Using Statistical Analysis Approach to Evaluate the Groundwater Quality in Jisser Diyala Area

*Ahmad A. Ramadhan[1], Tariq A. Hussian[2] and Kamal B. Nada[3]
[1]Department of Petroleum Technology, University of Technology, Baghdad-Iraq.
[2]Department of building and construction, University of Technology, Baghdad-Iraq.
[3]General Directorate for Environmental and Water Research, Ministry of Science and Technology, Baghdad-Iraq.
*E-mail: ahmad_2009_61@live.com

Abstract

Multivariable statistical approach were applied to evaluate groundwater quality in Jisser Diyala area, this technique was used to find the relation between physco-chemical properties of water and pattern of changing the groundwater quality in this area. Cluster and Factor analysis were applied in two type R-mode and Q-mode to identify the grouping of chemical properties and identify the factors controlling the change in quality. [DOI: 10.22401/JNUS.20.1.02]

Keywords: groundwater, quality, multivariable, cluster, factors analysis.

Introduction

The groundwater resources are one of the most important resources for human life, so it is important to study its quality and quantity to identify its suitability for different uses. The groundwater quality depend on geological situation, and the human activity are polluted the groundwater by adding some element to the natural water. The big growing in population and decreasing of water resources lead the researchers to find a new techniques for studying the water resources especially the groundwater.

The statistical approach or statistical analysis application for groundwater have been developed in the last decades because of numbers of programs which deal with the statistics, that can used in groundwater and hydrochemical data interpretation. The graphical interpretation of groundwater is limited because, there is a finite numbers of variables that can be considered, and the variables are generally limited by convention to major ions.

The multivariable approach is the more accurate in data interpretation from univariable approach because it avoid the limitation of major ions. This method was developed to identify the main trends in the hydrochemical evolution of a regional-scale system. [1]

Multivariate statistical methods were applicable for groundwater to extract high concentration from hydrochemical data sets in systems. These methods can give the solution to the hydrological factors like aquifer boundaries, ground water flow paths, or hydrochemical contents [2], identify geochemical controls on composition [3], and fell apart anomalies like artificial impacts from the background on a variety of scales [4]. These studies often use either or both R-mode (source of variation in a dataset from parameters) or Q mode (variation between samples) multivariate techniques [5].

The aim of this study are to apply the multivariable statistical methods to verify the hydrochamical situation and to find the relation between physco-chemical properties of water and pattern of changing the groundwater quality in this area.

a-Study area

The study area lies to the south of Baghdad and to the east of Diyala River, near the crossing with the Tigris River, and south of Al-Rustumia station for treatment of sewage water Fig.(1). This figure also shows the location of sampling point, which used for collect the groungwater samples from 6 wells distributed randomly in the study area, and one station for sampling river water.
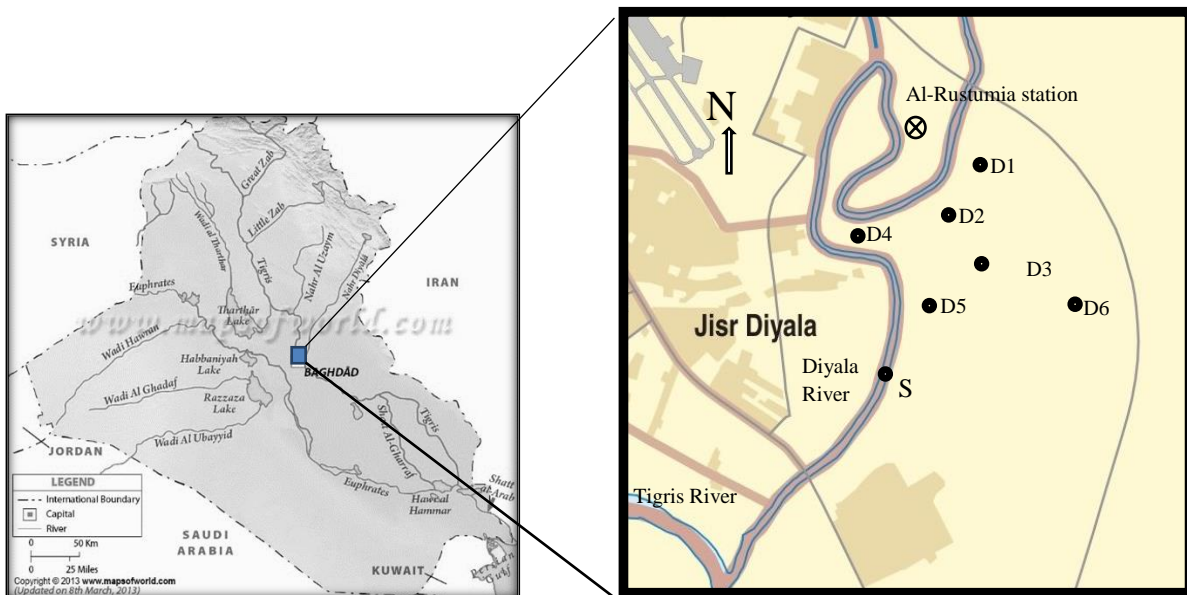
Fig. (1): Map of study area with location of sampling stations.

b-Geological and Hydrogeolgical Situation of Study Area

Study area lies within Baghdad Governorate which is a part of the sedimentary basin in the Mesopotamia Plain. It belongs to the Unstable Shelf and represents symmetrical syncline, filled with quaternary alluvial sediments, which belongs generally to Pleistocene and Holocene sediments. Those sediments composed mainly from clay, mud, silt and little bit of fine gravel, heterogeneously distributed, which cause unordered extent for ground water aquifers.

Clay and mud represent a confining bed at the upper part of quaternary sediments, its thickness almost between (10–20) m, while ground water available in the lower beds among fine sand and pebbles.

Ground water movement system in the study area represents part of the general movement system of Tigris Basin, considering Tigris and Diyala rivers as the discharge zone in general, The governorate ground water levels range between (> 25 − < 36) m above sea level, Groundwater where the movement from west to east and south-east, but the presence of the rivers (Tigris, Diyala and Euphrates), which represent important hydrogeological borders for aquifer systems specially in quaternary sediments, led to appearance of local movements due to the hydraulic connection between those rivers and ground water (recharge and discharge). Fig.(2).

Generally, the water table fluctuation depends on many natural and industrial factors, especially the amount and distribution of rainfall in the area, flood periods and high evaporation, which affect the shallow unconfined aquifers, in addition to digging of irrigation canals and the unwise draw from the wells. [6].
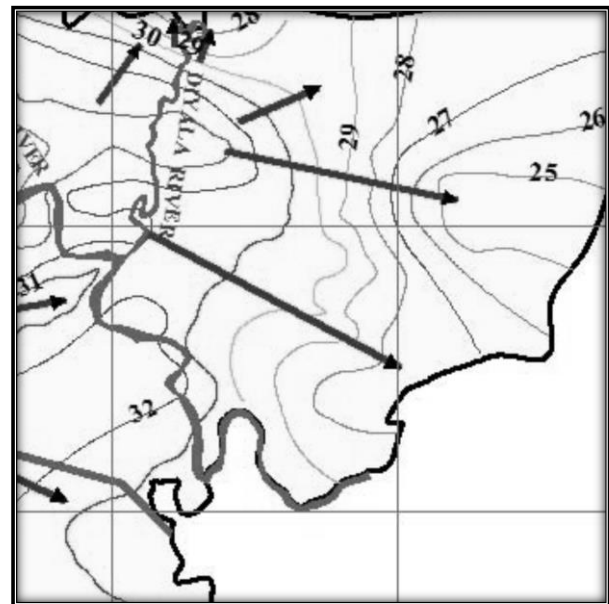

Fig.(2): flownet map of study area showing water level and flow direction. [6]

Materials and Methods
a-Field work

This work include collecting water sample from 6 wells (D1, D2, D3, D4, D5, D6) distributed in the study area and one station on Diyala river (S) as show in location map

Fig.(1) the sample collected from drilled well after pumping for 1 hour, and then collected the samples by 1 liter plastic bottle.

Three groups of sampling were carried out during 2008, the first group was carried out on March, and the second on June, and third group on August. The total number of samples are 21 samples.

b- Laboratory work

The laboratory work include the measurement of major anions and cations and some heavy metals according to APHA, ASTM standard methods. The chemical analysis includes:
1. Majer cation and anion (Ca, Na, Mg, Cl, HCO_3, SO_4).
2. Chemophysical element (pH, Ec, TDS, TSS, TH, Turb., DO, BOD).
3. Heavy metal or trace elements (Cd, Pb, Zn, Fe, Mn) for third group.

Results and Discussions

Table (1), show the result of laboratory analysis for the three group were sampled from the studied area, the table show the average for the three group. The pH ranging from 6.7 to 7.4 which indicate the groundwater and surface water as neutral. And the TDS for ground water ranging from 3012 to 4890 which reflect brackish type, while the surface water about 1860 which is slightly brackish water. For CaCO3 hardness is classified as very hard according to Freeze and Cherry, 1979. [7]

The major constituent of ground water ranging as shown in the table reflect the higher concentrations in relative to surface water. Which reflect the effect of geological processes on the groundwater quality.

Table (1)
Results of the average of chemical analysis in mg/l, 2008.

Physco-chemical variables	Wells No.						
	D1	D2	D3	D4	D5	D6	S
pH	7.3	7.4	7.3	6.7	7.3	7.2	7.3
Ec (µs/cm)	4890	4030	4810	4773	4139	3466	1498
DO	5.8	6.1	6.3	6.3	7.3	5.6	8.1
BOD	9.5	10.3	10.5	8.5	9.1	29	5.1
TDS	3966	3653	3807	3857	3949	3012	1860
TSS	13.3	22.3	40	45	43.3	88	43
T.H.	1446	1450	1435	1293	1177	1111	476
Ca	297	499	420	356	318	301	356
Mg	285	387	296	353	293	330	143
Na	537	529	544	566	528	478	338
Cl	620	432	492	482	466	366	181
SO4	1489	876	1665	1366	1203	913	620
HCO3	253	344	232	405	258	357	128
Cd	0.001	0.008	0.009	0.089	0.009	0.07	0.003
Pb	0.045	0.046	0.047	0.048	0.043	0.09	0.044
Zn	0.9	0.98	0.94	0.98	0.98	2.1	0.43
Fe	0.35	0.48	0.36	0.35	0.33	0.98	0.3
Mn	0.11	0.15	0.134	0.154	0.132	0.31	0.11

a-Statistical Evaluation

The statistical evaluation for the data of physco-chemical variables are used to interpret the water quality of the study area to give meaningful results. In this study Statistica software version 11 are used to carry out the statistical analysis, depending on using the cluster and factor analysis. The Table(2) shows the summery of basic statistical data for groundwater physco-chemical variables.

b-Basic Statistic

Table (2) shows the summery of statistics of physco-chemical data (variables) which are the iterance for Exploratory Data Analysis (EDA) which used to reflect the types of data and identify the shape of data through the distribution likes to get the normal distribution of data.

c-Normal Distribution

The Fig.(3) shows that the physco-chemical variables almost in shape to the normal distribution which reflect that the data center tendency are in normal distribution condition and data have a good or high confidence to make an interpretation of the quality of groundwater. [8]

Table (2)
The descriptive statistics of chemical analysis data in mg/l.

| Physco-chemical variables	Summery Statistics				
	Min.	Max.	Mean	Median	Std. Dev.
pH	6.7	7.4	7.21	7.3	0.234
Ec (µs/cm)	1498	4890	3943	4139	1196.4
DO	5.6	8.1	6.5	6.3	0.889
BOD	5.1	29	11.71	9.5	7.833
TDS	1860	3966	3443.4	3807	771.28
TSS	13.3	88	42.13	43	23.61
T.H.	476	1450	1198.3	1293	346.29
Ca	297	499	363.9	356	73.1
Mg	143	387	298.1	296	77.67
Na	338	566	502.9	529	77.41
Cl	181	620	434.1	466	135.4
SO4	620	1665	1161.7	1203	374.34
HCO3	128	405	282.4	258	93.3
Cd	0.001	0.089	0.027	0.009	0.0364
Pb	0.043	0.09	0.0519	0.046	0.0169
Zn	0.43	2.1	1.04	0.98	0.506
Fe	0.3	0.98	0.45	0.35	0.24
Mn	0.11	0.31	0.1571	0.134	0.06957
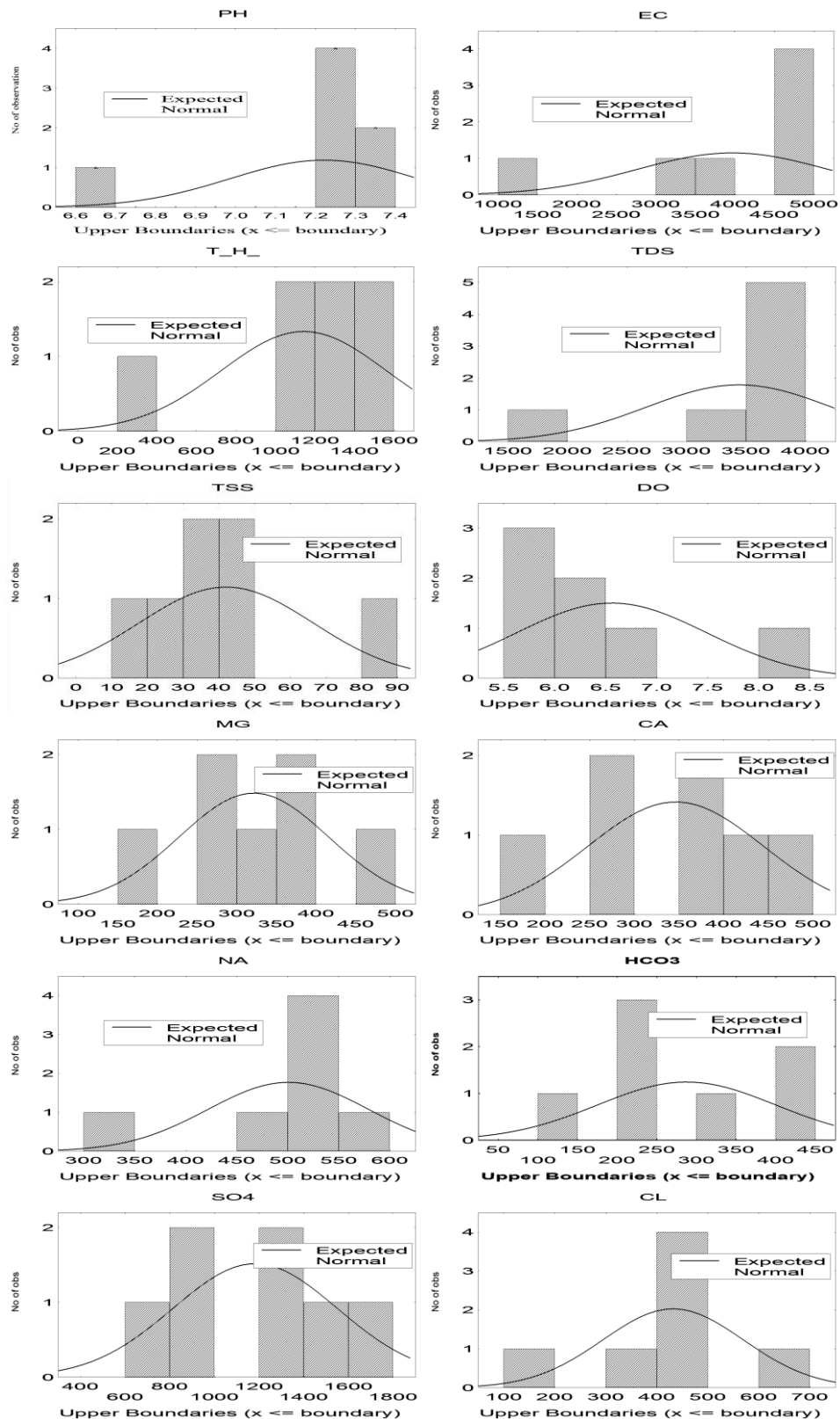
Fig.(3): The Normal distribution of physco-chemical variables.

d-Cluster analysis

Cluster analysis are used to grouping hydrochemical data, the grouping can be done in three trends, first observation clustering, second variable clustering, and K-means clustering. There are two types of K-means clustering. There are two types of cluster analysis, Q mode which used to grouping the

samples, and R mode which used to grouping variables, in this study the two modes were used as shown in Figures. (4, 5).
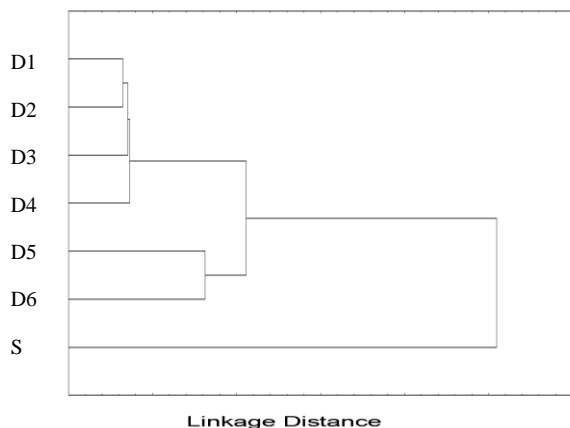

Fig.(4): Cluster analysis of samples showing the linkage distance.
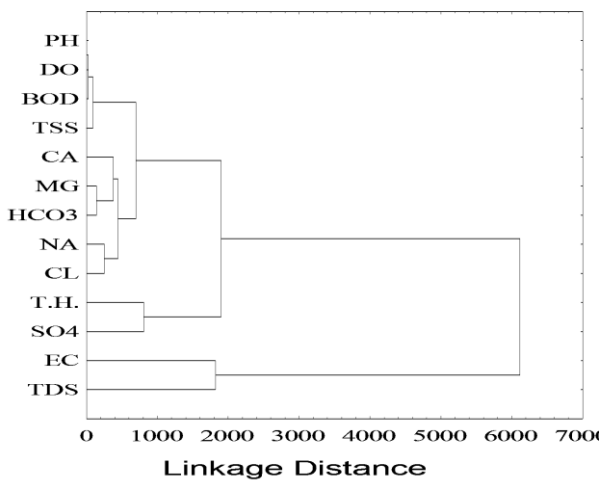

Fig.(5): The cluster analysis of chemical components of water samples.

From the cluster analysis (Q-mode) the result showing three groups, group A contains D1, D4, D3 and D5 have a high similarity in thier properties, group B contains D2 and D6, group C contains S alone as a group which represent the surface water.

Group A shows a similarity between D1 and D4 as a subgroup, this subgroup shows a similarity with D5 and then with D3 in close linkage distance, the linkage distance about 0.1, this close similarity which reflect the similarity in the types of water, which its Ca-Cl group as shown in Table (3), this group of groundwater was effected by the river water. The group B shows a less similarity between D2 and D6 because the linkage distance was bigger than its in group A, the linkage distance about 0.3 and its also have the same types of

groundwater (Na-HCO3) as shown in Table(3), and group C shows the nature of surface water which have a high dissimilarity with groundwater, the linkage distance about 1.0 this value refer to the maximum dissimilarity. Group A and B also show a similarity but at far linkage distance more than 0.4 of dissimilarity.

Table (3)
Hydrochemical formula and water quality of samples from this study.

Group	Well No.	Hydrochemical formula	Water type
Group 1	D1	$3980 \frac{So4-Cl-Hco3}{Na-Mg-Ca} 7.4$	Ca-Mg-sodium-Cl-Sulfate
	D2	$3600 \frac{So4-Cl-hco3}{Mg-Ca-Na} 7.3$	Na-Ca-Magnesium-Hco3-Cl-Sulfate
	D3	$3800 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.3$	Ca-Na-Magnesium-Cl-Sulfate
	D4	$3850 \frac{So4-Cl-hco3}{Mg-Na-Ca} 6.7$	Ca-Na-Magnesium-Cl-Sulfate
	D5	$3997 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.3$	Ca-Na-Magnesium-Cl-Sulfate
	D6	$3018 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.23$	Na-Magnisum-Hco3-Cl-Sulfate
	S	$1840 \frac{So4-Cl-hco3}{Ca-Na-Mg} 7.32$	Mg-Na-Calcium-Cl-Sulfate
Group 2	D1	$4000 \frac{So4-Cl-hco3}{Na-Mg-Ca} 7.2$	Ca-Mg-Sodium-Cl-Sulfate
	D2	$3800 \frac{So4-Cl-hco3}{Mg-Ca-Na} 7.2$	Na-Ca-Magnesium-Hco3-Cl-Sulfate
	D3	$3870 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.2$	Ca-Na-Magnesium-Cl-Sulate
	D4	$3890 \frac{So4-Cl-hco3}{Mg-Na-Ca} 6.5$	Ca-Na-Magnesium-Cl-Sulfate
	D5	$3999 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.2$	Ca-Na-Magnesium-Cl-Sulfate
	D6	$3020 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.1$	Ca-Na-Magnesium-Cl-Sulfate
	S	$1890 \frac{So4-Cl-hco3}{Ca-Na-Mg} 7.4$	Mg-Na-Calcium-Cl-Sulfate
Group 3	D1	$3920 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.3$	Ca-Na-Magnesium-Cl-Sulfate
	D2	$3560 \frac{So4-Cl-hco3}{Mg-Ca-Na} 9.7$	Na-Ca-Magnesium-Hco3-Cl-Sulphate
	D3	$3752 \frac{So4-Cl-hco3}{Mg-Na-Ca} 7.4$	Ca-Na-Magnesium-Cl-Sulfate
	D4	$3832 \frac{So4-Cl-hco3}{Mg-Na-Ca} 6.99$	Ca-Na-Magnesium-Cl-Sulfate
	D5	$3852 \frac{So4-Cl-hco3}{Na-Mg-Ca} 7.59$	Ca-Mg-Sodium-Cl-Sulfate
	D6	$3000 \frac{So4-Cl-hco3}{Ca-Na-Mg} 7.3$	Na-Calcium-Hco3-Cl-Sulfate
	S	$1852 \frac{So4-Cl-hco3}{Ca-Na-Mg} 7.45$	Mg-Na- Calcium-Cl-Sulfate

The multivariable approach is the more accurate in data interpretation from univariable approach because it avoid the limitation of the major ions. This method was developed to identify the main trends in the hydrochemical evolution of a regional-scale system in the studied area. [1]

In the last years, multivariate statistical methods were applied for groundwater to extract high concentration from hydrochemical data sets in systems. These methods can give the solution to the hydrological factors like aquifer boundaries, ground water flow paths, or hydrochemical contents [2], identify geochemical controls on composition [3], and fell apart anomalies like artificial impacts from the background on a variety of scales [4]. These studies often use either or both R-mode (source of variation in a dataset from parameters) or Q mode (variation between samples) multivariate techniques [5].

The cluster analysis (R-Mode) was carried out in this study as shown in Fig.(5); this cluster shows the similarity between the physco-chemicsl variables of the samples. The Fig.(5) show that the pH, DO, and BOD have the same similarity which means that the same behavior for all samples, in very close linkage

distance and then related with TSS in similarity as subgroup 1 reflected the physical properties of samples.

The Mg and HCO3 also have the same similarity and then related with Ca. Na and Cl have the same similarity too and related with HCO3, Mg, and Ca as subgroup 2, which reflect the geological process effect on groundwater on the variation of chemical content. The T.H. and SO4 are also reflect the hardness factor change in samples as subgroup 3. These three subgroup were also show similarity after a linkage distance about 2000, which reflect the maximum change in physco-chemical variables.

d-Factor Analysis

Before applied Factor Analysis on the data, or as iterance for using Factor Analysis, the applied of correlation matrix or correlation coefficient can point out the associations between variables, and that can show the coherent of data and refer to the situ of variables in the effected factors on the changing the quality of water. Table (3) shows the correlation matrix between the variables. The reference of coefficient value for significant change are (r>0.5). From this reference and from the Table (4) the Ec inspection related with TDS, T.H., Na, Cl, and SO4, this is a positive related. pH have a negative related with HCO3. Ca show no positive or negative correlation with any variables.

The Fig.(6). Show the regression relation between Ec and TDS, and the r-squared is about 95% and this value are compatible with value from correlation coefficient.

Table (4)
The correlation matrix of chemical variables of water samples.

	pH	Ec	DO	BOD	TDS	TSS	T.H.	Ca	Mg	Na	Cl	SO$_4$	HCO$_3$
pH	1.00												
Ec	-.25	1.00											
DO	.28	-.77	1.00										
BOD	.01	.01	-.44	1.00									
TDS	-.18	.98	-.75	-.05	1.00								
TSS	-.23	-.27	-.00	.81	-.38	1.00							
T.H.	-.25	.88	-.92	.18	.88	-.19	1.00						
Ca	-.01	.04	-.01	-.62	.12	-.64	.27	1.00					
Mg	-.27	.42	-.80	.76	.43	.46	.66	-.24	1.00				
Na	-.29	.97	-.84	-.00	.98	-.34	.94	.17	.51	1.00			
Cl	-.09	.93	-.79	.00	.92	-.42	.82	.01	.36	.92	1.00		
SO$_4$	-.14	.89	-.51	-.17	.81	-.34	.66	.06	.04	.77	.83	1.00	
HCO$_3$	-.55	.38	-.80	.69	.34	.47	.61	-.21	.92	.47	.33	.04	1.00

Ec = 1.51704 * TDS + -1249.69
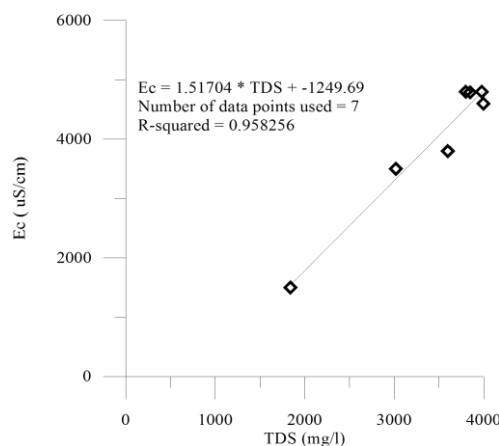Number of data points used = 7
R-squared = 0.958256

Fig.(6): The regression relation between Ec and TDS.

According to the correlation coefficient facts the factor analysis show effect of three factors as shown in Table (5). The factor 1 reflect 50% effect on groundwater change in quality and that reflect the geological factor (Ec, DO, TDS, T.H., and Cl. The factor 2 reflect 26% effect on changing the recharge effect on groundwater, agriculture and irrigation. And factor 3 reflect the others effect on quality of groundwater which was 9%.

Table (5)
The factor analysis (principal component analysis).

	Factor 1	**Factor 2**	**Factor 3**
pH	-0.307132	0.233936	-0.639007
Ec	0.930980	0.198654	-0.148462
DO	-0.908324	0.248987	-0.074797
BOD	0.194239	-0.901381	-0.261943
TDS	0.917054	0.258664	-0.110354
TSS	-0.162882	-0.895489	-0.084472
T.H.	0.941306	0.021198	0.135669
Ca	0.058302	0.608081	0.631545
Mg	0.625592	-0.682997	0.113042
Na	0.955208	0.185232	0.026325
Cl	0.890300	0.252078	-0.247545
SO_4	0.725905	0.405025	-0.298405
HCO_3	0.603708	-0.693078	0.312981
Expl.Var	6.568633	3.416434	1.202901
Prp.Totl	.505279	.262803	.092531

Conclusions and Recommendations

1. The data set were used in this study had a good normal distribution which reflect or refer to no error in chemical analysis and it is have a natural behavior.
2. Cluster analysis Q-mode show three groups of samples the first one include (D1,D3,D4.D5) represent the groundwater effected by surface water from Diyala river, the second group (D2, D7) represent the wells of natural groundwater, and the third group (S6) represent the surface water.
3. Cluster analysis R-mode show the relation between different variables of water.
4. Factor analysis produced three Factors effected the quality of groundwater, the first factor represent the geological factor and total effect of this factor about 50%, the second factor effected groundwater represent the recharge and irrigation factor, and its ratio about 26%, the third factor effect about 9% represent others process.
5. The recommendations are focus on using the statistical approach in others studies to verification of scientific trues resulting from traditional techniques.

References

[1] Güler C., and G.D. Thyne. "Hydrologic and geologic factors controlling surface and ground water chemistry in Indian Wells– Owens Valley area and surrounding ranges", California, USA. *Journal of Hydrology* 285, 177–198. 2004.
[2] Locsey K.L., and M. E. Cox. Statistical and hydrochemical methods to compare basalt- and basement rock-hosted groundwaters: Atheron Tablelands, northeastern Australia.

Environmental Geology. V. 43 (6), pp. 698 – 713. 2003.

[3] Adams S., Titus R., Pietersen K., Tredoux G., Harris C., Hydrochemical characteristics of aquifers near Sutherland in the western Karoo, South Africa. Journal of Hydrology 241 (1–2), 91–103. 2001.

[4] Briz-Kishore B.H., Murali G., Factor analysis of revealing hydrochemical characteristics of a watershed. Environmental Geology and Water Sciences 19 (1), 3–9. 1992.

[5] Dalton M.G., Upchurch S.B., "Interpretation of hydrochemical facies by factor analysis", ground water, Vol. 16, No. 4, pp. 228-233. 1978.

[6] Al-Basrawi N.H., Awad J.H., Hussian T.A., "Evaluation of the Ground Water in Baghdad Governorate / Iraq", Iraqi Journal of Science, Vol 56, No.2C, pp: 1708-1718. 2015.

[7] Freeze R.A., and Cheery R.A., Groundwater, Prentice Hall; 1st edition, pp. 604. 1979.

[8] Montgomery R.H., et. Al., Statistical Characterization of Groundwater Quality Variables, Ground Water Journal, Vol. 25. No. 2, pp. 176-184. 1987.

الخلاصة:

طبق الاسلوب الاحصائي متعدد المتغيرات في تقييم نوعية المياه الجوفية في منطقة جسر ديالى لغرض ايجاد علاقات الارتباط بين المتغيرات والتي تمثل الصفات الفيزيائية والكيميائية للمياه الجوفية لتحديد نمط التغير بنوعية المياه الجوفية وتحديد العوامل المؤثرة عليها. شمل التطبيق استخدام التحليل التجميعي (Cluster Analysis) والتحليل العاملي (Factor Analysis) ومن خلالهما تم تحديد المجاميع المرتبطة والعوامل المؤثرة على تغير نوعية المياه الجوفية.

الكلمات الدالة: المياه الجوفية، النوعية، متعدد المتغيرات، التحليل التجميعي، التحليل العاملي.