

# Hydrogeochemical characteristics prediction using multiple linear regression: case study on unconfined aquifer in northeastern Tunisia

Zouheira Ibn Ali<sup>1</sup> · Youssef Frika<sup>1</sup> · Lamia Laajili Ghzel<sup>2</sup> · Moncef Zairi<sup>1</sup>

Received: 5 April 2016 / Accepted: 21 August 2017 / Published online: 30 August 2017  
© Saudi Society for Geosciences 2017

**Abstract** The main factors and mechanisms controlling the groundwater chemistry and mineralization are recognized through hydrochemical data. However, water quality prediction remains a key parameter for groundwater resources management and planning. The geochemical study of groundwater of a multilayered aquifer system in Tunisia is recognized by measurements of the pH, EC, total dissolved solids (TDS), major ion concentration and nitrates of 36 samples from pumping wells covering the aquifer extension and analyzed using standard laboratory and field methods. The calcite precipitation, gypsum, anhydrite and halite dissolution, and direct and reverse ion exchange are the principal process of chemical evolution in the Nadhour-Saouaf aquifer system. Using stepwise regression, the concentration groups of (Ca, Cl, and NO<sub>3</sub>), (Cl, SO<sub>4</sub>, and Mg), and (Ca and Na) exhibit significant prediction of TDS in Plio-Quaternary, Miocene, and Oligocene aquifer levels, respectively. The highest values of  $R^2$  and adjusted  $R^2$  close to 1 revealed the accuracy of the developed models which is confirmed by the weak difference between the measured and estimated values varying between -12 and 8%. The important uncertainty parameters that affected the estimated TDS are assessed by the sensitivity analysis method. The concentration of (Cl), (Ca and Cl), and (Na) are the major parameters affecting the TDS sensitivity of the Plio-Quaternary, Miocene, and Oligocene

aquifer levels, respectively. Hence, the developed TDS models provide a more simple and easy alternative to other methods used for groundwater quality estimation and prediction as proven from external validation on groundwater samples unconsidered in the model construction.

**Keywords** Groundwater quality · Hydro-chemical · Multiple regression analysis prediction

## Introduction

Water scarcity is a common feature of our modern world and this threat is predicted to be worse in the future (Alcama et al. 1997; MED WS & D WG 2007). Groundwater makes up about 60% of the world's freshwater supply, which is about 0.6% of the entire world's water (EPA 2009). Groundwater also is recognized as one of the most valuable natural resources, immensely important and a dependable source of water in all climatic region all over the world (Todd and Mays 2005; Carreira 2010). This situation may induce severe water crisis (Varghese et al. 2012). The accurate prediction of groundwater quality is essential for sustainable utilization and management of vital groundwater resources either on local or regional scales. In addition, for an effective watershed management strategy, water quality assessment is essential in ensuring the protection of groundwater resources from an unavoidable climate change impact and other problems like industrial revolution, urbanization, agricultural increases, etc. (Subba Rao 2008; De Fraiture and Wichelns 2009; Liao et al. 2012). To that end, it is useful to provide information on water quality, classification of water for various purposes, identification of different groundwater aquifers, assessment of groundwater potential,

✉ Zouheira Ibn Ali  
zouheiramed@yahoo.fr

<sup>1</sup> Laboratory of Water, Energy and Environment, National School of Engineers of Sfax, 1173, 3038 Sfax, PB, Tunisia

<sup>2</sup> Higher School of Agriculture, Mograne, 1121, Zaghouan, Tunisia

and investigation of different chemical processes (Trabelsi et al. 2007; Liu et al. 2008; Chenini and Khemiri 2009). Thus, water-quality indicators must reflect mineralization process, integrate reservoir properties, and be sensitive to groundwater recharge rate and flow direction (Andre et al. 2005; Subba Rao 2008). The water chemistry is an important factor determining its use for domestic and irrigation purposes (WHO 1984; Memon et al. 2011). The chemical composition of groundwater is controlled by many factors that include the composition of precipitation water, climate, way of groundwater flow through the rock types, topography of the region, saline water intrusion in coastal areas, and human activities on the ground surface (Reghunath et al. 2002; Trabelsi et al. 2007; Subba Rao 2008; Liao et al. 2012; Arslan 2013; Bhat and Jeelani 2015). These factors can combine to create diverse water types that change in composition spatially and temporally. Different techniques have been used in an attempt to evaluate water quality, essentially based on chemical ions' correlation and some ions' reports (Pazand and Pazand 2014; Ben Alaya et al. 2014).

Therefore, the predictor models offer a cost-effective option to water-quality management, and they have the potential to be applied elsewhere. Multiple regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that a response or outcome variable can be predicted from the others. Predictor models can be used to supplement regular monitoring by identifying areas that need health warnings or more frequent monitoring and are useful between sampling periods (USEPA 2010). This methodology is widely used in the assessment of same-contaminant indicators: viral and bacterial water pollution (Yates et al. 1985; González-Ramón et al. 2012; Gonzalez and Noble 2014; Herrig et al. 2015), harmful substances (Ozekin 1994; Golfopoulos and Arhonditsis 2002; Yu et al. 2015), nitrate (Liao et al. 2012), metals (Manzoor et al. 2006; Kumaresan and Riyazuddin 2007; Ahsan et al. 2008), and rare earth elements (Janssen and Verweij 2003). Nevertheless, only a limited number of studies focused on the TDS parameter as a groundwater-quality indicator. WHO (1984) considered the TDS as the criteria for the classification of groundwater for domestic purposes.

Given the aforementioned context, the main objective of this work was to investigate the origins of mineralization and to test and validate the applicability of multiple regression analysis for estimating groundwater quality based on the TDS. The model developed was based on chemical analysis data from Nadhour-Saouaf aquifer samples. The data set was divided into a developing set, used for setting up the models, and an independent set, used for the validation. Then, a sensitivity analysis is performed to evaluate the reliability and uncertainty of the estimated TDS values.

## Materials and methods

### Study area

The Nadhour-Saouaf syncline, located in the southeastern Zaghuan City in Tunisia, extends over an area of 400 km<sup>2</sup> and lies between mountain ranges in the north and the north-west and the alluvial range in the south (Fig. 1). The topography varies from 78 to 923 m (Fig. 1). The climate is mostly semiarid, with hot dry summer and wet winter. The mean annual rainfall is about 400 mm; this is much lower than the potential evaporation which exceeds 1560 mm/year. The mean annual temperature is around 18 °C. This region has a rather unstable climate with irregular rainfall quantity and highly variable spatial distribution.

### Geology and hydrogeology

The geological outcrops in the study area range from Jurassic to Quaternary. The Quaternary deposits largely present to the south of Zaghuan area and the Plio-Quaternary formed the syncline fillings, whereas Miocene showed a thick sequence of lignite to the north and medium to coarse sandstone to the south (Hamza 1990; Kacem 2008). The upper Oligocene unit, situated in the bordering edges of infiltration, is mainly composed of coarse sandstone and clay. The Lower Eocene deposits are composed of limestone beds. The Upper Eocene deposits, a thick sequence of clay and marl, constitute the superficial aquifer substrate, outcropping near the borders of the syncline. The topographic heights of Mesozoic formations dominate the north of the studied area (Fig. 2a, b). The hydrogeological SW-NE cross section reveals an aquifer system featured by a synclinal structure and shows a depositional sequence of three main hydrostratigraphic series: Plio-Quaternary formed by conglomerate interstratified by sand and clay beds, the Miocene sandstone layer, and highly permeable Oligocene sandstones interstratified by sand and gravel beds (Fig. 3). The transmissivity ranged between 0.012 and 0.036 m<sup>2</sup>/s with a hydraulic conductivity of  $30 \times 10^{-6}$  to  $8.7 \times 10^{-3}$  m/s (SCET 2002). In the Nadhour-Saouaf watershed, the aquifer has been exploited since the 1980s. In the 2000s, the number of pumping wells increased rapidly, and consequently the extraction volume reached  $5.2 \times 10^6$  m<sup>3</sup> in 2007. Thus, since 1982, the groundwater level has continuously declined with a maximum drawdown of 4.5 m during the period of 1995–1996. As a direct impact of increased water abstraction, the groundwater level decline was spread over the entire aquifer. Consequently, the groundwater becomes overexploited as its natural recharge by rainwater cannot maintain a safe yield situation. To improve the natural recharge regime of the water table, an artificial groundwater recharge infrastructure was implemented

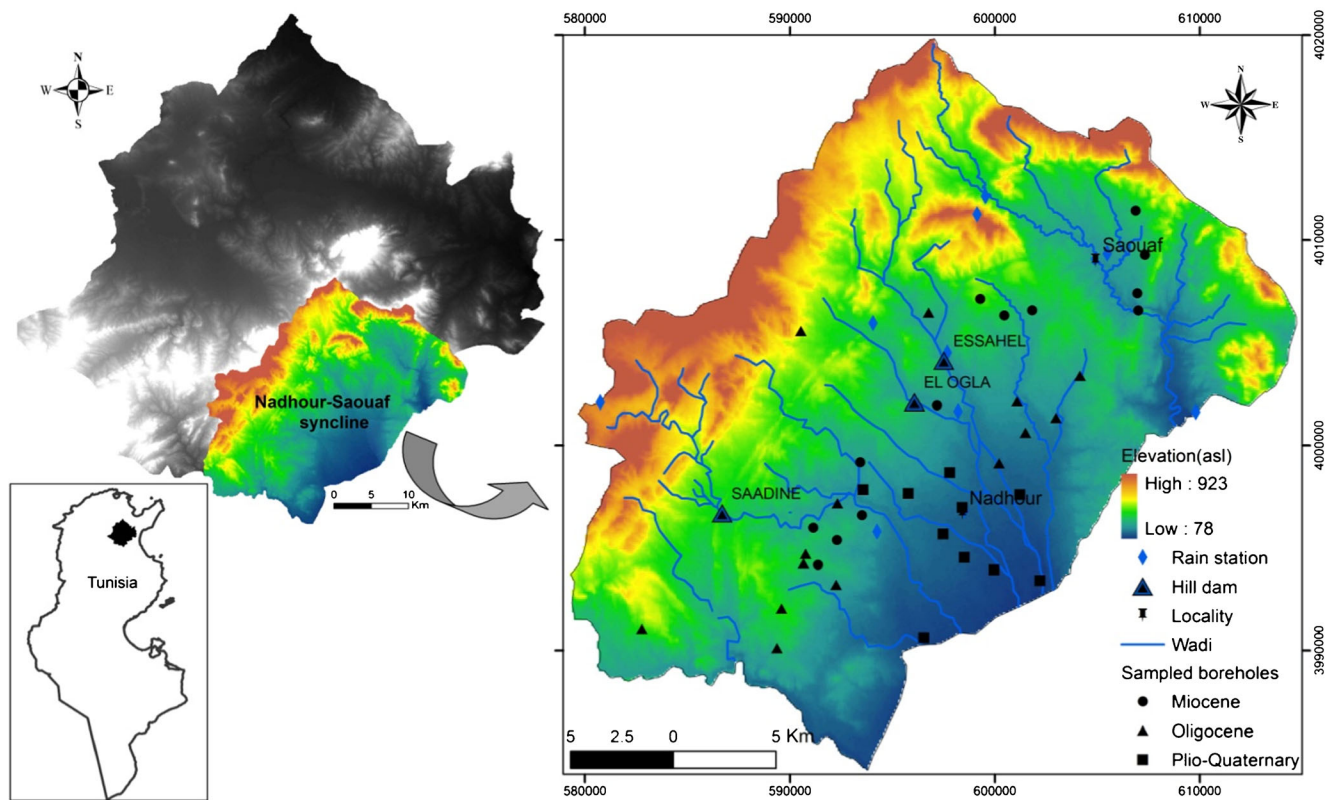


Fig. 1 Study area and groundwater sampling wells locations

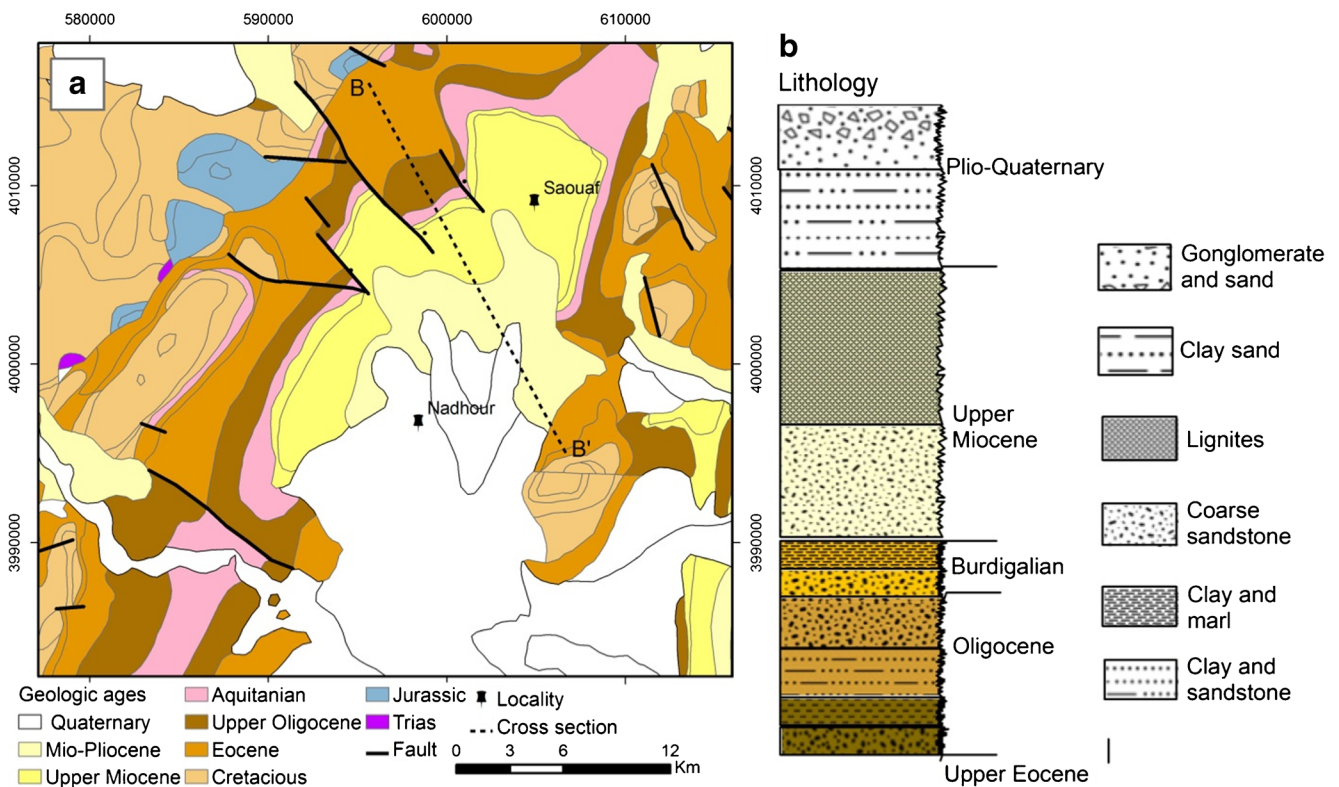
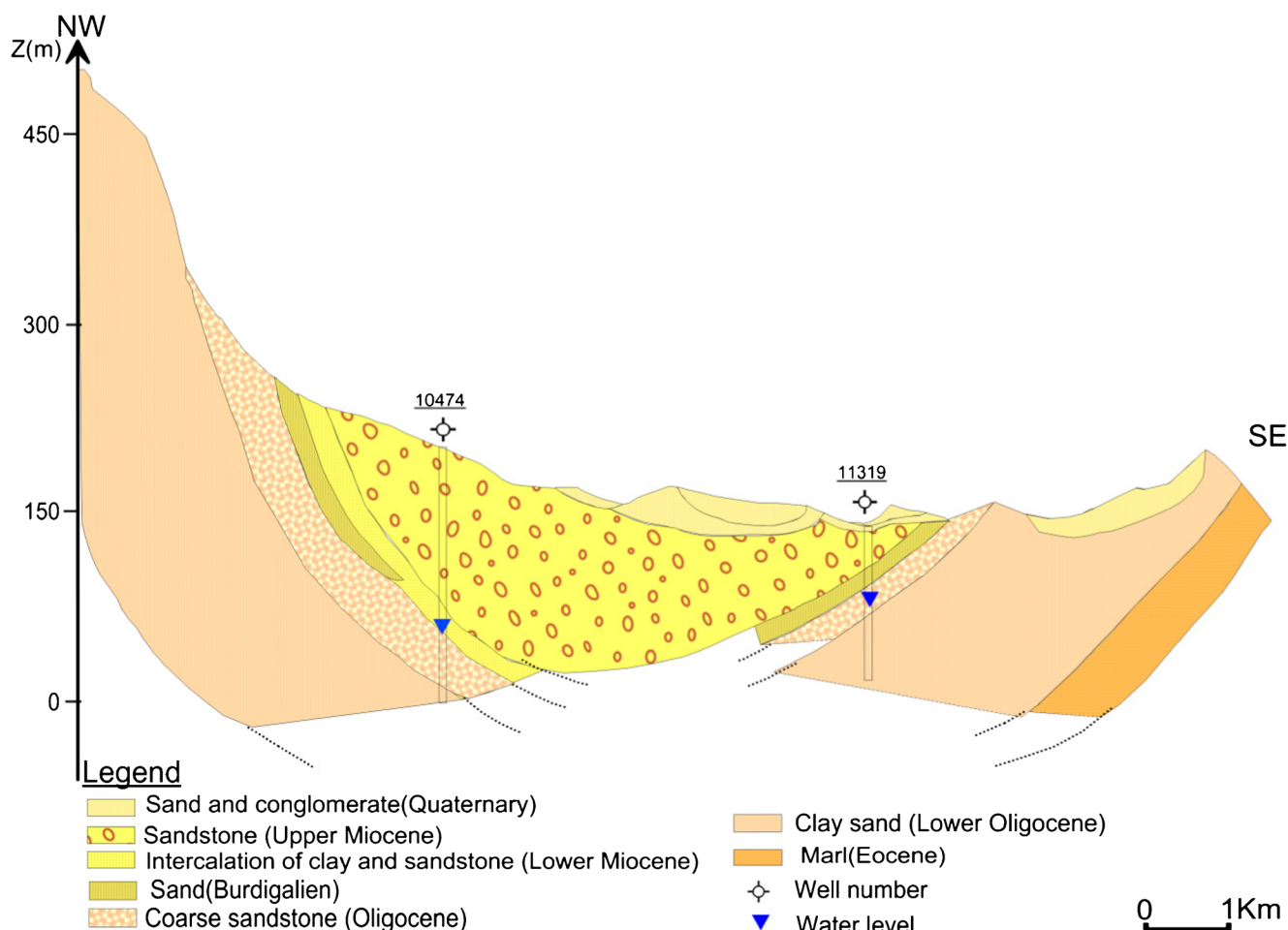


Fig. 2 Geological map (a) and synthetic litho-stratigraphic column of the study area (b)



**Fig. 3** Hydrogeological cross section along transect BB' (see Fig. 1 for location).

(Saadine Essahel and El Ogla hill dams) which consists of releasing water downstream the hill dams in the wadi bed.

Naturally, this aquifer is recharged by water flowing from the surrounding mountains, as well as by local infiltration from several rivers (Zammouri and Feki 2005). The groundwater flow direction is toward the over-pumping zone in the SE area (Fig. 4). The highest piezometric levels are at the NE of the study area, and the lowest ones are close to the Nadhour area. Water of the Nadhour-Saouaf aquifer is used erratically by different economic sectors. However, drinking-water supply and irrigation remain the primary use of groundwater resources.

### Groundwater sampling and analysis

Groundwater samples from 36 pumping wells in the Nadhour-Saouaf multilayered aquifer system were collected in 2010. All samples were obtained from existing water-supply boreholes used for domestic and agricultural purposes. The samples were divided into three main groups corresponding to three reservoir levels: the Plio-Quaternary, the Miocene,

and the Oligocene. An attempt was made to choose sampling locations that are uniformly distributed throughout the study area (Fig. 1). The samples were collected after 10 min of pumping and stored in polyethylene bottles. Immediately after sampling, temperature ( $T^{\circ}\text{C}$ ) and electrical conductivity (EC) were measured with a conductivity meter. pH was measured using a pH meter. Chemical analyses of major elements ( $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ , and  $\text{HCO}_3^-$ ) and nitrates ( $\text{NO}_3^-$ ) were carried out in the Laboratory at the National School of Engineers of Sfax (LARSEN) employing standard methods reported in Table 1. The total dissolved solids (TDS) in mg/l were determined gravimetrically.

The analytical data quality was ensured through careful standardization; the ionic charge balance of each sample was within  $\pm 5\%$ .

### Multiple linear regression

One of the classical problems in engineering investigation is to make predictions. Usually, such predictions require a formula to be found which explain a relationship between a

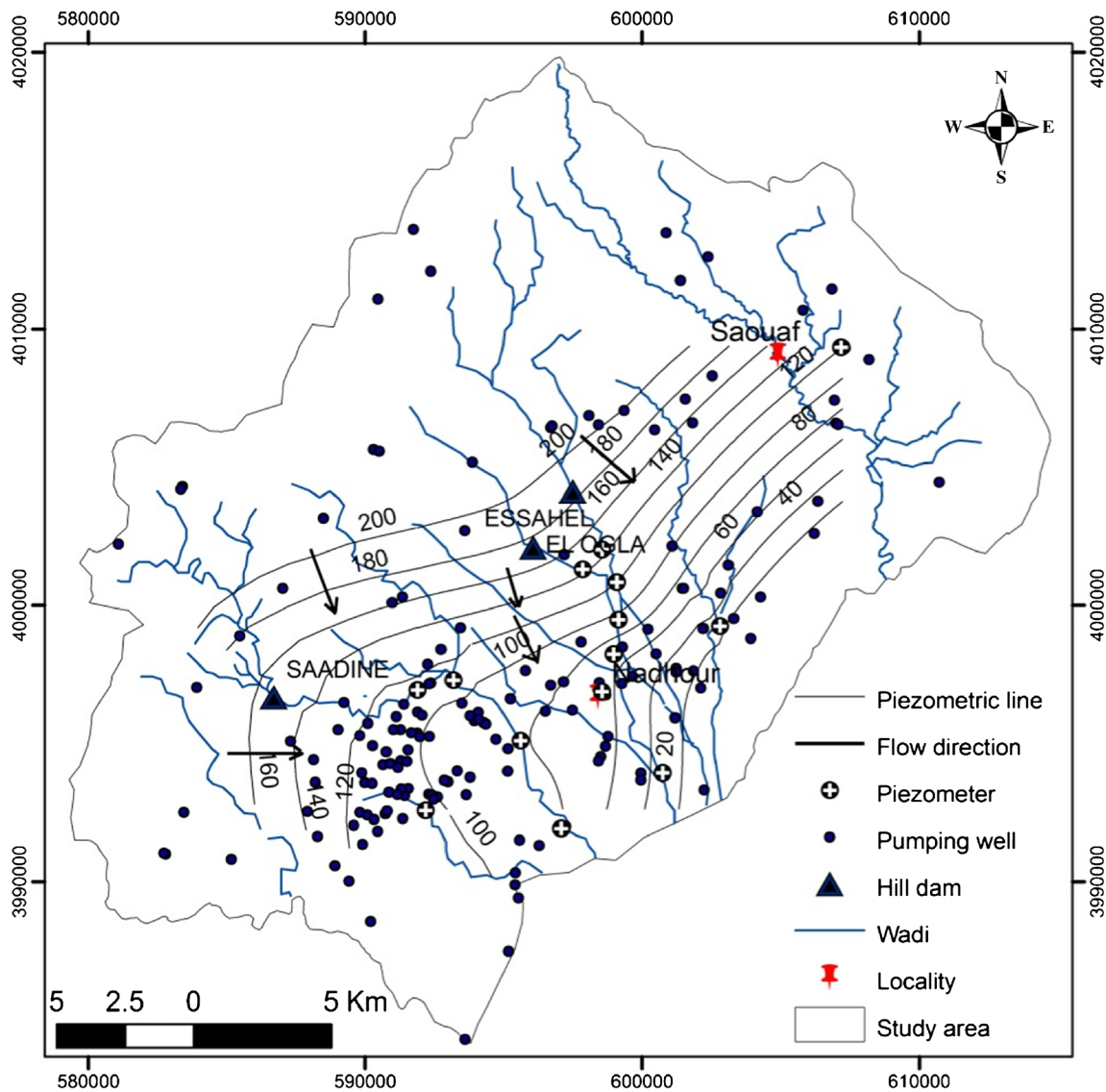


Fig. 4 Piezometric map of Nadhour-Saouaf aquifer in 2010

response variable and a set of regression variables (Tabari et al. 2011). Multiple linear regression (MLR) analysis is commonly used to describe quantitative relationships between the dependent variable and one or more independent variables (Shirsath and Singh 2010). This method is successfully used by different authors to establish statistical models (Ghasemi

and Saaidpour 2007). TDS is a common parameter in water-quality monitoring as it constitutes an excellent indicator of groundwater quality. Therefore, MLR method provides an equation linking the independent variables  $V_i$  ( $[Na]$ ,  $[Mg]$ ,  $[Ca]$ ,  $[K]$ ,  $[SO_4]$ ,  $[HCO_3]$ ,  $[Cl]$ , and  $[NO_3]$ ) to the dependent variable  $V_d$  (TDS) for this case using a relationship of the following type:

$$V_d = \beta_0 + \beta_i V_{i1} + \dots + \beta_n V_{in} \tag{1}$$

where the intercept ( $\beta_0$ ) and the regression coefficients of descriptors ( $\beta_i$ ) are determined by least square method (Green and Carroll 1996) and  $n$  is the number of groundwater samples.

Before establishing the statistical model, the selection of influential (significant) input variables has been applied. Generally, all of the potential input variables are not equally

Table 1 Used analytical methods

Element	Analytical method
Cl <sup>-</sup>	Titration with AgNO <sub>3</sub>
HCO <sub>3</sub> <sup>-</sup>	Titration with HCl
SO <sub>4</sub> <sup>2-</sup>	Chromatography liquid phase
NO <sub>3</sub> <sup>-</sup> , Na <sup>+</sup> , Mg <sup>2+</sup> , and K <sup>+</sup>	Atomic emission spectrophotometer
Ca <sup>2+</sup>	Titration with EDTA

informative, because some variables may be correlated, noisy, or have no significant relationship with the output to be modeled (Maier and Dandy 1998; Hassan et al. 2014). The selection of model inputs among chemical properties of sampled groundwater was carried out according to multiple variable correlations. The technique of multivariate data analysis gives the correlation coefficient  $r$  for each pair of variables. A correlation coefficient  $r$  is a number between  $-1$  and  $+1$ , which measures the strength of the linear relationship between two variables. For  $r > 0.7$ ,  $0.5 \leq r \leq 0.7$ , and  $r < 0.5$ , the variables were considered, respectively, strongly correlated, moderately correlated, and poorly correlated. The standardized skewness and standardized kurtosis were used to determine whether the sample comes from a normal distribution. Values of these statistics outside the range of  $-2$  to  $+2$  indicate significant departures from normality, which would tend to invalidate many of the statistical procedures normally applied to this data. A variable with standardized skewness and standardized kurtosis values outside the expected range is not influential. Then, variables were retained in the model on a significance level of  $P < 0.05$ .

The effectiveness of the developed MLR models was measured by a set of standard statistical indicators and the ANOVA table (Makridakis et al. 2008), namely, multiple correlation coefficient ( $R^2$ ), adjusted  $R^2$ ,  $P$  value, mean absolute error (MAE), and Durbin–Watson statistic (DWS). The

relationships between variables were established using the stepwise regression method (Bernstein 1988; Thayer 2002). Multiple variable analysis and multiple regression modeling were performed using STATGRAPHICS XVI.I statistical program (SPT 2009).

## Results and discussion

### Groundwater chemistry

A statistical summary of the total dissolved solids (TDS) and major ion and nitrate concentrations in the groundwater are presented in Table 2. The TDS values range from 0.9 to 1.8, 0.7 to 1.8, and 0.7 to 2.4 g/l, respectively, for the Plio-Quaternary, Miocene, and Oligocene levels. In order to identify the geochemical processes responsible for the groundwater salinization, the major elements versus TDS and ions versus saturation index was analyzed.

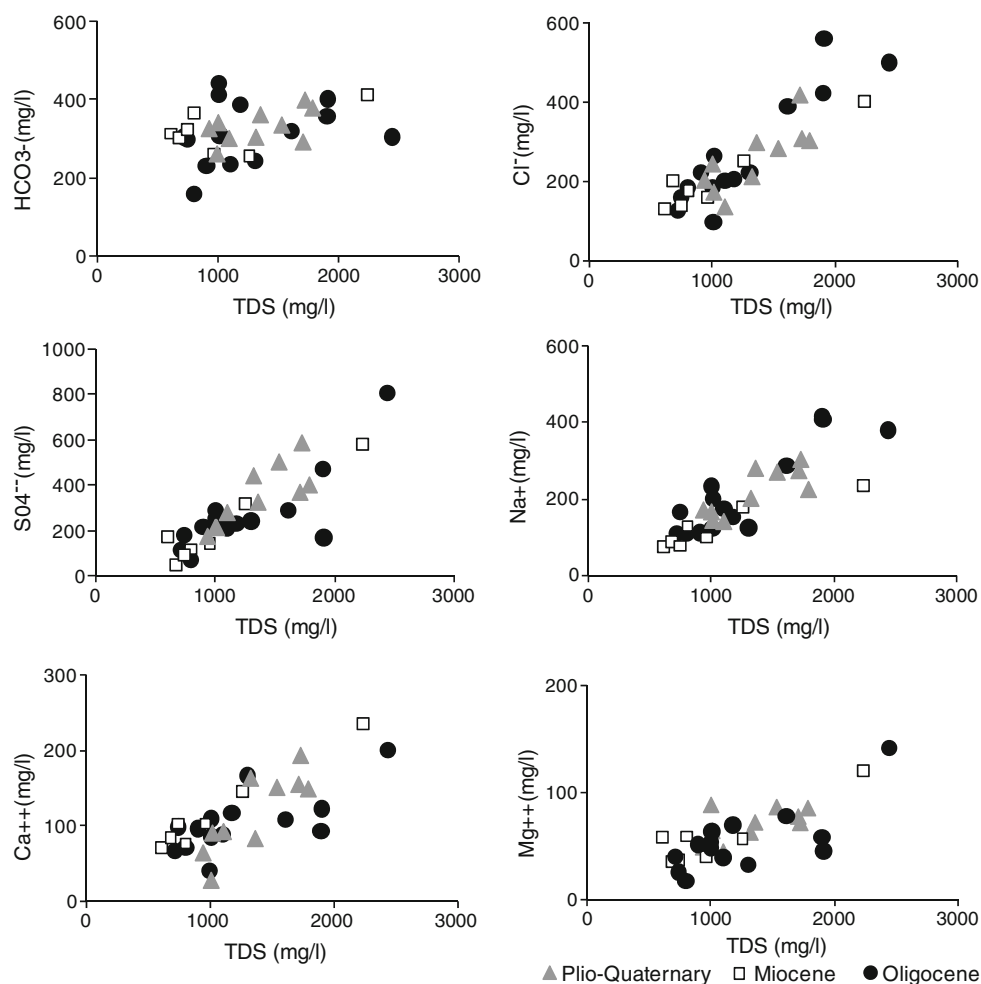
#### TDS versus major elements

In order to identify the geochemical processes contributing to the groundwater salinization in the multilayered Nadhour-Saouaf aquifer system, the relationship between concentrations of major elements and TDS was considered (Fig. 5).

**Table 2** Descriptive statistics of water-quality parameters

	TDS	Ca (mg/l)	Mg (mg/l)	Na (mg/l)	K (mg/l)	Cl (mg/l)	SO <sub>4</sub> (mg/l)	HCO <sub>3</sub> (mg/l)	NO <sub>3</sub> (mg/l)
Quaternary level ( $n = 10$ )									
Average	1.3	116.3	70.6	218.0	4.7	254.8	350.2	328.8	6.6
St dev	0.33	52	15.3	60.6	2.0	81.5	133.3	41.7	5.8
Min	0.9	28	45	142.0	2.03	133	176	261	0
Max	1.8	192	89	303	8.1	415	583	397	14.6
St skew	0.1	-0.3	-0.6	0.1	0.6	0.6	0.5	136.0	14.6
St kurtosis	-1.1	-0.6	-0.5	-1.1	-0.7	0.2	-0.5	0.1	0.1
Miocene level ( $n = 13$ )									
Average	1.04	253.68	245.6	11.0	209.1	84.8	60.3	154.4	4.9
St dev	0.34	108.8	92.2	16.8	147.3	60.4	30.3	66.3	2.2
Min	0.68	49.0	106.5	0	44.0	24.0	15.8	78.0	2.0
Max	1.81	414.0	400	54	581	236	120	297	9.1
St skew	1.7	-0.4	0.6	2.7	1.9	2.0	0.98	1.2	0.7
St kurtosis	0.5	-0.3	-0.4	2.1	1.7	1.6	0	0	-0.4
Oligocene level ( $n = 13$ )									
Average	1.3	104.6	56.4	218.6	5.1	274.5	272.7	314.3	7.8
St dev	0.5	41.8	30.0	116.5	2.8	145.5	184.6	82.9	15.8
Min	0.7	40	17	112	2.5	96	69.58	158.5	0
Max	2.4	200	141	415	10	560	800	440	44
St skew	1.6	1.4	2.8	1.2	1.2	1.3	3.2	-0.3	2.8
St kurtosis	0.3	1.0	3.9	-0.7	-0.5	-0.2	4.2	-0.5	1.7

**Fig. 5** Major elements concentration versus TDS relationships



For the Plio-Quaternary level, these diagrams indicate a well-defined correlation characterizing the relationship of  $\text{Ca}^{2+}$ ,  $\text{SO}_4^{2-}$ ,  $\text{Na}^+$ , and  $\text{Cl}^-$  versus TDS with correlation coefficients of 0.83, 0.82, 0.81, and 0.79, respectively. The relationships suggest these major elements' contribution to the water mineralization of the Plio-Quaternary level. Except the  $\text{HCO}_3^-$ , all major elements and TDS are well correlated for the Miocene level which indicates that the total salt content is mainly controlled by the concentrations of these elements, with correlation coefficients exceeding 0.85. Major elements  $\text{Cl}^-$ ,  $\text{Na}^+$ ,  $\text{SO}_4^{2-}$ , and  $\text{Mg}^{2+}$  are well correlated with TDS, with correlation coefficients of 0.9, 0.88, 0.8, and 0.7, respectively, which can further elucidate elements controlling mineralization in the Oligocene level.

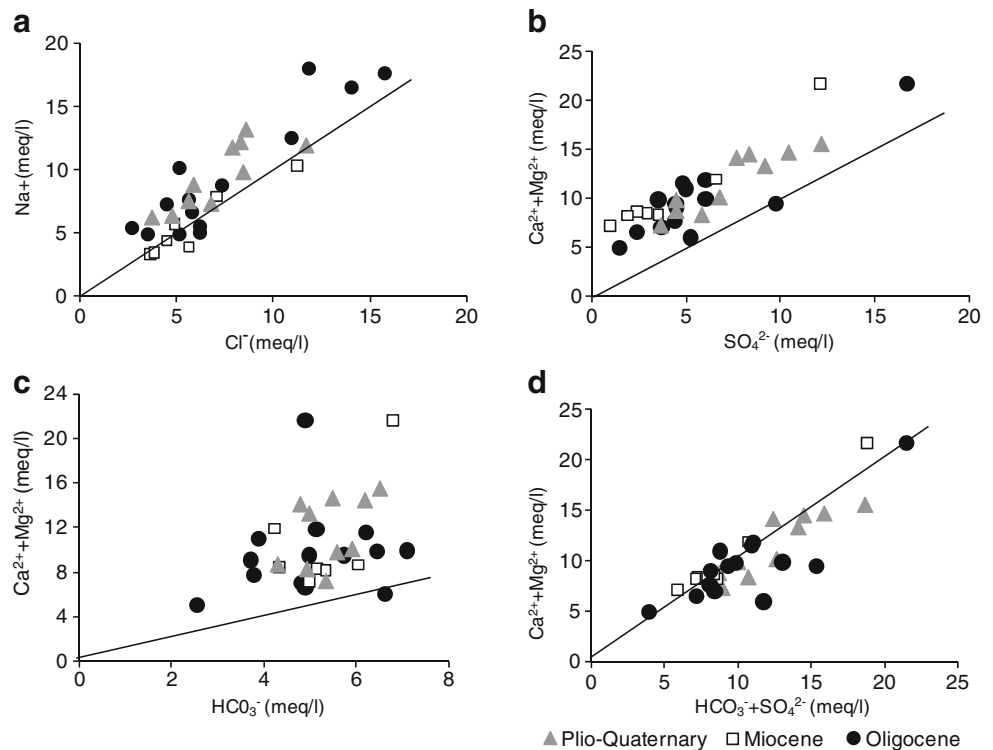
#### Relationship between ions

In order to underline the mechanisms of groundwater mineralization, relationships between the major elements were investigated. The  $\text{Na}^+/\text{Cl}^-$  relationship shows a high correlation with relatively high concentrations in ions  $\text{Na}^+$  compared to ions  $\text{Cl}^-$  as a consequence of halite dissolution as a

major process of mineralization (Fig. 6a). The excess of  $\text{Na}^+$  relative to  $\text{Cl}^-$  may be explained by the intervention of other processes like cation exchange between groundwater and the clay fraction of the aquifer material (Trabelsi et al. 2007; Kraiem et al. 2012).  $\text{SO}_4^{2-}$  versus  $(\text{Ca}^{2+} + \text{Mg}^{2+})$  shows a good positive correlation for all samples of the various aquifer levels (Fig. 6b) due to the probable cation exchange by generation of a  $\text{Ca}^{2+}$  deficiency relatively to  $\text{SO}_4^{2-}$  concentration. Figure 6c shows that dolomite dissolution is probably the source of  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$  in addition to other sources of these ions for all aquifer levels. The plot of  $(\text{Ca}^{2+} + \text{Mg}^{2+})$  versus  $(\text{HCO}_3^- + \text{SO}_4^{2-})$  shows that all Miocene samples and some samples from both Oligocene and Plio-Quaternary are placed close to the 1:1 line indicating that Nadhour-Saouaf groundwater mineralization is controlled by gypsum dissolution. A great number of Plio-Quaternary groundwater samples and a few ones of the Oligocene level are placed in the right due to an excess that happens to be of  $\text{HCO}_3^-$  and  $\text{SO}_4^{2-}$  indicating that groundwater mineralization is controlled by ion exchange in addition to mineral dissolution (Fig. 6d).

Furthermore, the referred exchange is confirmed through the plot of  $((\text{Ca}^{2+} + \text{Mg}^{2+}) - (\text{HCO}_3^- + \text{SO}_4^{2-}))$  versus  $(\text{Na}^+ +$

**Fig. 6** Ion relationship. **a** Na/Cl. **b** (Ca + Mg)/SO<sub>4</sub><sup>2-</sup>. **c** (Ca + Mg)/HCO<sub>3</sub><sup>-</sup>. **d** (Ca + Mg)/(HCO<sub>3</sub><sup>-</sup> + SO<sub>4</sub><sup>2-</sup>)



$K^+ - Cl^-$ ) as shown in Fig. 7, revealing an inverse proportional evolution with a slope of about  $-1$  (Mc Lean et al. 2000; Dassi 2004; Kamel et al. 2005; Kraiem et al. 2012). In the absence of these reactions, all data should plot close to the origin (Mc Lean et al. 2000). Figure 7 shows that the groundwater samples are distributed on both sides indicating that Nadhour-Saouaf groundwater mineralization is also controlled by ion exchange and reverse ion exchange process with clay minerals present in the aquifer material.

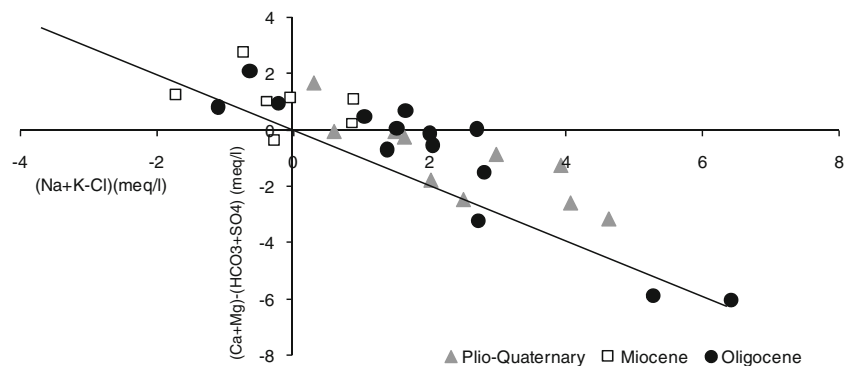
#### Saturation index

The saturation indices (SI) of calcite, dolomite, halite, anhydrite, and gypsum were calculated using the code Diagrammes 6.51. The results are presented in Fig. 8. SI evaluates the level of equilibrium between minerals of water and

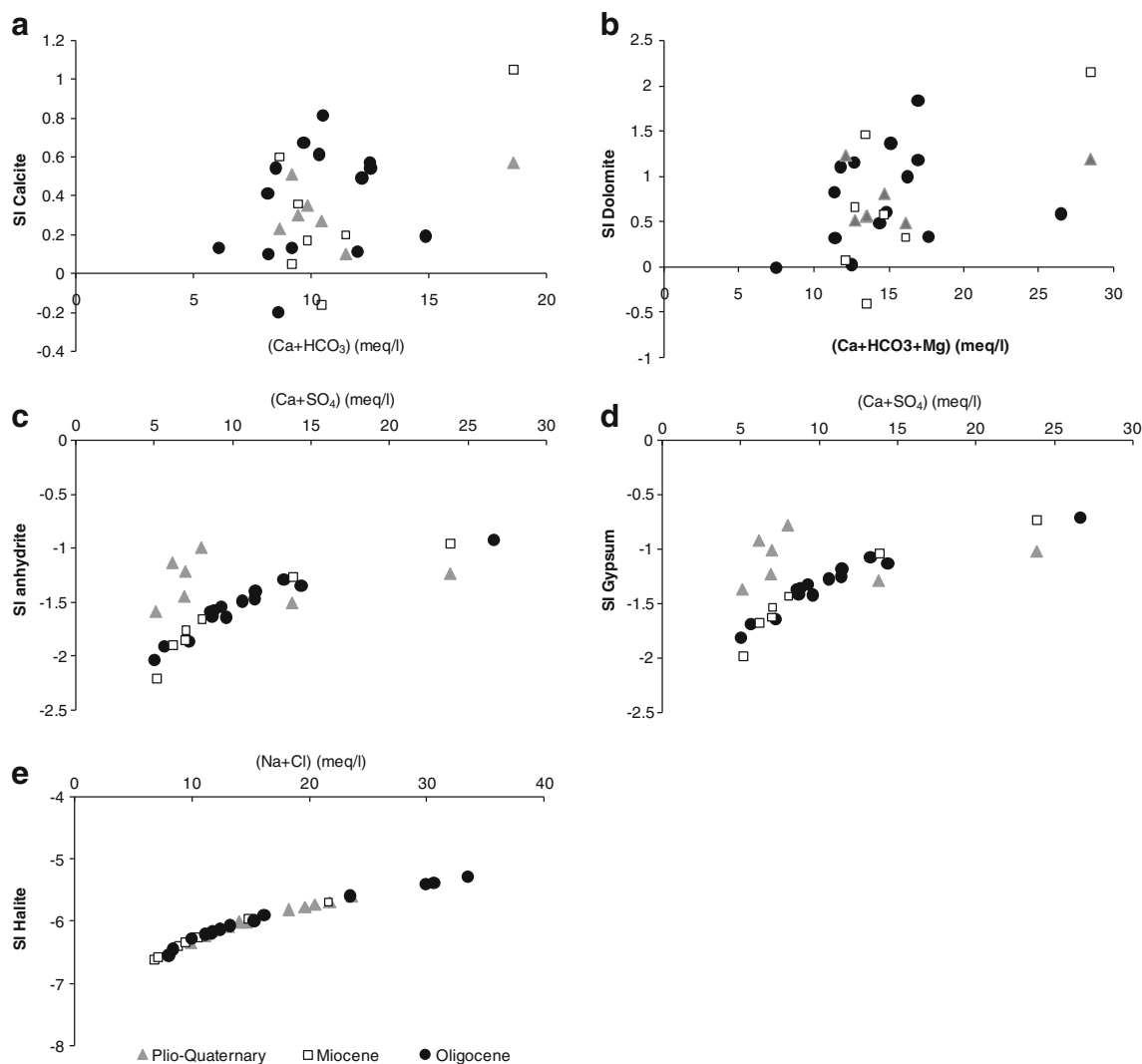
rocks (Elango et al. 2003). If SI is greater than zero ( $SI > 0$ ), the solution is saturated relative to the mineral, and then, precipitation from the groundwater is theoretically possible. When SI is below zero ( $SI < 0$ ), the solution is undersaturated and dissolution continues. If SI is equal to zero ( $SI = 0$ ), the mineral would be either precipitating or dissolving (Qiyan and Baoping 2002; Trabelsi et al. 2007).

As shown in Fig. 8a, b, all samples are supersaturated with respect to calcite and dolomite except for a few ones. In the cases of gypsum, anhydrite, and halite, SI values indicate undersaturation, suggesting that their soluble component Na, Cl, Ca, and SO<sub>4</sub> concentrations are not limited by mineral equilibrium (Guler et al. 2002). For the Miocene and Oligocene levels, the saturation indices of anhydrite and gypsum vary in inverse proportion to the sum of  $(Ca^{2+} + SO_4^{2-})$  (Fig. 8c, d).

**Fig. 7**  $(Ca + Mg) - (SO_4 + HCO_3)$ /(Na + K - Cl) relationship







**Fig. 8** Relationships between saturation indices and major elements in the analyzed groundwater samples. **a** IS (calcite)/(Ca + HCO<sub>3</sub><sup>-</sup>). **b** IS (dolomite)/(Ca<sub>2</sub> + Mg<sub>2</sub><sup>+</sup> + HCO<sub>3</sub>). **c** IS (anhydrite)/(Ca<sub>2</sub><sup>+</sup> + SO<sub>4</sub><sup>2-</sup> Na<sup>+</sup> + Cl<sup>-</sup>). **d** IS (gypsum)/(Ca<sub>2</sub><sup>+</sup> + SO<sub>4</sub><sup>2-</sup>). **e** IS (halite)/(Na<sup>+</sup> + Cl<sup>-</sup>)

The saturation indices of halite versus (Na<sup>+</sup> + Cl<sup>-</sup>) indicate that water evolves from a state close to undersaturated with respect to halite (Fig. 8e). Not only calcite precipitation but also gypsum, anhydrite, and halite mineral dissolution are the principal reactions that determine the chemical evolution in Nadhour-Saouaf aquifer system.

*Nitrates*

Nitrate is a very important parameter for assessing the contamination of groundwater (Einsiedl and Mayer 2006; Ameer et al. 2016). The limit of nitrate content in water is set by the World Health Organization (WHO) at 50 mg/l. High concentrations of nitrates can cause serious health problems. In the study area, this parameter range between 0 and 54 mg/l, with a nitrate average of 8 mg/l. Only one sample exceeds the nitrate concentration limit (50 mg/l), established by the WHO standards.

**MLR model development**

In developing the MLR model, an appropriate number of explicative variables required to provide a more precise simulation results were determined. This approach was used to extract related variables controlling groundwater quality. The range of the considered variables, the average, the standard deviation, the standard skewness, and the standard kurtosis are shown in Table 2. The standardized skewness and standardized kurtosis outside the range of - 2 to + 2 are in bold showing irrelevant variables in groundwater quality. For the Plio-Quaternary aquifer level, all variables can be considered as significant model inputs. The Na was deleted from the input variable list for the Miocene level and the Mg, Cl, and NO<sub>3</sub> were excluded from the list of variable considered in the Oligocene level.

The results of MLR analysis for the three aquifer levels in Nadhour-Saouaf syncline using a stepwise regression

technique are summarized in Table 3. The “beta” (standardized regression coefficients) values show the relative contribution of each independent variable in the prediction of groundwater quality. The “*P* value” refers to the significant variables, which are included in the regression equation.

After fitting the MLR models to the data set, an assessment is made of the adequacy of the fitting for each model. The value of  $R^2$  is 0.96, showing that about 96% of the total variations in the TDS can be accounted for the independent variables for the Plio-Quaternary aquifer level. The fitted model 1 retained three explicative variables (Ca, Cl, and  $\text{NO}_3$ ). For the Miocene water samples, MLR model 2 required other explicative variables (Cl,  $\text{SO}_4$ , and Mg). The high  $R^2$  shows that about 98% of the total variations in the TDS have been explained by these variables. However, for the Oligocene level, only two explicative variables (Ca and Na) appear obligatory to explain the TDS variation by the model 3. The value of  $R^2$  is 0.98, showing that about 98% of the total variations in the TDS can be accounted for the independent variables. Goodness of fit between measured and simulated TDS was assessed by the mean absolute error. A low MAE was recorded for three models. The autocorrelation of residuals was checked by using the Durbin–Watson statistic (Makridakis et al. 2008). The value of Durbin–Watson statistic for the residuals obtained at Plio-Quaternary, Miocene, and Oligocene levels range between 1.54 and 3.34, which is significant at a 95% confidence interval, thereby satisfying the condition of no autocorrelation at lag 1 in the groundwater samples of these aquifer levels. These diagnostic checks were employed for all of the three aquifer levels, and the results indicated that all MLR models satisfy the basic assumptions of MLR technique (Courville and Thompson 2001).

In order to test whether the dependent variable TDS is related to predictor variables, the ANOVA table was used (Williams 2015). Since *P* value is less than the significance level (5%), it indicates a TDS related to predictor variables for all aquifer levels. Table 4 shows that at the 5% significance level, whether it appears that any of the predictor variables can be removed from the full models is unnecessary. The entire coefficients for the models are

**Table 4** Statistical results obtained from the MLR analysis

Parameter	$\beta$	Standard error	<i>t</i> statistic	<i>P</i> value
Model 1				
Ca	0.00212499	0.000818763	2.59537	0.0409
Cl	0.00263781	0.000433984	6.07813	0.0009
$\text{NO}_3$	0.0194928	0.00650624	2.99602	0.0241
Model 2				
Cl	0.00132246	0.00019766	6.69057	0.0001
$\text{SO}_4$	0.00182834	0.000146889	12.4471	0.0000
Mg	-0.00158612	0.000593317	-2.67331	0.0255
Model 3				
Ca	0.00578409	0.000473929	12.2046	0.0000
Na	0.00327165	0.000170161	19.2268	0.0000

significant, i.e., *P* value of the *t* statistic for each coefficient is less than significance level (5%), so all the considered variables are useful as predictors of dependent variable TDS.

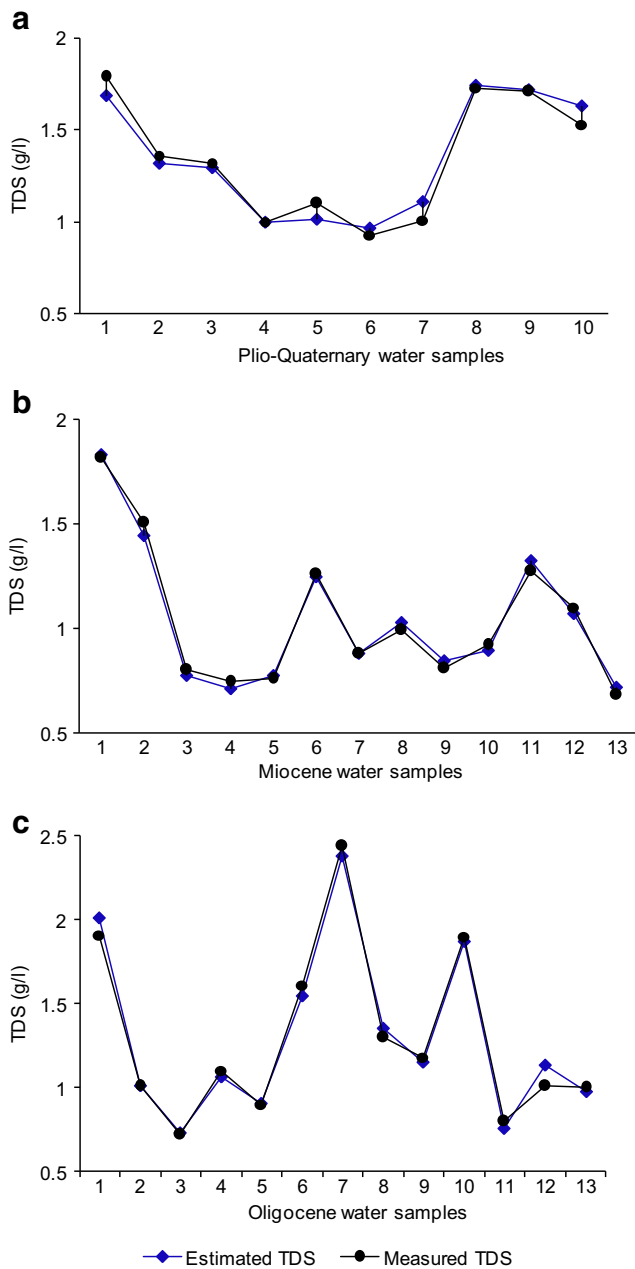
For the qualitative evaluation of the models’ performance, the results of graphical indicators are concerned; a comparison of measured and predicted TDS is shown in Fig. 9a, b, and c for the Plio-Quaternary, Miocene, and Oligocene aquifer levels, respectively. It is apparent from this figure and in Table 5 that all models showed reasonable correlation between measured and predicted TDS. Hence, the produced MLR models are viable tools for monitoring assessment of groundwater-quality status which can enhance groundwater resources management in the area.

### Sensitivity analysis

The groundwater-quality assessment was carried out using chemical data for each aquifer level. Unfortunately, there is a relatively great deal of the uncertainty in the process of chemical parameter analysis. Thus, the water quality uncertainties are compounded by uncertainties relating to the chemical parameters explaining TDS variation for each aquifer level. In the light of these probable uncertainties, a sensitivity analysis was conducted for the MLR models. The aim of this was to

**Table 3** MLR models and their performance

	<i>P</i> value	$R^2$	Adjusted $R^2$	MAE	DWS
Model 1: TDS = 0.299262 + 0.00212499 × Ca + 0.00263781 × Cl + 0.0194928 × $\text{NO}_3$	0.0002	95.74	93.61	0.05	2.23
Model 2: TDS = 0.429459 + 0.00132246 × Cl + 0.00182834 × $\text{SO}_4$ - 0.00158612 × Mg	0.0000	98.80	98.40	0.03	3.37
Model 3: TDS = - 0.0219974 + 0.00578409 × Ca + 0.00327165 × Na	0.0000	98.67	98.40	0.04	1.54



**Fig. 9** Comparison of measured and simulated TDS for **a** Plio-Quaternary level, **b** Miocene level, and **c** Oligocene level

investigate the relative importance of each input variable for accurately predicting groundwater quality. This analysis was carried out for all MLR models by imposing certain changes on individual inputs and observing their effects on the model output. According to the weak quantity of NO<sub>3</sub> in the Plio-Quaternary water samples, the effect of its uncertainty on the TDS assessment is not included in this analysis. The magnitude of the perturbation was ± 10% with respect to the original data, while keeping the other inputs at their original values and then calculating the change in the model output TDS. The uncertainties associated with the chemical parameters were

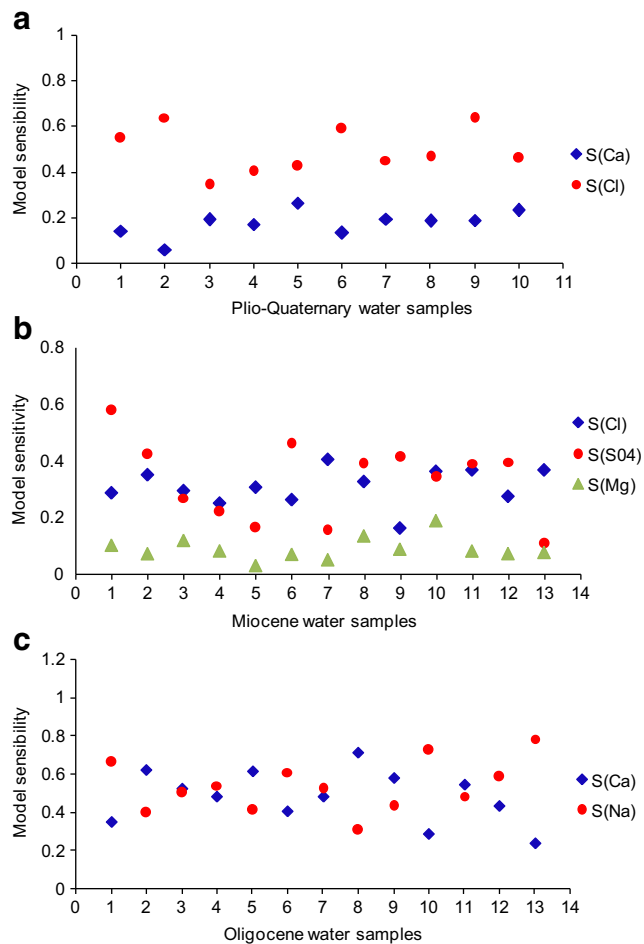
**Table 5** Statistical results of measured and predicted TDS in Nadhour-Saouaf aquifer system

	Min	Max	Average	St dev
Quaternary level (n = 10)				
Measured TDS	0.93	1.79	1.35	0.33
Estimated TDS	0.96	1.74	1.35	0.32
Residual	-0.10	0.10	0.00	0.07
Relative difference, %	-10.14	7.71	-0.26	5.41
Miocene level (n = 13)				
Measured TDS	0.68	1.81	1.04	0.34
Estimated TDS	0.71	1.83	1.04	0.34
Residual	-0.05	0.07	0.00	0.04
Relative difference, %	-5.54	4.78	-0.14	3.71
Oligocene level (n = 13)				
Measured TDS	0.72	2.44	1.30	0.52
Estimated TDS	0.73	2.38	1.30	0.51
Residual	-0.12	0.07	0.00	0.06
Relative difference, %	-11.93	5.61	-0.08	4.77

evaluated by computing a defined relative sensitivity (S) as AS/CP. The CP is the relative change of a given variable or parameter, defined as  $|V_s - V_m|/V_m \times 100$ , and AS is the relative change in the output TDS value, defined as  $|C_s - C_m|/C_m \times 100$ .  $V_s$  and  $V_m$  are variable values used for sensitivity and MLR model original value, respectively, and  $C_s$  and  $C_m$  are output data TDS computed in sensitivity and generated MLR model, respectively (Jiménez-Martínez et al. 2010). Figure 10a, b, and c present the results of the sensitivity analyses for the MLR models of Plio-Quaternary, Miocene, and Oligocene level, respectively. The variation of the sensitivity values with the inputs at each water sample are shown for all aquifer levels. It can be seen from Fig. 10a that in the Plio-Quaternary level groundwater, all the inputs have reasonably low values of sensitivity (the rank ranging between 0.05 and 0.64). Compared to Ca, the variation of the Cl affects significantly the TDS at this level. However, at the Miocene aquifer level, inputs Cl and SO<sub>4</sub> have an irregular value of sensitivity as compared to those for the Mg input (Fig. 10b). It has a low sensitivity value (< 0.2). Therefore, the greatest change in the TDS of Miocene water samples is not due to perturbations in the input Mg but also to variations in the Ca and Cl. For almost all the Oligocene water sample, the TDS sensitivity trend to Na was deduced (Fig. 10c). Therefore, the inputs having the highest sensitivity should be considered with greater accuracy so as to ensure reliable prediction of TDS by the MLR model.

**External validation of the TDS models**

The next task was to conduct a validation test of the calibrated models with data different from those used for model formulation (Table 3). The prediction was made for each

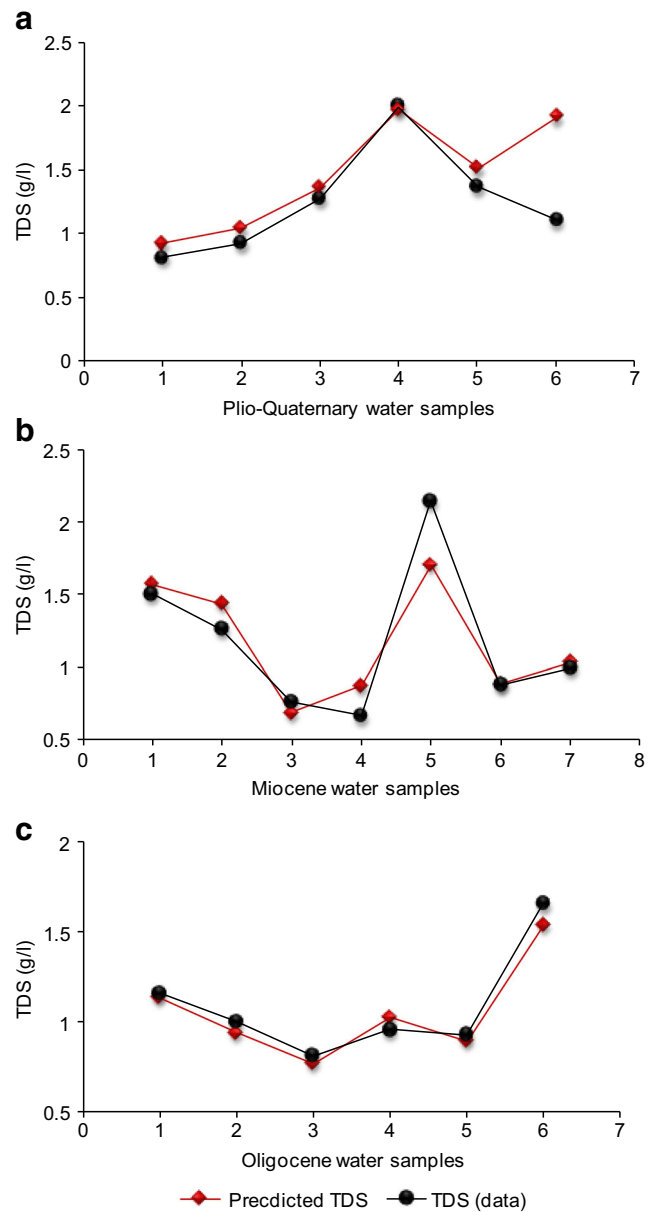


**Fig. 10** Sensitivity of the MLR models to the input variables for **a** Plio-Quaternary level, **b** Miocene level, and **c** Oligocene level

aquifer level separately based on the previously defined predictor variables. Figure 11a–c depicts the predicted TDS values and the TDS data derived from evaluation of the various groundwater samples based on the MLR models of Plio-Quaternary, Miocene, and Oligocene aquifer levels, respectively. Thus, Fig. 11 shows “perfect” simulation. TDS predicted for Nadhour-Saouaf aquifer are in reasonable agreement with the measured data.

## Conclusions

Investigations of groundwater quality have been based on two approaches: (i) hydrochemical characterization using major elements and relationships, (ii) statistical modeling with MLR for simulating and predicting of TDS using physico-chemical parameters (pH, EC, and TDS); major ions concentration [ $\text{HCO}_3^-$ ], [ $\text{Cl}^-$ ], [ $\text{SO}_4^{2-}$ ], [ $\text{Ca}^{2+}$ ], [ $\text{Mg}^{2+}$ ], [ $\text{Na}^+$ ], and [ $\text{K}^+$ ]; and nitrates. Calcite precipitation, gypsum, anhydrite, and halite dissolutions are the main reactions inducing the chemical evolution in Nadhour-Saouaf aquifer system, in addition to direct



**Fig. 11** Validation of the TDS-models for **a** Plio-Quaternary level, **b** Miocene level, and **c** Oligocene level

and reverse ion exchange with clay minerals. Then, for MLR approach, the standard protocols for MLR modeling, as well as all the pertinent and influential input variables, were used to achieve this goal. The performance of the MLR models developed for three aquifer levels was assessed both quantitatively and qualitatively by using appropriate statistical and graphical indicators. Analysis of the results indicated that the developed TDS models provide accurate prediction of TDS with considerably high values of  $R^2$  and lower values of MAE at a 95% significance level. The fitted MLR models are reasonably good for all aquifer levels, despite their low residuals. The MLR technique has important practical advantages such as the fact that implementation is much easier and less time-

consuming compared to other predictor methods. Consequently, the MLR technique can serve as an alternative and cost-effective tool for groundwater-quality prediction. The methodology presented in this study is very useful in groundwater management and protection and may be easily applied in other regions.

## References

- Ahsan FSM, Nazrul Islam M, Jamal Uddin M, Taj Uddin M (2008) Statistical evaluation of highly arsenic contaminated groundwater in south-western Bangladesh. *J Appl Quant Methods* 3:254–262
- Alcamo J, Döll P, Kaspar F, Siebert S (1997) Global change and global scenarios of water use and availability: an application of water GAP1.0, Center for Environmental Systems Research (CESR), University of Kassel, Germany
- Ameur M, Azaza FH, Gueddari M (2016) Nitrate contamination of Sminja aquifer groundwater in Zaghouan, northeast Tunisia: WQI and GIS assessments. *Desalin Water Treat* 57(50):1–11
- Andre L, Franceschi M, Pouchan P, Atteia O (2005) Using geochemical data and modeling to enhance the understanding of ground water flow in a regional deep aquifer, Aquitaine Basin, south-west of France. *J Hydrol* 305:40–62
- Arslan H (2013) Application of multivariate statistical techniques in the assessment of groundwater quality in seawater intrusion area in Bafla Plain, Turkey. *Environ Monit Assess* 185(3):2439–2452
- Ben Alaya M, Salwa S, Thouraya S, Zargouni F (2014) Suitability assessment of deep groundwater for drinking and irrigation use in the Djeffara aquifers (Northern Gabes, south-eastern Tunisia). *Environ Earth Sci* 71:3387–3421
- Bernstein IH (1988) Applied multivariate analysis. Springer, New York
- Bhat NA, Jeelani G (2015) Delineation of the recharge areas and distinguishing the sources of karst springs in Bringi watershed, Kashmir Himalayas using hydrochemistry. *J Earth Syst Sci* 124(8):1667–1676
- Carreira PM (2010) Groundwater assessment at Santiago Island (Caboerde): a multidisciplinary approach to a recurring source of water supply. *Water Res Manag* 24:1139–1159
- Chenini I, Khemiri S (2009) Evaluation of ground water quality using multiple linear regression and structural equation modeling. *Int J Environ Sci Tech* 6(3):509–519
- Courville T, Thompson B (2001) Use of structure coefficients in published multiple regression articles:  $\beta$  is not enough. *Educ Psychol Meas* 61:229–248
- Dassi L (2004) Etude hydrogéologique, géochimique et isotopique du système aquifère du bassin de Sbeitla (Tunisie Centrale). These, University Sfax, Tunisie, 180 pp
- De Fraiture C, Wichelns D (2009) Satisfying future water demands for agriculture. *J Agric Water Manage* 97:502–511
- Einsiedl F, Mayer B (2006) Hydrodynamic and microbial processes controlling nitrate in a fissured porous karst aquifer of the Franconian Alb, Southern Germany. *Environ Sci Technol* 40:6697–6702
- Elango L, Kannan R, Senthil K (2003) Major ion chemistry and identification of hydrogeochemical processes of groundwater in a part of Kancheepuram District, Tamil Nadu, India. *Environ Geol* 10(4):157–166
- EPA (2009) United States Environmental Protection Agency <http://www.epa.gov/ogwdw000/kids/wsb/pdfs/9124.pdf>. Accessed on 30 December 2009
- Ghasemi J, Saaidpour S (2007) Quantitative structure property relationship study of noctanol water partition coefficients of some of diverse drugs using multiple linear regression. *Anal Chim Acta* 604(2):99–106
- Golfinopoulos SK, Arhonditsis GB (2002) Multiple regression models: a methodology for evaluating trihalomethane concentrations in drinking water from raw water characteristics. *Chemosphere* 47(9):1007–1018
- Gonzalez RA, Noble RT (2014) Comparisons of statistical models to predict fecal indicator bacteria concentrations enumerated by qPCR- and culture-based methods. *Water Res* 48(1):296–305
- González-Ramón A, López-Chicano M, Rubio-Campos JC (2012) Piezometric and hydrogeochemical characterization of groundwater circulation in complex karst aquifers. A case study: the Mancha Real-Pegalajar aquifer (Southern Spain). *Environ Earth Sci* 67(3):923–937
- Green P, Carroll J (1996) Mathematical tools for applied multivariate analysis, Student edn. Academic Press, New York
- Guler C, Thyne GD, McCray JE, Turner AK (2002) Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol J* 10:455–474
- Hamza M (1990) Hydrogéologie du synclinal de Nadhour-Saouaf. Rapport DGRE, Tunisie, p 147
- Hassan M, Shamim MA, Hashmi HN, Ashiq SZ, Ahmed I, Pasha GA, Han D (2014) Predicting streamflows to a multipurpose reservoir using artificial neural networks and regression techniques. *Earth Sci Informatics*. <https://doi.org/10.1007/s12145-014-0161-7>
- Herrig IM, Böer SI, Brennholt N, Manz W (2015) Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. *Water Res* 15(85):148–57. <https://doi.org/10.1016/j.08-006>
- Janssen RPT, Verweij W (2003) Geochemistry of some rare earth elements in groundwater, Vierlingsbeek, The Netherlands. *Water Res* 37:1320–1350
- Jiménez-Martínez J, Candela L, Molinero J, Tamoh K (2010) Groundwater recharge in irrigated semi-arid areas: quantitative hydrological modelling and sensitivity analysis. *Hydrogeol J* 18(8):1811–1824
- Kacem A (2008) Management of water resources of Sisseb Elam basin: impact of hydraulic installations on water table recharge. Thesis of doctoral. 3<sup>rd</sup> cycle, faculty of sciences of Sfax
- Kamel S, Dassi L, Zouari K, Abidi B (2005) Geochemical and isotopic investigation of the aquifer system in the Djerid-Nefzaoua basin, southern Tunisia. *Environ Geol* 49(1):159–170
- Kraiem Z, Chkir N, Zouari K, Parisot JC (2012) Tomographic, hydrochemical and isotopic investigations of the salinization processes in the oasis shallow aquifers, Nefzaoua region, southwestern Tunisia. *J Earth Syst Sci* 121(5):1185–1200
- Kumaresan M, Riyazuddin P (2007) Factor analysis and linear regression model (LRM) of metal speciation and physico-chemical characters of groundwater samples. *Environ Monit Assess* 138(1–3):65–79
- Liao L, Green CT, Bekins B, Böhlke JK (2012) Factors controlling nitrate fluxes in groundwater in agricultural areas. *Water Resour Res* 48(2):1–18
- Liu CW, Jang CH, Chen CP, Lin CN, Lou KL (2008) Characterization of groundwater quality in Kinmen Island using multivariate analysis and geochemical modeling. *Hydrol Process* 22:376–383
- Maier HR, Dandy GC (1998) Understanding the behaviour and optimising the performance of back-propagation neural networks: an empirical study. *Environ Model Softw* 13:179–191
- Makridakis S, Wheelwright SC, Hyndman RJ (2008) Forecasting methods and applications, 3rd edn. Wiley, Singapore
- Manzoor S, Shah MH, Shaheen N, Khalique A, Jaffar M (2006) Multivariate analysis of trace metals in textile effluents in relation to soil and groundwater. *J Hazard Mater* 137(1):31–37

- MED WS & D WG 2007 Mediterranean water scarcity and drought report; Technical report on water scarcity and drought management in the Mediterranean and the Water Framework Directive, <http://www.emwis.net/topics/WaterScarcity>
- Mc Lean W, Jankowski J and Lavitt N (2000) Groundwater quality and sustainability in an alluvial aquifer, Australia In: Groundwater, past achievement and future challenges (eds) Sielilo et al., Balkema, Rotterdam, pp. 567–573
- Memon M, Soomro MS, Akhtar MS, Memon KS (2011) Drinking water quality assessment in Southern Sindh (Pakistan). *Environ Monit Assess* 177(1–4):39–50
- Ozekin K (1994) Modeling bromate formation during ozonation and assessing its control. University of Colorado, Boulder
- Pazand K, Pazand K (2014) Hydrogeochemical investigation using multivariate analytical methods in Esfadan basin, eastern Iran. *Environ Earth Sci* 72(2):483–489
- Qiyani F, Baoping H (2002) Hydrogeochemical simulation of water-rock interaction under water flood recovery in Renqiu Oilfield, Hebei Province, China. *Chin J Geochem* 21:156–162
- Reghunath R, Murthy TRS, Raghavan BR (2002) The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Res* 36:2437–2442
- Shirsath PB, Singh AK (2010) A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. *Water Resour Manag* 24:1571–1581
- SPT: Stat Point Technologies 2009 stat graphics
- Subba Rao N (2008) Factors controlling the salinity in groundwater in parts of Guntur district, Andhra Pradesh, India. *Environ Monit and Assess* 138(1–3):327–341
- Tabari H, Sabziparvar AA, Ahmadi M (2011) Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorol Atmos Phys* 110:135–142
- Thayer JD (2002) Stepwise regression as an exploratory data analysis procedure. Annual meetings of the American Educational Research Association, New Orleans, LA
- Trabelsi R, Zairi M, Ben Dhia H (2007) Groundwater salinization of the Sfax superficial aquifer, Tunisia. *Hydrogéologie J15*:1341–1355
- USEPA, 2010. Predictive tools for beach notification volume I: review and technical protocol [http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/P26-Report-Volume-I-Final\\_508.pdf](http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/P26-Report-Volume-I-Final_508.pdf)
- Varghese SK, Veetil PC, Speelman S, Buysse J, Huylenbroeck GV (2012) Estimating the causal effect of water scarcity on the groundwater use efficiency of rice farming in South India. *J Ecol Econ* 86:55–64
- WHO (1984) Guidelines for drinking water quality. World Health Organization, Geneva
- Williams R (2015) Review of multiple regression. University of Notre Dame <http://www3.nd.edu/~rwilliam>
- Yates MV, Gerba CP, Kelley LM (1985) Virus persistence in groundwater. *Appl Environ Microbiol* 49(4):778–781
- Yu J, Liu J, An W, Wang Y, Zhang J, Wei W, Yongjing W, Junzhi Z, Wei W, Ming S, Yang M (2015) Multiple linear regression model for bromate formation based on the survey data of source waters from geographically different regions across China. *Environ Sci Pollut Res* 22(2):1232–1239
- Zammouri M, Feki H (2005) Managing releases from small upland reservoirs for downstream recharge in semi-arid basins (Northeast of Tunisia). *J Hydrol* 314(1–4):125–138