

# Implementation and evaluation of multivariate analysis for groundwater hydrochemistry assessment in arid environments: a case study of Hajeb Elyoun–Jelma, Central Tunisia

Soumaya Hajji Kolsi · Salem Bouri ·  
Wafik Hachicha · Hamed Ben Dhia

Received: 10 June 2011 / Accepted: 1 March 2013 / Published online: 19 March 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** In most arid zones, groundwater (GW) is the major source of domestic, agriculture, drinking, and industrial water. Accordingly, the monitoring of its quality by different techniques and tools is a vital issue. The purpose of this paper is the evaluation of the combination of principal components analysis (PCA) and geostatistics as a technique for (1) identifying the processes affecting the groundwater chemistry of the detrital unconfined Middle Miocene Aquifer (MMA) of the Hajeb elyoun Jelma (HJ) aquifer (Tunisia) and (2) mapping the controlling variables for groundwater quality. This work is based on a limited database recorded in 22 wells tapping the aquifer and unequally distributed in the field. The proposed approach is carried out in two steps. In the first step, the application of PCA revealed that rock–water interaction, agriculture irrigation and domestic effluents could explain 85 % of the observed variability of the chemical GW quality of the MMA. As a result, two new variables are defined: V1 (rock–water interaction influence) and V2 (irrigation and domestic effluent influence). In the second step, the spatial variability of these variables over the extent of the MMA is mapped by applying a kriging

interpolation technique. The results of this study suggest that, while both natural and anthropogenic processes contribute to the GW quality of the MMA, natural impacts can be considered as the most important.

**Keywords** Geostatistics · Principal component analysis · Groundwater quality · Arid zone · Hajeb El Youn–Jelma · Tunisia

## Introduction

Especially in arid zones, groundwater resources worldwide are considered as major sources for freshwater supply worldwide, especially in arid zones. The weak rainfall as well as the intensive extraction of groundwater from shallow aquifers reduces freshwater budget and creates local water aquifer depression, causing a threat to groundwater budget. Furthermore, the increase of chemical fertilizers applications has been considered as an additional cause of groundwater pollution during the last decades. In arid regions, added to a severe water scarcity, water resources are characterized by a significant spatio-temporal variability (Zekai 2008).

Studies undertaken in arid and semi-arid showed the importance of the water assessment and management in any integrated development strategy. Accordingly, results could serve as an available scientific background for sustainable land use planning and groundwater management in the considered region.

The combination of statistics and geostatistics in a context of groundwater resources management has already been explored by many authors. The combination of principal component analysis and kriging was originally proposed by Espinosa et al. (1993) to characterize

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s12665-013-2377-0) contains supplementary material, which is available to authorized users.

---

S. H. Kolsi (✉) · S. Bouri · H. B. Dhia  
Water-Energy-Environment Laboratory (LR3E),  
National Engineering School of Sfax, University of Sfax,  
B.P. 1173, Sfax 3038, Tunisia  
e-mail: soumaya\_haji@yahoo.fr

W. Hachicha  
Mechanics, Modelling and Manufacturing Research Unit  
(U2MP), National Engineering School of Sfax,  
University of Sfax, B.P. 1173, Sfax 3038, Tunisia

anomalies in soil geochemical composition. The same approach was later used to characterize groundwater quality in a variety of situations (e.g. Sanchez-Martoz et al. 2001; Jiang et al. 2009). Kriging can also be used directly to map groundwater quality (e.g. Uyan and Cay 2010), or in combination with other techniques than PCA, such as cluster analysis (e.g. Yidana et al. 2010). Either both kriging and co-kriging or semi-variance analysis were applied for mapping GW spatio-temporal level fluctuations in arid and semi-arid regions (e.g. Ta'any et al. 2009).

Based on the previous overview, the combined use of PCA (and factor analysis) and kriging in hydrochemical analysis is well documented in the literature. In all these cases, ordinary kriging was used to interpolate factor scores which represented the weight of the respective processes affecting the hydrochemistry in the various basins. In this case study, PCA and geostatistics were combined and evaluated for GW quality mapping across an unconfined aquifer in Hajeb Elyoun–Jelma (HJ) basin under arid climate conditions and using limited database from 22 wells out of 35 wells belonging to the aquifer. In addition, the evaluation was made under an unequal geographic distribution of boreholes throughout the study. The selected wells have complete data sets and chemical analysis balance below 5 %.

The study is based on both the hydrogeochemical evolution in the aquifer and the physicochemical characteristics. Based on this multivariate and complex information, using PCA, it is intended to establish a series of factorial variables that summarize all the hydrogeochemical information. A geostatistical study of these derived variables allows working in a reduced multivariate space, and establishing their spatial distribution throughout the aquifer by the calculation of variograms and cross-validations. Likewise, it is intended to identify the spatial development of the principal processes acting on groundwater quality by mapping of groundwater quality using these factorial variables and ordinary kriging (OK) technique. In fact, OK is the most popular method which is based on the assumption of intrinsic stationarity and ergodicity of the data, and the availability of sufficient data to model spatial autocorrelation (Yidana et al. 2010). In this way, it is aimed in this case study to verify whether these new variables permit location of the zones where various physicochemical processes are superimposed, considering the hydrogeochemical and geological parameters.

## Study area

Located in the northern part of Central Tunisia (Fig. 1), the HJ Basin is characterized by an arid climate with large temperature and rainfall variations. The mean annual

temperature and rainfall are about 19 °C and 230 mm, respectively (Saidi et al. 2009). In addition, it is characterized by high evaporation of about 1,200 mm (Dassi et al. 2005). The precipitation/evaporation ratio (P/E) is about 0.19 classifying study area as an arid region (Maliva and Missimer 2012). The dry climate and low precipitation accentuate the drawdown of water resource and can also affect the water quality by salinization (Maoui et al. 2009). In addition, groundwater quality is highly influenced by an increasing exploitation, hydrogeological conditions, and human impacts (Saidi et al. 2009; Abid et al. 2010).

The HJ basin is part of Tunisian Atlas domain extending over 1,350 km<sup>2</sup>. It shows a complex geological situation characterized by numerous NE–SW and E–W anticlines associated with Triassic salt intrusions (J. Hamra and Kodiet El Halfa). The hydrogeologic basin is featured by a NE–SW synclinal structure developed through the major compressive movement (Atlasic event). The surrounding mountains are made up mainly of Cretaceous outcrops, The Eocene sediments are encountered only in the southern and the north eastern limits of the study area. In the central part of the region, the Quaternary crust (1–2 m), which covers the whole region, constitutes the superficial formation (Burrolet 1956; Zammouri 1988; Zouari and Mamou 1998; Koshel 1980).

Groundwater in the (HJ) synclinal basin is hosted in three main reservoirs namely the Cretaceous, the Middle Miocene, and Mio–Plio Quaternary aquifers. The second one is the most important. Indeed, it provides water for irrigation and drinking water for the whole region and part of the town of Sfax (900,000 inhabitants) located 170 km from these resources. Total groundwater abstraction in 2006 is estimated at 20 Mm<sup>3</sup>.

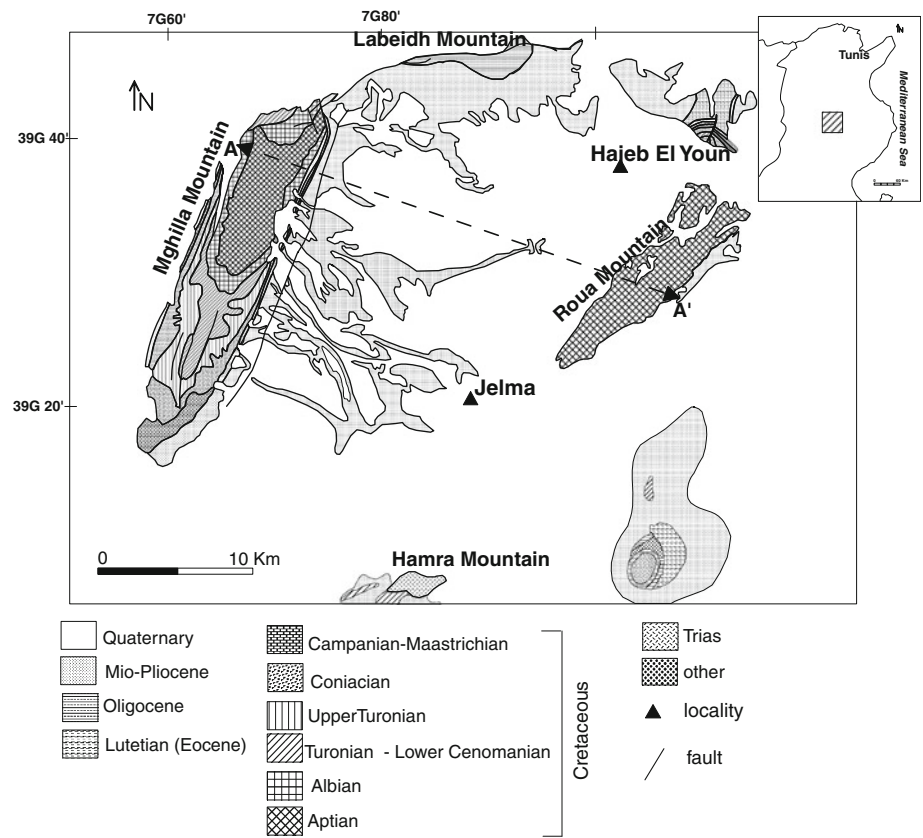
The cross-section AA' (Figure S1 ESM only) shows hydrostratigraphic units consisting of the three aquifer layers. Middle Miocene Aquifer with thickness ranging between 10 and 300 m is made up of coarse to medium grained sandstone. The MMAGW flows from the northwestern and northern highlands (Mghilla and Labeidh Mountains), where the MM outcrops to the southeast (Fig. 2).

## Materials and methods

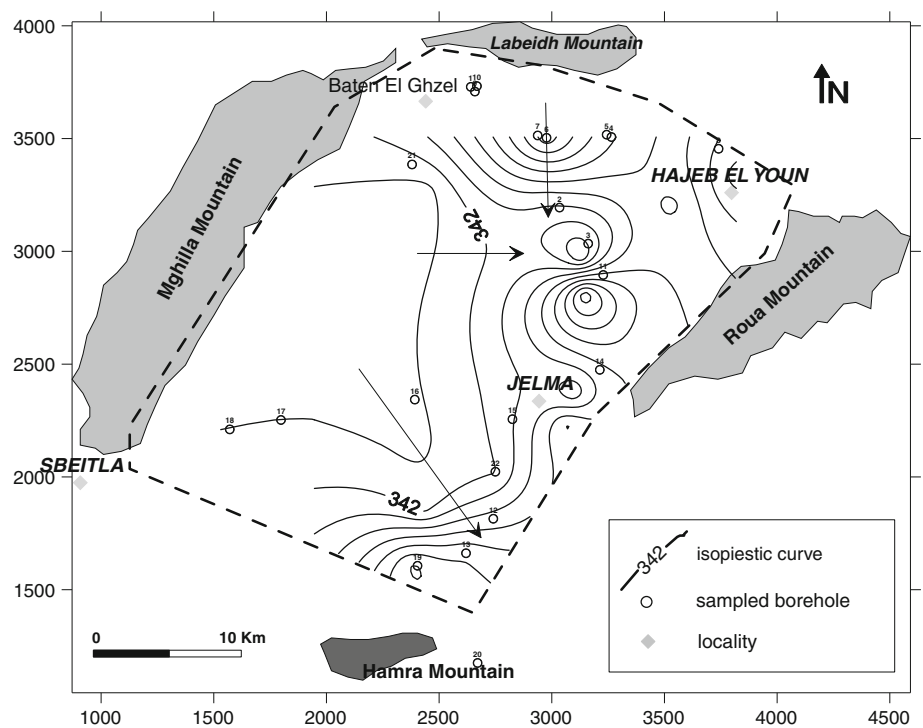
### Hydrochemical data

Hydrochemical data were obtained from a sampling network of 22 wells tapping the MMA. The density of wells is slightly higher in Hajeb and Jelma zones (East and South) than in Baten El Ghzel and Mghilla (Fig. 2). Sampling was carried out during the high-water period (May 2006). Electrical conductivity, temperature and pH were determined directly in situ using a field conductivity meter.

**Fig. 1** Location and geological map



**Fig. 2** Piezometric map (year 2006) and location of sampled boreholes

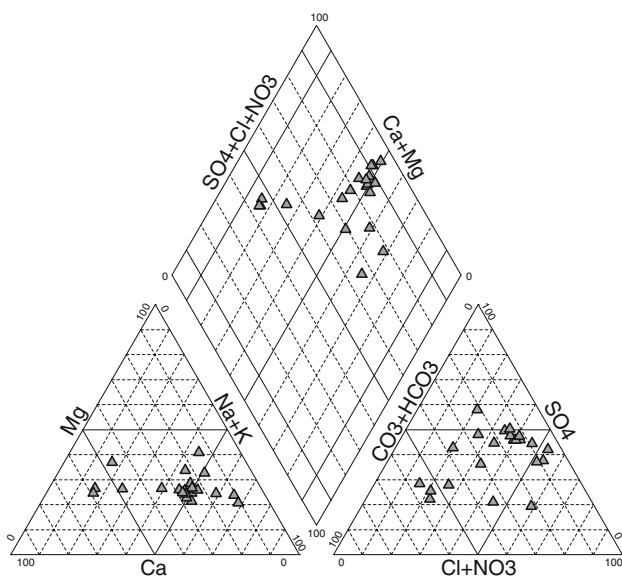


**Table 1** Used analytical methods

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

**Table 2** Descriptive statistics of chemical composition (*N* = 22)

	<i>N</i>	Minimum	Maximum	Mean	Standard deviation
TDS	22	225.00	2,494.00	812.8636	482.58059
EC	22	0.28	2.46	1.3877	0.64099
Ca	22	19.00	192.00	61.6818	38.43781
Mg	22	9.80	130.00	39.5273	24.99425
Na	22	10.40	442.00	134.7682	93.16731
K	22	1.90	25.00	6.3118	4.93593
HCO <sub>3</sub>	22	91.00	185.00	129.2727	25.43373
SO <sub>4</sub>	22	33.50	800.00	252.8091	173.58121
Cl	22	18.00	746.00	179.2273	158.60101
NO <sub>3</sub>	22	0.00	47.50	9.1364	14.70373

**Fig. 3** Piper diagram of MMA groundwater

Major elements were analysed in the LARSEN Laboratory at the National School of Engineers of Sfax. A list of analytical methods used in this study is reported in Table 1.

## Methodology

### Principal component analysis

The main problem solved by PCA is to transform a set of correlated variables into groups of elements, which could be interpreted in an ideal context as independent factors underlying the phenomenon (Espinosa et al. 1993; Wackernagel 1995; Sanchez-Martoz et al. 2001; Hachicha et al. 2008; Maoui et al. 2009; Hamzaoui et al. 2009). It is used to distinguish the contributions of natural processes and anthropogenic impacts to the chemical composition of MMAGW. PCA can provide information on the most meaningful parameters that can describe the whole data sets with minimal loss of original information (Wunderlin et al. 2001). The maths behind PCA was detailed in many references (e.g.: Davis 1986; Sanchez-Martoz et al. 2001; Yidana et al. 2010).

Data were standardized to their corresponding *z* scores (Eq. 1). Data standardization is essential in PCA because in the computation of the Euclidean distances, the parameters with the highest variances tend to have a greater influence over those with lower variances (Güler et al. 2002; Cloutier et al. 2008).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where, *x* is the data,  $\mu$ ,  $\sigma$  are respectively the mean and standard deviation of the datasets.

Moreover, PCA is generated using the rotation method Varimax with Kaiser Normalisation. Indeed, rotation method (Varimax) is the application of an orthogonal matrix to the factor matrix to maximize the differences among the factors which facilitate interpretation of the analysis the results (Yidana et al. 2010). Frequently, this method is applied to increase the participation of the variables with higher contribution and reduce those with lesser contributions. In this way, PCA was subjected to raw Varimax rotation. The so-called factor 1 is related to the largest eigenvalue and is able to explain the greatest amount of variance in the data set. The second factor (orthogonal and uncorrelated with the first one) explains most of the remaining variance, and so forth (Jiang et al. 2009).

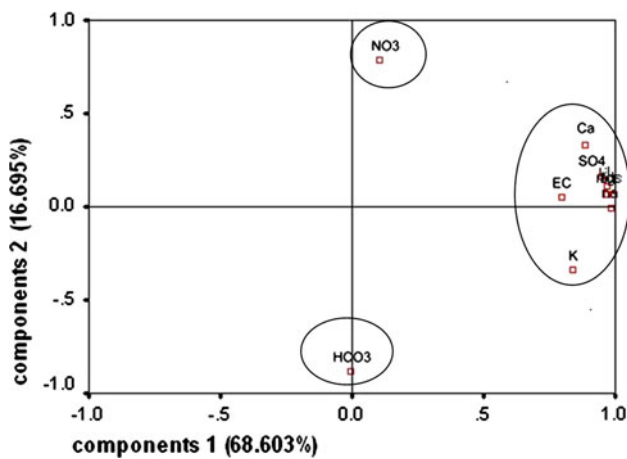
Kaiser's normalisation (Kaiser 1960) is applied in this study. In fact, this criterion is widely used in factor rotation for sizing down the number of factors that can be included in the final factor model. Factors selected are having eigenvalues at least equal to 1 (Ogasawara 1999; Yidana et al. 2010). The PCA is realized with SPSS software. After establishing the associations between the physico-chemical variables of the MMAGW, PCA factors are analysed by geostatistics to interpolation and mapping.

**Table 3** Loadings of principal component (PCA)

Components	Eigenvalue			Sum of factors squares selected			Sum of factors squares selected after rotation		
	Total	Variance explained (%)	Cumulative variance (%)	Total	Variance explained (%)	Cumulative variance (%)	Total	Variance explained (%)	Cumulative variance (%)
1	6.914	69.138	69.138	<b>6.914</b>	<b>69.138</b>	69.138	6.860	68.603	68.603
2	1.616	16.160	85.298	<b>1.616</b>	<b>16.160</b>	85.298	1.669	16.695	<b>85.298</b>
3	0.707	7.067	92.365						
4	0.460	4.600	96.965						
5	0.134	1.336	98.302						
6	$9.077 \times 10^{-02}$	0.908	99.209						
7	$5.739 \times 10^{-02}$	0.574	99.783						
8	$1.931 \times 10^{-02}$	0.193	99.976						
9	$2.211 \times 10^{-03}$	$2.211 \times 10^{-02}$	99.998						
10	$1.716 \times 10^{-04}$	$1.716 \times 10^{-03}$	100.000						

Bold is used to express the most significant Eigenvalues and Variance for PCA

Extraction method: principal component analysis



**Fig. 4** Scatter plot (PCA)

*Geostatistics components*

From variogram analysis through kriging cross-validation to mapping are performed by single integrated program: GS+ software (1988) (Ta’any et al. 2009; Yidana et al. 2010). The theoretical bases of geostatistics are described in several papers and textbooks (Matheron 1970; Goovaerts 1997; Ahmadi and Sedghamiz 2008). In addition, from a hydrochemical point of view there are widely used techniques (Sanchez-Martoz et al. 2001; Cambardella et al. 1994; Harmouzi 2010; Ta’any et al. 2009; Yidana et al. 2010)

The semi-variance analysis is used to measure spatial dependence or autocorrelation, depending on the distribution of the data and to check land isotropy (Yidana et al. 2010; Ahmadi and Sedghamiz 2008). The kriging provides a mean of interpolating values for points not physically

sampled using knowledge about the underlying spatial relationships in a data set. In this study we used the OK, which is considered the most commonly used method (Lefohn et al. 2005; Ta’any et al. 2009; Harmouzi 2010; Yidana et al. 2010). OK is a spatial estimation method where the error variance is minimized. This later is called the kriging variance. It is based on the configuration of the data and on the variogram, hence it is homoecedastic. It is not dependent on the data used to make the estimate. Yamamoto (2005) proved that the ordinary interpolation variance is a better measure of accuracy of the kriging estimate. The cross-validation analysis is a mean for evaluating effective parameters for kriging interpolations. In cross-validation analysis each measured point in a spatial domain is individually removed from the domain and its value estimated via kriging as though it were never there. In this way a graph can be constructed of the estimated vs. actual values for each sample location in the domain.

In others words, the basic process of cross-validation involves: (1) making estimates at all sampling locations one at a time, assuming no data value at the given location (i.e., ignoring its data value), and (2) comparing those estimates to the actual, known data values. In general, the two key goals of cross-validation are to compare the effectiveness and accuracy of two or more spatial estimation methods, and to evaluate the performance of any given spatial dependence model and/or search strategy in a kriging analysis.

The reliability of semi-variogram and goodness of kriging interpolation describing the linear regression equation is evaluated by the regression coefficient from cross-validation approach, which must be near to 1.



**Table 4** Correlation matrix

	TDS	EC	Ca	Mg	Na	K	HCO <sub>3</sub>	SO <sub>4</sub>	Cl	NO <sub>3</sub>
TDS	1.000									
EC	<b>0.774</b>	1.000								
Ca	<b>0.911</b>	0.683	1.000							
Mg	<b>0.974</b>	0.712	0.882	1.000						
Na	<b>0.982</b>	0.783	0.837	0.936	1.000					
K	<b>0.798</b>	0.493	0.606	0.792	0.835	1.000				
HCO <sub>3</sub>	−0.072	0.089	−0.342	−0.104	0.017	0.250	1.000			
SO <sub>4</sub>	<b>0.964</b>	0.762	<b>0.919</b>	0.959	0.913	0.659	−0.180	1.000		
Cl	<b>0.967</b>	0.718	0.877	0.927	<b>0.972</b>	0.835	−0.110	0.878	1.000	
NO <sub>3</sub>	0.143	0.225	0.254	0.112	0.132	−0.112	−0.439	0.132	0.230	1.000

Bold is used to express some significant correlation coefficients for PCA

**Table 5** Components matrix after rotation

	Components	
	1	2
TDS	0.995	$6.770 \times 10^{-2}$
EC	0.795	$5.270 \times 10^{-2}$
Ca	0.888	0.333
Mg	0.969	$6.867 \times 10^{-2}$
Na	0.985	$-1.015 \times 10^{-2}$
K	0.837	−0.336
HCO <sub>3</sub>	$-6.136 \times 10^{-3}$	−0.880
SO <sub>4</sub>	0.945	0.160
Cl	0.966	0.112
NO <sub>3</sub>	0.107	0.788

Extraction method: principal component analysis. Rotation method: Varimax with Kaiser normalisation

Rotation is converged after 3 iterations

**Table 6** V1 and V2 variogram's parameters and validation correlation coefficient (Rc)

Parameters	Symbols	V1	V2
Nugget effect	C <sub>0</sub>	0.001	0.14828
Sill	C <sub>0</sub> + C	0.404	1.36235
Range	A	658,179	124.01
Validation correlation coefficient	Rc	0.75	0.62

## Results and discussion

After hydrochemical parameter analysis, the ion balance between cations and anions is less than 5 % for all samples. According to WHO guidelines (1993), the range of pH value prescribed for drinking purposes is 6.5–8.5. The pH values of groundwater in the study area varied between 7.68 and 8.35, indicating slightly acidic to slightly basic water. These pH values were in the range of accepted values. The exceptional value of 8.82 may be originated from measurements error.

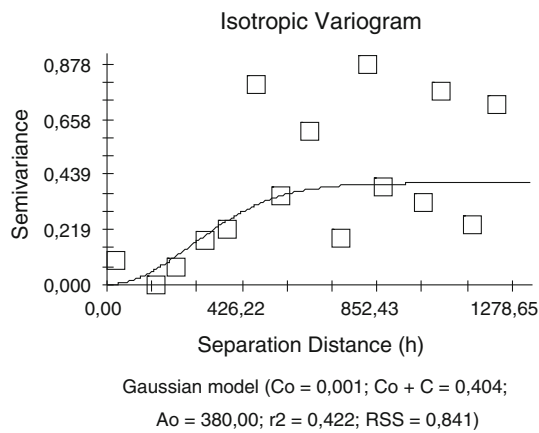
Electric conductivity (EC) shows the concentration of ionized substances in water. According to WHO, the maximum permissible value of EC for drinking water is 1,400  $\mu\text{S}/\text{cm}$ . EC values in MMAGW range between 280 and 2,460  $\mu\text{S}/\text{cm}$ . The lowest values are located at the northern and in the north-eastern parts of the study area.

### Descriptive statistics of chemical composition

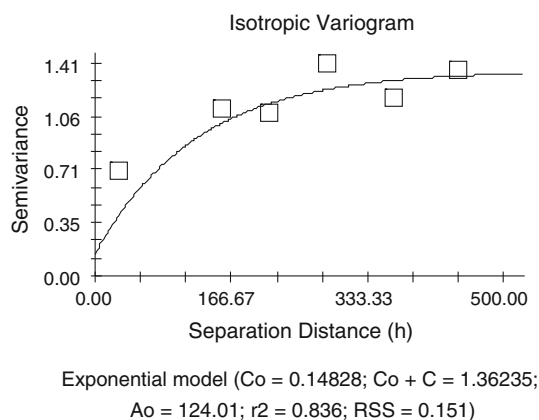
The ten hydrochemical parameters monitored in 22 boreholes are summarized in Table 2. It reflects a moderate to high variability (standard deviation) among samples of the variables. The piper diagram (Fig. 3) and Schoeller–Berkaloff diagram (Figure S2 ESM only) indicate that the water is featured by mixed composition. Indeed, the water of this aquifer has several types of facies, with graduation between Mg–Ca–SO<sub>4</sub>, Na–Mg and Cl–SO<sub>4</sub> and Na–Cl types. The enrichment with bicarbonate is due to leaching of the cretaceous limestones. The enrichment with sulphates can be due to a contribution of the triassic evaporitic rocks (southern part of the basin). To provide a better understanding of mineralisation mechanisms and spatial distribution, combination of PCA-geostatistics approaches was implemented.

### PCA results

Geostatistical methods are optimal when data are normally distributed and stationary (mean and variance do not vary significantly in space). Significant deviations from normality and stationarity can cause problems. The first step of this study is to plotting the histogram of data for checking normality and a posting of the data values in space to check for significant trends. Subsequently, PCA was applied to ten normalized variables data sets, separately (TDS, EC, Ca<sup>2+</sup>, Mg<sup>2+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Cl<sup>−</sup>, SO<sub>4</sub><sup>2−</sup>, HCO<sub>3</sub><sup>−</sup>, and NO<sub>3</sub><sup>−</sup>). According to the Eigenvalue (6.860 and 1.669, respectively), two principals components PC1 and PC2 were



**Fig. 5** Experimental isotropic variogram of V1



**Fig. 6** Experimental isotropic variogram of V2

selected, which explain 68.603 and 16.69 %, respectively, of the total variance, respectively (Table 3). The application of rotation Varimax method led to an increase in the PC1 and reduction in the PC2.

The scatter plot (Fig. 4) and correlation matrix (Table 4) indicates that the most relevant variables defining water quality are related to water dissolved salts ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ) and (EC, TDS), and the less relevant one are  $\text{HCO}_3$  and  $\text{NO}_3$ .

PC2 (16.695 %) of the variance is mainly driven by  $\text{HCO}_3$  and  $\text{NO}_3^-$  with correlation coefficient (loadings  $>0.77$ ). PC1 which explains 68.603 % of the cumulative variance is mainly driven by EC,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$  with correlation coefficient (loadings  $>0.8$ ) (Table 5), which are chemical variables. PC2 may be related to the evolution of the bicarbonates and contamination of water by organic fertilizers and manure and the transfer of pollution by domestic septic tanks. In fact, PC2 has high correlation with  $\text{NO}_3$  ( $r = 0.788$ ). The PC1 may be related to common sources of natural processes of dissolution of geological rocks components as following:

1. The high correlation between Mg–Ca ( $r = 0.882$ ) can be related to dolomitisation phenomenon;
2. The high correlation between Mg–Na and Mg–Cl ( $r = 0.936$ ,  $r = 0.927$ ) may be related to ionic exchange by interaction with clay level. This interrelationship indicates that the water hardness is permanent in nature (Hamzaoui et al. 2009);
3. The high correlation Mg– $\text{SO}_4$  ( $r = 0.959$ ) may be derived by the weathering of a magnesium sulphate mineral.
4. The high correlation between Na–Cl and K–Cl ( $r = 0.972$  and  $r = 0.835$  respectively) may be derived by the simultaneous Halite and Sylvite dissolution.
5. The calcium and magnesium sulphated type is produced by the presence of evaporitic formations in J. Hamra.
6. The calcium and magnesium bicarbonate type due to the scrubbing and dissolution of dolomitic limestone (J. Roua) and sandstone in the alimentation area.

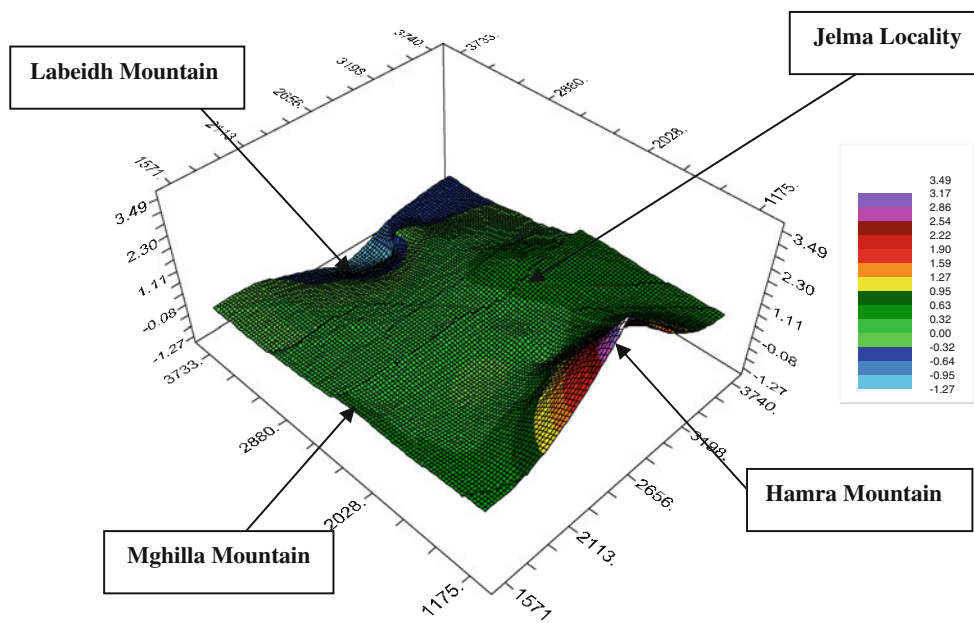
Following precedent, two main processes control the chemical composition of groundwater: (1) dissolution of saline materials; and (2) contamination of water by organic fertilizers and also (3) the transfer of pollution by domestic septic tanks.

Components selected: definitions of the new variables

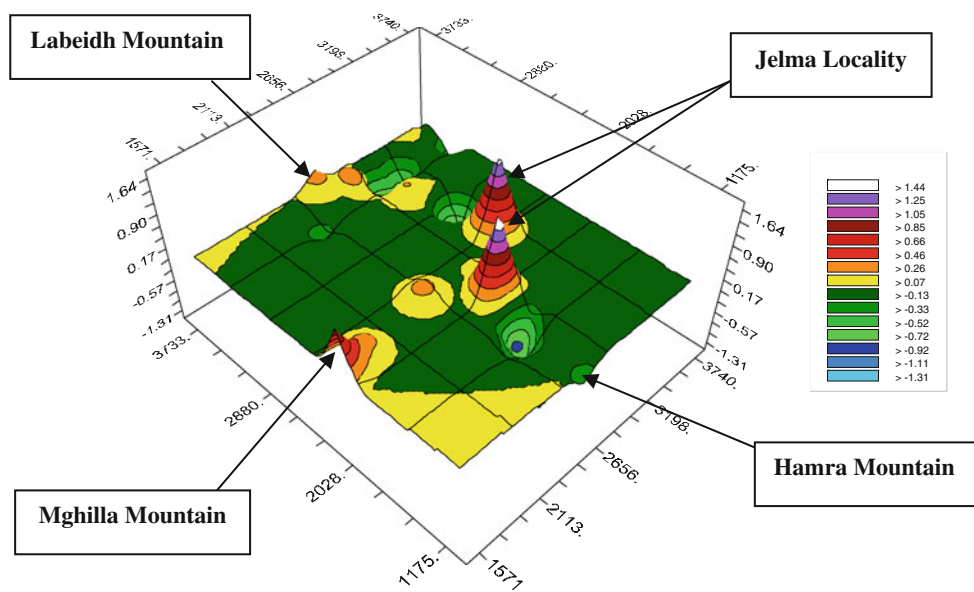
Based on the two components (PC1, PC2), two new variables (V1, V2) were established using the values of principal component scores of the samples, which project the  $n = 22$  observations into the two principal components (Table S1 ESM only). These new variables were used for the geostatistical study, and this enabled an analysis in an orthogonal multivariate space, which is more reduced than the ten original variables.

Mapping groundwater quality

The geostatistical study is performed by GS+ software. The principal aim of this step is to establish the spatial distribution of the two new variables V1 and V2 in the studied zone. For this, the spatial variability of V1 and V2 over the aquifer was defined by calculating their experimental isotropic variograms (Figs. 5, 6). Variograms parameters are showed in Table 6: these parameters are used to classify spatial dependence (sill and range) and to estimate recording errors (nugget effect) (Ta’any et al. 2009; Cambardella et al. 1994). In fact, the range values (A) indicate that the spacing between wells was suitable (Ta’any et al. 2009). The presence of nugget effect ( $\text{Co} = \text{variance at zero distance}$ ) implies inherited variability shorter than the spacing between observation wells.



**Fig. 7** Map of estimated V1 with point kriging method



**Fig. 8** Map of estimated V2 with point kriging method

For further study, the anisotropy of the aquifer was checked using omnidirectional experimental semivariograms. They were constructed in the four main directions E–W, NE–SW, N–S, and NW–SE for V1 and V2. There is a minor degree of anisotropy. Therefore, the detected anisotropies were disregarded, and, accordingly the isotropic models were adopted.

Then, cross-validation test is done. It refers to the reliability of the adopted models and accordingly to the reliability of the kriging estimates. Regression coefficient

obtained after cross-validation must be near to 1 reflecting uncertainty by doing kriging interpolation (Ta'any et al. 2009). Regression coefficients  $Rc1 = 0.95$  and  $Rc2 = 0.62$  are respectively obtained for V1 and V2 (Table 6). The two regression coefficients are above 0.5, besides  $Rc1 > Rc2$ . So variable V1 is more spatially significant than variable V2. Subsequently, mapping of V1 and V2 was performed by means of the point kriging method.

According to V1 map (Fig. 7) corresponding to dissolution of saline materials, the highest values are recorded in



southeast zone of the catchment corresponding to dissolution of triassic evaporitic material in J. Hamra. Mineralisation is also mobilized by groundwater flow from the North and the Northwest to the South of HJ Aquifer.

The region characterised by lower values is corresponding to J. labeidh situated in aquifer upstream.

Mapping of variable V2 (Fig. 8) shows highest values in Jelma and Mghilla. These are characterised by agriculture and also domestic activities. Moreover, high weight of V2 can be due to an intense exploitation from greater depth in these regions.

## Conclusions

The aim of this paper is to evaluate PCA and Geostatistics techniques for improving assessment of detrital MMAGW quality in HJ basin. First, based on PCA with variance of 85 %, two factors have been defined which are associated with (1) natural processes (lithology, rock-water interactions) leading to the enrichment mainly with Na, Cl, Ca and SO<sub>4</sub>; (2) anthropogenic activities (agriculture, industry, urban development): dissolution of the bicarbonates, contamination of water by organic fertilizers and manure and the transfer of pollution by domestic septic tanks. Second, these two components have enabled two new variables to be defined: VI and V2 geostatistics techniques constitute a useful tool for the study of spatial variability of these variables reflecting hydrogeochemical processes in MMAGW of HJ basin—since they enable the distribution of the variable throughout the aquifer to be analysed via its estimation by experimental isotropic variogram, then cross-validation test. Variogram's parameters indicate that the spacing between wells was suitable. Regression coefficients from cross-validation are above 0.5. Subsequent kriging interpolation technique enables us the mapping of variables VI, and V2 associated with the two principal processes implicated. Map1 (for V1) shows that mineralisation in MMAGW is derived by saline materials mainly by triassic evaporites sediments in J. Hamra in the southern zone and also mobilized by groundwater low from the North and the North-West to the South of HJ basin.

The Map2 (for V2) shows that contamination by agriculture and domestic activities is, locally, high in Jelma and Mghilla zone.

Finally, results suggest that both natural and anthropogenic processes contribute to MMAGW quality. Moreover, in terms of spatial distribution, natural impact is the most significant.

It is worth noting that results are satisfactory and demonstrate that combination PCA-Geostatistics can be applied in cases where the aquifer is complex, database set is limited and with unequal spatial distribution information in the studied field.

However, it must be recognized that these methods may not be applied in every situation. A minimum number of sample locations are required for any geostatistical analysis.

**Acknowledgments** The authors warmly thank the anonymous reviewers for their detailed and constructive criticisms, which were of great help in improving this manuscript. The authors also wish to express their thanks to Mr. Elhoucine Essefi from the National Engineering School of Sfax for carefully editing and proofreading English of the present study.

## References

- Abid K, Zouari K, Dulinski M, Chkir N, Abidi B (2010) Hydrologic and geologic factors controlling groundwater geochemistry in the Turonian aquifer (southern Tunisia). *Hydrogeol J* 19(2):415–427
- Ahmadi and Sedghamiz (2008) Application and evaluation of kriging and cokriging methods on groundwater depth mapping. *Environ Monit Assess* 138:357–368
- Burrolet PF (1956) Contribution à l'étude stratigraphique de la Tunisie centrale, *Annales des Mines et de la Géologie* (in French) No. 18, p 350
- Cambardella C, Moorman T, Novak J, Parkin T, Karlen D, Turco R (1994) Field scale variability of soil properties in Central Iowa soils. Vol. 58, No. 5, *Soil Science Society of America*, Madison pp 1501–1511
- Cloutier V, Lefebvre R, Therrien R, Savard M (2008) Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J Hydrol* 353:294–313
- Dassi L, Zouari K, Peter Seiler K, Faye S, Kamel S (2005) Flow exchange between the deep and shallow groundwaters in the Sbeitla synclinal basin (Tunisia): an isotopic approach. *Environ Geol* 47(4):501–511
- Davis JC (1986) *Statistics and data analysis in geology*. Wiley, New York
- Espinosa RJ, Sousa AJ, Chica-Olmo M (1993) Identification of geochemical anomalies using principal components analysis and factorial kriging analysis. *J Geochem Explor* 46:245–256
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford, New York
- Güler C, Thyne GD, McCray JE, Turner AK (2002) Evaluation of the graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology* 10:455–474
- Hachicha W, Masmoudi F, Haddar M (2008) Formation of machine groups and part families in cellular manufacturing systems using a correlation analysis approach. *Int J Adv Manuf Technol* 36(11–12):1157–1169
- Hamzaoui FA, Bouhlila R, Gueddari M (2009) Geochemistry of fluoride and major ion in the groundwater samples of triassic aquifer (South Eastern Tunisia), through multivariate and hydrochemical techniques. *J Appl Sci Res* 5(11):1941–1951
- Harmouzi O (2010) Reconnaissance détaillée de la partie nord-est du Bassin de Saïss (MAROC): interprétation de sondages électriques verticaux par combinaison des méthodes statistique, géostatistique et d'inversion. (in French) Thèse de doctorat. Univ Moulay Ismail-Univ Bordeaux 1
- Jiang Y, Wu Y, Groves Ch, Yuan D, Kambes P (2009) Natural and anthropogenic factors affecting the groundwater quality Nandong karst underground river system in Yunan, China. *J Contam Hydrol* 109:49–61

- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
- Koshel R (1980) Etude hydrogéologique de la nappe de Hajeb El Aïoun–Djelma–Ouled Asker. Edition spéciale de la Direction des Ressources en Eau et en Sol, Division des Ressources en Eau. Projet N° 6520/7 (in French) p 245
- Lefohn AS, Knudsen HP, Shadwick DS (2005) Using Ordinary Kriging to Estimate the Seasonal W126, and N100 24-h Concentrations for the Year 2000 and 2003. A.S.L. and Associates, 111 North Last Chance Gulch Suite 4A Helena, Montana 59601
- Maliva R, Missmer T (2012) Arid lands water evaluation and management. *Environmental Science and Engineering*. Springer-Verlag Berlin, Heidelberg. doi:10.1007/978-3-642-29104-3\_2
- Maoui A, Kherici N, Derradji F (2009) Hydrochemistry of an Albian sandstone aquifer in a semi arid region, Ain oussera, Algeria. *Environ Geol* 60(4):689–701
- Matheron (1970) La théorie des variables régionalisées et ses applications. les cahiers du centre de morphologie mathématique de Fontainebleau, fascicule 5, Ecole Nationale Supérieure des mines Paris, pp 211 (English version is available from the same publisher since 1971)
- Ogasawara H (1999) Standard errors for the direct oblique solution with Kaiser's normalization. *Shinrigaku Kenkyu* 70(4):333–338
- Saidi S, Bouri S, Ben Dhia H (2009) Groundwater vulnerability and risk mapping of the Hajeb-Jelma aquifer (Central Tunisia) using a GIS-based DRASTIC model. *Environ Earth Sci* 59(7):1579–1588
- Sanchez-Martoz F, Jiménez ER, Pulido BA (2001) Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, southeastern Spain. *Hydrol Sci J* 46(2): 227–242
- Ta'any RA, Tahboub AB, Saffarini GA (2009) Geostatistical analysis of spatiotemporal variability of groundwater level fluctuations in Amman–Zarqa basin, Jordan: a case study. *Environ Geol* 57(3):525–535
- Uyan M, Cay T (2010) Geostatistical methods for mapping groundwater nitrate concentrations. 3rd International conference on cartography and GIS 15–20 June, 2010, Nessebar, Bulgaria
- Wackernagel H (1995) *Multivariate geostatistics: an introduction with applications*. Springer, Berlin
- World Health Organization (WHO) (1993) *Guidelines for drinking water quality, recommendations*, vol. 1, Geneva, p 1308
- Wunderlin DA, Diaz MP, Ame MV, Pesce SF, Hued AC, Bistoni MA (2001) Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba-Argentina). *Water Res* 35:2881–2894
- Yamamoto JK (2005) Comparing ordinary kriging interpolation variance and indicator kriging conditional variance for assessing uncertainties at unsampled locations. In: Dessureault S, Ganguli R, Kecojevic V, Dwyer JG (eds) *Application of computers and operations research in the mineral industry*, Balkema
- Yidana SM, Banoeng-Yakubo B, Akabzaa TM (2010) Analysis of groundwater quality using multivariate and spatial analyses in the Keta basin, Ghana. *J Afr Earth Sci* 58(2):220–234
- Zammouri M (1988) *Modèle de simulation des nappes de Hajeb El Youn-Jelma-Ouled Asker* Report (in French). DGRE Tunis
- Zekai C (2008) Wadi Hydrology, Chapter 6. In: *Groundwater management methods for arid regions* (ed.) Taylor and Francis, London, pp 243–301
- Zouari K, Mamou A (1998) *Etude isotopique et hydrochimique des systèmes multicouches de Gafsa et de Hajeb El Youn*. (in French) Rapport final de Projet (AIEA)