TJES

ISSN: 1813-162X
Tikrit Journal of Engineering Sciences
available online at: http://www.tj-es.com

Classification of Selected Ground Water Wells within and around Mosul City according to their Water Quality Using Factor and Cluster Analysis

Abdulmuhsin S. Shihab[a], Abeer Hashim[b]

[a]Center of Environmental Research and Pollution Control
[b]Environmental Engineering Department, College of Engineering, University of Mosul
Nenawa Province, Republic of Iraq.

ABSTRACT

The research aimed to classify 66 wells within and around Mosul city according to their water quality using cluster analysis. Water samples were collected and analyzed for pH, total dissolved solids, conductivity, calcium, magnesium, chloride, sulphate and bicarbonate using standard methods. The data were analyzed statistically using factor and cluster analysis. The results of factor analysis show four groups. Conductivity, total dissolved solids, sulphate and calcium represents the first group with the highest percent of variation (30.55%) between wells. Cluster analysis divided the wells into four homogenous clusters. The first cluster represents 15(22.7%) of the wells, most of the wells of this cluster are distributed along Tigris river with lowest pH, highest sulphate and bicarbonate concentration. The second cluster includes the largest number of wells 33(50%) with the lowest salinity since it had the lowest conductivity, total dissolved solids, calcium, magnesium and chloride. The third cluster with 4(6.1%) wells, had the highest salinity since it had the highest conductivity, total dissolved solids, calcium, magnesium and chloride. The fourth cluster includes 14(21.2%) of less acidity wells with highest pH and highest bicarbonate concentration. The research concluded that cluster analysis could be used as an efficient statistical grouping tool according to water quality parameters. Additionally, factor analysis can be used to analyze a large number of data and study the variation in water quality.

Keywords: Wells, Water Quality Parameters, Cluster Analysis, Factor Analysis.

تصنيف نوعية المياه الجوفية لمجموعة من الآبار داخل مدينة الموصل وحولها باستخدام التحليل العاملي والعنقودي

الخلاصــة

تم استخدام التحليل العنقودي لتصنيف 66 بئرا خلال وحول مدينة الموصل حسب نوعية المياه ونتائج هذا التصنيف قد تكون نافعة عند التخطيط لاستخدام المياه الجوفية في المنطقة . تم جمع عينات مياه الآبار وإجراء التحاليل الفيزياوية والكيمياوية والمتمثلة بالرقم الهيدروجيني والمواد الصلبة الذائبة الكلية والتوصيل الكهربائي والكالسيوم والمغنيسيوم والكلوريدات والكبريتات والبيكاربونات باستخدام الطرق القياسية المتبعة في فحص عينات مياه الآبار . تم تحليل البيانات إحصائيا باستخدام التحليل العاملي والعنقودي . استخلص التحليل العاملي أربعة عوامل حمل التوصيل الكهربائي والمواد الصلبة الذائبة الكلية والكبريتات والكالسيوم على العامل الأول والذي شكل اعلى نسبة تغاير في نوعية المياه بين الابار بمقدار (30.55%) . أما نتائج التحليل العنقودي فقد قسمت الآبار إلى أربعة عناقيد متجانسة. اشتمل العنقود الأولى على عدد من الابار بواقع 15(22.7%) التي توزعت على امتداد نهر دجلة وامتازت بأوطأ قيم للرقم الهيدروجيني وأعلى تراكيز للكبريتات والبيكاربونات. وتضمن العنقود الثاني العدد الأكبر من الآبار بواقع 33(50%) وامتازت بأوطأ قيم للملوحة ضمن الابار المدروسة من خلال اقل قيم للتوصيل الكهربائي والمواد الصلبة الذائبة الكلية والكالسيوم والمغنيسيوم والكلوريدات. أما العنقود الثالث فقد اشتمل على 4(6.1%) من الابار والتي امتازت بأعلى قيم للملوحة من خلال اعلى قيم للتوصيل الكهربائي والمواد الصلبة الذائبة الكلية والكالسيوم والمغنيسيوم والكلوريدات . واشتمل العنقود الاخير على 14(21.2%) بئراً امتازت بانخفاض قيم الحامضية من خلال ارتفاع قيم الدالة الحامضية وارتفاع تراكيز البيكاربونات بالنسبة للآبار المدروسة. واستنتج البحث ان نتائج التحليل العنقودي يمكن ان تكون ذات فائدة كبيرة في مجال ادارة نوعية المياه ، وان التحليل العنقودي هو اداة كفوءة في تصنيف

نوعية المياه حسب عوامل نوعية المياه مجتمعة ، فضلا عن ان التحليل العامل ذات قدرة على التعامل مع عدد كبير من البيانات ودراسة التغاير في نوعية المياه .

INTRODUCTION

Conventional studies of ground water have placed a heavy emphasis on the variations in the chemical characteristics of ground water in time and space[1]. Therefore, many researchers have performed multiple ground water sampling and subsequent chemical analyses. The main tools for interpretation of chemical analysis results are graphical methods combined with basic statistics (e.g. average, frequency, correlation)[2-5]. Other researchers studied the suitability of ground water for specified use by comparing the results with the standards or classify ground water quality by Piper diagram [6-10].

For a better understanding of ground water quality system, multivariate analyses can be performed. As indicated in Suk and Lee [11], spatial or temporal measurements of chemical or physical properties usually do not directly reveal the underlying governing processes in the ground water system of interest. Factor and cluster analyses have been employed to reveal the most important governing process and hydrogeochemical similarities between observation points through data reduction and classification. Several researchers have applied factor and/or cluster analyses of ground water chemical data in order to understand ground water systems [12-14].

The interest of the present study is to classify the studied wells and grouped them according to their water quality using cluster analysis and analyze ground water quality relationships within and around Mosul city.

MATERIALS AND METHODS

The samples of groundwater collected from 66 wells of studied area within and around Mosul . It is bounded by Mosul city dam lake in the North to Al-Hatra in the South and from Bashiqa in the East to Ba'aj in the West with an area of about 232 km^2. The number of the studied wells are 66. Water samples were collected from the wells within Mosul city. The depth of the wells ranged from 5-14 meter in Mosul city and from 15-90 meter around the city. pH, total dissolved solids, conductivity, calcium, magnesium, chloride, sulphate and bicarbonate were tested according to the standard methods [15]. In addition, to the data obtained from the record of Water Wells, Drilling Company in Mosul city are also used.

The statistical analysis includes descriptive statistics of water quality parameters presented by the mean, standard deviation and range. Correlation analysis was conducted to show the relationships between the measured parameters. Factor analysis with varimax rotation was conducted on the standardized data [16, 17], and the factor loadings were obtained. Hierarchical cluster analysis was used to group studied wells. Ward method, which is considered as anefficient, was applied since it uses the analysis of variance approach to evaluate the distances between clusters. This method also minimizes the sum of squares of any two clusters that can be formed at each step [18].

RESULTS AND DISCUSSION

The descriptive statistics for the measured water quality parameters of groundwater wells are shown in Table (1). There is a wide variation in water quality, which is clear from the minimum, maximum, and standard deviation. For example total dissolved solids ranged between 116 to 6886 mg/l with a standard deviation of 1327.7. These wide variations are due to the distribution of the studied wells along large area of different feeding sources and geological formations, which is more suitable to conduct cluster analysis and classify the wells.

A bivariate correlation between measured parameters for 66 wells, shows that these parameters are significantly correlated with each other, except for chloride with bicarbonate and pH; Mg with $SO_4^{=}$ and pH (Table 2).

Factor analysis with rotation show four groups from standardized data depending on scree plot (Fig. 1). The purpose of factor rotation is to yield a factor structure that is simpler. Also it is needed when factor loadings plot highly on more than one axis [19].

The extracted factors explained 86.25% of the variance in water quality among the studied wells (Table 3). The factor loadings which reflect the correlations between the variables and the extracted factors are shown in Table (3).

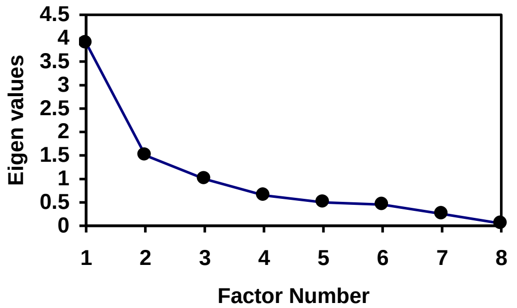


Fig. 1. Scree plot for selection of number of factors.

Table 1. Descriptive statistics of water quality of the measured parameters in the studied wells.

Par.	Mean	SD	Min.	Max.
pH	7.66	0.47	6.65	8.40
EC (μhos/cm)	2609	1657.61	345.0	8800.0
TDS (mg/l)	2036	1323.70	116.0	6886.0
Ca^{+2} (mg/l)	271	244.23	13.0	1122.0
Mg^{+2} (mg/l)	189	236.59	2.10	1775
Cl^{-} (mg/l)	322	791.67	18.0	5610.0
$SO_4^{=}$ (mg/l)	1000	704.18	20.0	2800.0
HCO_3^{-} (mg/l)	288	143.26	3.0	620.0

The loadings greater than 0.6 are bolded.

Factor I, accounts for 30.55% of the variation in water quality of the ground water of the studied wells. It is dominated by salinity, which is represented by electrical conductivity, total dissolved solids, sulphate as anion and calcium as cation. The parameters of this factor are significantly correlated between them. This factor reflects the variation in dissolution of rocks encountered in the studied area.

Lesser variation (19.995%) was observed in pH and bicarbonate, which is explained by factor II. The parameters of this factor have significant negative correlation between them.

Factor III is dominated by magnesium. It represents 18.61% of variation in water quality, which represents the variation of dolomite rocks in the studied area. While factor IV accounts for 17.09% of the variation, which is dominated by chloride.

The results of factor analysis reflect the variation in groundwater quality within the studied area. These variations are attributed to different abilities of rocks dissolution and the variation in the type of rocks and salts encountered in the geological formation within the studied area, in addition to the different feeding sources.

Cluster analysis is a method used to combine the studied wells into homogenous groups according to their water quality. In this analysis, the percentage of variation explained for the number of clusters 2,3 and 4 were 48.0%, 68.0% and 92.9% respectively, i.e. as the number of clusters increased, the percentage of variation explained also increased. When four clusters are formed (Fig. 2), the first cluster includes 15(22.7%) wells, with no sub-clusters. Most of the wells of this cluster are within Mosul city and distributed along the two sides of Tigris river as shown in Fig. (3). The water of these wells had the lowest mean pH values, highest bicarbonate concentrations and highest sulphate concentration with means 7.12, 445.53 mg/l and 1694.33 mg/l; and ranges 6.65-7.48, 250-620 and 965-2200 respectively, (Table 4 and Figs. 4-6). These results coincided with the results of Kalander and Al-Joboury [20] who found an increase in bicarbonate concentration and decrease in pH values in the ground water wells east Tigris river where sandstone is found.

The second cluster includes the largest number of wells 33(50%) with two sub-clusters and 4 sub-sub-clusters (Fig. 2). Most of the wells of this cluster are distributed around Mosul city with few wells in this city. The wells of this cluster has the

Table 2. Correlation matrix for the studied parameters.

	EC	**TDS**	**Ca**	**Mg**	**Cl**	**$SO^{4=}$**	**HCO_3^-**	**pH**
EC	1							
TDS	0.932**	1						
Ca	0.743**	0.716**	1					
Mg	0.475**	0.547**	0.303**	1				
Cl^-	0.356**	0.334**	0.595**	0.234*	1			
$SO_4^=$	0.560**	0.511**	0.508**	0.166	0.248*	1		
HCO_3^-	0.416**	0.358**	0.258*	0.306**	0.052	0.309**	1	
pH	0.398**	0.296**	0.298**	-0.09	0.045	0.464**	-0.549	1

* Significant at $p<0.05$, ** Significant at $p<0.01$

Table 3. Factor loading for the water quality parameters of the studied wells.

Parameters	**Factor**			
	1	**2**	**3**	**4**
pH	-0.366	**-0.816**	0.110	0.040
EC (_hos/cm)	**0.788**	0.209	0.449	0.208
TDS (mg/l)	**0.771**	0.114	0.540	0.169
Ca^{++} (mg/l)	**0.648**	0.147	0.193	0.595
Mg^{++} (mg/l)	0.131	0.100	**0.902**	0.115
$Cl^{□}$ (mg/l)	0.163	-0.087	0.098	**0.955**
$SO_4^{□}$ (mg/l)	**0.790**	0.298	-0.124	0.119
$HCO_3^{□}$ (mg/l)	0.088	**+0.866**	0.328	-0.008
% Variation	30.550	19.995	18.610	17.090
Cumulative %	30.550	50.545	69.16	86.25

Table 4. Mean water quality parameters for the clusters constructed.

Cluster No. / Parameters	**Mean ± SD**			
	1	**2**	**3**	**4**
pH	7.1 ± 0.3	7.8 ± 0.4	7.6 ± 0.6	8.0 ± 0.2
EC (□hos/cm)	3755.3 ± 747.6	1459.6 ± 679.6	6550.0 ± 2531.8	2967.5 ± 856.3
TDS (mg/l)	2769.3 ± 487.3	1081.8 ± 557.1	4974.3 ± 2136.5	2663.7 ± 821.7
Ca (mg/l)	364.3 ± 153.4	146.6 ± 131.1	869.0 ± 374.8	294.1 ± 195.6
Mg (mg/l)	262.3 ± 96.0	98.9 ± 92.0	770.5 ± 671.0	157.8 ± 115.4
Cl (mg/l)	190.1 ± 71.6	92.9 ± 55.7	2439.5 ± 2273.9	400.9 ± 622.6
SO_4 (mg/l)	1694.3 ± 409.0	489.5 ± 401.9	1259.0 ± 870.6	1387.8 ± 571.2
HCO_3 (mg/l)	445.5 ± 96.1	249.1 ± 119.3	367.0 ± 160.2	190.1 ± 83.7

lowest salinity in the studied wells since it has the lowest EC, TDS, Ca^{++}, Mg^{++} and Cl^{-} concentration, with means of 1459.6, 1081.8, 146.61and 98.93; ranges 345-2900 ms/cm, 116-2000, 13-505 and 2.1-523 respectively (Table 4 and Fig. 7-11).

On the other hand,3rd cluster had the worst ground quality since it has the highest salinity represented by EC, TDS Ca^{++}, Mg^{++} and Cl^{-} with means 6550, 4974.25, 869, 770.5 and 2439.5; ranges 3800-8800, 2117-6880, 312-1122, 390-1775 and 200-5610 respectively (Table 4 and Figs. 7-11). This cluster was the smallest one, it includes only 4(6.67%) of wells as shown in Fig. (3).

The fourth cluster includes 14(21.2%) wells. It has high pH values and lowest bicarbonate concentration with means 8.0 and 190.1; ranges of 7.6-8.25 and 89-350 respectively (Table 4 and Figs. 4-5). These well are distributed around Mosul city (Fig. 3)

CONCLUSIONS

1. The highest variation in ground water quality for the studied area was in conductivity, total dissolved solids and sulphate which reflect the variation in the abilities of rocks dissolution and the variation in the type of rocks and salts encountered in the geological formation within the studied area, in addition to the different feeding sources.
2. Cluster analysis efficiently divides the wells included in this research into four homogenous clusters. The first cluster with lower pH and highest bicarbonate and sulphate concentrations along the sides of Tigris River, the second cluster has lowest salinity distributed around Mosul city, the third cluster had the higher salinity and the fourth cluster was more acidic and had the higher pH.
3. Cluster analysis was found as an efficient statistical grouping tool.

REFERENCES

1. Kennedy L.G., Everett, J.W., Dewers, T., Pickins W. and Edwards D. (1999). Application of mineral iron and sulfide analysis to evaluate natural attenuation at fuel contaminated site. J. Environ. Eng. 125(1): 47-56.
2. Montgomery, R.H., Loftis J.C. and Harris J. (1987). Statistical characteristics of ground water quality variables. Ground Water 25(2): 176-184.
3. Hem, J.D. (1992). Study and interpretation of the chemical characteristics of natural water. US Government Printing Office, Washington, D.C.
4. Frapporti, G., Vriedn S.P. and Gaans P.F.M. (1993). Hydrochemistry of the shallow Dutch groundwater. Interpretation of the national groundwater quality monitoring network. Water Resour. Res. 29(9): 2993-3004.
5. Al-Rawi S.M. and Shihab A.S. (2005). Application of factor analysis as a tool for water quality management of Tigris river within Mosul city. Raf. J. Sci. 16(1): 56-64.
6. Jin-Yong Lee, Jung-Chan Choi & Kang-Kun Lee. (2005) Variations in heavy metal contamination of stream water and groundwater affected by an abandoned lead–zinc mine in Korea. Environmental Geochemistry and Health 27: 237–257.
7. Ramkumar, T., Venkatramanan, S., Anitha Mary, I., Tamilselvi, M. and Ramesh, G. (2010) Hydrogeo-chemical Quality of Groundwater in Vedaraniyam Town, TamilNadu, India. Research Journal of Environmental and Earth Sciences 2(1): 44-48.
8. Al-Layla M.A., Al-Rawi S.M. and Al-Kawaz H.A. (1990). Physico-chemical evaluation of groundwater around Mosul lake used for drinking and domestic purposes. 2nd Scientific Conference of Mosul Dam Centre/Mosul University, pp. 173-188.
9. Habib R.H., Al-Saigh N.H. and Hassan Z.M. (1990). Geochemistry of underground water in Erbil city-Iraq. 2nd Scientific Conference of Mosul Dam Centre/Mosul University, pp. 173-188.
10. 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. 2nd Scientific Conference of Mosul Dam Centre/Mosul University, pp. 173-188.
11. Suk, H. and Lee K.K. (1999). Characterization of a ground water hydrochemical system through multivariate analysis: clustering into

ground water zones. Ground Water 37(3): 358-366.

12. Usunoff E.J. and Guzman-Guzman A. (1989). Multivariate analysis in hydrochemistry: an example of the use of factor and corresponding analyses. Ground Water 27(1): 27-34.
13. Ritzi R.W., Wright S.L., Mann B. and Chen M. (1993). Analysis of temporal variability in hydrogeochemical data used for multivariate analyses. Ground 31(2): 221-229.
14. Ochsenkuhn, K.M. Kontoyannakos, J. and Ochsenkuhn-Petroppulu M. (1997). A new approach to a hydrochemical study of groundwater flow. J. Hydrol. 194: 64-75.
15. APHA, AWWA and WPCF (1992). Standard Methods for the Examination of Water and Wastewater. 18th ed., APHA, Washington, D.C.
16. Kaiser, H.F. 1958. The varimax criterion for analytic rotation in factor analysis. Pyrometrical 23: 187-200.
17. Davis, J.G. 1973. Statistics and data analysis in geology. John Wiley and Sons, Inc. New York,
18. Ward J.H. 1963. Heirarchical grouping to optimize an objective function. J. Am. Statistical Assoc. 58: 236-244.
19. Thurstone L.L. 1947. Multiple factor analysis. Chicago: University of Chicago Press.
20. Qalinder, L. 1989. Hydrology of left bank of Mosul city – East Tigris river. Iraqi J. of Hydrology 22(1): 65-74 (In Arabic).

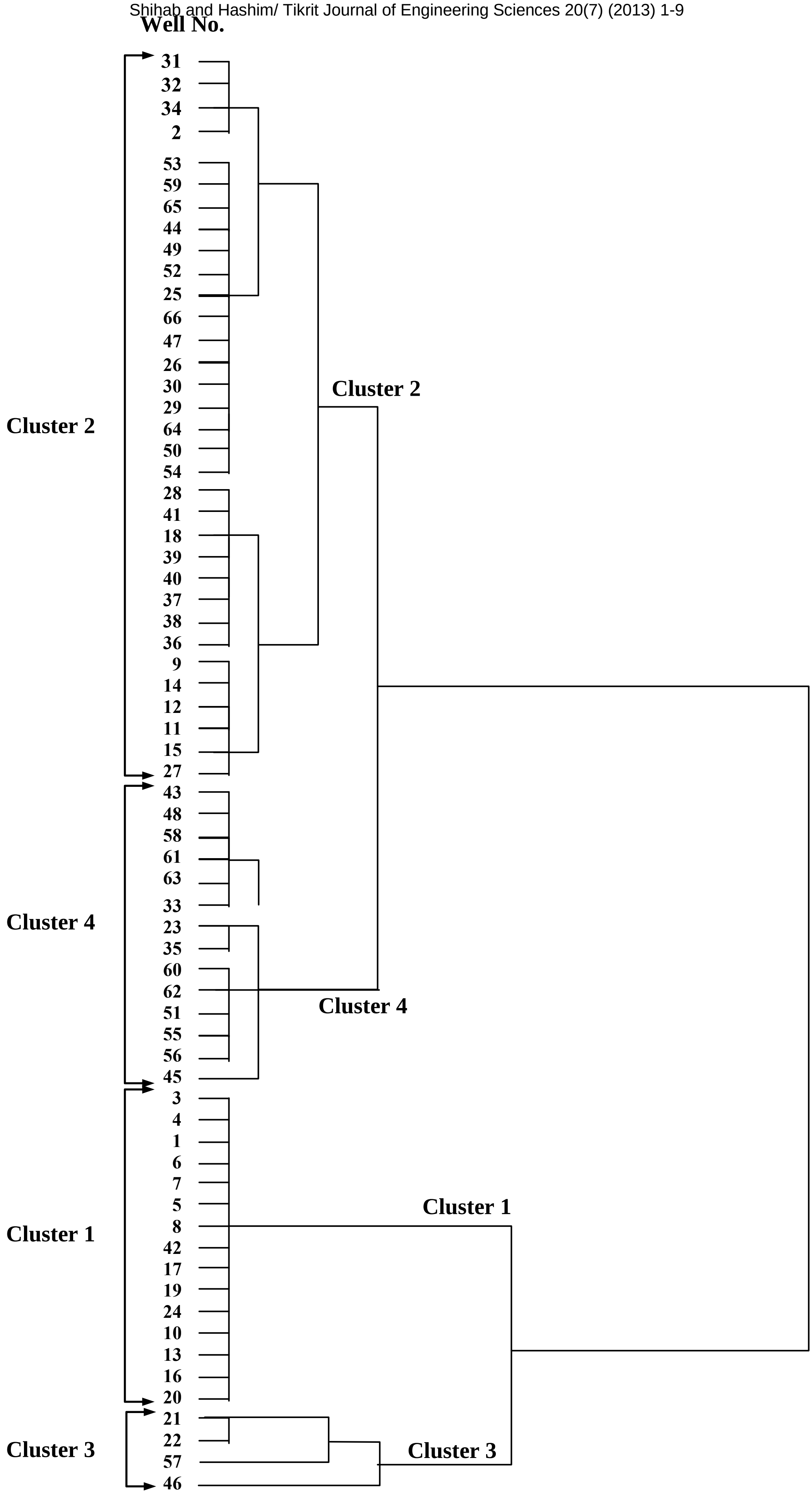


Fig. 2. Hierarchical cluster analysis results for 66 wells according to Ward method.

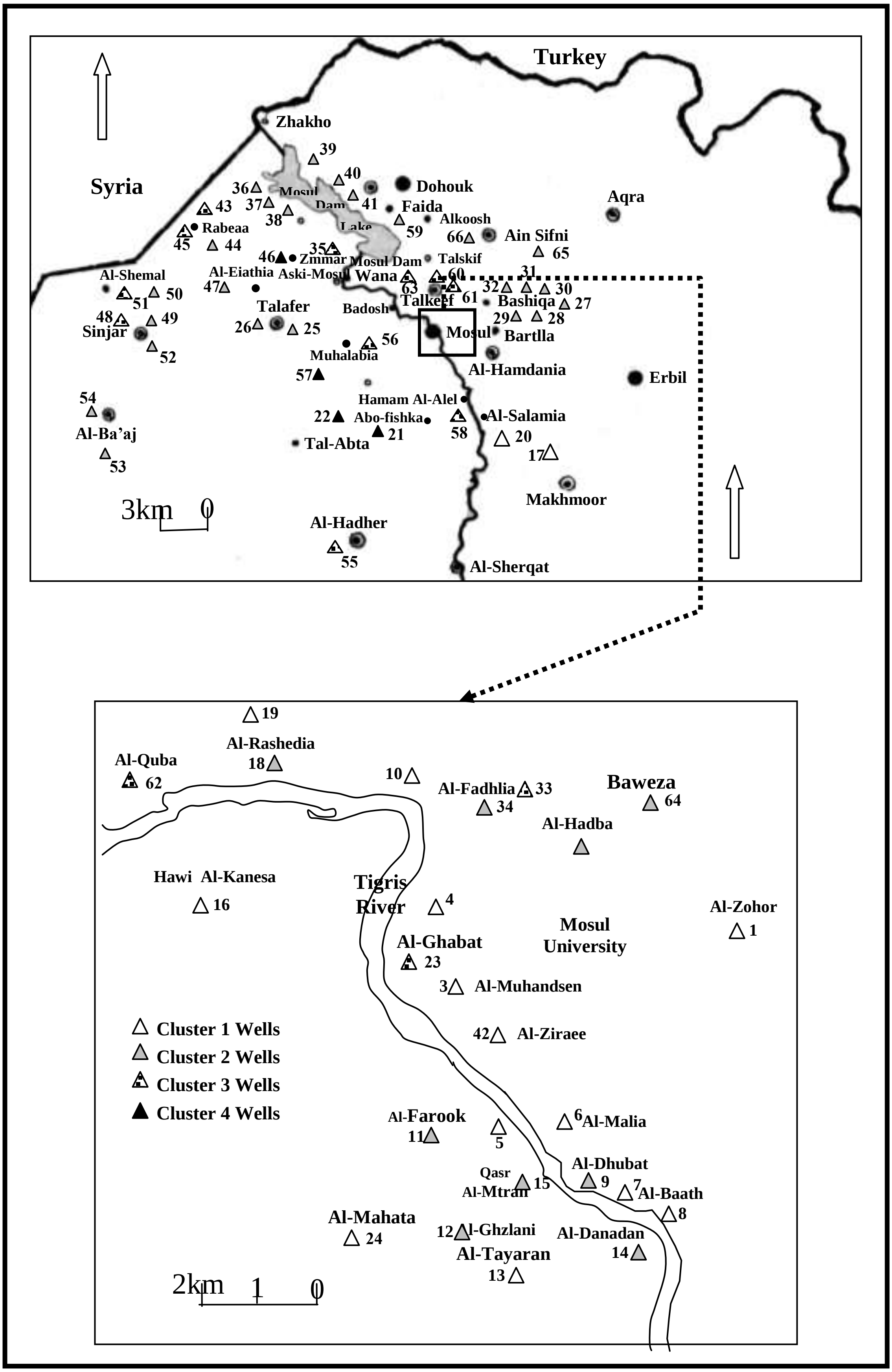
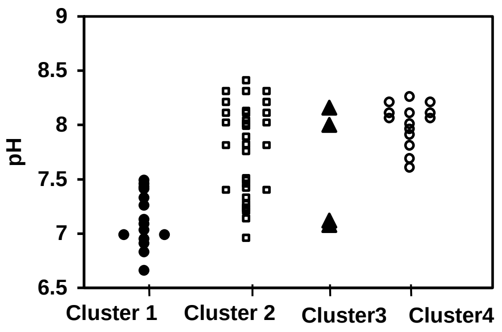


Fig. 3. Location map showing the wells included in the study and their classification according to cluster analysis.



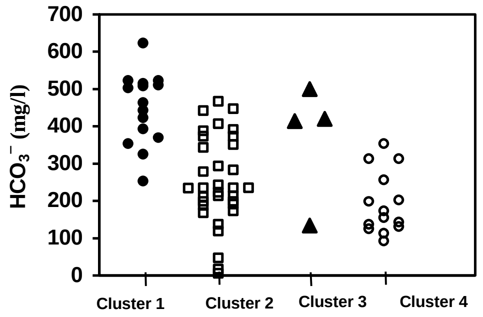
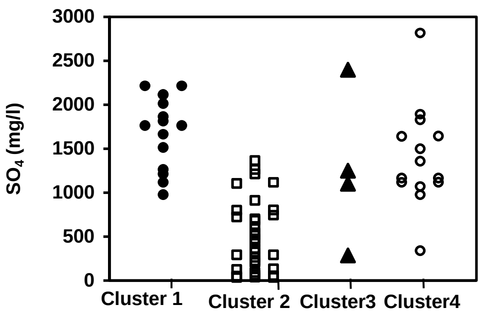


Fig. (4) Distribution of extracted clusters according to pH. **Fig. (5)** Distribution of extracted clusters according to HCO3-



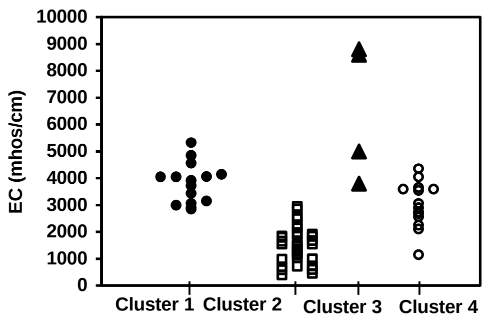
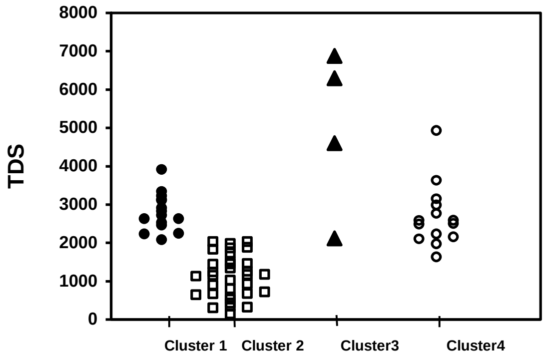


Fig. (6) Distribution of extracted clusters according to sulphat **Fig. (7)** Distribution of extracted clusters according to EC.



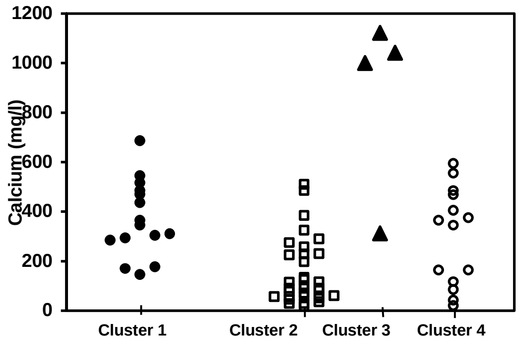
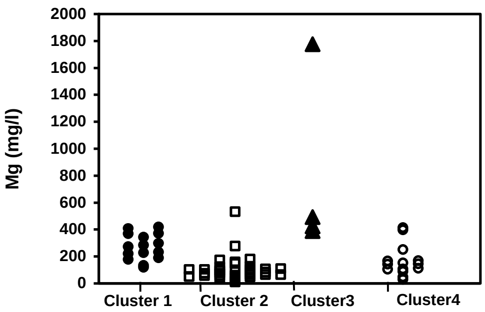


Fig. (8) Distribution of extracted clusters according to TDS. **Fig. (9)** Distribution of extracted clusters according to Ca.



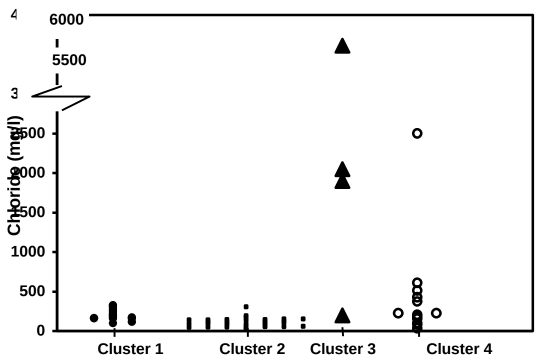


Fig. (10) Distribution of extracted clusters according to Mg. **Fig. (9)** Distribution of extracted clusters according to Cl.