JOURNAL OF
ENVIRONMENTAL HYDROLOGY

The Electronic Journal of the International Association for Environmental Hydrology
On the World Wide Web at http://www.hydroweb.com

VOLUME 14 2006

CLUSTER ANALYSIS CLASSIFICATION OF GROUNDWATER QUALITY IN WELLS WITHIN AND AROUND MOSUL CITY, IRAQ

Abdulmuhsin S. Shihab[1]
Abeer Hashim[2]

[1]Environmental Research Centre and Pollution Control, Mosul Univ., Mosul, Iraq
[2]Civil Engineering Dept., College of Engineering, Mosul Univ., Mosul, Iraq

Cluster analysis was used to classify 66 wells within and around Mosul city according to groundwater quality. This classification may provide useful results in the planning for groundwater use in this area. Water samples were collected and analyzed for pH, total dissolved solids, conductivity, calcium, magnesium, chloride, sulfate and bicarbonate using standard methods. The data were analyzed statistically using factor and cluster analysis. The results of factor analysis extracted four factors. Conductivity, total dissolved solids, sulfate and calcium represents the first factor 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 sulfate 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 included 14(21.2%) of less acidity wells with highest pH and lowest bicarbonate concentration. The research concluded that cluster analysis can provide useful information in water quality management as it is an efficient statistical grouping tool for water quality parameters. Additionally, factor analysis can be used to analyze a large number of data and study variations in water quality.

INTRODUCTION

Conventional studies of groundwater have placed a heavy emphasis on the variations in the chemical characteristics of groundwater in time and space (Kennedy et al., 1999). Therefore, many researchers have performed multiple groundwater 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) (Montgomery et al., 1987; Hem, 1992; Frapporti et al., 1993; Al-Rawi and Shihab, 2005). Other researchers study the suitability of groundwater for specified use by comparing the results with the standards or classify groundwater quality by Piper diagram (Al-Layla et al., 1990; Habib et al., 1990; Al-Rawi et al., 1990).

For a better understanding of groundwater quality systems, multivariate analyses can be performed. As indicated in Suk and Lee (1999), spatial or temporal measurements of chemical or physical properties usually do not directly reveal the underlying governing processes in the groundwater system of interest. Factor and cluster analyses have been employed to reveal the most important governing processes and hydrogeochemical similarities between observation points through data reduction and classification. Several researchers have applied factor and/or cluster analyses to groundwater chemical data in order to understand groundwater systems (Usunoff and Guzman-Guzman, 1989).

The present study analyzes groundwater quality relationships within and around Mosul city, classifies the studied wells and groups them according to their water quality using cluster analysis. The results of this study will be useful in the planning for groundwater use in this area.

MATERIALS AND METHODS

Sixty-six wells were included in this study. These wells lie within and around Mosul city from Mosul dam lake in the North to Al-Hatra in the South and from Bashiqa in the East to Ba'aj in the West, covering an area of about 232 km^2. Water samples were collected from the wells within Mosul city. The depth of the wells ranged from 5-14 m in Mosul city and from 15-90 m outside the city. The samples were tested for pH, total dissolved solids, conductivity, calcium, magnesium, chloride, sulfate and bicarbonate according to the standard methods (APHA, AWW and WPCF, 1992). In addition, data were obtained from Water Wells Drilling Company Records in Mosul city.

The statistical analyses include descriptive statistics of water quality parameters presented by mean, standard deviation and range. Correlation analysis was also conducted to show the relationships between the measured parameters. Factor analysis with varimax rotation was conducted on the standardized data (Kaiser, 1958; Davis, 1973), and the factor loadings were obtained. Hierarchical cluster analysis was used to group the studied wells. The Ward method, which is considered to be efficient, was applied since it uses the analysis of variance approach to evaluate the distances between clusters. The Ward method also minimizes the sum of squares of any two clusters that can be formed at each step (Ward, 1963).

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 the total dissolved solids ranged between 116 to

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

Parameters	Mean	SD	Min.	Max.
pH	7.66	0.47	6.65	8.40
EC (μhos/cm)	2609.73	1657.61	345.0	8800.0
TDS (mg/l)	2036.79	1323.70	116.0	6880.0
Ca^{+2} (mg/l)	271.14	244.23	13.0	1122.0
Mg^{+2} (mg/l)	189.25	236.59	2.10	1775
Cl^{-} (mg/l)	322.53	791.67	18.0	5610.0
$SO_4^{=}$ (mg/l)	1000.5	704.18	20.0	2800.0
HCO_3^{-} (mg/l)	288.0	143.26	3.0	620.0

6886 mg/l with a standard deviation of 1323.7. These wide variations are due to the distribution of the wells over a large area of different groundwater sources and geological formations, which makes it mandatory to conduct cluster analysis for classifying the wells.

A bivariate correlation between the measured parameters for the 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 extracted four factors from the standardized data depending on the scree plot (Figure 1). The purpose of factor rotation is to yield a factor structure that is simple. Also it is needed when factor loadings plot highly on more than one axis (Thurstone, 1947).

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.

Factor I accounts for 30.55% of the variation in water quality of the groundwater in the studied wells. It is dominated by sulfate as anion and calcium as cation and salinity, which is represented by electrical conductivity and total dissolved solids (bolded parameters in Table 3). The parameters

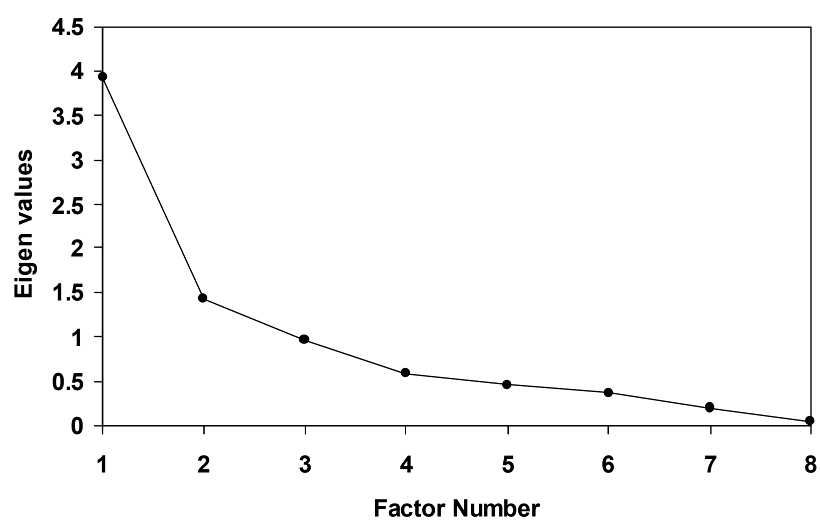

Figure 1. Scree plot for selection of number of factors.

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 $p<0.01$

of this factor are significantly correlated among each other. This factor reflects the variation in dissolution of rocks encountered in the study 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 each other.

Factor III is dominated by magnesium. It represents 18.61% of the variation in water quality, which represents the variation of dolomite rocks in the study area. 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 Mosul and surrounding areas. These variations are attributed to different geological and geochemical conditions and different groundwater sources.

Cluster analysis is a method used to combine the wells into homogenous groups according to their water quality. In this analysis, the percentage of variation explained for the number of clusters from 2 to 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 (Figure 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 Figure 3. The

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 (μhos/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	69.155	86.245

The loadings greater than 0.6 are bolded.

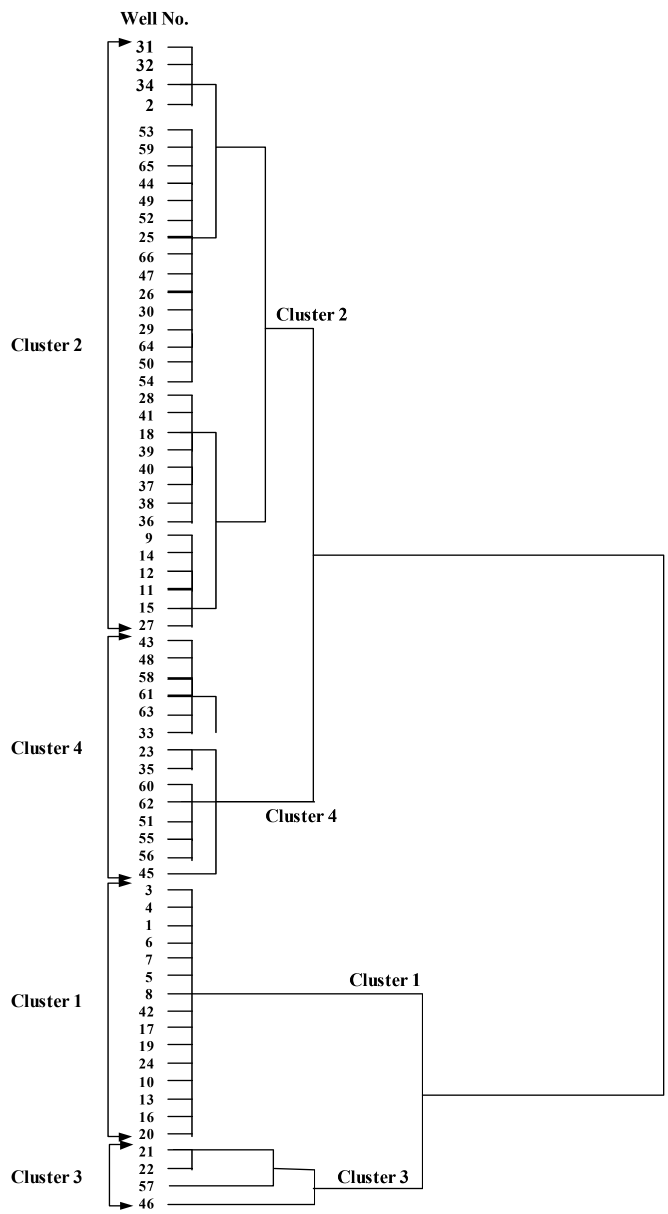

Figure 2. Hierarchical cluster analysis results for 66 wells according to the Ward method.

water of these wells had the lowest mean pH values, highest bicarbonate concentrations and highest sulfate concentration with means 7.12, 445.53 and 1694.33 mg/l and ranges 6.65-7.48, 250-620 and 965-2200 respectively, (Table 4 and Figures 4-6). These results coincided with the results of Kalander and Al-Joboury (1989) who found an increase in bicarbonate concentration and decrease in pH values in the groundwater wells east of the 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 (Figure 2). Most of the wells of this cluster are distributed around Mosul city with few wells in this town. The wells of this cluster have the lowest salinity in the studied wells since it has the lowest EC, TDS, Ca, Mg and chloride concentration, with means of 1459.6, 1081.8, 146.61 and 98.93 and ranges 345-2900, 116-2000, 13-505 and 2.1-523 respectively (Table 4 and Figure 7-11).

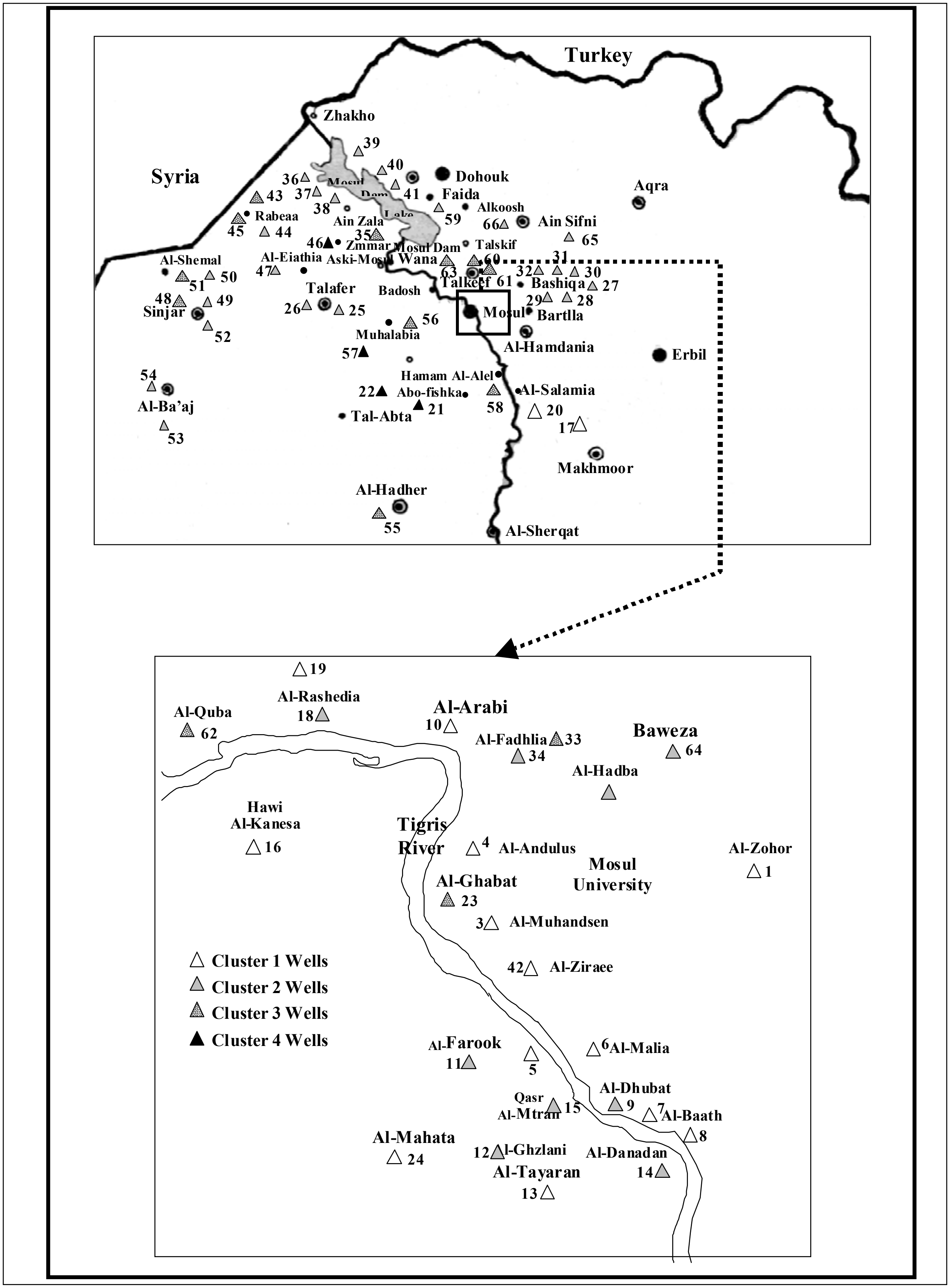

Figure 3. Location map showing the wells included in the study and their classification according to cluster analysis.

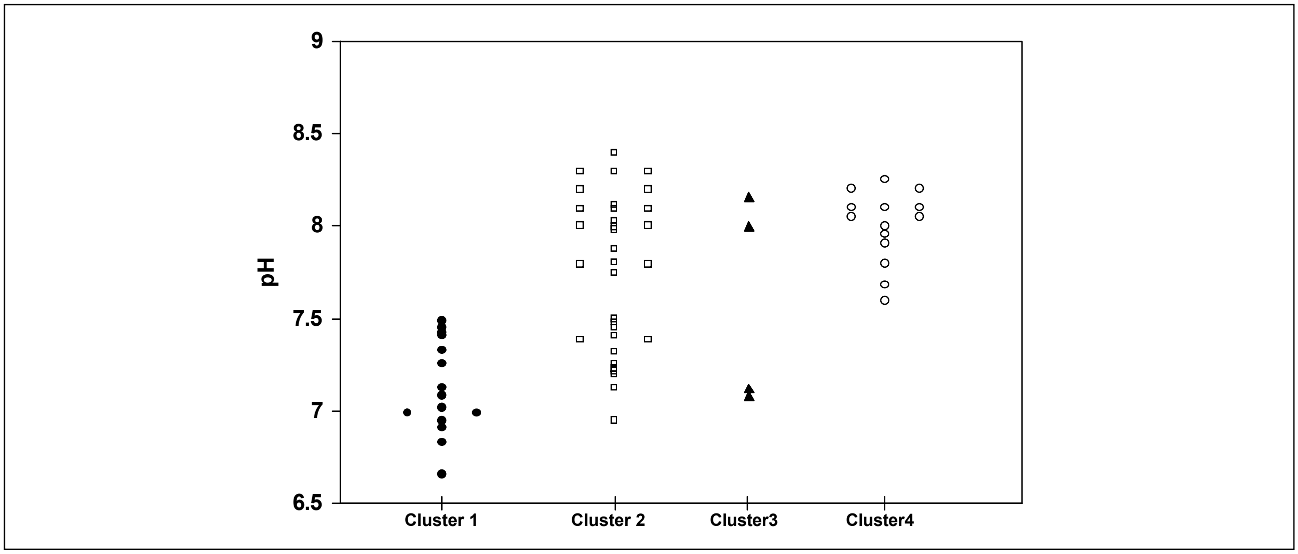

Figure 4. Distribution of extracted clusters according to pH.

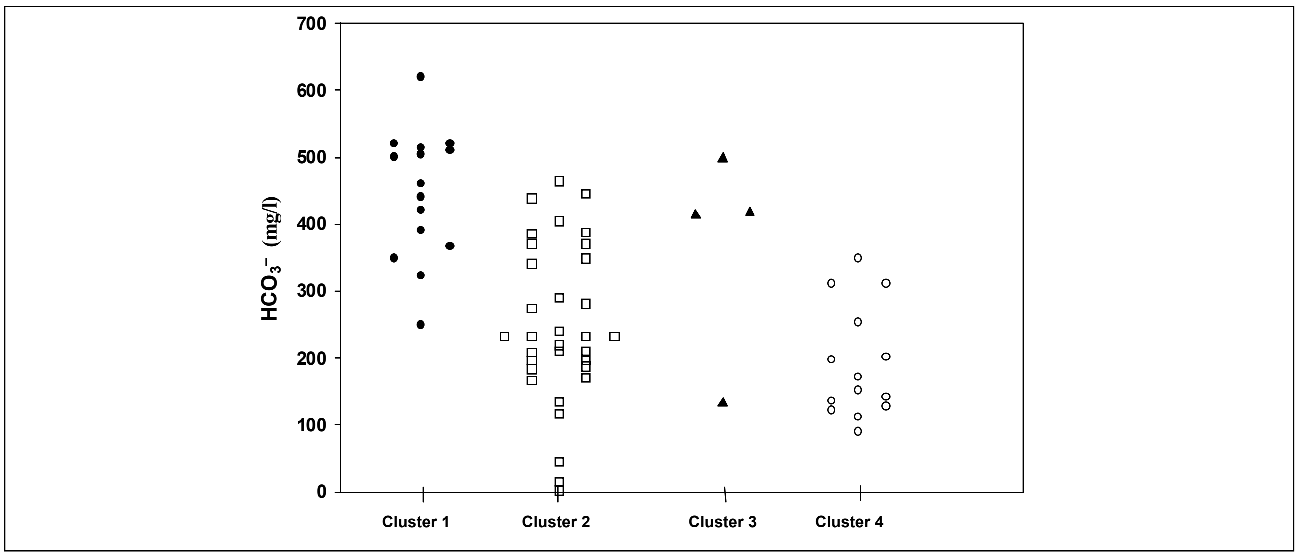

Figure 5. Distribution of extracted clusters according to bicarbonate concentration.

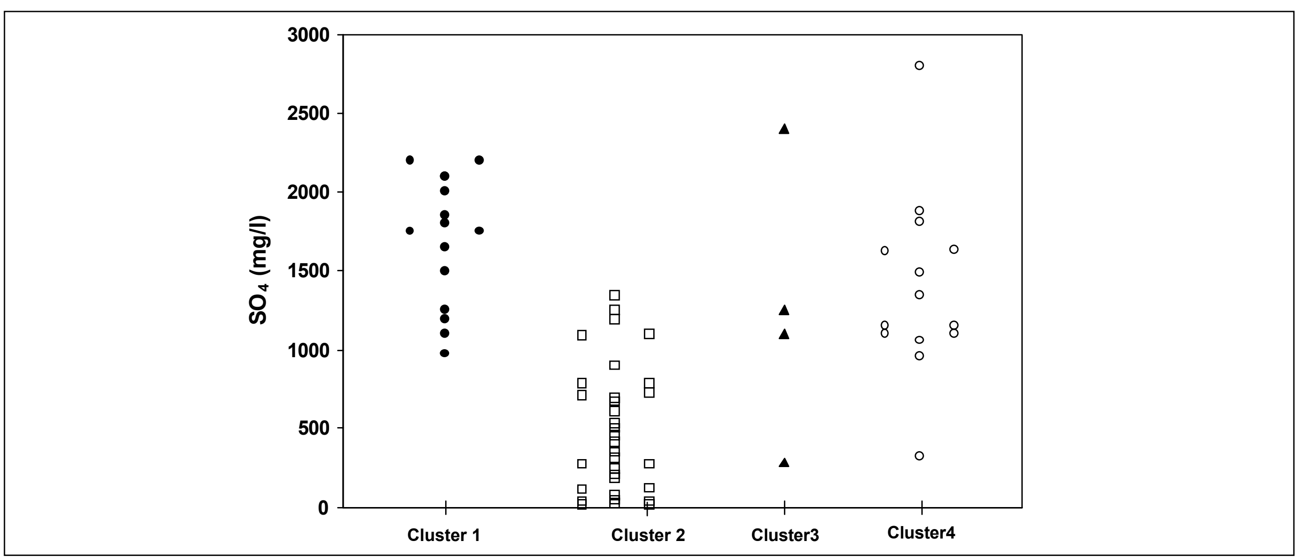

Figure 6. Distribution of extracted clusters according to sulfate concentration.

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 (μhos/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

On the other hand, cluster three 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 and ranges 3800-8800, 2117-6880, 312-1122, 390-1775 and 200-5610 respectively (Table 4 and Figures 7-11). This cluster was the smallest one; it includes only 4(6.67%) wells as shown in Figure (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 and ranges of 7.6-8.25 and 89-350 respectively (Table 4 and Figures 4-5). These wells are distributed around Mosul city (Figure 3).

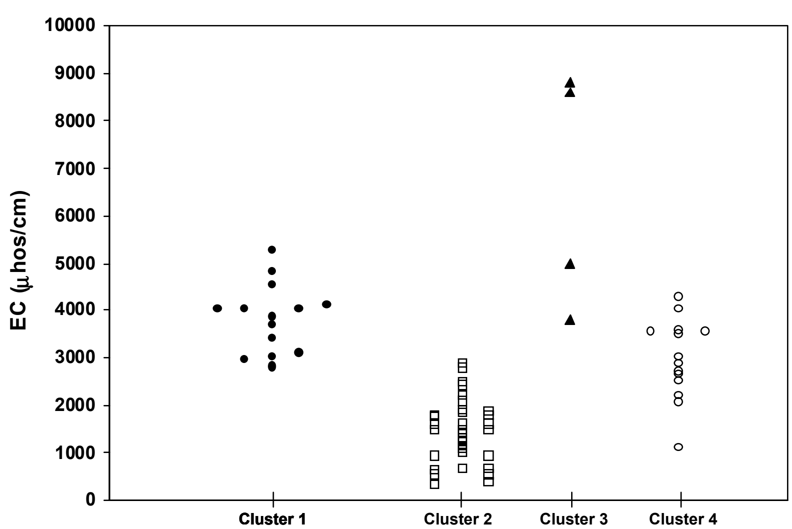

Figure 7. Distribution of extracted clusters according to EC.

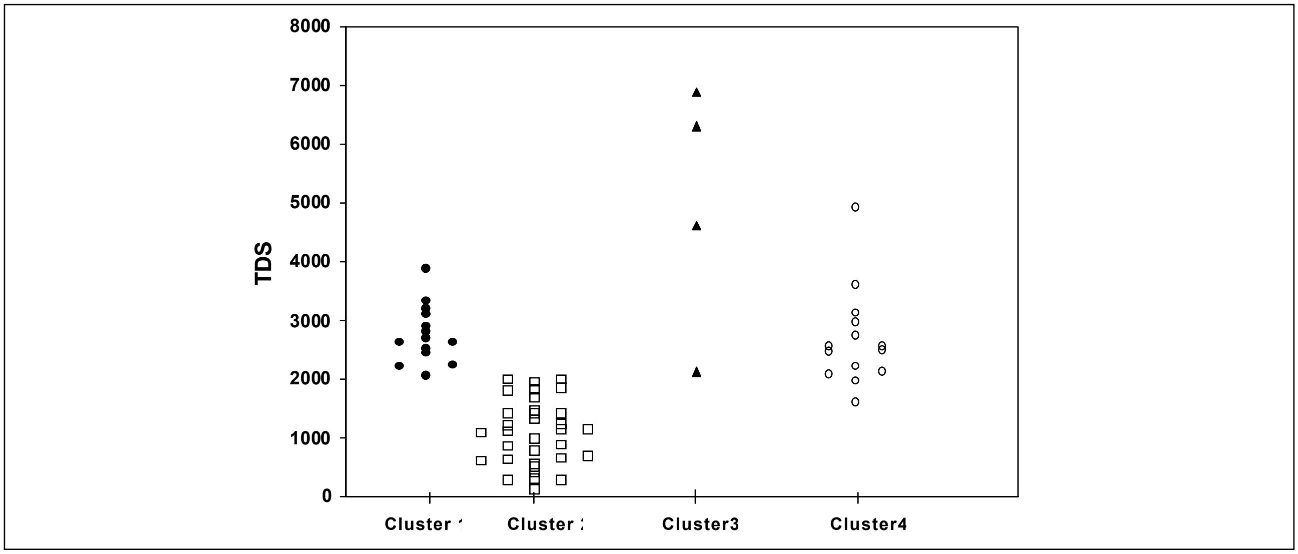

Figure 8. Distribution of extracted clusters according to TDS.

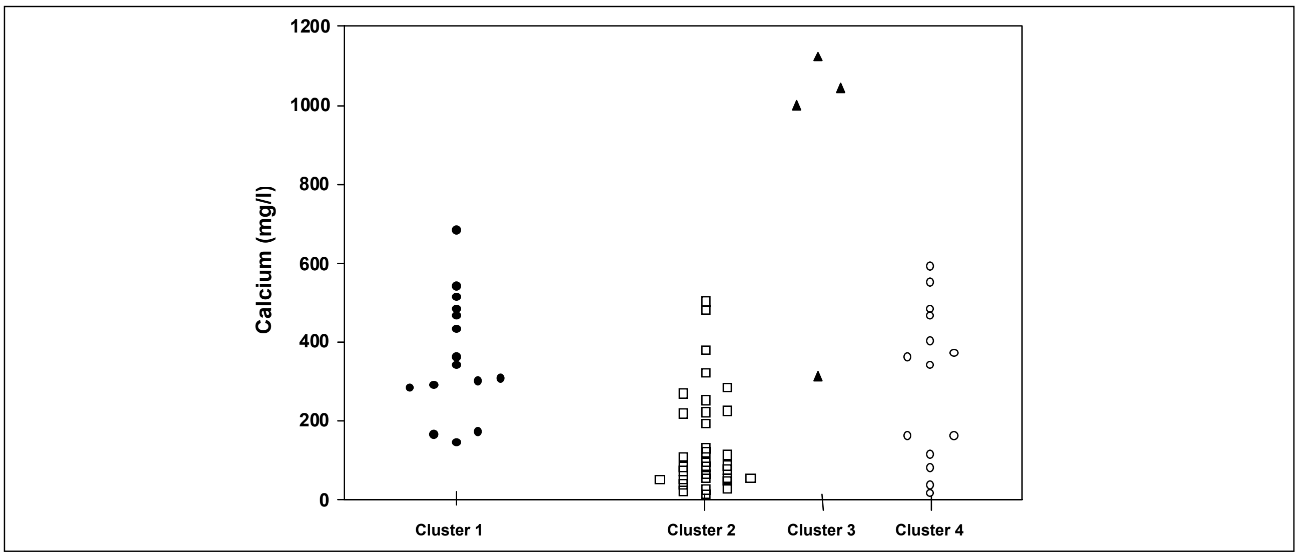

Figure 9. Distribution of extracted clusters according to calcium.

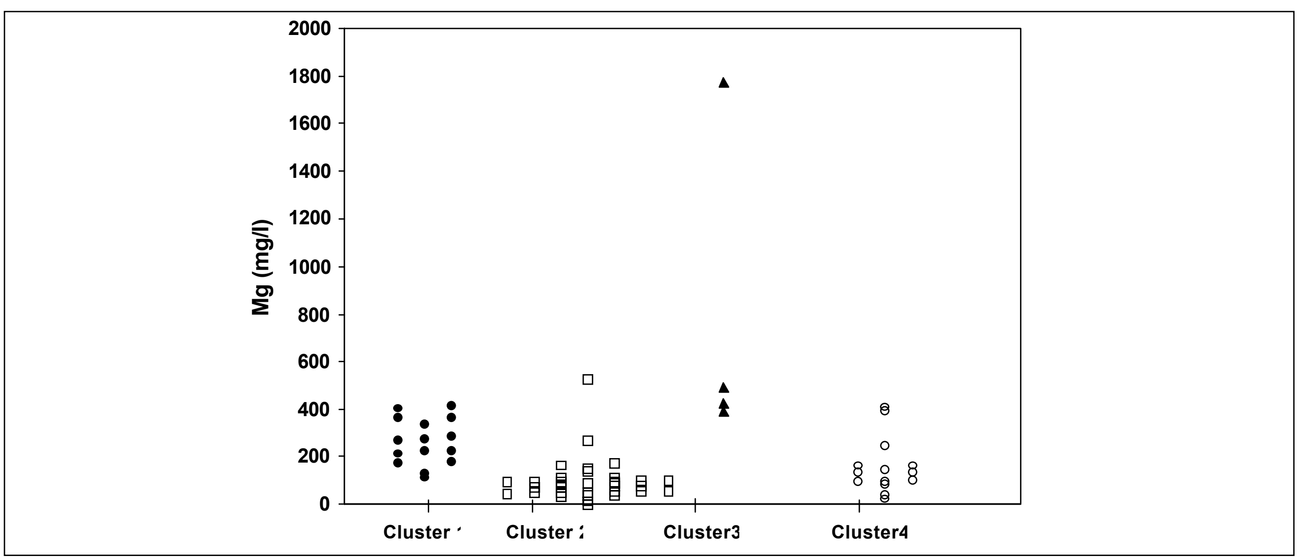

Figure 10. Distribution of extracted clusters according to Mg concentration.

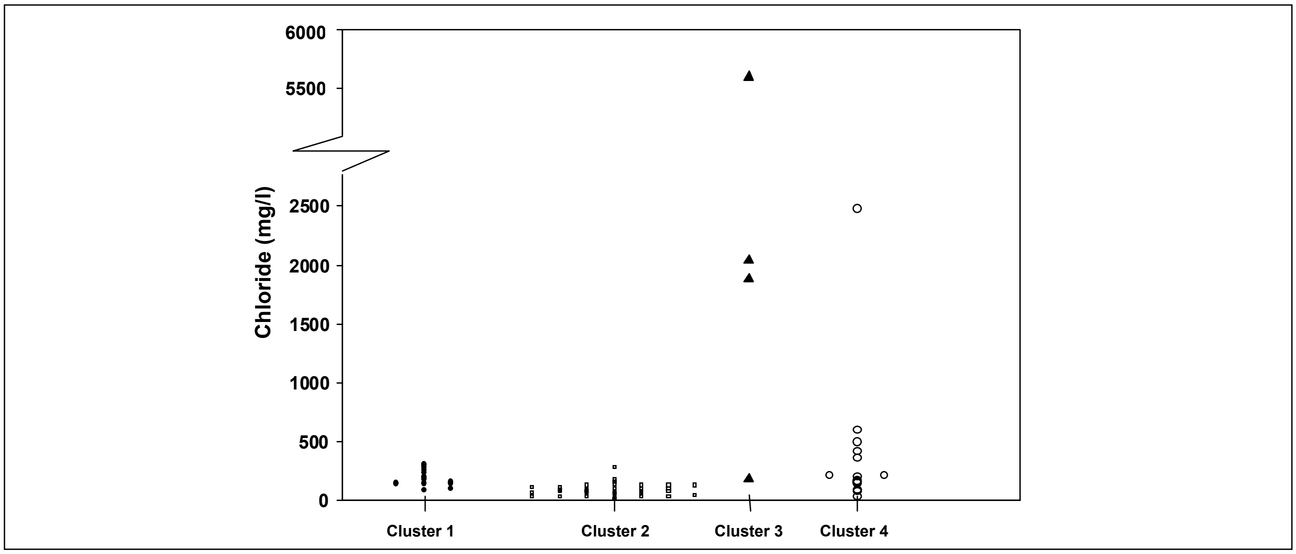

Figure 11. Distribution of extracted clusters according to chloride concentration.

CONCLUSIONS

1. The highest variation in groundwater quality for the study area was in conductivity, total dissolved solids and sulfate which reflects the variation in the geological and geochemical conditions and different groundwater sources.

2. Cluster analysis efficiently divides the wells included in this research into four relatively homogenous clusters. The first cluster with lower pH and highest bicarbonate and sulfate concentrations is located mostly along the sides of Tigris River; the second cluster has lowest salinity and is 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 to be a useful and efficient tool for water quality management.

REFERENCES

Al-Layla, M.A., S.M. Al-Rawi, and H.A. Al-Kawaz. 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.

Al-Rawi S.M., and A.S. Shihab. 2005. Application of factor analysis as a tool for water quality management of Tigris river within Mosul city. Raf. J. Sci. v. 16, no. 1, pp. 56-64.

Al-Rawi, S.M., S.I. Al-Azzo, and S.A. Abbawi. 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.

APHA, AWWA and WPCF. 1992. Standard Methods for the Examination of Water and Wastewater. 18th ed., APHA, Washington, D.C.

Davis, J.G. 1973. Statistics and data analysis in geology. John Wiley and Sons, Inc. New York, pp. 473-524.

Frapporti, G., S.P. Vriedn, and P.F.M. Gaans. 1993. Hydrochemistry of the shallow Dutch groundwater: Interpretation of the national groundwater quality monitoring network. Water Resour. Res. v. 29, no. 9, pp. 2993-3004.

Habib, R.H., N.H. Al-Saigh, and Z.M. Hassan. 1990. Geochemistry of underground water in Erbil city-Iraq. 2nd Scientific Conference of Mosul Dam Centre/Mosul University, pp. 173-188.

Hem, J.D. 1992. Study and interpretation of the chemical characteristics of natural water. US Government

Printing Office, Washington, D.C.

Kaiser, H.F. 1958. The varimax criterion for analytic rotation in factor analysis. Pyrometrical v. 23, pp. 187-200.

Kalander, L.A, and M.A. Al-Joboury. 1989. Hydrology of left bank of Mosul city, East Tigris river. J. Geol. Soc. v. 22, no. 1, pp. 146-159. (in Arabic)

Kennedy, L.G., J.W. Everett, T. Dewers, W. Pickins, and D. Edwards. 1999. Application of mineral iron and sulfide analysis to evaluate natural attenuation at fuel contaminated site. J. Environ. Eng. v. 125, no. 1, pp. 47-56.

Montgomery, R.H., J.C. Loftis, and J. Harris. 1987. Statistical characteristics of groundwater quality variables. Groundwater v. 25, no. 2, pp. 176-184.

Ochsenkuhn, K.M., J. Kontoyannakos, and M. Ochsenkuhn-Petroppulu. 1997. A new approach to a hydrochemical study of groundwater flow. J. Hydrol. v. 194, pp. 64-75.

Ritzi, R.W., S.L. Wright, B. Mann, and M. Chen. 1993. Analysis of temporal variability in hydrogeochemical data used for multivariate analyses. Groundwater v. 31, no. 2, pp. 221-229.

Suk, H., and K.K. Lee. 1999. Characterization of a groundwater hydrochemical system through multivariate analysis: clustering into groundwater zones. Groundwater v. 37, no. 3pp. 358-366.

Thurstone, L.L. 1947. Multiple factor analysis. Chicago: University of Chicago Press.

Usunoff, E.J., and A. Guzman-Guzman. 1989. Multivariate analysis in hydrochemistry: an example of the use of factor and corresponding analyses. Groundwater v. 27, no. 1, pp. 27-34.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. J. Am. Statistical Assoc. v. 58, pp. 236-244.

ADDRESS FOR CORRESPONDENCE
Abdulmuhsin S. Shihab
Environmental Research Centre and Pollution Control
Mosul University
Mosul, Iraq

Email: muhsin_shehab@yahoo.com