Water Quality Assessment of Al-Gharraf River, South of Iraq Using Multivariate Statistical Techniques

Salam Hussein Ewaid[1*] and Salwan Ali Abed[2]
[1]Technical Institute of Shatra, Southern Technical University, Iraq.
[2]Department of Environment, College of Science, University of Al-Qadisiyah, Iraq.
*Corresponding Author: Salamalhelai@stu.edu.iq

Abstract

The multivariate statistical methods (principal component analysis "PCA" and hierarchical cluster analysis "CA"), were used to evaluate the water quality of Al-Gharraf River, the largest branch of Tigris River, south of Iraq. Water quality parameters were calculated monthly at each of the ten stations along the river during the year 2014-2015, to investigate the potential pollution sources and samples were analyzed for 18 physicochemical and biological parameters. The PCA supported in extracting and recognizing the factors that are responsible for river water quality variance over the year. Four factors identified that are responsible for 70% of the total variance in water quality of the river. The first factor is a multi-source factor that explains 31.6% of the total variation. The second factor is organic source factor that explains 14.6% of the total variance. The third factor is soil erosion source factor that explains 12.8% of the total variance. The fourth factor is caused by sulfates and calcium and explains 11% of the total variance. The hierarchical CA classified the 10 monitoring sites into three differentiated groups showed similar water quality features. The natural hydro-geochemical and anthropogenic activities such as domestic wastewater influents were the significant factors affect the water quality. The results of this work can be used to decrease the number of samples analyzed in both space and time to help the decision makers in recognizing priorities to improve water quality. [DOI: 10.22401/JNUS.20.2.16]

Keywords: Al-Gharraf River, water quality, multivariate statistical methods.

Introduction

Water is an essential natural resource for human life, developing economics and society regarding agriculture, industry and various facilities. Rivers are not supplying water for human consumption only, but also to receive wastes discharged from human activities [1]. Constant polluting sources are formed by municipal and industrial wastewater ejection, whereas surface overflow is a seasonal phenomenon, largely affected by climate within the surface water basin [2].

The Seasonal variation of precipitation, surface runoff, groundwater flow, and anthropogenic transfers have an intensive effect on the river discharge and, subsequently, on the amount and concentration of pollutants in water. Due to these complexities, water quality specialists and decision-makers often are confronted with significant challenges in their efforts to control water pollution [3, 4]. Through identifying spatial and temporal patterns in the river water quality, an improved understanding of the environmental conditions may help the stakeholder and decision-maker to establish priorities for sustainable of water management [5, 6, 7].

Previous studies have demonstrated that Iraq currently faces serious water problems; not only over-exploitation and uneven in the spatial distribution of water resources, but also severe water pollution in Iraqis main rivers, which give contribution for the scarcity of water of adequate quantity and quality [8,9,10]. Growing municipal wastewater discharges due to urbanization, agriculture, and industrial practices, along with the limitation of wastewater treatment facility and capacity, considered the principal drivers of water pollution. In Iraq, all wastewater discharged into rivers, and most of that is untreated [11, 12, 13].

Application of the multivariate statistical methods, like principal component analysis (PCA) and cluster analysis (CA), helps in the elucidation of complex data to improve the understanding of the water quality. Such tools facilitate the identification of possible factors

that influence water quality and can aid in the reliable management of water resources as well as a solution to pollution problems [14, 15, 16].

Studies investigated the changeability of water quality have informed that water quality is highly depending on watershed runoff and urbanization, these studies have maintained the pollution sources with the possible influences of natural practices and anthropogenic activities on the variation of water quality [17, 18, 19]. In this study, water quality data set of Al-Gharraf River during 2014-2015 was analyzed using CA and PCA to identify the significant parameters explaining the variation of pollution sources and an attempt to recognize and clarify the main factors which explain the structure of water quality dataset.

Materials and Methods
A. Description of the Study Area
Al-Gharraf is the main branch of Tigris River, south of Iraq. Its water type and discharge in conjunction with the hydrological characterizations of Tigris River in Al-Kut barrage [20]. It suffers from natural and human problems such as pollution, accumulation of clay, the growth of plants like the water hyacinth (*Eichhornia crassipes*), water-squandering, and reduction of water levels [21]. In Wasit and Dhi-Qar governorates it passes through many towns and green heavy populated villages. In Dhi-Qar it has two branches before its end in the marshes [22].

The length of the main river is about 230 km, 50- 80 m width and 3-7 m depth. Its basin populated by more than million people using about 432000 m³/year of refined water and passing through an agricultural area of about 215019 hectares in the southwest of Iraq within the sediment plain [20].

The geographical position of the study stations is shown in Table (1). This position gives the area climate features like; the high rate of sun radiation, the high temperature in summer, few rain occasions in winter and spring, low moisture, and high rate of evaporation [21].

Table (1)
The coordinates of the study area.

No.	Stations	Coordinates	
		N	E
1.	S-1	32°30'4.93"	45°48'31.27"
2.	S-2	32°21'21.60"	45°56'47.33"
3.	S-3	32°11'35.91"	46° 0'29.16"
4.	S-4	32° 2'59.63"	45°59'18.10"
5.	S-5	31°53'15.59"	46° 2'59.91"
6.	S-6	31°45'5.16"	46° 6'39.06"
7.	S-7	31°34'33.46"	46° 7'26.40"
8.	S-8	31°26'54.12"	46°10'21.78"
9.	S-9	31°23'29.56"	46°10'48.38"
10.	S-10	31°18'9.25"	46°14'18.46"

There are large yearly fluctuation in water level and average annual discharge, in summer and autumn (the dry seasons) water comes from the reservoirs full with organic materials, planktons, algae and plants with dark green color causing decrease in pH and dissolved oxygen, increase in turbidity, total dissolved solids and affecting the water quality [22, 23]. Ten sampling stations were chosen before, between and after the main towns along the river to represent its situation Fig.(1).
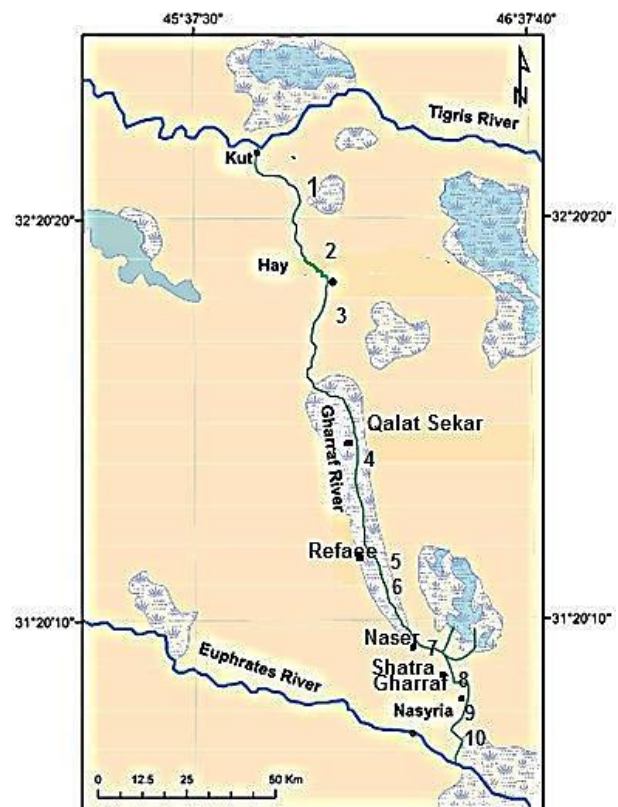

Fig.(1): The Map of the study area and the sampling stations. Modified from [24].

B. Field sampling and analytical measures

Water samples were gathered monthly from winter 2014 to autumn 2015, at each sampling station on the river. Samples were preserved and examined according to the standard techniques of the American Public Health Association (APHA) [25].

Physical, chemical and biological parameters including water temperature (T) (ºC), dissolved oxygen (mg/L) (DO), electrical conductivity (EC) (µS/cm), and pH, were measured *in situ* using a WTW multi-meter model Multi 340i. Turbidity (Tur) (NTU) was determined *in situ* using WTW portable turbidity meter model TURB 355 IR/T. Total dissolved solids (TDS) (mg/L) by the temperature controlled oven. Alkalinity (Alk) (mg/L) as $CaCO_3$ measured by titration method, total hardness (TH), Ca^{+2} and Mg^{+2}(mg/L) were measured by EDTA complex metric titration. Na^+ and K^+ (mg/L) concentration was measured by flame photometer model M410 ٬UK, and Cl^- (mg/L) concentration was measured by silver nitrate titration method. Sulfate (SO_4^{-2}) (mg/L) concentration was determined spectrophotometrically using the barium sulfate turbidity method. Nitrate (NO_3) and phosphate (PO_4) concentrations (mg/L) were measured by cadmium reduction and molybdate-ascorbic acid methods, respectively. Biological oxygen demand (BOD_5) (mg/L) by azide modification at 20ºC (5 days) and fecal coliform (FC) (CFU/100 ml) by membrane filtration technique [24]. The positions of stations were determined by the Global Positioning System, Geko 201, Garmin Ltd. /Taiwan.

C. Data treatment and multivariate statistical methods

The Windows statistical software package SPSS 21, was used for descriptive analysis, correlation coefficient and multivariate statistical analysis of the data. The Kolmogorov–Smirnov (K-S) experiment was used to examine and evaluate the goodness of proper data to a log-normal distribution. According to the K-S experiment, all the variables were log-normally distributed with 95% or more certainty [26].

Kaiser-Meyer-Olkin (KMO) and Bartlett's test were used to inspect and check the fittingness of data for PCA. KMO is a measure of sampling competence, which indicates the proportion of variance (common variance), that might be caused by underlying factors [27]. A high value (close to 1) points out that PCA may be useful. In this study, KMO = 0.6. Bartlett's test of sphericity indicates whether correlation matrix is an identity matrix, which would mean unconnected variants. In this study, the significance level which is 0 (less than 0.05) indicates that there are significant relationships among variables, Table (2).

Table (2)
KMO and Bartlett's Test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0.615
Approx. Chi-Square Bartlett's Test of Sphericity	254.116
df	66
Sig	0.00

D. Principal component analysis

This way of investigation can identify the most significant parameters which describe all data set and data reduction, providing the correlation among water components without changing the original data [27]. In the current study, normalized variables were extracted the significant PCs and decrease the contribution of variables with minor importance by varimax rotation generating factors [27, 28]. The actors that have more than 1 variance are included only because any component should explain more variance than any single variable [28]. The PCA was applied by using varimax rotation with Kaiser Normalization and by extracting the eigenvalues from the correlation matrix, the number of the significant factors and the percent of variance that explained by each of them was calculated.

E. Cluster analysis

It is a technique that grouped similar objects into homogeneous groups respect to certain properties and different from or unrelated to the objects in the other groups [29].

By hierarchical clustering, the levels of similarity, at which observations are combined, are used to build a dendrogram. The agglomerative hierarchical clustering was used with Ward's linkage method and squared Euclidean distances as a measure of similarity [30].

Results and Discussion

The essential statistics of all water quality variables, which were measured during the sampling period of one year at the ten stations on the river, are summarized in Table (3).

A. Correlation analysis

Pearson correlation of the ten sampling stations, which is combined to calculate the correlation matrix of the 18 variables, showed that the parameters in water samples collected from Al-Gharraf River were weak and moderately correlated to each other at $p \leq 0.01$ and $p \leq 0.05$ levels, Table (4), this is because they are affected by the same spatial and temporal variations. However, some clear hydrochemical relationships were found.

A strong significant positive of correlation was found between FC and T, a moderate significant positive of correlation was found between DO and Tur, DO-NO_3, Cl - T, EC-T, EC-TDS, EC-BOD, T-TH, BOD-T, Mg-T, FC-Cl, FC-TH, FC-Mg, FC-EC, FC-BOD.

A moderate significant negative correlation was found between DO and T, DO-TH, pH-T, NO_3-T, NO_3-Mg, FC-CO3, FC-DO, Table (4).

Table (4)
Pearson correlation between water quality parameters of the study sites.

	Temp	TDS	Tur	CO3	DO	Cl	Ca	TH	Mg	SO4	PO4	pH	NO3	Na	K	EC	BOD	FC
Temp	1																	
TDS	.43**	1																
Tur	-.39-*	.01-	1															
CO3	-.48-**	-.08-	-.18-	1														
DO	-.69-**	-.22-	.60**	.16	1													
Cl	.64**	.43**	-.25-	-.40-*	-.49-**	1												
Ca	-.29-	-.27-	.084	.02	.181	-.181-	1											
TH	.55**	.24	-.35-*	-.130-	-.53-**	.397*	.103	1										
Mg	.61**	.38*	-.19-	-.24-	-.34-*	.427**	-.074-	.488**	1									
SO4	-.32-*	-.36-*	-.20-	.11	-.183-	-.037-	.426**	.100	-.256-	1								
PO4	.07	.19	-.4-**	.20	-.196-	.114	-.110-	.062	.097	.098	1							
pH	-.60-**	.04	.126	.49**	.398*	-.34-*	.174	-.246-	-.37-*	.127	.091	1						
NO3	-.79-**	-.19-	.50**	.21	.604**	-.48-**	.306	-.43-**	-.52-**	.286	.018	.477**	1					
Na	-.17-	.320*	-.09-	.48**	.020	-.037-	-.42-**	-.096-	-.167-	-.079-	.027	.273	.065	1				
K	.11	.289	-.22-	.32*	-.095-	.076	-.271-	.039	-.015-	-.097-	.374*	.139	-.271-	.370*	1			
EC	.62**	.752**	-.24-	-.15-	-.52-**	.563**	-.41-**	.453**	.318*	-.138-	.246	-.104-	-.294-	.337*	.230	1		
BOD	.58**	.247	-.6-**	.014	-.66-**	.387*	-.018-	.491**	.247	-.009-	.317*	-.103-	-.45-**	.159	.142	.565**	1	
FC	.91**	.495**	-.30-	-.50-**	-.65-**	.689**	-.196-	.608**	.513**	-.252-	.109	-.43-**	-.65-**	-.135-	.111	.714**	.656**	1

**. Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

B. Application of cluster analysis

CA gathered sampling stations in the river into 3 clusters of similar water quality features. Fig.(2) illustrates the dendrogram output by using Ward's linkage method and square Euclidean distances. Cluster 1 (stations 1 and 2), Cluster 2 (stations 3, 4, 5 and 6) and Cluster 3 (stations 7, 8, 9 and 10) correspond to the relatively low pollution, moderate pollution, and high pollution regions respectively.

In Cluster 1, which contains relatively less polluted sites, this could be accredited to the fact that fewer human activities were taking place at stations 1 and 2 upstream the river; they are far from the discharge of effluent in the beginning of the river. The river here surrounded with agricultural land, characterized by wide riverbed, fast flow rate, high water level and has the same features of Tigris River which characterize with no cities or large pollution sources for a long distance after it left Baghdad [21].

Cluster 2 is made up of moderately polluted sites, the four stations, 3rd, 4th, 5th, and 6th are located after and between Al-Hay, Qalat Sekar, and Al-Refaee towns, the river here characterized by salt areas (Sabkha) with saline water drainage, agricultural fields on both sides, incidence under the influence of human activity of these towns as the wastewater is added directly, the presence of some slaughterhouses and car workshops which discharge untreated domestic wastewater to the river [20]. Cluster 3 contains the four stations 7th, 8th, 9th and 10th which are located downstream the river among Al-Nasr, Al-Shatra and Al-Gharraf towns where the river region characterized by significant reduction in water level, high population density, under the influence of streams and the remnants of the town's markets, sewage flowing and large quantity of solid and liquid waste [21].

It is worth mentioning that there are four drinking water treatment plants on the right side near station 8 before the river has divided into its two main branches, these plants provide drinking water to Dhi-Qar and Basrah provinces, Fig.(1).
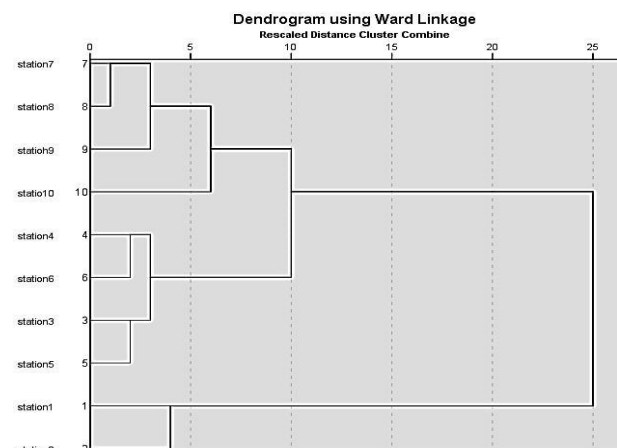

Fig. (2): Dendrogram of the CA of water quality parameters.

C. Application of the principal component analysis

PCA is employed in this study, for the 18 variables from the 10 sampling stations during 2014-2015, to identify the most significant water quality parameters. The eigenvalues, which provide the highest measure of factors significance, are the most significant eigenvalues. Eigenvalues of 1.0 or greater are considered as significant Fig.(3) and Table (5) [28]. Classification of PCs (strong, moderate and weak) is corresponding to absolute loading values of > 0.75, 0.75-0.50 and 0.50-0.30, respectively[27]. Table (5), which summarizes the PCA, is including the loadings and eigenvalues of each PC, explaining the total variance and the cumulative variance. It must be noted that the rotation, which resulted in an increase in the number of factors, is necessary to explain the same amount of variance in the original dataset [6].

In the obtained PCA data, five components were extracted, explaining 70% of the total variance which helps in the elucidation of the results and in the identification of sources of pollution from the water quality data. PC 1 showed that 31.6% of the variance, and loaded heavily on T, FC, NO_3, DO, pH, Cl, Mg, CO_3, and TH. Water temperature, which reflects the atmospheric temperature and the negative strong loadings on pH, NO_3 and DO, represents the contribution of non-point source pollution and refers to the decayed organic matters from agricultural areas. The strong positive loading on FC bacteria represents the contribution of point source pollution of human activities and the positive correlation with Alk, Cl^- TH, and Mg^{2+}denotes the contribution of point pollution and the physical-chemistry of the river.

PC 2 was loaded chiefly by Tur, BOD, and PO4 and explained 14.6% of the total variance. This factor represents the contribution of the non-point source of organic pollution from agricultural areas and points source pollution from domestic sewage, the presence of PO_4 is most probably due to agricultural runoff such as livestock waste and fertilizers and the municipal sewage because PO_4 is an important component of detergents [17].

PC 3 was responsible for 12.8% of the variance and was best represented by TDS and EC. This factor contains the contribution of non-point source pollution from hydro-geochemical and mineralization processes of the geological components of soil.

Additional 10.975% of the variance was explained in PC 4; this factor contains SO_4 and the hydro-geochemical variable Ca^{2+} originating from mineralization of soil.

Table (5)
Loadings of 18 variables on 4 significant Principal components, varimax rotated component matrix.

Water Quality Variables	Component			
	1	**2**	**3**	**4**
T	.941	.113	.124	-.211-
FC	.926		.217	
NO_3	-.755-	-.240-		.365
DO	-.711-	-.450-	-.135-	-.129-
pH	-.697-	.114	.310	.176
Cl	.662		.369	
Mg	.660		.147	
CO_3	-.605-	.552		
TH	.581	.209	.283	.373
Tur	-.357-	-.796-	.127	
BOD	.536	.667	.264	.133
PO_4		.646	.113	
K		.480	.245	-.470-
TDS	.200		.859	-.231-
EC	.420	.190	.815	-.107-
Na	-.358-	.332	.543	-.397-
Ca	-.107-		-.201-	.781
SO_4	-.179-	.322		.723
Eigenvalue	5.687	2.633	2.319	1.975
% of Variance	31.593	14.626	12.881	10.975
% Cumulative of Variance	31.593	46.219	59.1	70.075

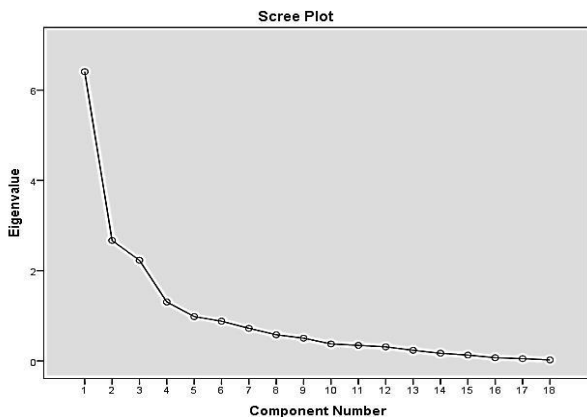*Bold values in dicate strong and moderate loadings.*

Fig.(3): Scree plot of the Eigenvalue and component number.

Conclusions

In this research, the results of multivariate statistical methods suggest that natural and anthropogenic pollution sources had significant effects on water quality of the river. The PCA analysis produces four factors to show 70% of the total variance in water quality. The monitoring stations are grouped by Cluster analysis into 3 clusters of similarity based on water quality characteristics at different stations. These consequences show that agricultural, domestic and hydro-geochemical sources are responsible for the pollution in Al-Gharraf River.

The outcomes, which were attained from CA, point out that CA method is very helpful in the categorization of the areas of surface water into a different category such as comparatively less pollute, moderately pollute and highly polluted area, depending on the extent of pollution. This means that for quick assessment of water quality, only one site in each cluster may serve as a good and appropriate site in the the spatial assessment of the water quality of the whole network.

References

[1] ICMM International Council on Mining and Metals "Metals Environmental Risk Assessment Guidance", London, UK, 2007.

[2] Singh, K.P., Malik, A., Mohan, D., Sinha, S., "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, 2004.

[3] Vega, M., Pardo, R., Barrado, E., Deban, L., "Assessment of seasonal polluting effects on the quality of river water by data analysis". Water Research 32, 3581–3592. 1998.

[4] Xu, H., Z. Xu, W. Wu and F. Tang, 2012. "Assessment and Spatiotemporal Variation Analysis of Water Quality in the Zhangweinan River Basin, China. Procedia Environmental Sciences, 13: 1641-1652.

[5] Kowalkowski, T., R. Zbytniewski, J. Szpejna and B. Buszewski "Application of chemometrics in river water classification". Water research, 40(4): 744-752. 2006.

[6] Razmkhah, H., A. Abrishamchi and A. Torkian "Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran)" J. of Env. Management, 91(4): 852-860. 2010.

[7] Shrestha, S., Kazama, F. "Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River basin, Japan". Env. Modeling and Software 22, 464–475. 2007.

[8] Issa, I. E., Al-Ansari, N. A. Sherwany, G. and Knutsson S. "Trends and future challenges of water resources in the Tigris–Euphrates Rivers basin in Iraq". Hydrol. Earth Syst. Sci. Discuss., 10, 14617–14644. 2013.

[9] Hassan, F. M., AL-Zubaidi, N. A. and Al-Dulaimi, W. A. "An ecological assessment for Tigris River within Baghdad, Iraq" J. of Babylon Univ. Special Issue - Proceeding of 5th Inter. Conference of Env. Science Univ. of Babylon/ Env. Research Center 3-5. 2013.

[10] Al-Obaidy, A. M. J. and Al-Khateeb, N. "The Challenges of Water Sustainability in Iraq" Eng. & Tech. Journal, Vol.31, Part (A), No.5. 2013.

[11] Ministry of Environment, "System Maintenance of River and Public Water Pollution in Iraq," 1998. http://www.moen.gov.iq

[12] Rabee, A. M., Abdul-Kareem B. M. and Al-Dhamin A. S. "Seasonal Variations of Some Ecological Parameters in Tigris River Water at Baghdad Region, Iraq", J. of Water Resource and Protection, Vol. 3, No. 4, 2011, pp. 262-267.

[13] Flaieh, H. M., Mohammed-Ridha, M. J. and Abdul-Ahad, M. Y. "Assessing Tigris River Water Quality in Baghdad City Using Water Quality Index and Multivariate Statistical Analysis" Inter. J. of Eng. Sciences & Res. Tech. [687-699] 3(7): July, 2014].

[14] Yang, L., X. Linyu and L. Shun, "Water quality analysis of the Songhua River Basin using multivariate techniques" J. of Water Resource and Prot., 1: 110. 2009.

[15] Shrestha S. and Muang S. "Assessment of surface water quality of Songkhram River using environmetric techniques" Inter. J. of River Basin Manag. 12(4), 2014, pp. 341-356.

[16] Zhou, F., Liu Y. and Guo H. C. "Application of multivariate. statistical methods to the water quality assessment of the watercourses in the northwestern New Territories" Hong Kong. Environmental Monitoring and Assessment 132(1-3), 2007, pp. 1-13.

[17] Al-Shujairi, S. O. H. "Develop and Apply Water Quality Index to Evaluate Water Quality of Tigris and Euphrates Rivers in Iraq" Inter. J. of Modern Eng. Res. (IJMER) Vol. 3, Issue. 4, Jul - Aug. 2013 pp-2119-2126.

[18] Kannel, P.R., S. Lee, S.R. Kanel and S.P. Khan "Chemometric application in classification and assessment of monitoring locations of an urban river system" Analytica Chimica Acta, 582(2): 390-399. 2007.

[19] Boyacioglu H. and H. Boyacioglu H. "Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey" Envir. Geology 54, 2007, pp. 275-282.

[20] Al-Gizzy H. S. N. "Hydrology of Shatt Al-Gharraf and its investment" M.Sc Thesis, Univ. of Basrah, Basrah, (in Arabic). 2005.

[21] Ewaid S. H. "Investigation of Some Heavy Metals in Water, Sediment and Some Biota of Al-Gharraf River, South of Iraq" M. Sc Thesis, Univ. of Baghdad. 2011.

[22] Iraq Ministries of: Water Resources, Municipalities and Public Works and Environment, "New Eden Master Plan For Integrated Water Resources Management in the Marshland Area" Main Water Control Structures (Dams and Water and Reservoirs). Annex III. 2006.

[23] Ministry of Agriculture and Irrigation Republic of Iraq (MOAI) "East Gharraf Project AL-Hai South Zone, Sector (A) irrigation & drainage system of the main canal and drainage pumping station. State Commission for Irrigation and Reclamation Projects, Al-Furat Center for Studies and Designs of Irrigation Projects".1991.

[24] Al-Abadi A (2014) Modeling of stage–discharge relationship for Gharraf River, southern Iraq using back propagation artificial neural networks, M5 decision trees, and Takagi-Sugeno inference system technique. Appl Water Sci. doi:10.1007/s13201-014-0258-7

[25] American Public Health Association (APHA) "Standard Methods for the Examination of Water and Wastewater" 27th ed. Washington, DC. 2012.

[26] Field, A P."Discovering statistics using SPSS" (2nd ed.), London, Sage. (Ch. 15). 2005

[27] Everitt B & Hothorn T. "Introduction to Multivariate Analysis with R." Springer. 2011

[28] Tabachnich, B. G. & Fidell, L. S. ; "Using Multivariate Statistics" (3rd ed), New York, Harper and Row,1996.

[29] Shrestha, S., Kazama F. and Nakamura T. "Use of principal component analysis, factor analysis and discriminate analysis to evaluate spatial and temporal variations in water quality of the Mekong River" J. of Hydroinformatics 10, 2008, pp. 43–54.

[30] Simeonov, V., Simeonova, P. & R. Tsitouridou. "Chemometric quality assessment of surface waters: two case studies" Chem. Eng. Ecology 11(6), 2004, pp. 449-469.