Assesement of the Spatial Variation of River Tigris Chemistry in Iraq by Discriminant Analysis

Sawsan M. Ali | University of Baghdad, College of Science, Geology dept. Jadiriya, Baghdad, Iraq.

ABSTRACT *Multivariate cluster analysis and discriminant analysis were applied to investigate and assess the spatial variations of some hydrochemical parameters along eight stations on the River Tigris in Iraq. These parameters are:pH,TDS,TH,Ca,Mg,Na,K,SO4,HCO3,Cl, NO3, COD and DOB. They are used as an independent variables in the application of the discriminant analysis. Cluster analysis results show presence of three distinct groups of similar water quality parameters behavior. Application of the stepwise selection of the spatial discriminant analysis explained that pH, TDS, EC, TH and Ca are the most discriminating variables among the previously identified groups. Discriminant analysis shows that 100% and 76% of the original data and cross-validated groups data were correctly classified. These results confirm that there no significant effects of anthropogenic factors along all the selected stations could be detected , where the hydrologic and geologic factors are can be regarded as the main responsible factors controlling the River Tigris water chemistry in Iraq*

INTRODUCTION

Tigris River arises from the southeastern parts of Turkey with elevation ranging from 1000-2000 meters above sea level. The catchment area of the river covers Turkey, Iraq, Syria and Iran and it has an area of approximately 235000 km². River Tigris length is 1415Km inside Iraq and it represents one of the most famous rivers in the world, it flows throughout the whole country starting from Feshkhabour village at the extreme northern parts of the country and ends at Basrah area (in the south of Iraq) to make a confluence with Euphrates river,(Fig-1). Tigris River is the main source for using the water for the different purposes such as agricultural, domestic, industrial and power generation throughout the whole country. The natural gradient of the river exhibits great variations since it starting in mountainous area and ends up in a flat area at the southern parts of Iraq, which characterized by the elevation of few meters above sea level (Al-Ansari and Toma, 1984).

Figure 1: The River Tigris selected stations of the present study.

The geologic formations croping out in the catchment area are back to different geological era where they belong to the Paleozoic era composing from different sedimentary, igneous and, at sometimes, metamorphic rocks, especially at the extreme northern parts of the river catchment area, (Jassim and Goff,2006). The River Tigris passes through the Tertiary formations of carbonate and evaporates with some clastic lithology in the central and southern parts of the country. The flood plain and river terraces of the Quaternary age can also be noticed within the same above parts of the catchment area,(Fig-2).
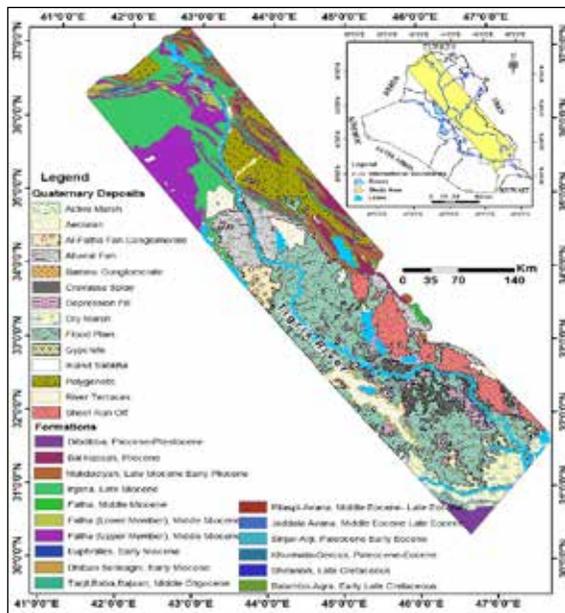
Figure 2: Geologic map of the River Tigris catchment area.

Several studies dealing with various geological, hydrological and hydrochemical aspects by many researchers were conducted. Al-Ansari, et al, 1986; Al-Jebori, 997; Auad, 1997 and Ahmed, 1999 studies are among others. Hydrochemical studies previously, achieved show significant spatial and temporal changes in the chemical parameters along the River Tigris, therefore it is of critical importance to continuously monitor these parameters to explain the main changes and the cause of these changes using different methods and techniques.

The main objective of the present study is to assess the nature of resemblance among selected stations along the River Tigris and determine the prime water quality parameters thought to be responsible for the spatial variations of the river stations by using of the multivariate cluster and discriminate analyses. The stations used in this study are: Feshkhabour, Mosul, Sharkat, Sammara, Baghdad, Kut, Amara and Ali Gharbi (Fig-1), whereas the targeted parameters are: pH, TDS, TH,EC,Ca,Mg,Na,K,SO₄,HCO₃,Cl, NO₃,BOD and COD.

STATISICAL ANALYSIS AND DATA USED

Eight stations and fourteen water quality parameters were used to study the spatial variation behavior of the Tigris River water chemistry. These data represent the mean monthly values of the hydrochemical parameters for the period 2009-2012.

It is provided by the National Center for Water Resources Management, Ministry of Water Recourses. The present data were statistically treated by using the correlations and multivariate cluster and discriminant analyses.

Testing the normality of the present data:
The main assumption in applying of the various multivariate analysis techniques is the normality of the variable data, therefore, checking the multivariate normal distribution assumption is the beginning step in conducting this analysis. The normality can be checked by applying of different statistical approaches such as skewness, kurtosis, Q-Q plot as well as other tests such Kolmogorov-Smirnov and Shapiro-Wilk. (Dillon and Goldstein, 1984).

Cluster analysis:
The objective of this analysis is the grouping of the multivariable data and reduction of variables. The most common approach is the Hierarchal Cluster Analysis, (HCA) which begins with separate clusters and join them together step by step until only one cluster remains. The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between transformed values of the samples (Isah et al, 2013, Davis, 2002)

Discriminant analysis
Discriminant Analysis (DA) is a statistical technique commonly used to build a predictive descriptive model of group of discrimination based on the observed predictor variables and to classify each observation into one of the groups (Fernandez, 2006). The common objectives of DA are :i) to investigate difference between groups, ii) to discriminant groups effectively, iii) to identify important discriminate variables, iv) to perform hypothesis testing on the difference between the expected groups ; v) to classify new observation into pre-existing groups,(Shrestha and Kazama,2007) ..

DA operates on original data and the method constructs a discriminant function for each group as follows (Shrestha et al, 2013):

$$F\ (G_i) = K_i + \quad --- (1)$$

Where (i) is the number of groups (G), K_i is an intercept inherent to each group, n is the number of parameters; W_i is the weight coefficient assigned by DA to a given parameters (P_i). Discriminant analysis depends on the data matrix of pre-classified groups which can be obtained by various statistical techniques such as the cluster analysis,(Manoj,et al,2013).

RESULTS AND DISCUSSION:
The descriptive statistics of the present study parameters for the study period are presented in Table -1. As can be seen, all of these parameters reveal wide ranges for the selected stations along the river course throughout the country.

Table 1: Mean and standard deviation of the water quality parameters of the River Tigris stations.

| Parameter Mean St. Dev.	Feshkhaboor	Mosul	Sharkat	Samarra	Baghdad	Kut	Amara	Ali Gharb
pH	7.55 .26	7.48 .15	7.48 .09	7.58 .15	7.60 .20	7.55 .26	7.68 .22	7.70 .24
TDS	285.8 30.12	380.08 169.82	300 34.43	324.95 22.10	600.25 99.72	712 68.09	1022.3 91.59	1150 103.71
EC	420 .03	530 .18	500 .005	500 .10	880 .14	1250 .30	1530 .15	1830 .29
TH	220.73 8.181	256.90 73.53	236.45 17.90	240.48 3.71	352.25 68.72	409 20.76	462.50 43.42	500.00 19.87
Ca	48.20 3.35	54.45 11.94	50.18 4.12	52.63 6.82	67.13 15.05	82.88 7.92	90.30 12.27	91.08 7.07
Mg	24.10 2.83	27.98 11.40	26.25 4.85	29.78 9.07	42.50 5.80	45.50 1.29	52.33 4.93	62.58 5.25
Na	17.78 2.79	26.80 12.26	21.80 2.94	22.60 2.65	63.98 14.34	85.00 18.70	167.75 37.69	199.50 47.54
K	3.0 .31	3.15 .26	3.15 .31	3.05 .31	3.88 .75	4.20 .69	5.00 .80	5.38 .61
SO4	103.33 7.08	144.70 82.20	117.40 27.07	119.80 8.53	236.25 58.73	302.0 26.62	362.75 57.13	387.25 30.06
HCO3	154.63 9.42	150.18 8.67	154.70 11.28	156.95 6.73	155.00 16.24	160.00 15.12	160.25 13.17	157.75 13.22
Cl	15.28 .88	21.38 8.23	21.70 2.46	21.83 .56	74.50 15.81	93.25 6.89	218.25 25.78	272.50 55.73
NO3	3.35 .67	2.53 .75	2.33 .56	2.58 .58	3.45 1.20	3.20 .31	4.43 .78	5.53 2.55
BOD	1.45 .38	1.48 .45	1.68 .56	1.35 .19	1.50 .56	1.98 .46	1.60 .17	1.75 .43
COD	.99 6.31	1.09 .29	1.25 .20	1.50 .24	1.10 .17	1.13 .17	1.45 .26	1.45 .18

As previously stated the best way for studying of these variations is the use of the multivariate statistical techniques. Therefore, the present study focused only on the spatial variations among the study stations. Concerning the normality, all of the selected pa-

rameters show non–multivariate normal distribution (Fig-3). The testing values of the Kolmogorov-

Smirnov and Shapiro-Wilk as well as the values of Kurtosis and Skeweness (not presented here) support the non-nor-mality distribution of the present data. Therefore, these data should be log transformed before conducting the multivariate cluster and discriminant analysis, (Papatheodoru et al,2006).
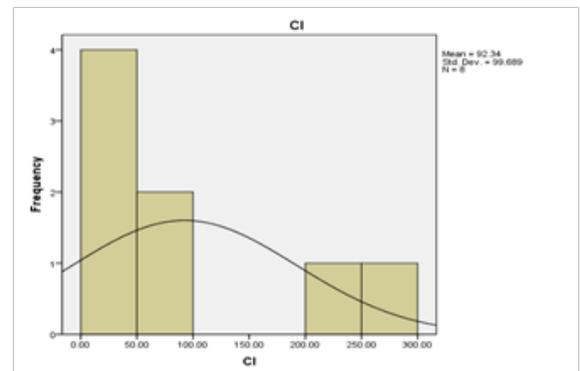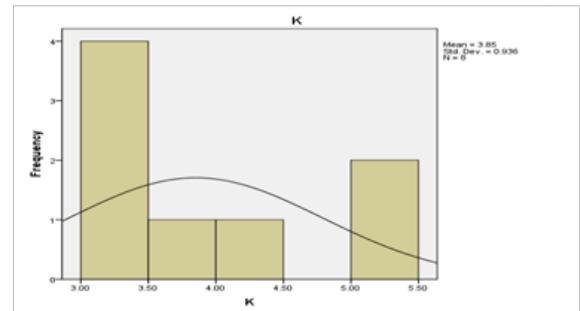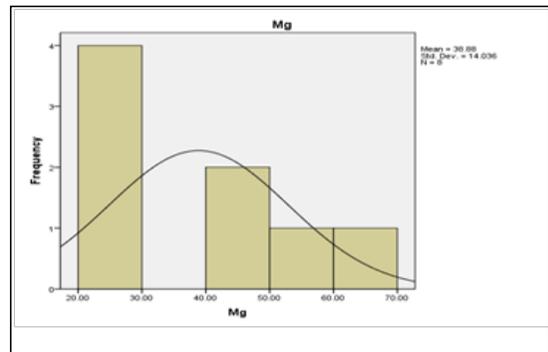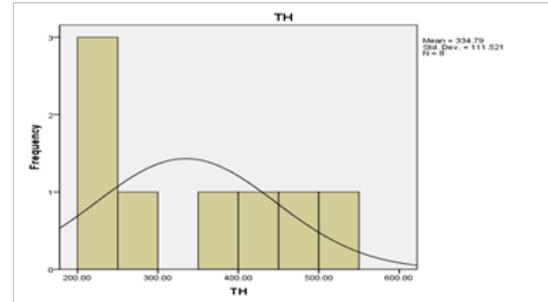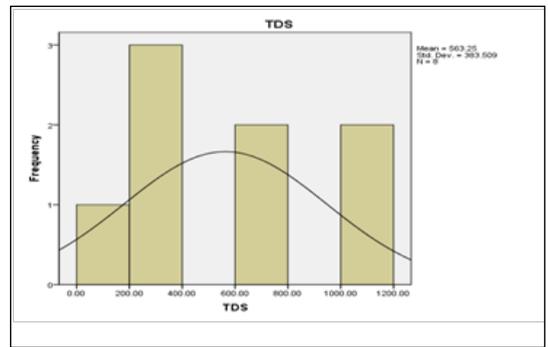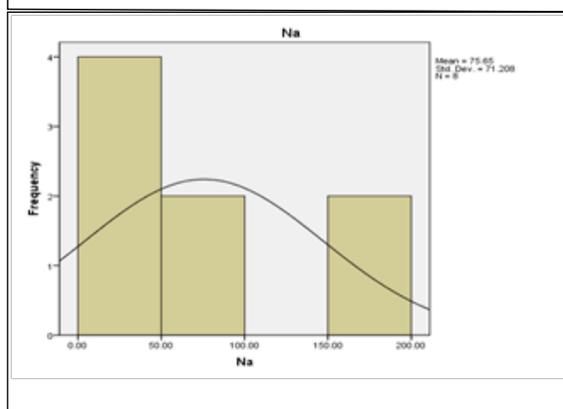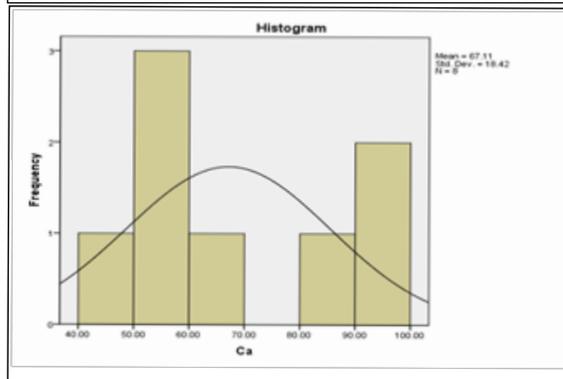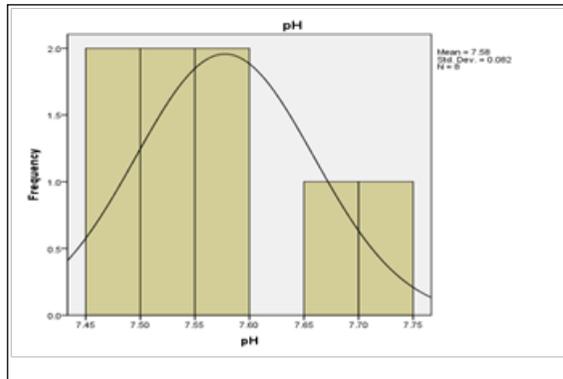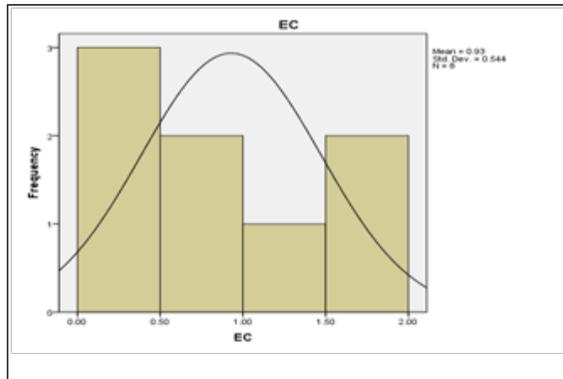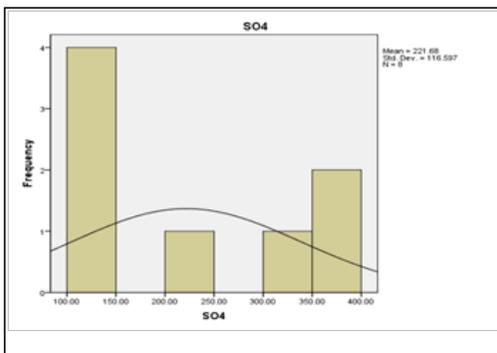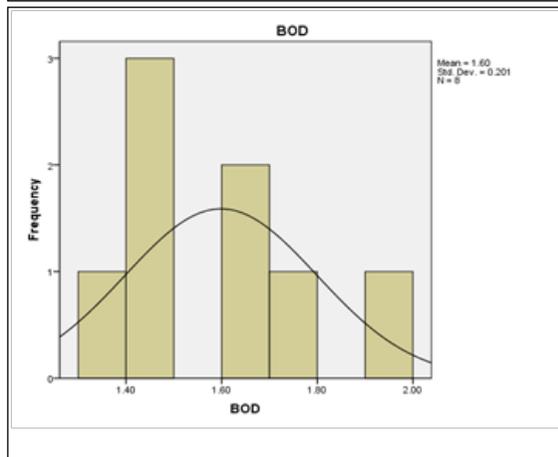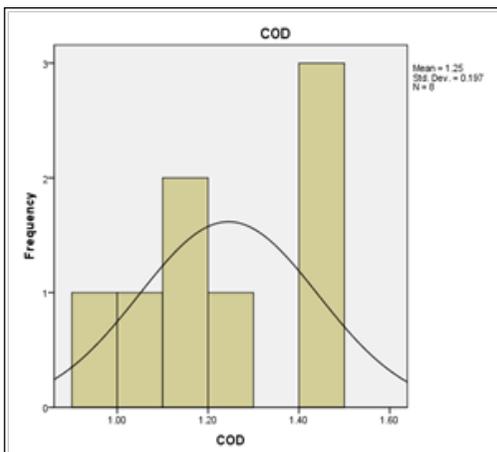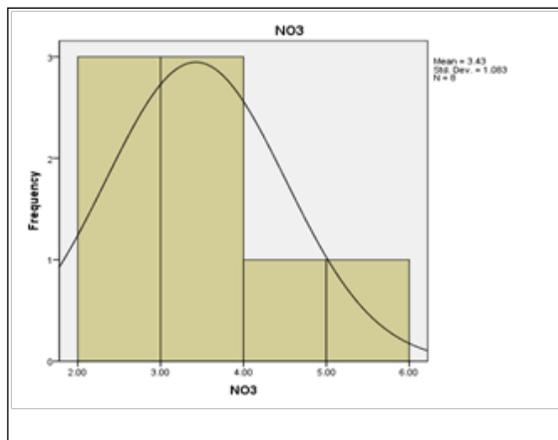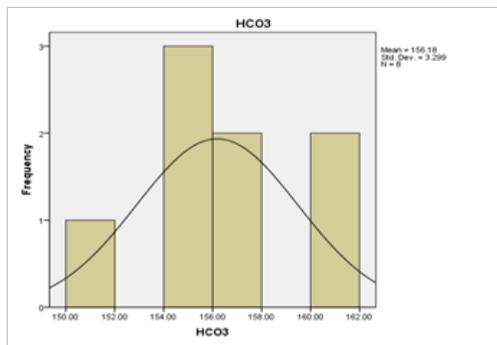
Figure 3: Frequency distribution of the River Tigris water quality parameters.

Cluster analysis:
Table (2) shows the correlation coefficients matrix of the selected hydrochemical parameters of the present study. According to this table, most of these variables show significant correlations coefficients among each other except of BOD and COD which correlate insignificantly with the remaining parameters. This shows that the major elements are the key factors controlling the behavior of the River Tigris hydrochemistry.

Table 2: Correlation matrix of the River Tigris parameters.

	pH	TDS	EC	TH	Ca	Mg	Na	K	SO4	HCO3	Cl	NO3	BOD	COD
pH	1													
TDS	.737*	1												
EC	.787*	.715*	1											
TH	.786*	.741*	.997**	1										
Ca	.770*	.750*	.994**	.994**	1									
Mg	.834*	.748*	.986**	.989**	.976**	1								
Na	.822*	.733*	.994**	.995**	.987**	.987**	1							
K	.833*	.696	.990**	.986**	.975**	.977**	.995**	1						
SO4	.755*	.747*	.992**	.998**	.993**	.981**	.990**	.975**	1					
HCO3	.667	.435	.695	.671	.715*	.673	.662	.670	.649	1				
Cl	.838**	.706	.991**	.991**	.980**	.989**	.997**	.995**	.983**	.686	1			
NO3	.915**	.74	.830*	.827*	.803*	.836**	.855**	.881**	.800*	.548	.856**	1		
BOD	.125	.139	.641	.615	.632	.547	.567	.586	.616	.511	.561	.321	1	
COD	.589	.181	.437	.405	.419	.479	.453	.461	.375	.524	.480	.320	.013	1

*Refers to significant values.

Agglomerated Hierarchial Cluster (AHC) analysis, (Fig-4), shows that there are three main clusters (groups), the first group includes Feshkhabour, Mosul, Sharkat and Samarra station, (northern stations) whereas the second group comprise from Baghdad and Kut stations (central stations). The third group includes: Amarah and Ali Gharbi (southern stations). As can be noticed from the above grouping, sub-grouping is not strong that one can distinguish them significantly.
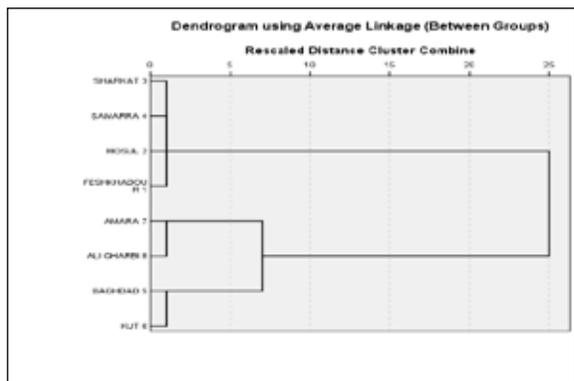

Figure 4: Dendrogram of the River Tigris stations.

Discriminant analysis
Discriminate analysis was applied on the already identified group by cluster analysis. The purpose of DA is to find the best combination of the predictors (independent parameters) differentiating the above three groups, Results of the present study show two latent discriminate functions, since that there are three potential groups, so only two functions was displayed and it is beilved that they gave an index of overall model fit which interpreted as being the proportion of the total variance explained (R^2). The above two functions revealed that, pH, TDS, EC, TH and Ca, (Table-3) differentiated well the grouping of the River Tigris stations in Iraq.

Table - 3: Classification function coefficients (Fisher's linear discriminant functions).

Parameters	Groups		
	1	2	3
pH	852978.008	861598.722	864757.094
TDS	-66940.552	-67589.360	-67678.227
EC	-153367.533	-154870.108	-155014.893
TH	396100.302	400330.661	401041.089
Ca	46034.979	46317.882	46257.598
Constant	-824778.430	-841838.225	-846101.865

Wilks Lambda values, (Table 4), also confirms that the mentioned predictors were significant in supporting the predictors importance of the discriminate functions.

Table 4: Outlines of the canonical discriminant functions.

Eigenvalues				
Function	Eigenvalue	Variance %	Cumulative %	Correlation
1	120.533	97.8	97.8	.996
2	2.691	2.2	100.0	.854
Wilks' Lambda				

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.002	18.318	10	0.050
2	.271	3.917	4	.417
Discriminant function coefficients				

	function	
	1	2
pH	601.504	15.347
TDS	-39.805	32.487
EC	-89.943	89.030
TH	264.345	-182.213
Ca	13.720	-36.550
Const.	-1116.742	429.669

Table (5) shows the leave-one-out cross validation classification procedure and indicating that 100% of the original groups cases were correctly classified. The relative importance of DFI compared to DF2 in explaining the differences among the three groups (Table-4), show that DF1 have eigenvalue of 97.8% of the total data variance of the three groups. The degree of relationship between the selected predictors and groups is also higher for DF1 as compared with DF2 which explained by the respective canonical correlation values of 0.996 and 0.854, Table (4).

Table 5: Classification results of the present study groups.

1			Group 2	Predicted group membership			Total
				3			
original	count	1	1	4	0	0	4
		2	2	0	2	0	2
		3	3	0	0	2	2
	%	1	1	100	0	0	100
	%	2	2	.0	100	0	100
	%	3	3	.0	0	100	100
Cross-validated	count	1	1	3	0	1	4
		2	2	1	0	1	2
		3	3	0	1	1	2
	%	1	1	75	0	25	100
	%	2	2	50	0	50	100
	%	3	3	.0	50	50	100

Correctly classified original grouped cases = 100.00%
Correctly classified cross-validated grouped cases = 76%

Box and whisker plots
Box and whisker plots of the discriminating parameters were used to show the main responsible parameters for discriminating the obtained three groups of the Tigris River. The concentrations mean and standard deviations of all the selected water quality parameters are significantly variable among the identified three groups, (Fig-5).
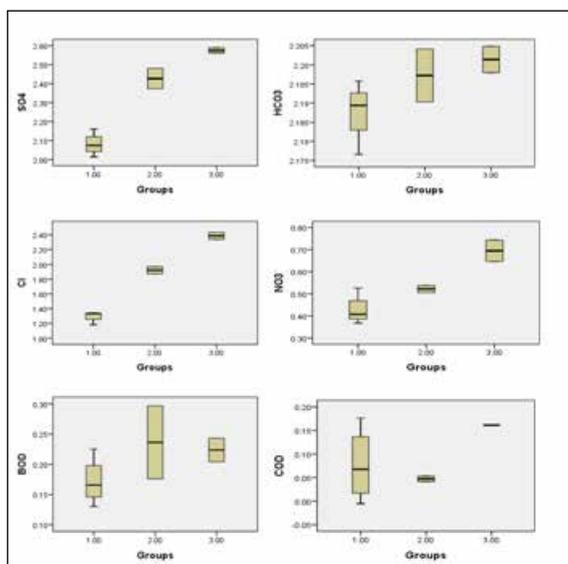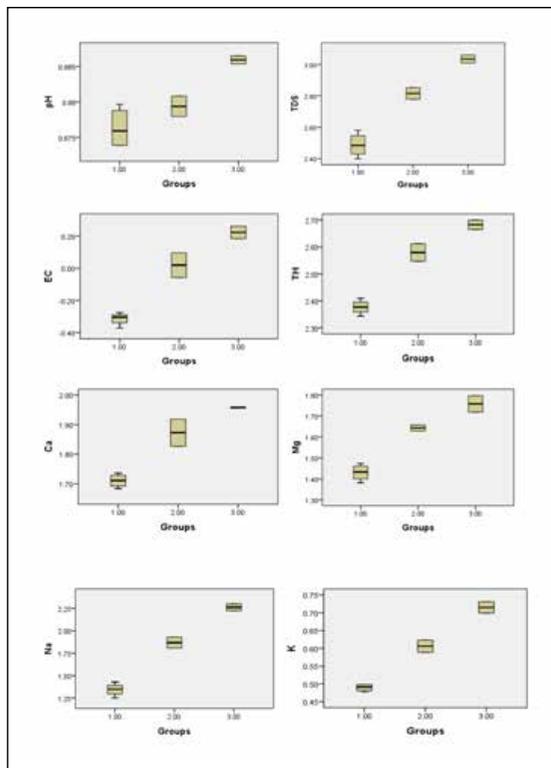
According to the above results, pH, TDS, EC, TH and Ca discriminate very well the three groups of the River Tigris. As NO_3, BOD and COD seems to be having small values and still within the normal ranges, therefore they could not be used for differentiating the present study groups and hence they show no significant indications of pollution along the selected stations.

The above findings explain that the natural factors such as the geologic and hydrologic factors still the most affecting factors on the River Tigris hydrochemistry. Salinity, alkalinity and pH are the most controlling parameters on the river chemistry, whereas the anthropogenic effects could not be detected among any of the present study stations.

CONCULSIONS:

Cluster analysis results show that River Tigris stations in Iraq can be classified into three groups which are very useful for any suggested future monitoring program for the spatial variations of the hydrochemical and environmental parameters. The most important and significant parameters for the discriminating the present study three groups as obtained by discriminant analysis were pH, TDS, EC, TH and Ca where most of the present data variance is explained by the variations of these parameters. According to this analysis 100% and 76% of the original data and cross-validation cases groups were correctly classified respectively. Therefore, DA analysis proved to be successful tool in reducing the dimensionality of large data set and delineate a few parameters responsible for the water quality variations interpretation. No anthropogenic effects were detected, hence the River Tigris still affected by the natural conditions on regional scale level and a local scale studies between the river stretches are necessary for the continuous monitoring of the river quality parameters.

Figure 5: Box and Wihsker's plots of the selected parameters of Tigris River.

REFERENCE Ahmed, L.M. (1999). Hydrochemistry of Tigris and its change at selected ststions through the period 1979-1996, M.Sc. thesis, College of Science, University of Baghhdad. | Al-Ansari, N.A. and A. Toma (1984).Bed characteristics of the River Tigris within Baghhdad.J. Water Resources, | 3(1), 1-23. | Al-Ansari, N.A., A. Sayfy, G.T Al-Sinawi and Ovanessian A.(1986). Evolution of the water quality for the lower reach of River Tigris using multilane analysis, J. Water Research, 5(2), 173-187. | Al-Jebori, J. H. (1997). Hydrochemisrty and sediments pollution of Tigris River between Hammam Al-Aleel and Baghdad, M.Sc. thesis, College of Science, University of Baghdad, 93p,(in Arabic). | Auad, J.H. (1997). The hydrological and hydrochemistry of Tigris River for the period 1979-1993, Unpublished Ph.D. thesis, College of Science, Univ. of Baghdad,181, (in Arabic). | Davis, G., 2002; Statistic and Data analysis in geology; John Wiely and Sons Inc, New York. | Dillon, W.R. and M. Goldstein, 1984, Multivariate analysis methods and applications, John Wiley and Sons, New York. | Fernandez, G., 2010; Discriminate analysis, A powerful classification technique in data Mining, Statistic and Data analysis. | Jassim, S.Z. and Goff, J.C. (2006): Geology of Iraq, Published by Dolin, Prague Moravian Museum, Brno, p341. | Isah, A. UAbdulla and M. Ndamitso; 2013;Application of multivariate method for assesaments of variations in river streams water quality in Nager state, Nigeria, Amer. Jour. of Theoretical and Applied Statistics,.2(6), 176-183. | Manoj, K., S. Ghosh and Plc Pandy; 2013; Characterization and Classification of hydrochemistry using multivariate graphical and hydrostatical techniques, Research Jour. Of Chemical Sciences, (5), 32-42 | National Center for Water Resources Management,2013, Monthly records of some selected chemical characteristics of the River Tigris, IRAQ. Ministry of Water Resources. | Papatheodoru, G., G. demopoulou and N. Lambrakis; 2006; A long-term study of temporal hydrous chemical data in a shallow lake using multivariate statistical techniques, Ecological modeling, 193,:759-776 | Shrestha, S. and F. Kazama,2007, Assessment of water quality usin Multivariate statistical techniques: A case study of the Fuji river basin, Japan, Environmental Modelling an Software, 22, 464-475. | Shrestha, S.; F. Kazama and T. Nakamura; 2008; Use of principle component analysis, factor analysis ad discriminate analysis to evaluate spatial and temporal variations in water quality of Mekong River; Jour of Hydroinformatics, 1(10), 4.-56. |