Research Journal of Environmental and Earth Sciences 4(6): 665-667, 2012
ISSN: 2041-0492

Submitted: April 20, 2012 Accepted: May 14, 2012 Published: June 30, 2012

Multivariate Statistical Analysis of Geochemical Data of Groundwater in El-Bahariya Oasis, Western Desert, Egypt

Ali M. Hamdan
Geology Department, Faculty of Science, Aswan University, Egypt

Abstract: The aim of the present study is to study the application of multivariate statistical analyses of hydrochemical data using the chemical analyses for 125 groundwater samples with 18 parameters include the hyrdrochemical compositions (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, (HCO_3), $(SO_4)^{2-}$ and Cl^{-}) and the physicochemical parameters (EC, TDS, TH, SAR, RSBC, PI, KR, SSP, MAR, RSC and Na%). The linear regression is an approach to modeling the relationship between two variables using a set of individual data point and used to explain or predict the behavior of a dependent variable. Two variables were used to develop a relationship between TDS as an independent variable and different hyrdrochemical data as a dependent variable. Using these equations, by known TDS value, the equation tries to predict any unknown other variables. The linear regression equations used also between the EC as an independent variable and all different water quality variables. The correlation matrix performed for the groundwater using the hyrdrochemical compositions (r varies from 0.84 to 0.08). All data have positive relations reflecting direct relationship with all hydrochemical data. Good correlation observed between TDS and each of other variables, while weak positive relation detected between (HCO_3) and Ca^{2+}, $(SO_4)^{2-}$, Ng^{2+} and Cl^{-}. Two clusters were performed, the first use TDS, Ca, Mg, Na, K, HCO_3, SO_4, Cl, EC and TH while the second use PI, TH, MAR, EC, SAR, KR, Na%, RSBC, RSC and SSP as variables. Skewness and kurtosis are calculated for all data to describe the shape and symmetry of the distribution of geochemical data along the study area. Skewness values vary from 3.22 to -1.36. Positive skewness were notice in most parameters indicates that the shape of their statistical distribution diagrams show the tail on the right side (direction of high values) is longer than the left side and the bulk of the values (possibly including the median) lie to the left of the mean for each parameter. Kurtosis values vary from 18.17 (for SO_4) to -0.65 (for RSBC). Positive Kurtosis characterize most parameters indicates a peaked distribution relative to a normal distribution of the data, while the other are negative (indicates a flat distribution). The SO_4, KR, MAR and TH have high kurtosis values, indicates tend to have a distinct peak near the mean and have heavy tails.

Keywords: Cluster analysis, Egypt, El-bahariya oasis, hydrochemistry, skewness and kurtosis, statistical analysis, western desert

INTRODUCTION

El-Bahariya Oasis is a natural topographic depression located in the Western Desert of Egypt. It is located between latitudes 27º48′ and 28º30′ N and between longitudes 28º35′ and 29º10′ E (Fig. 1), about 370 km southwest of Cairo. It covers an approximate area of 1800 km^2. The groundwater is the essentially water source not only in El-Bahariya Oasis, but also in the Western Desert. With the increasing demands for water due to increasing population, urbanization and agricultural expansion, groundwater resources are gaining much attention.

Multivariate statistical analysis generally refers to a range of statistical techniques/methods which primarily involves data with several variables, with the objective of investigating the dependence relations between the involved variables. In present study, statistical analysis was used to study the variations, relations, distributions of the hydrogeochemical data along the study area. The large database was subjected to different multivariate statistical techniques with a view to extract information about the similarities or dissimilarities between the sampling sites.

In recent times, multivariate statistical tools have been successfully applied widely in processing hydrogeochemical data (Cameron, 1996; Gupta and Subramanian, 1998; Laaksoharju *et al*., 1999; Duffy and Brandes, 2001; Anazawa *et al*., 2003; Güler and Thyne, 2004; Anazawa and Ohmori, 2005; Shedid *et al*., 2005; Kim *et al*., 2009; Cheng-Shin, 2010). The combined use of principal component analysis and cluster analysis enabled the classification of water samples into distinct groups on the basis of their hydrochemical characteristics. Regression equations were successfully used in different hydrogeochemical processes (Voudouris *et al*., 2000; Hartmann *et al*., 2005; Papatheodorou *et al*., 2006; Chenini and Khemiri, 2009).

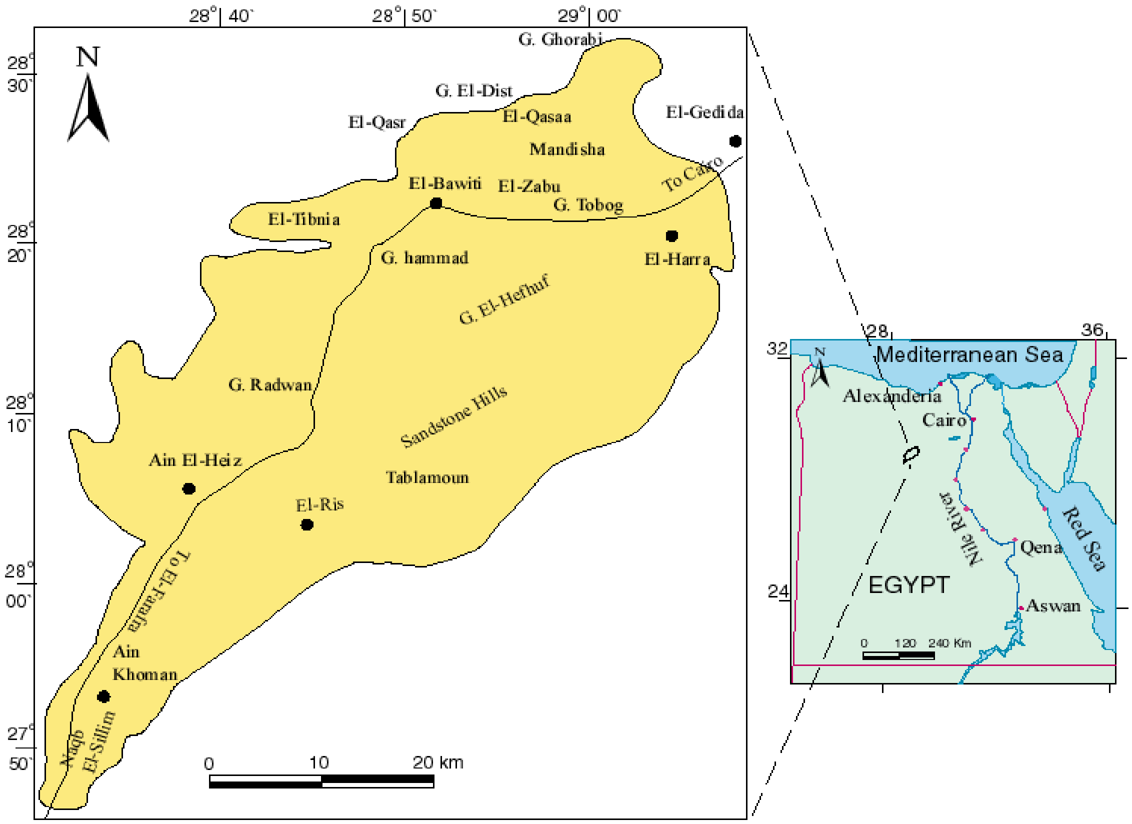


Fig. 1: Location map of El-bahariya oasis

The main objective of this study is to study multivariate statistical analysis for the hydrogeochemical of ground waters in El-Bahariya Oasis area using many hyrdrochemical parameters. Through this study:

- The linear regression was applied using two variables to develop a relationship between TDS as an independent variable and different hyrdrochemical data as a dependent variable. Also it applied to determine the relationship between Electrical Conductivity (EC) and the physicochemical characteristics of groundwater resources.
- The correlation matrix performed for the groundwater using the hyrdrochemical compositions.
- Two clusters were performed using hydrochemical composition (TDS, Ca, Mg, Na, K, HCO_3, SO_4, Cl, EC and TH) and using physiochemical parameters (PI, TH, MAR, EC, SAR, KR, Na%, RSBC, RSC and SSP) as variables. It used to provide a useful means of detecting the existence of groups of similar objects in a high dimensional space.
- Skewness and kurtosis are calculated for all data to describe the shape and symmetry of the distribution of geochemical data along the study area.

Study area:

Geologyical setting: The general geology of the area is relatively well known and has been studied by several investigators (El-Akkad and Issawi, 1963; Said and Issawi, 1964; Basta and Amer, 1970; Issawi, 1972; El-Bassyouny, 1978; Soliman and El-Badry, 1980; Issawi and Labib, 1996; Issawi *et al*., 1999; El-Aref *et al*., 1999).

The following lithological units were distinguished in the study area (Fig. 2):

- **Cretaceous rocks:** Lie on top of the Cambrian deposits of undifferentiated sandstones, sands and clays intercalated with each other. Cretaceous Rocks range in age from Cenomanian to Oligocene: Bahariya Formation (*L. cenomanian*), El-Heiz Formation (*U. cenomanian*), El-Hefhuf Formation (Campanian-Turonian), Ain Giffara Formation (Campanian), Khoman Chalk (Maastrichtian), Plateau Limestone (Eocene) and Radwan Formation (Oligocene)
- **Eocene rocks or the eocene limestone rocks:** They form the eroded plateau surface surrounding El-Bahariya depression and some of the isolated hills

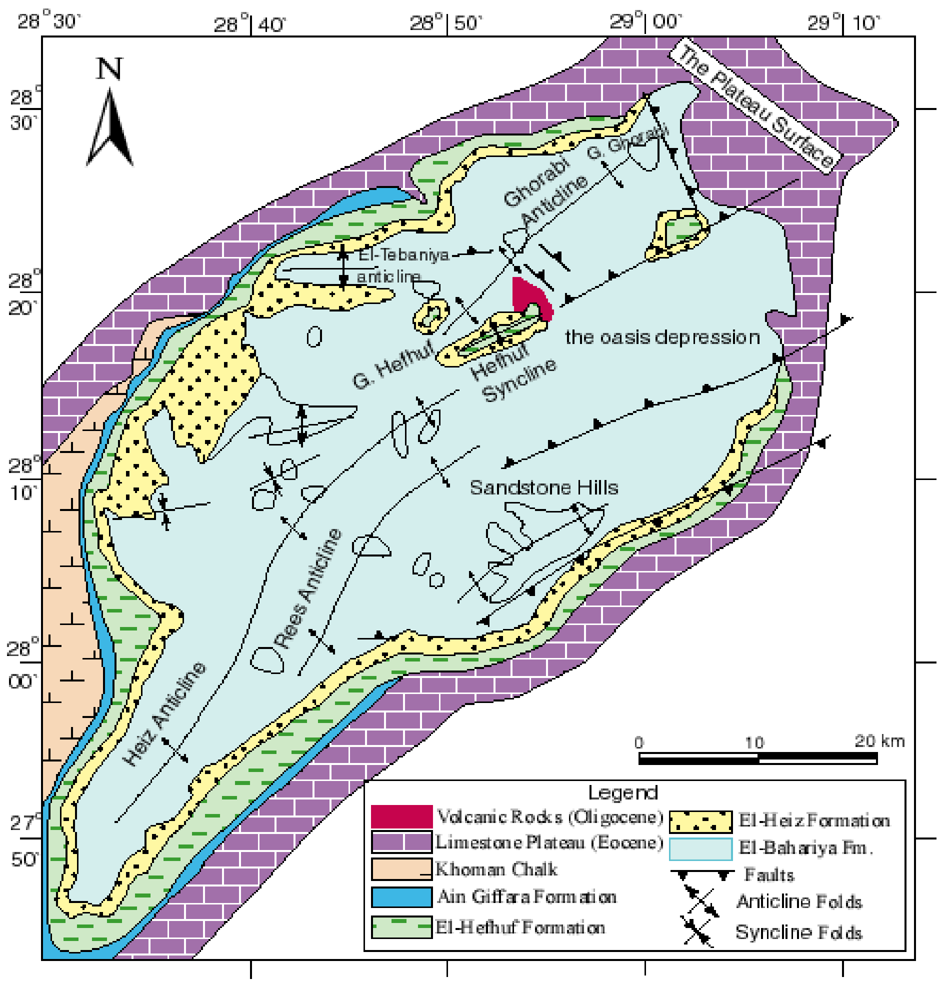


Fig. 2: Map showing geological units, geomorphological and structural features at El-bahariya oasis (El-Akkad and Issawi, 1963; El-Bassyouny, 1978)

within it. The Eocene strata rest unconformably on the Upper Cretaceous rocks. It consists of Farafra Formation; Naqb Formation (Lower Middle Eocene); Qazzun Formation (Upper Middle Eocene); and El-Hamra Formation (Middle-Upper Eocene).

- **Tertiary rocks:** The sedimentary formations are capped with sheets of volcanic rocks (of Oligocene age), mostly extrusive basalt and dolerite.
- **Quaternary rocks:** they vary and represented by aeolian sands (form the scattered sand dunes within the depression and on the plateau surface); sabkhas and salt deposits (Distributed around the cultivated lands in El-Qasaa, El-Harra and El-Heiz localities and produced due to the seepage of the water from natural flowing springs and wells through the clay horizons on the depression floor); and playa deposits (composed of fine sand, silt and dark brown clay mixed with gypsum and halite).

Structural setting: El-Bahariya Oasis is considered to be a major doubly plunging anticline with a NE-SW trend, a typical structure of the Syrian Arc belt. The axis of this great anticline runs in a southwest trend from Gebel Ghorabi in the north, passing to the central hills of the depression to the southern part of the oasis and seems to continue south to include El-Farafra structure. The major Folds in the study area (Fig. 2) comprise the following:

- Folds of NE-SW Trend
- Ghorabi Plunged Anticline
- El-Heiz Plunged Anticline
- The Sandstone Hill Anticline
- El-Ris Anticline
- El-Hefhuf Syncline
- El-Tebaniya anticline

El-Bahariya Oasis characterized by three different striking major fault systems. The NE-SW faults which running parallel to El-Bahariya major anticline represent the most common fault trend with a throw ranges between 40 and 50 m, respectively. The NW-SE faults are the second common trend at El-Bahariya area and have relatively low throws compared with NE-SW faults (30-40 m). The E-W trending faults are the least common in El-Bahariya with throw reach about 40 m (Abdel Ati, 2002).

Geomorphology: El-Bahariya Oasis is one of the naturally excavated depressions located in the Western Desert of Egypt (in the Eocene limestone plateau); it is believed to be of tectonic origin, started during the Lower Eocene times. It differs from the other oases in being entirely surrounded by escarpments and in having a large number of isolated hills within the depression.

Three morphological features (Fig. 2) are distinguished in El-Bahariya depression. *The plateau surface* (its surface is ragged with a northward slope and dissected by long to short dry wadies draining into the excavated depression); *the bounded escarpments* (having different modes of formation and run in most irregular manner to form well marked embayment and promontories and its face takes the shape of a questa); and *the depression* (the floor of the depression is excavated in the soft clastics of El-Bahariya Formation). Several landforms are well developed on the depression floor area, where some of them are of structural origin, while the other is of depositional nature. Among them are the isolated hills and sand dunes.

Hydrogeological setting: The Nubian Sandstone represents the main water-bearing horizon in the studied area. It consists of continental elastic sediments, mainly sandstone alternating with shale and clays. The groundwater system in the studied area is hydraulically connected with the surrounding and the underlying aquifers through a good pathways or channels that permit upward leakage. Accordingly, the Nubian sandstone aquifer in El-Bahariya Oasis is described as a multilayered artesian aquifer that behaves as one hydrogeologic system. The result of the detailed hydrogeologic studies of the Nubian Sandstone Aquifer system in El-Bahariya Oasis revealed that the aquifer attains a total thickness of about 1,800 m (Parsons, 1962; Ezzat, 1974; Euro and Pacer, 1983).

The groundwater bearing horizons in the investigated area follow two aquifer systems (Khalifa, 2006). The first is Post-Nubian sandstone aquifer System (occurs to the north of latitude 26º in the Western Desert of Egypt (CEDARE, 2001). The second is The Nubian sandstone aquifer system which represents the main water-hearing horizon in the studied area consists of continental clastic sediments mainly sandstone alternating with shale and clays (Himida, 1964; Diab, 1972).

The transmissivity of the main aquifer vary from 236 to 3045 m^2/day (high to moderate potentiality aquifer) with gradual increase from the southern to northeastern direction. The storage coefficient values range between 1.04×10^{-4} and 5.22×10^{-3}, which ensure that the Nubian sandstone aquifer is classified as semi-confined to confined aquifer type. The hydraulic conductivity values vary from to 0.46 m/day in the northern part to 10.88 m/day in the southern part with an average of 5.67 m/day (Hamdan and Sawires, 2011).

METHODOLOGY OF STUDY

During the present study, the chemical analyses for 125 groundwater samples (collected during the period from May 2003 to 2008) were used to apply the multivariate statistical analysis in El-Bahariya Oasis, Western Desert of Egypt. Eighteen parameters include hyrdrochemical compositions (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, $(HCO_3)^{-}$, $(SO_4)^{2-}$ and Cl^{-}) and physicochemical parameters (Electric Conductivity (EC), TDS, Total Hardness (TH), Sodium Adsorption Ratio (SAR), Residual Sodium Bicarbonate (RSBC), Permeability Index (PI), Kelly's Ratio (KR), Soluble Sodium Percentage (SSP), Magnesium Adsorption Ratio (MAR), Residual Sodium Carbonate (RSC) and Sodium percent Na %) were used in the statistical analysis.

Minitab 16.1 and SPSS computer programs were applied to calculate statistical analysis during this study.

To find the relationship between TDS and electrical conductivity with different water quality variables, the linear regression equation were applied. Regression equations were successfully used to study the hydrogeochemical processes (Voudouris *et al*., 2000; Hartmann *et al*., 2005; Papatheodorou *et al*., 2006; Chenini and Khemiri, 2009). Some statistical parameters as the mean, median, maximum, minimum and the standard deviation were determined.

Because some of these parameters are not symmetrically distributed, each parameter was examined normality based on skewness and kurtosis. To test the normality of considered variables (Reimann and Filzmoser, 2000; Kim *et al*., 2009), the skewness for each variable was calculated as follows:

$$skewness = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^3}{(N-1)S^3}$$

For univariate data X_1, X_2, ..., X_N, the formula for kurtosis is:

$$kurtosis = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^4}{(N-1)S^4}$$

where, $\bar{X}$ is the mean, S is the standard deviation and N is the number of data points.

The standard deviation can be calculated using the following formula:

$$S = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}{N-1}}$$

where, S is the standard deviation, $\bar{X}$ is the mean, X_i is value of sample (i) and N is the number of data points.

The correlation matrix and the correlation coefficient were performed for the groundwater using the hyrdrochemical compositions TDS, Ca^{2+}, Mg^{2+}, Na^+, K^+, $(HCO_3)^-$, $(SO_4)^{2-}$ and Cl^-. The correlation coefficient for each variable was calculated as the following equation:

$$r = \frac{\sum_{i=1}^{N}\left(\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right)}{\sqrt{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}}$$

where, r is the correlation coefficient, $\bar{X}$ and $\bar{Y}$ is refer to the mean of the X and Y values, respectively and N is the number of data points.

Cluster analysis groups a system of variables into clusters on the basis of similarities (or dissimilarities) such that each cluster represents a specific process in the system. In this study, the cluster analysis was applied to the raw data of groundwater samples using Minitab 16.1 Software program. Cluster analysis is a powerful tool for analyzing water chemistry data (Meng and Maynard, 2001; Yidana, 2010; Belkhiri *et al*., 2011). A classification scheme using the Euclidean distance for similarity measures and the Ward's method for linkage produces the most distinctive classification where each member within a group is more similar to its fellow members than to any member outside of the group (Güler *et al*., 2002).

RESULTS AND DISCUSSION

Statistical parameters: Along the study area, some statistical parameters as the mean, median, maximum, minimum and the standard deviation give a general view for the distribution of different chemical components. The calculated parameters were sited in Table 1.

Standard deviation is a measure of how spreads out the data points are. A set with a low standard deviation has most of the data points centered around the average. A set with a high standard deviation has data points that are not so clustered around the average.

The median of a set is another way of calculating a sort of middle value for a data set. In fact, the median is the actual middle number when the data put in order.

Statistical results of the physical parameters (TDS, TH, pH and Temperature):

- The SD of pH is very low indicating most of data points centered around the average.
- The SD of EC is very high and median equals 271 μmhos/cm reflecting many variations exist between EC values along the area.
- The median of the TDS is 203.28 ppm. This reflect that the area characterized by very low TDS values.
- The SD of the TDS is relatively high (96.25), indicates presence of some locations have high spreads from the mean and have relatively high TDS values.
- The SD, median and mean values of the temperature pointed a low spreads between different t-values.

Table 1: Statistical analysis results of ions, physical and important irrigation parameters in groundwater samples of the study area

Parameters	Units	Mean	SD	Min.	Median.	Max.
pH	-------	7.900	0.380	6.0500	8.000	8.7000
EC	μmhos/cm	317.380	148.700	105.000	271.000	972.000
T	ºC	29.800	5.670	15.000	28.500	45.000
TDS	mg/L	232.300	96.250	122.000	203.280	712.060
TH	mg/L	81.330	28.000	50.900	71.400	230.000
Na^+	mg/L	35.720	25.160	9.910	25.500	145.000
K^+	mg/L	13.430	4.700	8.000	13.000	35.000
Ca^{2+}	mg/L	13.820	4.530	4.000	12.240	35.000
Mg^{2+}	mg/L	11.380	4.880	0.810	9.900	34.710
Cl^-	mg/L	53.280	24.890	28.720	42.080	131.650
HCO_3	mg/L	71.700	36.200	17.500	68.420	170.800
SO_4	mg/L	35.990	28.930	6.000	30.000	190.000
SAR	meq/L	1.724	1.200	0.531	1.339	7.108
RSC	meq/L	-0.449	0.697	-2.919	-0.601	1.272
Na^+ %	%	51.160	10.510	35.490	50.730	83.130
SSP	%	51.160	10.510	35.490	50.730	83.130
PI	%	3.993	0.768	1.943	3.961	5.625
MAR	%	56.810	8.410	4.260	57.140	83.360
KR	meq/L	0.979	0.742	0.326	0.791	4.652
RSBC	meq/L	0.485	0.618	-0.574	0.364	1.985
NaCl	mg/L	73.840	41.030	27.010	55.260	208.400

Min: Minimum; Max: Maximum

Statistical results of the cations (Ca^{2+}, Mg^{2+}, Na^{+} and K^{+}):

- The SD of Ca^{2+}, Mg^{2+} and K^{+} are low indicating all groundwater samples have nearly homogeneous distribution along the study area.
- The SD of Na is high comparing with the other cations reflecting presence of some locations have high concentration of Na.
- The median of Ca^{2+}, Mg^{2+}, Na^{+} and K^{+} are 12.2, 9.9, 25.5 and 13.0 ppm, respectively.

Statistical parameters and the anions ($(HCO_3)^-$, $(SO_4)^{2-}$ and Cl^-):

- The SD values of the anions in the study area are high (24.9, 28.9 and 36.2 for Cl^-, $(SO_4)^{2-}$ and $(HCO_3)^-$, respectively). This reflects many variations exist between the concentrations of anions distributed along the area with great spread from the average.
- The median of $(SO_4)^{2-}$, Cl^- and $(HCO_3)^-$ are 30, 42.08 and 68.4 ppm, respectively.
- The mean values of $(SO_4)^{2-}$, Cl^- and $(HCO_3)^-$ are 35.9, 53.3 and 71.7 ppm, respectively. Indicating that the study area has low concentrations of the major anions.

Applying the regression equation: It is a statistical technique used to discover a mathematical relationship between two variables using a set of individual data point and used to explain or predict the behavior of a dependent variable.

The linear regression is an approach to modeling the relationship between a scalar variable Y and one or more variables denoted X. The data are modeled using linear functions and unknown model parameters are estimated from the data.

Generally, a regression equation takes the form:

$$Y = a + bX$$

where,

Y = The dependent variable that the equation tries to predict

X = The independent variable that is being used to predict Y

(a) = The intercept point of the regression line and the y axis

(b) = The slope of the regression line

The values of (a) and (b) are known values for each equation. (a) can be calculated through the equation (a = Y-bX).

In the study area, this equation was applied for the groundwater samples. Two variables were used to develop a relationship between TDS as an independent variable and different water quality variables such as each cation (Ca^{2+}, Mg^{2+}, Na^{+} and K^{+}), each anion ($(HCO_3)^-$, $(SO_4)^{2-}$ and Cl^-), TH and EC as a dependent variable. The data obtained are shown in Table 2. Using these equations, by known TDS value, the equation tries to predict any unknown other cations, anions, EC, or TH values.

The electric conductivity as a physical property of groundwater samples can measures easily in the field by the electric conductivity meter. It represents other independent variable and can be used to predict any water quality variables. The obtained regression equations between EC and the other variable are shown in Table 3.

Table 2: Regression equations used to develop a relationship between TDS (ppm) as an independent variable and different water quality variables

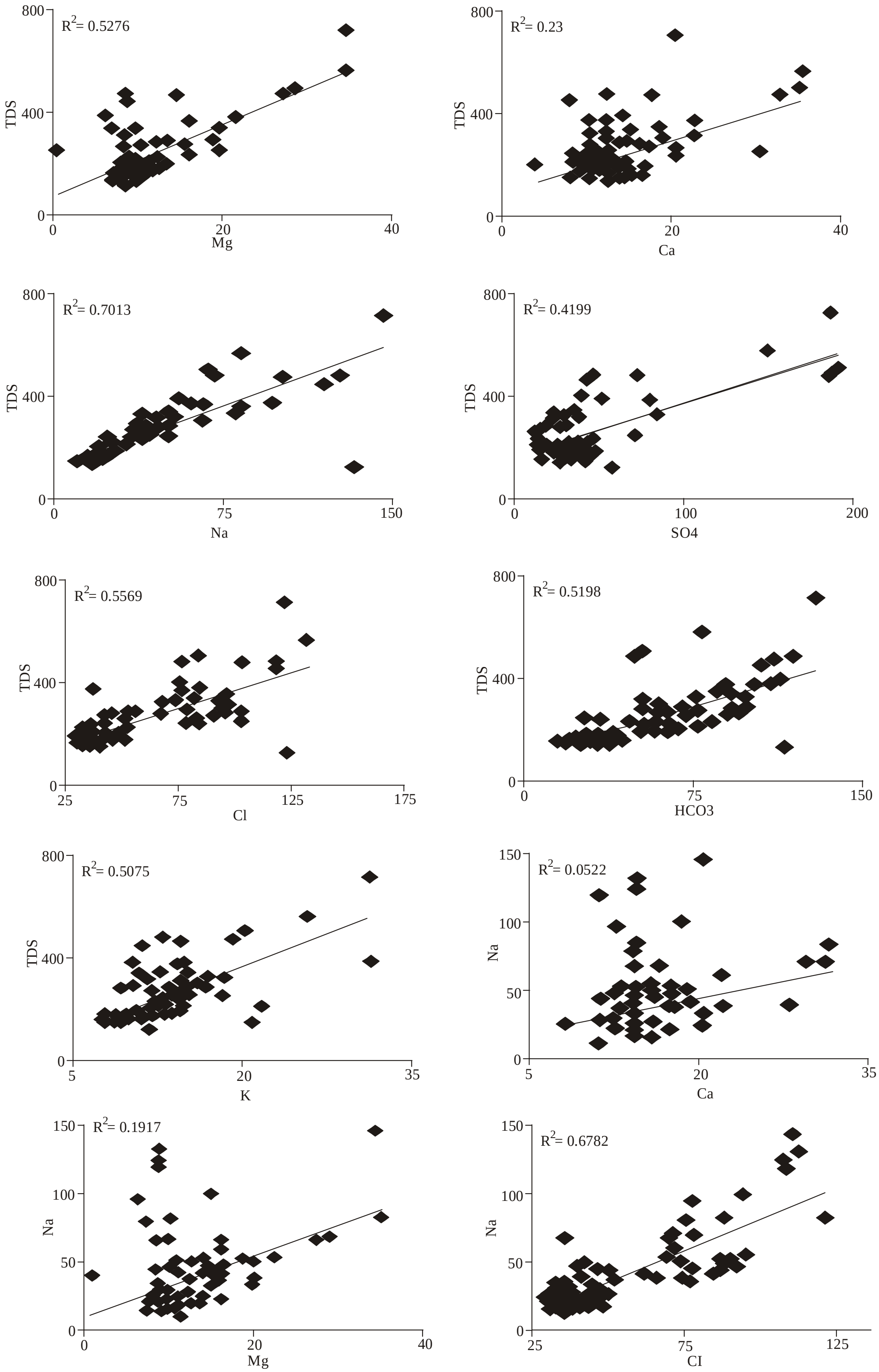
Dependent variable	Regression equation
Ca^{2+}	Ca = 0.036 TDS + 2.819
Mg^{2+}	Mg = 0.036 TDS + 2.819
Na^{+}	Na = 0.218 TDS - 15.13
K^{+}	K = 0.034 TDS + 5.341
(HCO_3)	(HCO_3) = 0.271 TDS + 8.7
(SO_4)	(SO_4) = 0.186 TDS - 7.354
Cl	Cl = 0.193 TDS + 8.458
TH	TH = 0.207 TDS + 33.04
EC	EC = 1.271 TDS + 21.94
SAR	SAR = 0.036 TDS + 2.819
PI	PI = 4.869 - 0.0038 TDS
KR	KR = 0.0039 TDS + 0.0823
MAR	MAR = 0.0247 TDS + 51.152
RSBC	RSBC = 0.0024 TDS - 0.0628
SSP	SSP = 0.0411 TDS + 41.716
RSC	RSC = -0.0001 TDS - 0.4191

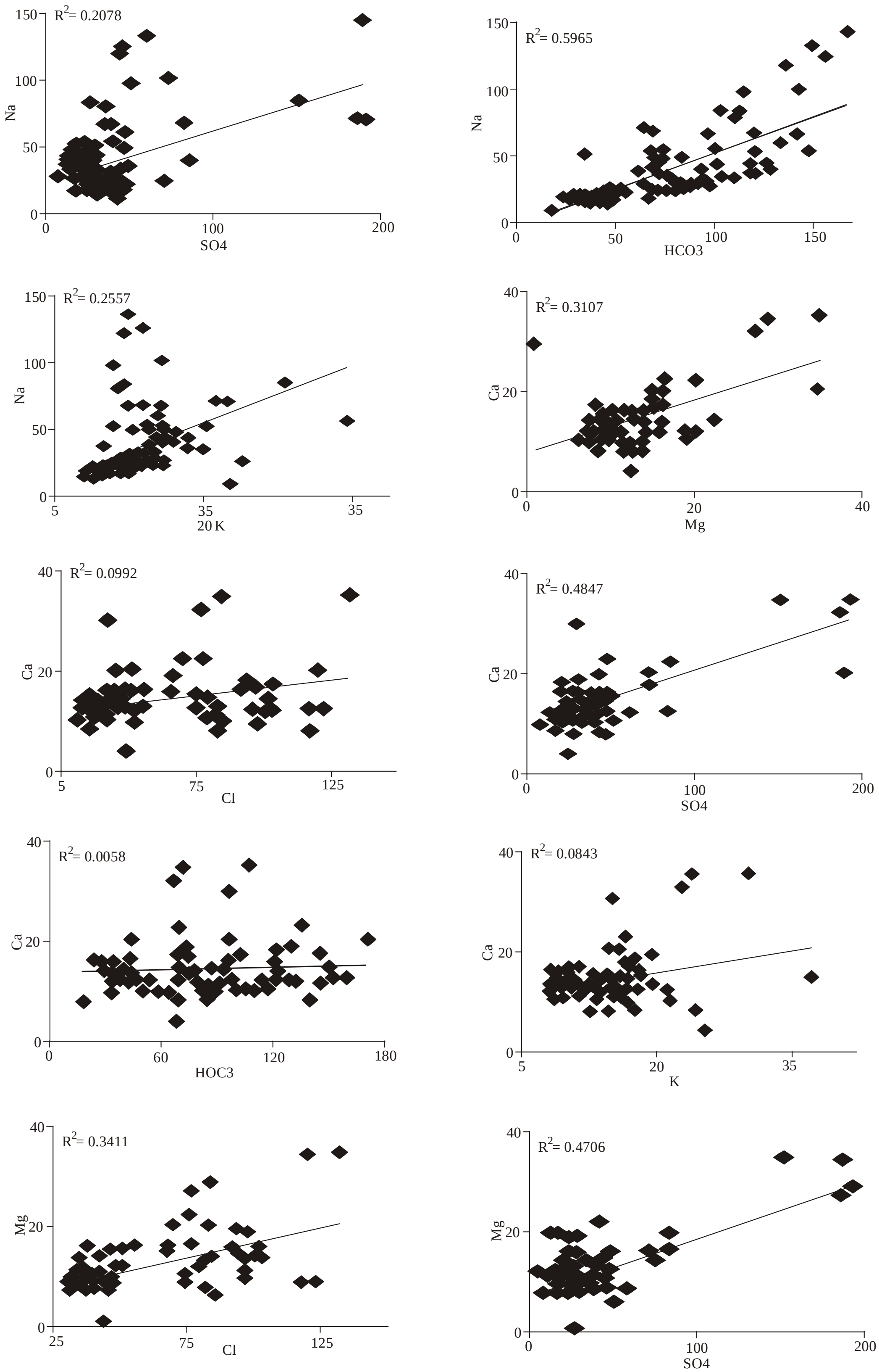
Table 3: Regression equations used relationship between EC (μmhos/cm) as an independent variable and different water quality variables in the study area

Dependent variable	Regression equation
Ca^{2+}	Ca = 0.012 EC + 9.874
Mg^{2+}	Mg = 0.036 EC + 2.819
Na^{+}	Na = 0.137 EC - 7.949
K^{+}	K = 0.019 EC + 7.246
(HCO_3)	(HCO_3) = 0.152 EC + 23.2
(SO_4)	(SO_4) = 0.116 EC - 0.973
Cl^-	Cl = 0.121 EC + 14.78
TH	TH = 0.118 EC + 43.75
TDS	TDS = 0.533 EC + 63.10
SAR	SAR = 0.036 EC + 2.819
PI	PI = 4.8313 - 0.0026 EC
KR	KR = 0.0018 EC + 0.4206
MAR	MAR = 0.0121 EC + 53.038
RSBC	RSBC = 0.0014 EC + 0.0461
SSP	SSP = 0.0246 EC + 43.467
RSC	RSC = 8E-05 EC - 0.4725

Correlation matrix: The correlation matrix were performed for the groundwater using the hyrdrochemical compositions TDS, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, $(HCO_3)^-$, $(SO_4)^{2-}$ and Cl^- (Fig. 3). The results of the correlation matrix are shown in Table 4.

From the correlation matrix the following points were revealed:


R2= 0.5276
TDS
Mg
R2= 0.23
TDS
Ca
R2= 0.7013
TDS
Na
R2= 0.4199
TDS
SO4
R2= 0.5569
TDS
Cl
R2= 0.5198
TDS
HCO3
R2= 0.5075
TDS
K
R2= 0.0522
Na
Ca
R2= 0.1917
Na
Mg
R2= 0.6782
Na
CI


R2= 0.2078
Na
SO4
R2= 0.5965
Na
HCO3
R2= 0.2557
Na
20 K
R2= 0.3107
Ca
Mg
R2= 0.0992
Ca
Cl
R2= 0.4847
Ca
SO4
R2= 0.0058
Ca
HOC3
R2= 0.0843
Ca
K
R2= 0.3411
Mg
Cl
R2= 0.4706
Mg
SO4

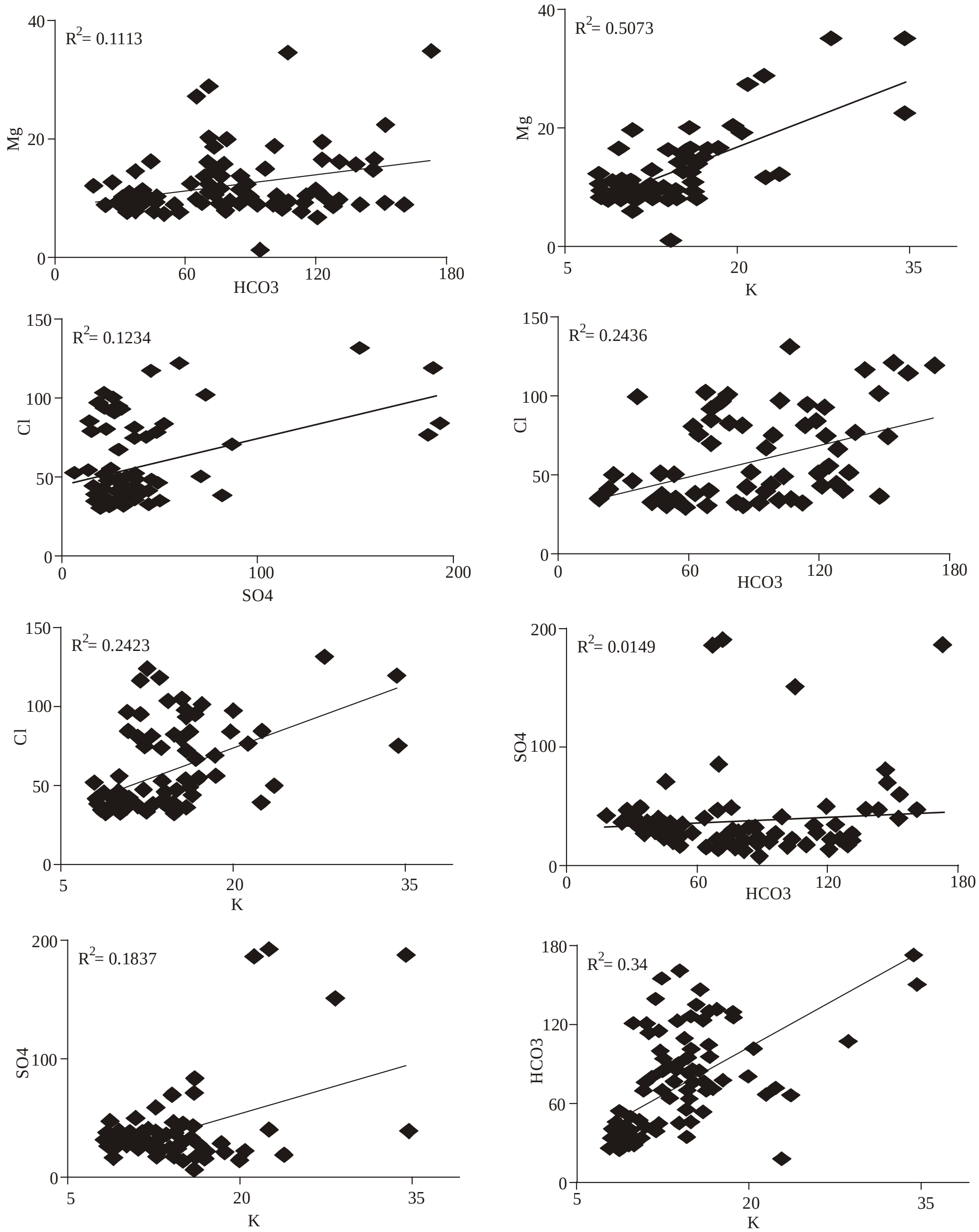


Fig. 3: Graphical correlation matrix of hydrochemical data in the study area

Table 4: Correlation matrix of the different hyrdrochemical compositions along the study area

Variables	TDS	Na	Ca	Mg	Cl	SO_4	HCO_3	K
TDS	1							
Na	0.84	1						
Ca	0.48	0.23	1					
Mg	0.73	0.44	0.56	1				
Cl	0.75	0.82	0.31	0.58	1			
SO_4	0.65	0.46	0.70	0.69	0.35	1		
HCO_3	0.72	0.77	0.08	0.33	0.49	0.12	1	
K	0.71	0.51	0.29	0.71	0.49	0.43	0.58	1

- All data have positive relations which reflect a direct relationship with all the hydrochemical data.
- The Correlation coefficient values vary from 0.08 (weak positive relation as detected between (HCO3) and Ca) to 0.84 (strong positive relation as found between TDS and Na).
- Good correlation between TDS and each of Na^+, Cl^- and $(SO_4)^{2-}$ (0.84, 0.75 and 0.67, respectively) indicate mixing between water of different geneses,

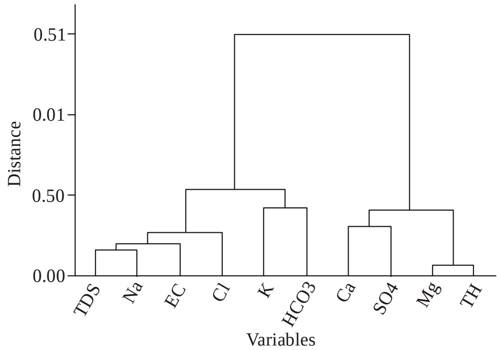


Fig. 4: The dendrogram showing the clustering of hydrochemical constitute of the groundwater samples in the study area

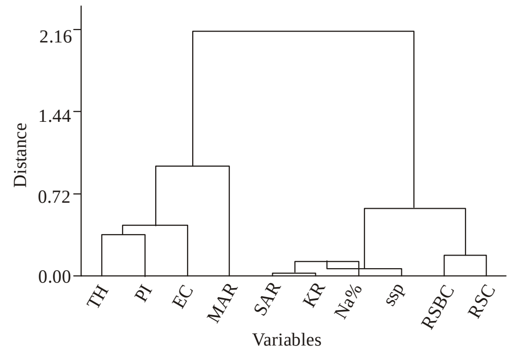


Fig. 5: The dendrogram showing the clustering of physiochemical parameters of the groundwater samples in the study area

leaching of lacustrine sediments which rich in Cl^- and Ca ions and percolation of meteoric water which rich in Na^+ and $(SO_4)^{2-}$.

- Good correlation between Na^+ and Cl^- is revealed, which confirms mixing between water of different geneses.
- Weak correlation notes between HCO_3 and Ca^{2+}, $(SO_4)^{2-}$, Ng^{2+} and Cl^- (0.08, 0.12, 0.33 and 0.49, respectively).

Cluster analysis: It provides a useful means of detecting the existence of groups of similar objects in a high dimensional space. The aim of cluster analysis is to divide the data into subsets (clusters) such that the similarity of the objects within any subset is greater than their similarity with the other subsets.Results were reported in the form of dendrograms. On the basis of the connecting distances between parameters and their positions on the dendrograms, distinctive clusters of the variables were defined. Though this procedure is subjective, the distinction between clusters in this analysis is quite clear from the dendrograms.

Two cluster analysis was performed in the study area for groundwater samples using hydrochemical composition (TDS, Ca, Mg, Na, K, HCO_3, SO_4, Cl, EC and TH) (Fig. 4) and using physiochemical parameters (PI, TH, MAR, EC, SAR, KR, Na%, RSBC, RSC and SSP) as variables (Fig. 5).

From the cluster analysis of the hydrochemical constitute (Fig. 4), two main groups are visible:

- Group 1 comprises TDS, Na, EC and Cl. It contain cluster of TDS and Na with two independents variables (EC and Cl) at linkage distance of 0.25. It indicates that TDS in the water has a dominant contribution from the Na, EC and Cl in the groundwater.
- Group 2 is made up of three clusters. HCO_3-K cluster, $CaSO_4$ cluster and Mg-TH cluster, at linkage distance 0.45, 0.35 and 0.1, respectively. Three main groups are visible from the results of the cluster analysis of the physiochemical parameters (Fig. 5).
- Group 1 contain two independents variables (MAR and EC) with cluster of TH and PI (at linkage distance of 0.35)
- Group 2 is made up of SAR, KR, Na% and SSP physiochemical parameters. It contain SAR-KR cluster at very low linkage distance (0.05) with Na% and SSP as independents variables, indicates good contribution between these parameters along the study area.
- Group 3 represents by RSBC-RSC cluster at linkage distance of 0.25.

Skwenss and kurtosis: Skewness and kurtosis are terms that describe the shape and symmetry of the distribution of geochemical data along the study area. Unless you plan to do inferential statistics on your data set skewness and kurtosis only serve as descriptions of the distribution of your data. In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable.

In the study area, Skewness values for the different hydrochemical data and parameters (Table 5) vary from 3.22 to -1.36. Positive skewness were notice in TDS, TH, T, EC Ca, Mg, Na, K, HCO_3, SO_4, Cl, SAR, KR, RSBC, SSP and Na%. This indicates that the shape of their statistical distribution diagrams show the tail on the right side (direction of high values) is longer than the left side and the bulk of the values (possibly including the median) lie to the left of the mean for each parameter.

The skewness values of pH, PI, MAR and RSC are negative skew indicates that the tail of their distribution diagrams on the left side of the probability density

Table 5: Skewness and kurtosis values for the hydrochemical data in the study area

Variable	Skewness	Kurtosis
TDS (ppm)	2.04	5.59
Ca (ppm)	2.57	9.28
Mg (ppm)	2.52	8.51
Na (ppm)	2.17	5.38
K (ppm)	1.95	6.29
HCO_3 (ppm)	0.70	-0.47
SO_4 (ppm)	4.04	18.17
Cl (ppm)	1.37	0.76
TH (ppm)	3.07	11.35
EC	2.20	5.89
Temp	1.10	1.68
pH	-1.20	4.01
SAR	2.63	8.06
PI	-0.23	-0.03
KR	3.22	11.74
MAR	-1.36	11.93
RSBC	0.57	-0.65
SSP	0.94	0.77
Na%	0.94	0.77
RSC	-0.40	2.32
NaCl (ppm)	1.52	1.49

function is longer than the right side and the bulk of the values lie to the right of the mean.

In probability theory and statistics, kurtosis is a measure of the peakedness of the probability distribution of a real-valued random variable, although some sources are insistent that heavy tails and not peakedness, is what is really being measured by kurtosis. Higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. Kurtosis values (Table 5) vary from 18.17 (for SO_4 values) to -0.65 (for RSBC values). Positive Kurtosis characterize the statistical distribution diagrams of TDS, Ca, Mg, Na, K, SO_4, Cl, TH, EC, T, pH, SAR, KR, MAR, SSP, RSC, NaCl and Na% from the geochemical data along the study area. This indicates a peaked distribution relative to a normal distribution of the hydrochemical data. The values of HCO_3, PI and RSBC are negative Kurtosis indicates a flat distribution. The SO_4, KR, MAR and TH characterized by high kurtosis values, indicates tend to have a distinct peak near the mean and have heavy tails.

CONCLUSION

During the present study, the chemical analyses for 125 groundwater samples were used to apply the multivariate statistical analysis. Eighteen parameters include hyrdrochemical compositions and the physicochemical parameters were used in the statistical analysis. The SD of cations (Ca^{2+}, Mg^{2+} and K^+) are low, while SD values of the anions (Cl^-, $(SO_4)^{2-}$ and $(HCO_3)^-$) are high. This reflects many variations exist between the concentrations distributed along the area with different spread from the average.

The linear regression used to explain or predict the behavior of a dependent variable. Two variables were used to develop a relationship between TDS as an independent variable and different hyrdrochemical data as a dependent variable. Using these equations, by known TDS value, the equation tries to predict any unknown other variables. The linear regression equations used also between the EC as an independent variable and all different water quality variables.

The correlation matrix performed for the groundwater using the hyrdrochemical compositions. All data have positive relations between each others reflect a direct relationship with all the hydrochemical data. The correlation coefficient values vary from 0.84 to 0.08. Good correlation observed between TDS and each of other variables, while weak positive relation detected between (HCO_3) and Ca^{2+}, $(SO_4)^{2-}$, Ng^{2+} and Cl^-.

Two cluster analysis was performed in the study area for groundwater samples using hydrochemical composition (TDS, Ca, Mg, Na, K, HCO_3, SO_4, Cl, EC and TH) and using physiochemical parameters (PI, TH, MAR, EC, SAR, KR, Na%, RSBC, RSC and SSP) as variables. Two main groups are visible from the first cluster. Group 1 (comprises TDS, Na, EC and Cl) contain cluster of TDS and Na with two independents variables (EC and Cl) at linkage distance of 0.25. It indicates that TDS in the water has a dominant contribution from the Na, EC and Cl in the groundwater. Group 2 is made up of three clusters HCO_3-K, Ca-SO_4 and Mg-TH clusters, at linkage distance 0.45, 0.35 and 0.1, respectively.

Three main groups are visible from the results of the second cluster. Group 1 contain two independents variables (MAR and EC) with cluster of TH and PI (at linkage distance of 0.35); group 2 is made up of SAR, KR, Na% and SSP physiochemical parameters. It contain SAR-KR cluster at very low linkage distance (0.05) with Na% and SSP as independents variables, indicates good contribution between these parameters along the study area; and group 3 represents by RSBC-RSC cluster at linkage distance of 0.25.

Skewness and kurtosis are calculated for all data in the study area to describe the shape and symmetry of the distribution of geochemical data along the study area. Skewness values vary from 3.22 to -1.36. A positive skewness were notice in TDS, TH, T, EC, Ca, Mg, Na, K, HCO_3, SO_4, Cl, SAR, KR, RSBC, SSP and Na% indicates that the shape of their statistical distribution diagrams show the tail on the right side (direction of high values) is longer than the left side and the bulk of the values (possibly including the median) lie to the left of the mean for each parameter. Kurtosis values vary from 18.17 (for SO_4) to -0.65 (for RSBC). A positive Kurtosis detected in TDS, Ca, Mg, Na, K, SO_4, Cl, TH, EC, T, pH, SAR, KR, MAR, SSP, RSC, NaCl and Na% indicates a peaked distribution relative to a normal distribution of the data, while the other are negative (indicates a flat distribution). The SO_4, KR, MAR and TH characterized by high kurtosis values, indicates tend to have a distinct peak near the mean and have heavy tails.

REFERENCES

Abdel Ati, A.A., 2002. Hydrogeological studies on the Nubian Sandstone Aquifer in Bahariya and Farafra depressions, Western Desert, Egypt. Ph.D. Thesis, Fac. Sci., Ain Shams Univ., pp: 164.

Anazawa, K. and H. Ohmori, 2005. The hydrochemistry of surface waters in Andesitic Volcanic area, Norikura volcano, central Japan. Chemosphere, 59: 605-615.

Anazawa, K., H. Ohmori, T. Tomiyasu and H. Sakamoto, 2003. Hydrochemistry at a volcanic summit area, Norikura, central Japan. Geochim. Cosmochim. Ac., 67(18S): 17.

Basta, E.Z. and H.I. Amer, 1970. Geological and Petrographic studies on El Gedida area. Bahariya Oasis, UAR. Bull. Fac. Sci. Cairo Univ., 43: 189-215.

Belkhiri, L., A. Boudoukha, L. Mouni and T. Baouz 2011. Statistical categorization geochemical modeling of groundwater in Ain Azel plain (Algeria). J. African Earth Sci., 59(1): 140-148.

Cameron, E.M., 1996. The hydrochemistry of the fraser river, british columbia: Seasonal variation in major and minor components. J. Hydrol., 182: 209-215.

CEDARE-Center for Environment and Development for Arab Region and Europe, 2001. Regional strategy for the utilization of the Nubian Sandstone Aquifer System. Cairo, V.I, II and III.

Cheng-Shin, J., 2010 Applying scores of multivariate statistical analyses to characterize relationships between hydrochemical properties and geological origins of springs in Taiwan. J. Geochem. Explor., 105: 1-2.

Chenini, I. and S. Khemiri, 2009 Evaluation of ground water quality using multiple linear regression and structural equation modeling. Int. J. Environ. Sci. Tech., 6(3): 509-519.

Diab, M.S., 1972. Hydrogeological and Hydrochemical studies of the Nubian Sandstone water-bearing complex in some localities in United Arab Republic. PhD. Thesis, Assiut University, Egypt.

Duffy, C.J. and D. Brandes, 2001. Dimension reduction and source identification for multispecies groundwater contamination. J. Contam. Hydrol., 48: 151-165.

El-Akkad, S. and B. Issawi, 1963. Geology and iron ore deposits of the Bahariya Oasis: Geological Survey and Mineral Research Department, Egypt, Paper 18, pp: 301.

El-Aref, M.M., M.A. Sharkawi and M.M. Khalel, 1999. Geology and genesis of the stratabound and strataform Cretaceous-Eocene iron ore deposits of El-Bahariya Region, Western Desert, Egypt. 4th International Conference on Geology of the Arab World, Cairo University, pp: 450-475.

El-Bassyouny, A.A., 1978. Structure of the northeastern plateau of the Bahariya Oasis, Western Desert, Egypt. Geologie en Mijnbouw, 57: 77-86.

Euro and C. Pacer, 1983. Regional development plan for New Valley. Arab Republic of Egypt. Vol. 1 (Main Report) and Vol. 2 (Soils and Groundwater).

Ezzat, M.A., 1974. Groundwater series in the Arab Republic of Egypt; exploitation of groundwater in El-Wadi El-Gedid project area. General Desert Development Authority. Ministry of Irrigation, Cairo.

Güler, C. and G.D. Thyne, 2004. Hydrologic and geologic factors controlling surface and Groundwater chemistry in Indian wells-Owens Valley area, southeastern California, USA. J. Hydrol., 285: 177-198.

Güler, C., G.D. Thyne, J.E. McCray and A.K. Turner, 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeol. J., 10: 455-474.

Gupta, L.P. and V. Subramanian, 1998. Geochemical factors controlling the chemical nature of water and sediments in the Gomte River, India. Env. Geolog., 36: 102-108.

Hamdan, A. and R. Sawires, 2011. Hydrogeological studies on the Nubian sandstone aquifer in El-Bahariya Oasis, Western Desert, Egypt. Arab Journal of Geosci., pp: 1-15, doi:10.1007/s12517-011-0439-8.

Hartmann, J., Z. Berner, S.D. Doris and N. Henze, 2005. A statistical procedure for the analysis of seismotectonically induced hydrochemical signals: A case study from the Eastern Carpathians. Rom. Tectonophys. , 405: 77-98.

Himida, I.H., 1964. Artesian water of the oases of Libyan Desert in U.A.R. Ph.D. Thesis, M.G.R.U., Moscow (Russian Language).

Issawi, B., 1972. Review of Upper Cretaceous-Lower Tertaiary stratigraphy in the central and northern Egypt. Am. Assoc. Petr. Geol. B., 56: 1448-1463.

Issawi, B. and S. Labib, 1996. A guidebook for an excursion to Bahariya, Farafra, Dakhla and Kharga oases. Egypt. Geol. Surv. Centrnnial, pp: 60.

Issawi, B., M. El-Hinnawi, M. Francis and A. Mazhar, 1999 The Phanerozoic geology of Egypt. A geodynamic approach. Egypt. Geol. Surv., Spec. Publ., 76: 462.

Khalifa, R.M., 2006. Study of Groundwater Resources Management in El-Bahariya Oasis. Ph.D. Thesis, Fac. Sci. Alexandria Univ., pp: 226.

Kim, K., S. Yun, B. Choi, G. Chae, Y. Joo, K. Kim and H. Kim, 2009. Hydrochemical and multivariate statistical interpretations of spatial controls of nitrate concentrations in a shallow alluvial aquifer around oxbow lakes (Osong area, central Korea). J. Contam. Hydrol., 107: 114-127.

Laaksoharju, M., I. Gurban, C. Skarman and E. Skarman, 1999. Multivariate mixing and mass balance (M3) calculations, a new tool for decoding hydrogeochemical information. Appl. Geochem., 14: 861-871.

Meng, S.X. and J.B. Maynard, 2001. Use of multivariate analysis to formulate conceptual models of geochemical behavior: Water chemical data from the botucata aquifer in sao paulo state, brazil. J. Hydrol., 250: 78-97.

Papatheodorou, G., D.G. Gerasimoula and N. Lambrakis, 2006. A long-term study of temporal hydrochemical data in a shallow lake using multivariate statistical techniques. Ecol. Modell., 193: 759-776.

Parsons, R.M., 1962. Bahariya and Farafra areas, New Valley project, Western Desert of Egypt. Final Report, Egyptian Desert Development Organization, Cairo.

Reimann, C. and P. Filzmoser, 2000. Normal and lognormal data distribution in geochemistry: Death of amyth: Consequences for the statistical treatment of geochemical and environmental data. Environ. Geol., 39: 1001-1014.

Said, R. and B. Issawi, 1964. Geology of northern plateau, Bahariya Oasis, Egypt. Geol. Surv. Egypt, Pap., 29: 41.

Shedid, A., M. Yehia, S. Selim and A. Mohamed, 2005. Multivariate Statistical Analysis and chemical model of water resources at Kom Ombo area, upper Egypt. Eng. Res. J., 100: C90-C122.

Soliman, S.M. and O.A. El-Badry, 1980. Petrology and tectonic framework of the Cretaceous, Bahariya Oasis, Egypt. J. Geolog., 24: 11-51.

Voudouris, K., A. Panagopoulos and J. Koumantakis, 2000. Multivariate statistical analysis in the assessment of hydrochemistry of the northern korinthia prefecture alluvial aquifer system (Peloponnese, Greece). Nat. Res. Res., 9(2): 135-146.

Yidana, S., 2010. Groundwater classification using multivariate statistical methods: Southern Ghana. J. Afri. Earth Sci., 57: 455-469.