Interpretation of Water Quality parameters for Tigris River within Mosul City by Using Principal Components Analysis

Mazin N. Al-Sanjari[1] , Mus'ab A. Al-Tamir[2]
[1]*College of Environmental Science and Technology, [1]University of Mosul, Mosu , Iraq*
College of Engineerin/ Civil Department , [2]University of Mosul , Mosu , Iraq

(Received 16 / 2 / 2008 , Accepted 29 / 5 / 2008)

Abstract

Principal components analysis (PCA) were applied on the physical, chemical and biological water quality parameter from Tigris river within Mosul city, to give simpler and more easily interpretation results for the evaluation of these parameters. The PCA produced five significant main components explain more than 78% of the variance, Namely, seasonal effects, Mosul reservoir effects, agricultural and storm water effects, geological effect, and disposed wastewater effect; that represent 31.19%, 18.6%, 12.56%, 9.62% and 6.12% respectively of the total variance of water quality in Tigris river. Moreover, each parameter described with more than 50% in these five components. Finally the results of PCA reflected a good look on the water quality monitoring and interpretation of the surface water.

Key words: Principle components analysis, Tigris river, Water quality, Mosul, Iraq.

Introduction

The quality of water is identified by its physical, chemical and biological properties. The particulate problem in case of water quality monitoring is the complexity associated with analysis a large number of measured variables (Bayacioglu, 2006). The data sets contain rich information about the behavior of the water body. The classification, modeling and interpretations of monitoring data are the most important steps in the assessment of water quality.

Water quality parameters interact with each other to define the resource water quality. Many researchers treated water quality parameters individually by describing the seasonal variability and their causes. While it is difficult to interpret all the parameters patterns in combinations (Shihab, 1993).

In recent years many studies have been done using principal components analysis in the interpretation of water quality parameters, (Lohani and Todino, 1984) utilized principal components technique to provide a quick analytical method for the water quality of Chao Phraya river in Thailand. (Shihab, 1993) used this technique in order to describe the variation in water quality in Al_Mosul dam reservoir and the regulating lake to reduce the number of water quality parameters needed for monitoring the lake water. He noted that (90%)of the seasonal variation in the water quality of the reservoir are exhibited in only three parameters are Algae, Electrical conductivity (EC.), and Total Solids (TS). He referred that these parameters responsible for the main variability in water quality, While the periodically monitoring for insignificant parameters can be reduced; or special studies can be conducted on them when needed.

(Mazlum et al., 199٩) determined factors that caused variations in water quality at the Ağaçköy monitoring stations on the Porsuk Tributary in the Sakarya river basin by using principal component analysis (PCA). Researchers referred that PCA is more reliable than factor analysis and it is a pure mathematical technique without any assumption.

Also PCA has been successfully applied to sort out hydrogeological and hydrogeochemical processes from commonly collected ground water quality data (Jayakumar and Siraz, 1997), (Salman and Abu Ruka'h, 1998), (Praus, 2005), (Olobaniyi and Owoyeni, 2006).

(Iyer et al., 2003) developed a statistical, model which is based on the PCA for coastal water quality data from the Cochin coast in south west India, which explain the relationships between the various physicochemical variables that have been monitored and environmental conditions effect on the coastal water quality.

In this study multivariate principal components analysis (PCA) are being used in order to interpret and describe the variation in water quality of Tigris river within Mosul city.

Methodology

Samples were collected from five stations through Tigris river within Mosul city (Fig.1); monthly samples were collected from these positions from Sep. 2000 to Jun. 2001, So 50 samples were taken through this period, The samples were collected from just under water surface, for each sample 17 tests (parameters) were measured these parameters including: Temperature, pH, Dissolved Oxygen (DO), Total Hardness (TH), CO_2, Na^{+1}, K^{+1}, Ca^{+2}, Mg^{+2}, (NO_3^{-1}), (PO_4^{-3}), Chlorophyll(A), Chlorophyll(B), Chlorophyll(C), Biological Oxygen Demand (BOD5), Fecal Coliform (FC) and Total Plate Count (TPC), So the number of total parameters tested in laboratory work is 850 parameter.

All tests were done according to standard method presented by the American Public Health Association APHA (1985). Descriptive statistics of the data set are presented in (Table 1).

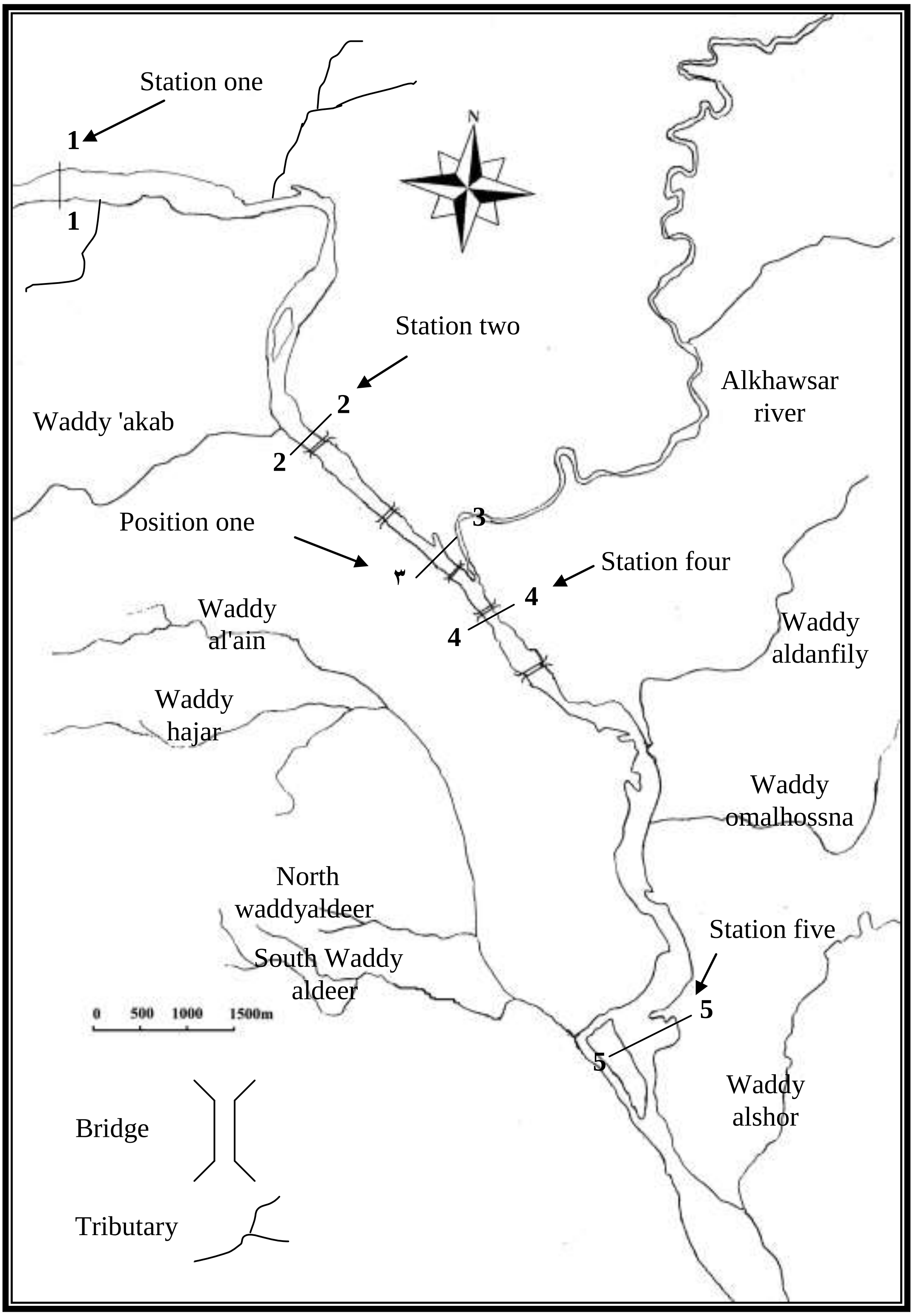

Fig. 1: Tigris River map within Mosul city with sampling stations.

Table 1: Descriptive statistics of the data.

Description Parameter	No.	Minimum	Maximum	Mean	Std. Deviation	Variance
Temp.(^oC)	50	11.9	20.05	15.949	2.680304	7.18403
pH	50	7.12	8.05	7.686	0.20056	0.040224
DO(mg/l)	50	5.4	9.8	7.589	1.125552	1.266866
CO_2(mg/l)	50	1.2	13.64	6.0112	2.920033	8.526594
TH mg/l as $CaCO_3$	50	207	413	290.426	53.64477	2877.762
Na^{+1} (mg/l)	50	13.43	37.67	28.8506	7.066165	49.93069
K^{+1} (mg/l)	50	2.05	9.11	5.4766	1.927956	3.717015
Ca^{+2}(mg/l)	50	55.7112	112.8252	80.67703	15.13946	229.2032
Mg^{+2}(mg/l)	50	12.15	41.553	21.65422	6.903418	47.65719
NO_3^{-1}(µg/l)	50	0.09275	1.15799	0.210174	0.144425	0.020859
PO_4^{-3} (µg/l)	50	0.03	0.25774	0.109632	0.069768	0.004868
BOD_5(mg/l)	50	0.2	4.3	1.703	0.936963	0.877899
TPC(Cell/ml)	50	155	1875	944.38	508.3257	258395.1
FC (Cell/ml)	50	14	1960	458.97	555.3805	308447.5
Chlorophyll A (mg/l)	50	0.13	1.86	0.5382	0.358191	0.128301
Chlorophyll B (mg/l)	50	0.12	1.84	0.5272	0.408392	0.166784
Chlorophyll C (mg/l)	50	0.16	5.43	1.2944	1.250132	1.562829

Data Processing

The 850 parameters obtained from the laboratory analysis were used as variables inputs for Principal Components Analysis (PCA), PCA performed by using the SPSS 10 computer program. Since water quality parameters had different magnitudes and scales of measurements so the data were standardized to produce a normally distribution of all variables (Davis, 1973).

In the standardization, the raw data were converted to unit less form of zero mean and a variance of one, by subtracting from each variable the mean of data set and dividing by standard deviation.

From the standardized covariance or correlation matrix of the data the initial factor solution were extracted by the multivariate principal components extraction, then a number of PC were selected from the initial according to their Eigenvalues and scree diagram. Orthogonal rotation of the selected initial components to terminal factor solutions was done by Kaiser's equamax schemes that attempt to achieve simple structure with respect to both the rows and columns (Kleinbaum, 1988).

DISCUSSION

Principal Components Analysis was performed on standardized matrix of the raw data in which a water samples is described by seventeen physical, chemical and biological parameters. This technique aims to transform the observed variables to a new set of variables (PC) which are uncorrelated and arranged in decreasing order of importance so that to simplify the problem. Table 2 represented the determined initial PC and its Eigenvalues and percent of variance contributed in each PC, Fig. 2 Show the scree plot of the Eigenvalues for each component.

Table 2: Explains initial components

Components	Initial Eigenvalues		
	Total	% of Variance	Cumulative%
1	5.30281	31.193	31.193
2	3.16363	18.60959	49.80259
3	2.135629	12.56253	62.36512
4	1.636821	9.628358	71.99348
5	1.050948	6.182045	78.17552
6	0.939958	5.529162	83.70468
7	0.708137	4.165512	87.8702
8	0.562182	3.306956	91.17715
9	0.529713	3.115957	94.29311
10	0.278359	1.637408	95.93052
11	0.239071	1.406301	97.33682
12	0.192579	1.132815	98.46963
13	0.108652	0.639127	99.10876
14	0.077947	0.458514	99.56727
15	0.053336	0.313738	99.88101
16	0.015217	0.089514	99.97053
17	0.00501	0.029472	100

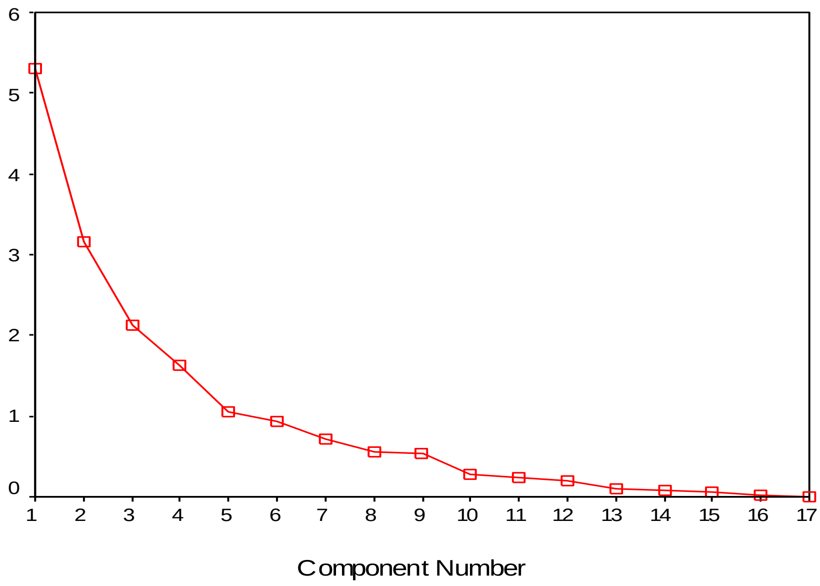

Fig. 1: Scree plot graph for components with its eigenvalues

Eigenvalues accounts and scree plot showed that the first five PC is the most significant components which represent 78.175% of the variance in water quality of Tigris river 31.19% by PC1, 18.61% by PC2, 12.58% by PC3, 9.62 by PC4, and 6.18 by PC5; in addition it have Eigenvalues of more than one.

Components loading and communalities for each variable in five selected components before equamax rotation were explained in Table 3a; and after equamax rotation in Table 3b.

Table 3a: Unrotated component matrix with its communalities.

	Components					
Variables	1	2	3	4	5	Communalities
Chlorophyll B	0.785721	0.322125	0.159084	0.079446	-0.16408	0.966
Chlorophyll C	0.752144	0.260402	0.409416	0.011045	-0.2868	0.629
PO_4^{-3}	0.72107	0.159218	-0.1295	0.057167	0.258021	0.934
Mg^{+2}	-0.71429	0.295779	0.243878	0.193624	0.196497	0.716
TH	-0.68509	0.091413	0.631027	-0.1997	0.231894	0.785
K^{+1}	0.664329	0.532701	0.047506	-0.22361	0.194894	0.970
CO_2	-0.63009	0.513602	-0.24528	0.007887	-0.2539	0.924
NO_3^{-1}	0.578434	0.082823	-0.25157	-0.15583	0.322904	0.733
pH	-0.43087	0.243221	0.373606	0.350056	-0.34938	0.533
TEMP.	-0.16805	0.930989	-0.26416	-0.0075	-0.03733	0.632
TPC	-0.26492	0.749088	-0.07742	0.109144	0.338428	0.778
DO	0.604011	-0.66115	0.307262	0.118767	0.151812	0.815
FC	-0.49209	-0.50827	-0.20659	0.147183	-0.19628	0.845
Ca^{+2}	-0.43491	-0.08154	0.718081	-0.41741	0.195406	0.884
CLOR.A	0.470197	0.331926	0.578059	-0.28586	-0.3123	0.780
Na^{+1}	-0.00693	0.018871	0.314276	0.740467	0.360698	0.603
BOD_5	0.382031	0.129527	0.203823	0.70774	-0.10348	0.764

Table 3b: Rotated component matrix with its communalities.

	Components					
Variables	1	2	3	4	5	Communalities
TEMP.	0.958645	0.183338	-0.08732	0.07597	-0.01424	0.966
DO	-0.88921	0.142454	-0.07397	0.243799	0.240042	0.629
TPC	0.790882	-0.02518	0.199483	0.211372	0.230555	0.934
CO_2	0.752583	-0.17027	0.040766	-0.40395	-0.15901	0.716
Chlorophyll A	-0.06541	0.900393	0.154495	0.015249	-0.07463	0.785
Chlorophyll B	-0.17403	0.879001	-0.18976	0.131219	0.16544	0.970
Chlorophyll C	-0.05445	0.730429	-0.35096	0.286562	0.194626	0.924
FC	-0.20476	-0.56209	-0.0533	-0.4808	-0.10709	0.733
TH	0.148647	-0.07886	0.934672	-0.25855	0.027978	0.533
Ca^{+2}	-0.12919	0.090769	0.92751	-0.13793	-0.14008	0.632
Mg^{+2}	0.463831	-0.26906	0.538903	-0.29387	0.262625	0.778
NO_3^{-1}	-0.06606	0.13438	-0.26602	0.66179	-0.04612	0.815
pH	0.268481	0.133036	0.214462	-0.63404	0.302019	0.845
K^{+1}	0.206575	0.602089	-0.11474	0.629985	0.010552	0.884
PO_4^{+31}	-0.082	0.30267	-0.34505	0.616438	0.185789	0.780
Na^{+1}	-0.04263	-0.10783	0.12389	0.006213	0.865296	0.603
BOD_5	-0.06729	0.299673	-0.31685	-0.08951	0.716335	0.764
Extraction Method: Principal Component Analysis. Rotation Method: Equamax with Kaiser Normalization.						

Communalities provide an index to the efficiency of the reduced set of components and degree of contribution of each variable in the selected five components. In our case communalities showed that all the variables have been described to an acceptable levels in the selected components with more than 50% for the less one of them which is the pH.

Table 4 shows the correlation matrix of components of equamax rotated five PC. It note that there are no correlation between components, this is, each components represent a discrete unit from others.

Table 4: Components score covariance matrix

Components	1	2	3	4	5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

Components loading (correlation coefficients), Which measure the degree of closeness between the variables and the PC. The largest loading either positive or negative, suggests the meaning of the dimensions; positive loading indicates that the contribution of the variables increases with the increasing loading in dimension; and negative loading indicates a decrease (Jayakumar and Siraz, 1997).

In general, component's loading larger than 0.6 may be taken into consideration in the interpretation, in other words, the most significant variables in the components represented by high loadings have been taken into consideration in evaluation the components (Mazlum et al.,1999).

Principal Components Interpretations

Principal Component I, has a high loading of temperature, DO, TPC and CO_2 and explain 31.19% of the total variance (Table 3) This component can be ascribed to the variation of natural atmospheric conditions, from temperature and rainfall that effect to the gases concentrations and bacteria growth etc., so this component represent seasonal effects upon water quality.

Principal Components II, explain 18.61% of the total variance, including Chlorophyll (A), Chlorophyll (B) and Chlorophyll (C). This component reflect the effect of Mosul reservoir in Tigris river, because of its direct and main effect upon the algae growth (Al-Tayar, 1988) and (Al-Nima et al., 2000).

Principal Component III, have 12.56% of the total variance in Tigris river water and includes total hardness, Ca^{+2}, and Mg^{+2}. Since the concentrations of TH, Ca^{+2} and Mg^{+2} in ground water in the region is much greater than in Tigris river water and the flow of ground water is toward the river. Also the geological strata of the river basin is constituted cyclic deposition of marl, limeston, and gypsum with sandston and siltston at its upper part that represent the major sources of Ca^{+2} and Mg^{+2} (Davis and Dewiest, 1966), (Al-Rawi, et al., 1990) and (Al-Tamir, 2005). So this component can be ascribed to the intrusion of ground water into river water system and the effect of geological information of the river basin.

Principal component V, This component has high loading in the NO_3^{-1}, pH, K^{+1} and PO_4^{-3}, Represent 9.628% of the total variance. It can indicate to the effect of drainage of agricultural and storm water from the river basin that washing out the top soil with its different impurities with its chemical fertilizer that contain the nitrate, phosphate and potassium on its structure.

Principal Component IV, account for 6.81% of the total variance. This component is highly correlated with Na^{+1} and BOD_5; this component can be reflect the wastewater effluents from the domestic and industrial and its organic load disposed to the river from Mosul city. As it well known many wastewater effluents disposing directly into Tigris river within Mosul city without any treatment. These effluents affects upon the variation of the BOD_5 and many parameters of water quality like Na^{+1} and Cl^{-1} and etc.(Environmental research center, 2002) and (Al-Jahssany, 2003).

In summary the five extracted Principal Components representing five different processing are:

- Seasonal effects.
- Mosul reservoir effects and Algae growth.
- Agricultural drainage and storm water effects.
- Geological and ground water effects of the river basin.
- Wastewater pollution from domestic and industrial and its organic load.

Conclosion

From the 17 components in Table 2 the first five components are sufficient to explain the monitoring area. These components explain more than 78% of the total variance of the original data set in Tigris river. Moreover, the first five selected principle components explained more than 50% of the variance of each quality variable (see communalities in Table 3a and 3b).

Principal components analysis of water quality data for Tigris river showed that seasonal effects, Mosul reservoir effects, agricultural wastes and storm water effects, geological and ground water of the river basin effects and domestic and industrial wastewater discharges and its organic loads are caused the main variation in water quality of the Tigris river in Mosul city. The results of PCA reflected a good look on the water quality monitoring and interpretation of the surface water.

References

Al-Jahssany, N. K., 2003. The negative reflections of domestic and industrial water discharges of Mosul city upon Tigris river water quality, M. Sc. Thesis, College of biological science, University of Mosul, Mosul, Iraq, 116p.

Al-Nima, B. B., Nassouri, G. A. and Dabbagh, A.G., 2000. Impact of rainfall scarcity on Tigris river water quality, Journal of Environmental Devlopment research, Vol. 3, no. 2, pp. 78-94.

Al-Rawi, S. M., Al-Azzo, S.I. and Abbawi, S. A., 1990. Hydrogeochemical evaluation of groundwater in some parts of Mosul city and suitability for irrigation, SDRC conference 18-20 March 1990, University of Mosul, Iraq, pp.224-238.

Al-Tayar, T. A., 1988. Effect of Saddam Dam construction on water quality and water treatment plants, M.Sc. Thesis College of Engg., Civil Dept., University of Mosul, Mosul, Iraq, 113p.

Al-Tamir, M. A., 2005. water quality Index for a group of wells in northwest of Mosul city, Journal of Sci. and Tech. Vol.16, No.2, Geology special issue, pp.27-40. (Arabic)

Bayacioglu, H.,2006. Surface water quality assessment using factor analysis, Water SA, Vol.32, No. 3, July, PP. 389-394.

Davis, S. N. and Dewiest, R. J. M., 1966. Hydrogeology, John wiley and sons, new york, 463p.

Davis, J.C., 1973. Statistics and data analysis in geology, John Wiley and Sons, inc., Newyork, 550p.

Environmental research center, 2002. A study submitted to the Mosul government, Unpublished.

Iyer, C.S.; Sindhu, M.; Kulkarni, S.G.; Tambe, S.S. and Kulkarni,B.D., 2003. Statistical analysis of the physico-chemical data on the coastal water of cochin, J. Environ. Monit., 5, pp.324-327.

Jayakumar, R. and Siraz, L., 1997. Factor analysis in hydrogeochemistry of Coastal aquifers-preliminary study, Environmental Geology, 31(3/4)June, pp.174-177.

Khattab, M.F., 2003. Factor analysis of water quality I Hit-Cubaisa area, west of Iraq, Iraqi Journal for Earth science, Vol. 4, No. 1, pp. 110-98.

Kleinbaum, D.G.; Kupper, L.L. and Muller, K.E., 1988. Applied Regresion Analysis and Other Multivariable Methods, PWS-KENT Publishing Comp., 2nd ed., USA., 718p.

Lohani, B.N. and Todino, G., 1984. Water Quality Index for Chao Pharya River. Journal of ASCE, Environmental Division, Vol. 110, No. 6, pp. 1163-1176.

Mazlum, N; Ozer, A. and Mazlum, S., 1999. Interpretation of Water Quality Data by Principal Components Analysis, J. of Engineering and Environmental Science, 23, pp. 19-26.

Olobaniyi, S.B. and Owoyemi, F.B., 2006. Chemical Facies of groundwater in the Deltaic Plain Sands Aquifer of Warri, |Western Niger Delta, Nigeria, AJST, Vol. 7, No. 1, June, pp. 73-81.

Praus, P.,2005. Water Quality Assessment Using SVD-based Principal Component Analysis of Hydrological Data, Water SA, Vol. 31, No. 4, pp. 417-422.

Salman, S.R. and AbuRuka'h, Y.H., 1998. Multivariate and principle component Statistical Analysis of contamination in urban and agricultural Soils from north Jordan, Environmental Soils from north Jordan, Environmental Geology , 38(3) September, pp.265-270.

Shihab, A.S., 1993. Application of Multivariate Method in the Interpretation of Water Quality Monitoring Data of Saddam Dam Reservoir, Confidential.

تفسير متغيرات نوعية مياه نهر دجلة في مدينة الموصل باستخدام تحليل المكونات الأساسية

مازن نزار السنجري[1] و مصعب عبد الجبار التمر[2]

[1] *كلية علوم البيئة وتقاناتها ، جامعة الموصل ، الموصل ، العراق*

[2] *قسم الهندسة المدنية ، جامعة الموصل ، الموصل ، العراق*

(تاريخ الاستلام: ١٧ / ٢ /٢٠٠٨ ، تاريخ القبول: ٢٩ / ٥ / ٢٠٠٨)

الملخص

أجري تحليل المكونات الأساسية على المعاملات الفيزيائية والكيميائية والبيولوجية لخصائص نوعية ماء نهر دجلة وذلك لتبسيط تفسير هذه الخصائص وبقيمها. نتج عن عملية تحليل المكونات الأساسية خمس مكونات رئيسة تمثل أكثر من ٧٨% من التغاير في مياه نهر دجلة هذه المكونات هي: التأثيرات الفصلية وتأثير الخزن في بحيرة سد الموصل وتأثير المطروحات الزراعية والسيح وتأثير جيولوجية حوض النهر وتأثير مياه الفضلات المطروحة من المدينة إلى النهر وبنسب مقدارها ٣١,٩% و ١٨,٦% و ١٢,٥٦% و ٩,٦٢% و ٦,١٢% على الترتيب. هذا وتشارك المعاملات النوعية في هذه المكونات الخمسة بنسبة لا تقل عن ٥٠% لكل منها. ويمكن الاعتماد على تقنية تحليل المكونات الأساسية في عملية متابعة وتفسير نوعية المياه السطحية.

الكلمات الدالة: تحليل المكونات الأساسية، نوعية الماء، نهر دجلة، الموصل، العراق.