. Otto and H. M. Widmer, Eds., *Analytical Chemistry*, Wiley-VCH, Weinheim, 1998.

[26] B. Tabachnick and L. Fidell, "Using Multivariate Statistics," Harper Collins College Publishers, New York, 1996.

[27] C. Liu, K. Lin and Y. Kuo, "Application Factor Analysis in the Assessment of Groundwater Quality in the Blackfoot Disease Area in Taiwan," *Science of the Total Environment*, Vol. 313, No. 1-3, 2003, pp. 77-89. doi:10.1016/S0048-9697(02)00683-6

[28] E. Al-Heety, A. Turky and E. Al-Othman, "Physico-Chemical Assessment of Euphrates River between Heet and Ramadi Cities, Iraq," *Journal of Water Resource and Protection*, Vol. 3, No. 11, 2011, pp. 812-823. doi:10.4236/jwarp.2011.311091

[29] K. Singh, A. Malik, D. Mohan and S. Sinha, "Water Quality Assessment and Apportionment of Pollution Sources of Gomti River (India) Using Multivariate Statistical Techniques: A case Study," *Analytica Chimica Acta*, Vol. 538, No. 1-2, 2005, pp. 355-374. doi:10.1016/j.aca.2005.02.006

[30] Reghunath, T. Murthy and B. Raghavan, "The Utility of Multivariate Statistical Techniques in Hydro-geochemical Studies: An Example from Karnataka, India," *Water Research*, Vol. 36, No. 10, 2002, pp. 2437-2442. doi:10.1016/S0043-1354(01)00490-0

