23. McCloskey, D., Palsson, B. O. and Feist, A. M., Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol*., 2013, **9**, 661; doi:10.1038/msb.2013.18.
24. Varma, A., Boesch, B. W. and Palsson, B. O., Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.*, 1993, **59**, 2465–2473.
25. Fischer, E., Zamboni, N. and Sauer, U., High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived ^{13}C constraints. *Anal. Biochem*., 2004, **325**, 308–316; doi:http://dx.doi.org/10.1016/j.ab.2003.10.036.
26. Le Novere, N. *et al*., BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*., 2006, **34**, D689–D691; doi:10.1093/nar/gkj092.
27. Mattam, A. J., Clomburg, J. M., Gonzalez, R. and Yazdani, S. S., Fermentation of glycerol and production of valuable chemical and biofuel molecules. *Biotechnol. Lett*., 2013, **35**, 831–842; doi:10.1007/s10529-013-1240-4.
28. Jin, P., Li, S., Lu, S. G., Zhu, J. G. and Huang, H., Improved 1,3-propanediol production with hemicellulosic hydrolysates (corn straw) as cosubstrate: impact of degradation products on *Klebsiella pneumoniae* growth and 1,3-propanediol fermentation. *Bioresour. Technol*., 2011, **102**, 1815–1821; doi:10.1016/j.biortech.2010.09.048.
29. Lin, H., Bennett, G. N. and San, K. Y., Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. *Metab. Eng*., 2005, **7**, 116–127; doi:10.1016/j.ymben.2004.10.003.
30. Vemuri, G. N., Eiteman, M. A. and Altman, E., Effects of growth mode and pyruvate carboxylase on succinic acid production by metabolically engineered strains of *Escherichia coli*. *Appl. Environ. Microbiol*., 2002, **68**, 1715–1727; doi:10.1128/aem.68.4.1715-1727.2002.
31. Cheng, K.-K. *et al*., Global metabolic network reorganization by adaptive mutations allows fast growth of *Escherichia coli* on glycerol. *Nature Commun*., 2014, **5**; doi:10.1038/ncomms4233.
32. Hildebrand, A., Schlacta, T., Warmack, R., Kasuga, T. and Fan, Z., Engineering Escherichia coli for improved ethanol production from gluconate. *J. Biotechnol*., 2013, **168**, 101–106; doi:10.1016/j.jbiotec.2013.07.033.
33. Yazdani, S. S. and Gonzalez, R., Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Curr. Opin. Biotechnol*., 2007, **18**, 213–219; doi:10.1016/j.copbio.2007.05.002.
34. Dharmadi, Y., Murarka, A. and Gonzalez, R., Anaerobic fermentation of glycerol by Escherichia coli: a new platform for metabolic engineering. *Biotechnol. Bioeng*., 2005, **94**, 821–829; doi:10.1002/bit.21025.
35. Stols, L. and Donnelly, M. I., Production of succinic acid through overexpression of NAD+-dependent malic enzyme in an *Escherichia coli* mutant. *Appl. Environ. Microbiol*., 1997, **63**, 2695–2701.
36. Hong, S. H. and Lee, S. Y., Metabolic flux analysis for succinic acid production by recombinant *Escherichia coli* with amplified malic enzyme activity. *Biotechnol. Bioeng*., 2001, **74**, 89–95.
37. Shams Yazdani, S. and Gonzalez, R., Engineering *Escherichia coli* for the efficient conversion of glycerol to ethanol and co-products. *Metab. Eng*., 2008, **10**, 340–351; doi:10.1016/j.ymben.2008.08.005.
38. Wang, Q., Chen, X., Yang, Y. and Zhao, X., Genome-scale *in silico* aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. *Appl. Microbiol. Biotechnol*., 2006, **73**, 887–894; doi:10.1007/s00253-006-0535-y.

Received 15 May 2014; revised accepted 10 December 2014

The combination of principal component analysis and geostatistics as a technique in assessment of groundwater hydrochemistry in arid environment

Yousef Nazzal[1,]*, Faisal K. Zaidi[2], Izrar Ahmed[3], Habes Ghrefat[2], Muhammad Naeem[2], Nassir S. N. Al-Arifi[2], Saeed A. Al-Shaltoni[2] and Khaled M. Al-Kahtany[2]

[1]Department of Mathematics and Applied Sciences, College of Arts & Sciences, Abu Dhabi University, PO Box 59911, Abu Dhabi, UAE
[2]Department of Geology and Geophysics, King Saud University, PO Box 2455, Riyadh 11451, Saudi Arabia
[3]Colleges of Engineering, King Saud University, Riyadh 11451, Saudi Arabia

Central Saudi Arabia is one of the most arid regions of the world with very little precipitation and extreme climatic conditions. In the absence of available surface water supplies, the non-renewable groundwater resources stored in the Palaeozoic and Mesozoic sedimentary formations form the most important source for irrigation and domestic water requirements. The present study deals with 97 groundwater samples collected from Saq aquifer, which is the major aquifer in the region. The study involves the use of principal component analysis (PCA) and variogram analysis for groundwater quality mapping. PCA helped in establishing a series of factorial variables that summarize all the hydrochemical information. Efforts have been made to identify the spatial development of the principal process acting on groundwater quality by mapping it using factorial variables and ordinary kriging techniques. Two principal components (PCs) were extracted revealing that the chemical characteristics of groundwater in the region were acquired through rock–water interactions and anthropogenic influences. Finally, by applying kriging interpolation technique on the factor distribution values for the two PCs in the area under investigation, the factor distribution maps were prepared. The results concluded that both natural and anthropogenic processes contribute to the groundwater quality, but anthropogenic impacts are more important and may result in further deterioration of groundwater quality if relevant protection methodologies are not adopted.

Keywords: Arid region, geostatistics, groundwater quality, kriging, principal component analysis.

Groundwater resources worldwide are considered as precious sources for meeting the agricultural, domestic and industrial demands. This is especially true for arid

*For correspondence. (e-mail: yousef.nazal@gmail.com)

regions where rainfall is scanty and surface water supplies are practically negligible. At the same time the excessive extraction of groundwater from shallow aquifers and their minimal recharge results in overall groundwater depletion and a negative water budget. Furthermore, the increase of chemical fertilizers for improving agricultural yields has led to groundwater pollution during the last decade[1]. In the arid regions, other than the severe scarcity, water resources are also characterized by a significant spatio-temporal variability[2].

Geostatistical methods were used for mathematical modelling of spatial correlation structures with a variogram as the quantitative measure of this spatial correlation. The variogram is commonly used in geostatistics and the interpolation technique known as kriging provides the best unbiased linear estimate of a regionalized variable in an unsampled location, where best is defined in a least squares sense. The emphasis is on local accuracy, i.e. closeness of the estimate to the actual but unknown value without any regard for the global statistical properties of the estimates. The kriging estimation variances are independent of the value being estimated and are related only to the spatial arrangement of the sample data and to the model variogram[3].

Studies undertaken in arid and semi-arid regions showed the importance of the groundwater assessment and management in any integrated development strategy. Accordingly, results could serve as an available scientific background requirement in the considered regions[1]. Many authors have used statistics and geostatistics to study groundwater resource management to obtain good results. The combination of principal component analysis (PCA) and kriging was originally proposed by Espinosa *et al.*[4] to characterize anomalies in soil geochemical composition. The same approach was later used by many authors[5,6] to characterize groundwater quality in a variety of situations. To map groundwater quality, kriging can also be used in combination with other techniques than PCA, such as cluster analysis[7]. Kriging, co-kriging or semi-variance analysis have been applied for mapping spatio-temporal fluctuations in groundwater levels in arid and semi-arid regions[8].

The present study is based on hydrochemical evaluation using multivariate and complex information of factorial variables that summarize all the hydrochemical information. The integrated use of PCA and geostatistics helps in spatial evaluation of groundwater quality mapping. It is also intended to identify the spatial development of the principal process acting on groundwater quality.

The study area is located between lat. 25°N and 26.5°N, and long. 43.25°E and 46.25°E and forms a part of NW Riyadh and Qassim provinces of Saudi Arabia (Figure 1). The study area represents a typical arid region with very low average annual rainfall (<150 mm), which mostly occurs between November and March. The rainfall is torrential and may cause small run-off to wadi channels and low-lying areas. The average annual evaporation is about 3000 mm. The region is characterized by a high diurnal range of temperatures, which averages from 43°C to 28°C during summer and 21°C to 9°C during winter. Temperatures falling up to 0°C are common in the area during winter. The study area hosts significant agricultural farming with groundwater serving as the major source of irrigation. Over the past three decades cultivation has developed significantly in the area.

A total of 103 groundwater samples were collected from Saq aquifer, which is the major aquifer in the region, from the different agricultural farms lying in the study area (Figure 2); however, only 97 samples were used for the interpretation. Samples were collected in polyethylene bottles of 1 litre capacity. Prior to their filling with sampled water, these bottles were rinsed to minimize the chance of any contamination. The sample preservation and the used analytical techniques were in accordance with the standard methods provided by the American Public Health Association[9]. Unstable parameters such as hydrogen ion concentration (pH), total dissolved solids (TDS) and electrical conductivity (EC) were determined at the sampling sites with the help of a pH-meter, a portable EC-meter and a TDS-meter (Hanna Instruments, Michigan, USA).

The sodium (Na^+), potassium (K^+), magnesium (Mg^{2+}), and calcium (Ca^{2+}) ions were determined by atomic absorption spectrophotometer (AAS). Bicarbonate (HCO_3^-) and chloride (Cl^-) were analysed by volumetric methods. Sulphate (SO_4^{2-}) was estimated by the colorimetric and turbid metric methods. Nitrate (NO_3^-) was measured by ionic chromatography.

The statistical analysis used in the present study comprises of PCA. This is basically a variation reduction procedure, wherein a number of observed/measures parameters can be transformed into a small number of artificial variables known as principal components (PCs). The extracted PCs account for most of the variance in the observed parameters and can be interpreted as an independent factor governing a given phenomenon[10].

PC1 accounts for the greatest variability[3] which can be seen on the scree plot. The factor loading or the PC score associated with each of the variables in a given PCs are the correlation between the original variable and the factor, and gives an idea about the processes which control the data variability[11]. A factor loading close to ±1 indicates a strong correlation between the given variable and the factor. The variables which show loadings greater than 0.5 are generally considered to be significant. The detailed mathematics behind PCA is available in numerous published workss[5,7,12]. The statistical software used in the present study was the SPSS 17 software package.

In the present study the data are standardized to their corresponding Z scores (eq. (1)). Data standardization is essential in PCA because in the computation of the

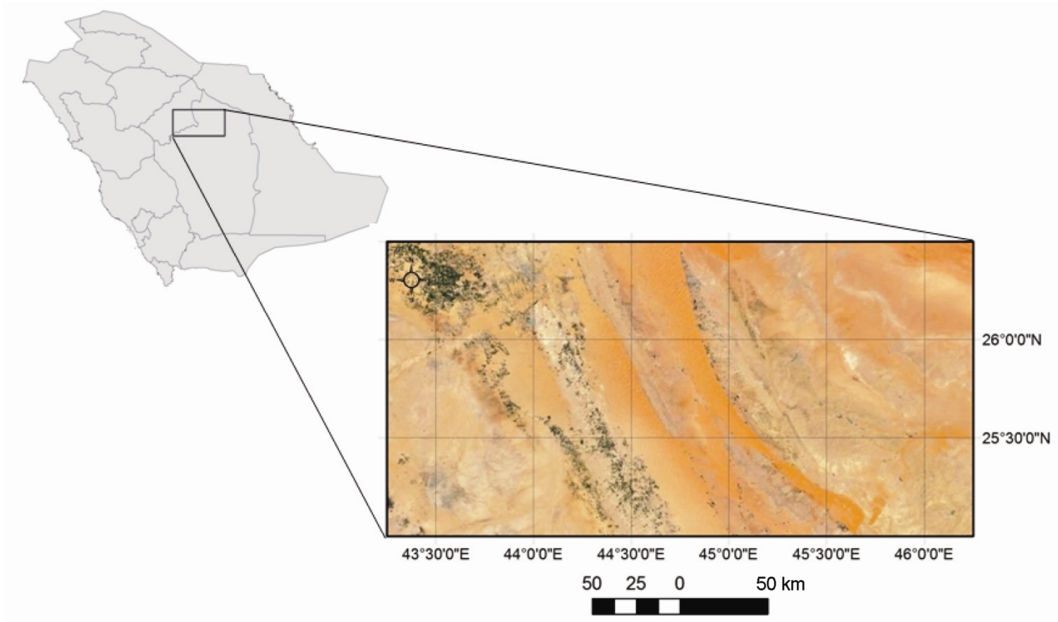

Figure 1. Location of the study area.

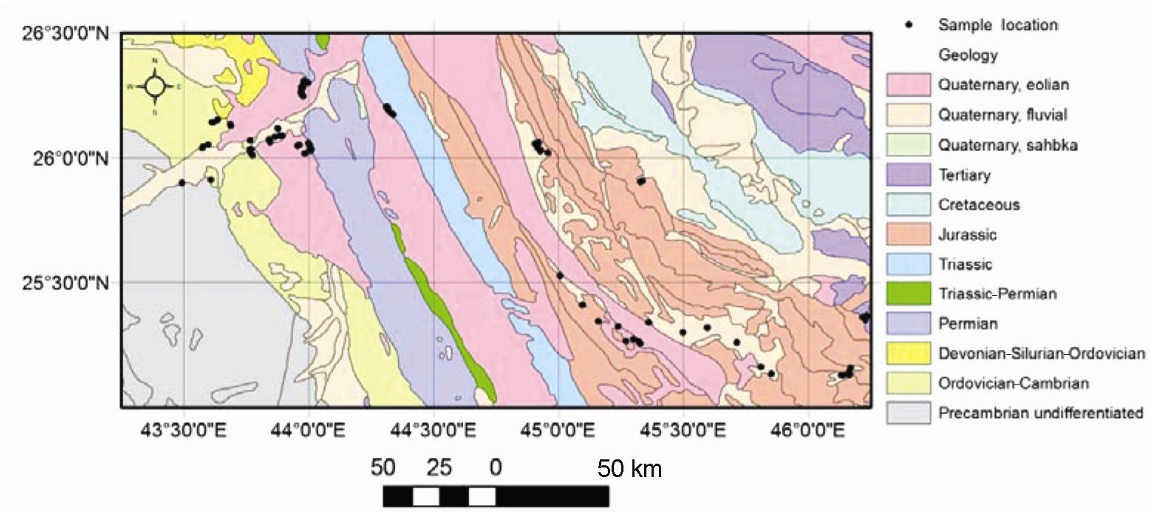

Figure 2. Geology of the study area and location of collected groundwater samples.

Euclidean distances, the parameters with the highest variance tend to have greater influence over those with lower variance[1,13,14]

$$Z = (X - \mu) \div \sigma, \quad (1)$$

where X are the data, while μ and σ are respectively, the mean and standard deviation of the datasets. In the present study Kaiser's normalization[15] is applied. 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 have eigenvalues >1 (refs 7, 16). Varimax rotation is generally applied to all the extracted PCs to reduce the contribution of the variables which are not significant[17].

PCA was done with SPSS software. PCA factors were analysed by geostatistical methods of interpolation and mapping. The single integrated program GS + Software (1988) was used to carry out the variogram analysis, kriging, cross validation and mapping[1,7,8]. The theoretical basis of geostatistics has been described in the literature[18,19]. In addition, from a hypothetical point of view there are widely used techniques that have appeared in different studies[5,7,8,20,21].

The collected samples were analysed for physico-chemical and major ions. Out of the total 103 samples, only 97 were used for interpretation as mentioned earlier. The remaining six samples were removed on account of high ion balance error (>5%). According to WHO guidelines[22], the pH for drinking water ranges from 6.5 to 8.5. In the analysed samples, the pH value ranged from 6.68 (slightly acidic) to 8.00 (slightly basic), with an average of 7.15. The average EC value in the analysed groundwater samples was 5843 μS/cm and fell within the very high salinity hazard range. According to WHO, the maximum permissible value of EC for drinking water is 1400 μS/cm. The TDS content of the samples ranges from 352 to 8812.0 mg/l with an average of 2863.49 mg/l, which is much beyond the maximum permissible limit of 500 mg/l. Descriptive statistics of chemical composition of the ten hydrochemical parameters monitored in 97 boreholes is summarized in Table 1. The major ions were plotted on the piper diagram to understand the groundwater classification and main groundwater facies present in the region (Figure 3).

Table 1. Descriptive statistics of chemical composition (N = 97)

	Minimum	Maximum	Mean	Standard deviation
EC (μS/cm)	716.00	17980.00	5843.85	4127.45
TDS (mg/l)	352.00	8812.00	2863.75	2023.08
T alkal (mg/l)	40.00	305.00	208.24	56.947
HCO_3 (mg/l)	49.00	372.00	255.52	68.89
Cl (mg/l)	192.00	4652.00	1318.78	887.33
NO_3 (mg/l)	4.00	49.00	30.844	11.29
SO_4 (mg/l)	96.00	7968.00	3238.78	2378.23
Ca (mg/l)	132.20	3152.00	1235.83	771.36
K (mg/l)	9.00	544.00	142.27	138.95
Mg (mg/l)	12.00	352.00	96.67	80.87
Na (mg/l)	19.00	2370.00	821.99	665.66

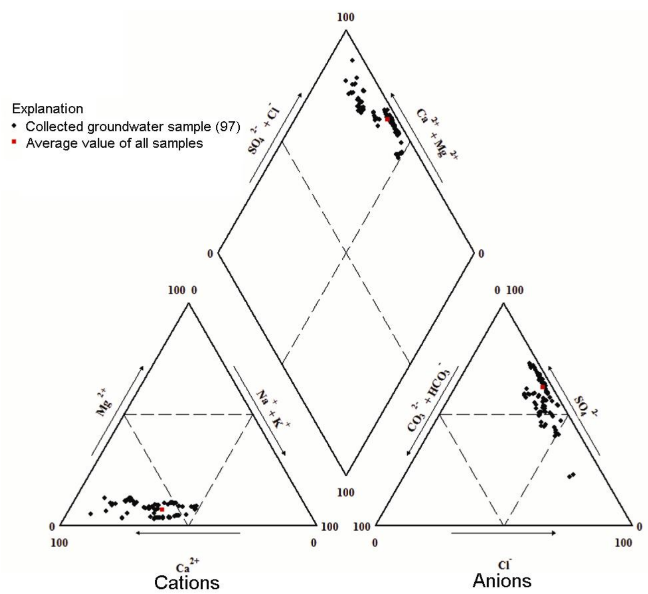

Figure 3. Piper plot of the analysed groundwater samples.

The cationic triangle is mainly dominated by calcium (Ca) with a few samples falling in the 'no dominant type of cations'. On the other hand, the anions fall within the segment of the triangle dominated by sulphate (SO_4). A few samples fall in the segment not dominated by any of the anionic species and two samples fall within the ionic species dominated by chloride (Cl). The water in the study area can be classified as $Ca–SO_4^{2-}$ type. All collected samples fall in the zone of permanent hardness on the piper plot.

The rate of evaporation, rock composition and chemical composition of rainwater control the overall chemistry of the groundwater in a given area[23]. The log of TDS versus $Na^+/Na^+ + Ca^{2+}$ and $Cl^-/Cl^- + HCO_3^-$ of the analysed samples from the study area was plotted on the Gibbs diagram. Though the Gibbs plots (Figure 4) indicate that evaporation is the major dominating factor controlling the water chemistry of the region, but rock–water interaction also plays a major role which was highlighted in PCA.

PCA is one of the frequently used procedures for the multivariate statistical analysis of groundwater quality data, which helps in inferring the natural or anthropogenic processes controlling the groundwater chemistry of a given area[6,10,24–27]. Geostatistical methods are optimal when data are normally distributed and stationary (mean and variance do not vary significantly in space)[1]. Significant deviations from normality can cause problems. The study was initiated with normality check through histogram plot and posting of the data values in space to check for significant trends. Subsequently, PCA was applied to ten normalized variable sets, including TDS, EC, Mg^{2+}, Na^+, K^+, HCO_3^-, SO_4^{2-}, Cl^-, NO_3^- and Ca^{2+}. Based on the eigenvalues of 8.14 and 1.32 respectively, two principal components PC1 and PC2 were selected, which explained 74.04% and 12.01% of the total variance respectively (Table 2). The application of rotation matrix method led to an increase in PC1 and reduction in PC2 (Table 3). The scatter plot (Figure 5) and correlation matrix (Table 4) indicate that the wide range of the variables defining the groundwater quality are related to the dissolved salts. PC1 with 74.04% variance shows positive loading of elements like Ca^{2+}, Mg^{2+}, Na^+, K^+, SO_4^{2-} and Cl^-, thus covering almost the entire range of groundwater evolution processes. The main process involved here is the water–rock interaction, high aridity and salinity due to long resident time, etc.[28,29]. PC2 (12.01% of the total variance) is mainly driven by NO_3^-, and HCO_3 with factor loading of 0.68 and 0.79. With this unique characteristic, PC2 may be related to anthropogenic activities with geochemical reactions at shallow groundwater levels.

The high factor loadings for the variables in PC1 can be attributed to the natural processes of dissolution of geological rocks components as explained below:

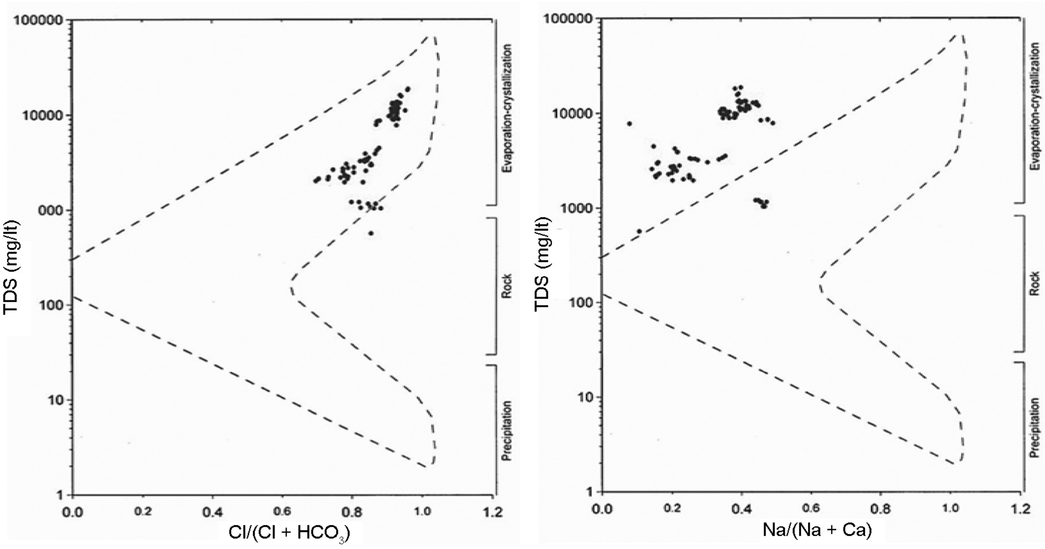

Figure 4. Gibbs plot showing the dominant factor controlling groundwater chemistry.

Table 2. Loading of principle component (PCA)

	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings[a]
Component	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)	Total
1	8.11	74.04	74.04	8.14	74.04	74.04	7.07
2	1.32	12.01	86.09	1.32	12.01	86.06	2.39
3	0.73	3.83	92.17				
4	0.42	1.75	96.03				
5	0.19	0.86	98.36				
6	0.09	0.53	99.22				
7	0.06	0.17	99.62				
8	0.02	0.09	99.94				
9	0.01	0.04	99.98				
10	0.004	0.02	100.00				

Extraction method: PCA.
[a]When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 3. Rotation matrix[a]

	Component	
	1	2
EC	0.949	0.180
TDS	0.949	0.180
T alkal	0.591	0.686
HCO_3	0.610	0.687
Cl	0.904	0.273
NO_3	–0.140	0.791
SO_4	0.946	0.240
Ca	0.910	0.331
K	0.817	–0.122
Mg	0.890	0.043
Na	0.922	0.230

Extraction method: PCA.
Rotation method: Varimax with Kaiser normalization.
[a]Rotation converged in three iterations.

- The high correlation between Mg and Ca ($r = 0.87$) can be related to silicate weathering and dolomitization phenomenon.
- The high correlation between Mg and Cl ($r = 0.86$) may be related to reverse ion exchange taking place in the area.
- The high correlation between Mg and SO_4 ($r = 0.90$) may have its source in weathering of $MgSO_4$ mineral[1], gypsum dissolution and evaporites.
- The high correlation between Na and Cl ($r = 0.91$) as well as Ca and Cl ($r = 0.98$) may be derived by the simultaneous halite or silvite dissolution.

Based on two components (PC1 and PC2) two new variables, i.e. V1 and V2 were established using the values of principal component scores of the samples included in the present study, which project the 97 observations into the two PCs. The geostatistical study was performed using GS+ software. The spatial distribution of

Table 4. Correlation matrix

	EC	TDS	T alkal	HCO_3	Cl	NO_3	SO_4	Ca	K	Mg	Na
EC	1.000										
TDS	1.000	1.000									
T alkal	0.643	0.643	1.000								
HCO_3	0.664	0.664	0.976	1.000							
Cl	0.898	0.898	0.696	0.718	1.000						
NO_3	0.031	0.031	0.253	0.260	0.063	1.000					
SO_4	0.948	0.948	0.672	0.696	0.907	0.107	1.000				
Ca	0.912	0.912	0.738	0.763	0.974	0.114	0.957	1.000			
K	0.677	0.677	0.405	0.406	0.626	–0.049	0.728	0.643	1.000		
Mg	0.801	0.801	0.569	0.581	0.767	0.004	0.821	0.762	0.877	1.000	
Na	0.960	0.960	0.630	0.653	0.907	0.088	0.972	0.939	0.631	0.738	1.000

Table 5. V1 and V2 variogram parameters and validation correlation coefficient

Parameter	Variogram parameter and correlation	
	PCA factor-1	PCA factor-2
Variogram model	Spherical	Gaussian
Nugget effect	0.05	0.1
Sill	0.6	0.45
Range	8000	5000

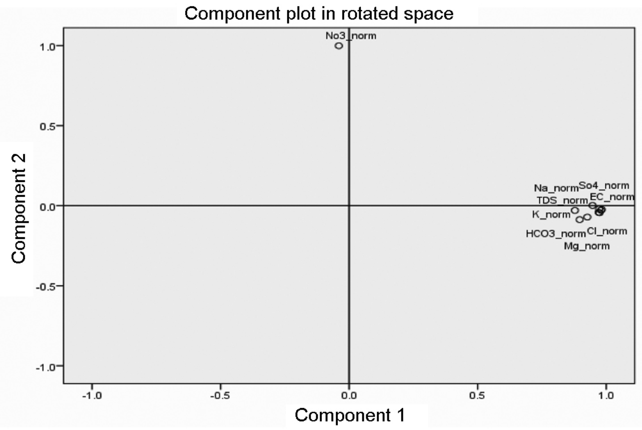

Figure 5. Scatter plot of PC1 and PC2.

the two variables V1 and V2 over the aquifer by calculating their experimental isotropic variograms is shown in Figure 6.

Variogram parameters are shown in Table 5. The parameters sill and range were used to classify spatial dependence; however, the nugget effect shows recording errors[1,8,20].

The present study shows that the range value (A) of spacing between 97 wells was suitable (V1 12,000, V2 1,000). The presence of nugget effect (Co = variance of zero distance) implies inherited variability shorter than the spacing between observation wells[1]. The anisotropy of the aquifer was also checked. For this, unidirectional experimental semi-variogram were used. Further, the anisotropy of the aquifer was constructed in the four main directions, i.e. E–W, NE–SW, N–S and NW–SE for both V1 and V2.

Cross validation test which helps in checking the reliability of the adopted models and reliability of kriging estimates was performed. After cross validation which must be near 1, regression coefficients were obtained[8]. Regression coefficients RC1 = 0.82 and RC2 = 0.76 respectively, were obtained for V1 and V2. The two regression coefficients were above 0.5. This shows that V1 is spatially more significant than V2. Furthermore, mapping of V1 and V2 was done using the point kriging method (Figure 7). V1 plot corresponds to the dissolution of saline materials with high values recorded in most of the selected or analysed samples.

The areas corresponding to high values in V1 and V2 maps (Figure 7) are characterized by extensive agricultural activities and human settlements. Moreover, high values of V2 may be due to an intense exploitation from both shallow and greater depth in these regions. Moreover rock–water interaction as well as evaporation have played an important role in controlling the groundwater chemistry.

The present study is based on both the hydrochemical evaluation in the aquifer and the physio-chemical characteristics. Based on this multivariate and complex information, using PCA, the present study aims to establish a series of factorial variables that summarize all the hydrochemical information. The study also intends to identify the spatial development of the principle process acting on groundwater quality by mapping it using these factorial variables and ordinary kriging techniques.

Based on PCA, the study came out with two important variables – V1 showing the influence of rock–water interactions and V2 showing anthropogenic influences. By applying kriging interpolation technique, the spatial variability of these variables over the extent of the study area was mapped. The study results concluded that both natural and anthropogenic processes contribute to the groundwater quality, but anthropogenic impacts can be

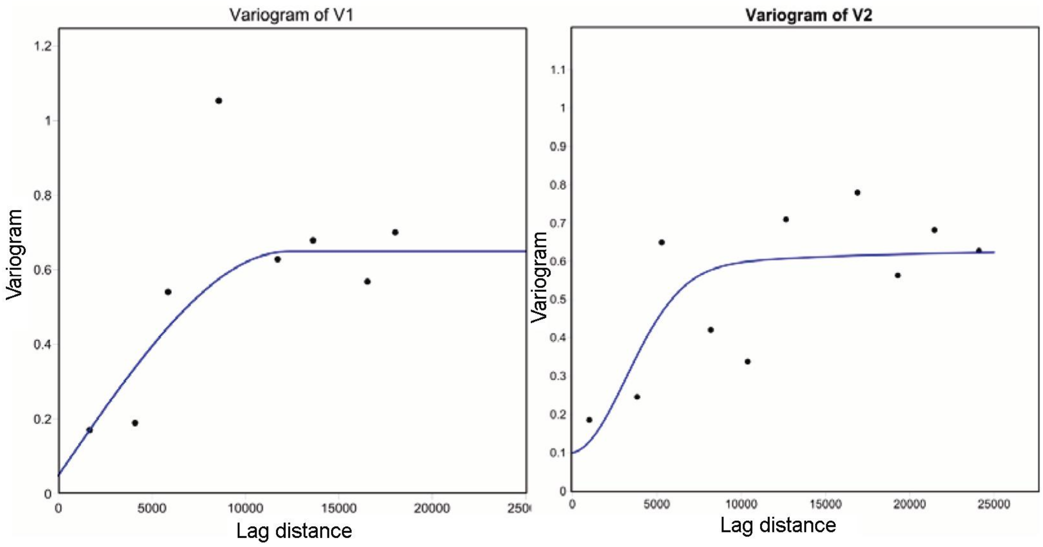

Figure 6. Variogram of V1 and V2.

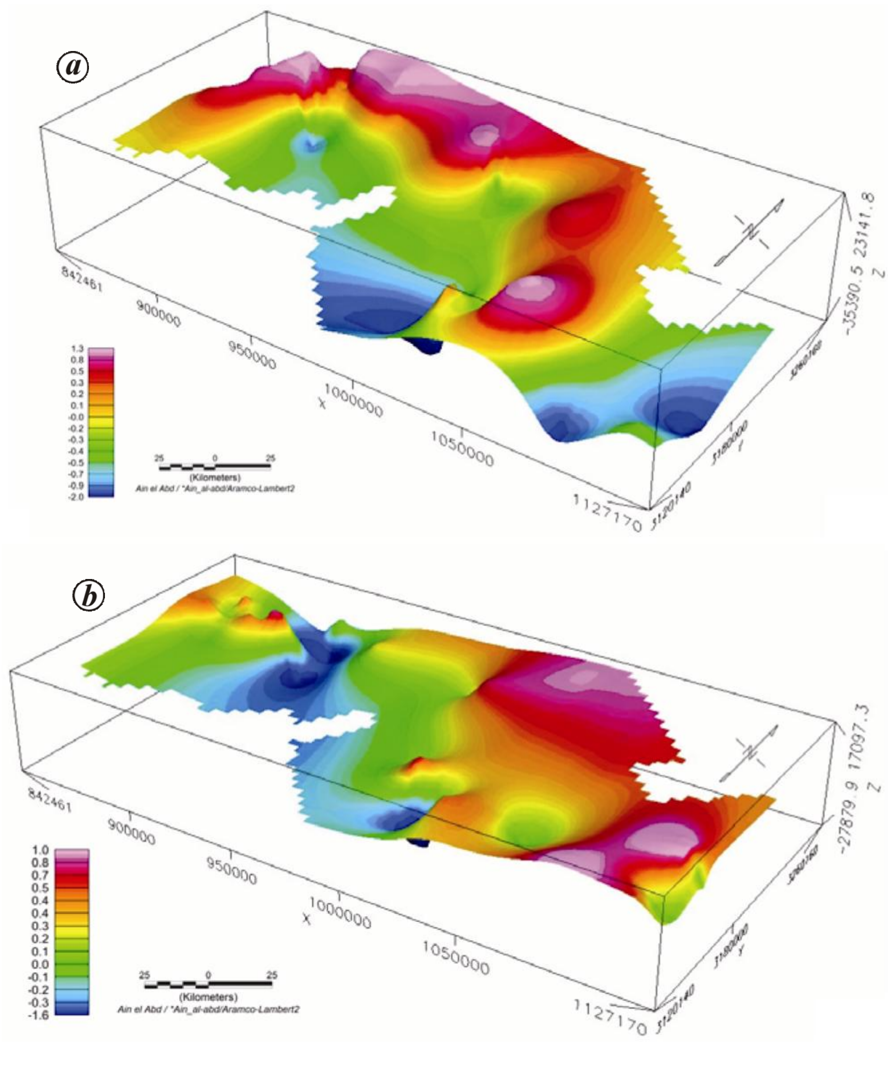

Figure 7. Map showing distribution of (***a***) estimated V1 and (***b***) estimated V2 using point kriging.

considered as the most important and influential. The study demonstrates that the combination of PCA and geostatistics can be applied in cases where the aquifer is complex, database set is limited and with unequal spatial distribution of information.

1. Soumaya, H., Kolsi, S., Bouri, W., Hachicha, H. and Ben, D., Implementation and evaluation of multivariate analysed for groundwater hydrochemistry assessment in arid environment: a case study of Hajeb Elyoun-jelma, Central T Unisia. *Environ. Earth Sci.*, 2013, **70**, 2215–2224.
2. Zekai, C. (ed.), Wadi hydrology, In *Groundwater Management Methods for Arid Regions*, Taylor and Francis, London, 2008, pp. 243–301.
3. Webster, R. and Oliver, M. A., Statistics to support soil research and their presentation. *Eur. J. Soil Sci.*, 2001, **52**(2), 331–340.
4. Espinosa, R. J., Sousa, A. J. and Chica-Olmo, M., Identification of geochemical anomalies using principal components analysis and factorial kriging analysis. *J. Geochem. Explor.*, 1993, **46**, 245–256.
5. Sanchez-Martoz, F., Jiménez, E. R. and Pulido, B. A., Mapping groundwater quality variables using PCA and geostatistics: a case study of Bajo Andarax, southeastern Spain. *Hydrol. Sci. J.*, 2001, **46**(2), 227–242.
6. Jiang, Y., Wu, Y., Groves, Ch., Yuan, D. and Kambesis, P., Natural and anthropogenic factors affecting the groundwater quality in Nandong karst underground river system in Yunan, China. *J. Contam. Hydrol.*, 2009, **109**, 49–61; *Environ Earth Sci.*, 2013, **70**, 2215–2224.
7. Yidana, S. M., Banoeng-Yakubo, B. and Akabzaa, T. M., Analysis of groundwater quality using multivariate and spatial analyses in the Keta basin, Ghana. *J. Afr. Earth Sci.*, 2010, **58**(2), 220–234.
8. Ta'any, R. A., Tahboub, A. B. and Saffarini, G. A., Geostatistical analysis of spatiotemporal variability of groundwater level fluctuations in Amman–Zarqa basin, Jordan: a case study. *Environ. Geol.*, 2009, **57**(3), 525–535.
9. American Public Health Association, Standard methods for the examination of water and wastewater, APHA, Washington DC, USA, 1995, 19th edn.
10. Kolsi, S. H., Bouri, S., Hachicha, W. and Dhia, H. B., Implementation and evaluation of multivariate analysis for groundwater hydrochemistry assessment in arid environments: a case study of Hajeb Elyoun–Jelma, Central Tunisia. *Environ. Earth Sci.*, 2013, 1–10.
11. Hu, S., Luo, T. and Jing, C., Principal component analysis of fluoride geochemistry of groundwater in Shanxi and Inner Mongolia, China. *J. Geochem. Explor.*, 2012.
12. Davis, J. C., *Statistics and Data Analysis in Geology*, Wiley, 2002, 3rd edn, p. 656; ISBN-9780471172758.
13. Giiler, C., Thyne, G. D., McCray, J. E. and Turner, A. K., Evaluation of the graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology*, 2002, **10**, 455–474.
14. Cloutier, V., Lefebvre, R., Therrien, R. and Savard, M., Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J. Hydrol.*, 2008, **353**, 294–313.
15. Kaiser, H. F., The application of electronic computers to factor analysis. *Educ. Psychol. Meas.*, 1960, **20**, 141–151.
16. Ogasawara, H., Standard errors for the direct oblimin solution with Kaiser's normalization. *Shinrigaku Kenkyu*, 1999, **70**(4), 333–338.
17. Closs, L. G. and Nichol, I., The role of factor and regression analysis in the interpretation of geochemical reconnaissance data. *Can. J. Earth Sci.*, 1975, **12**(8), 1316–1330.
18. Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford, New York, 1997.
19. Ahmadi and Sedghamiz, Application and evaluation of kriging and corking methods on groundwater depth mapping. *Environ Monit Assess.*, 2008, **138**, 357–368.
20. Cambardella, C., Moorman, T., Novak, J., Parkin, T., Karlen, D. and Turco, R., *Field Scale Variability of Soil Properties in Central Iowa Soils*, Soil Science Society of America, Madison, 1994, vol. 58, No. 5, pp. 1501–1511.
21. Harmouzi, O., Reconnaissance détaillée de la partie nord-est du Bassin de Saïss (MAROC): interpre´tation de sondages électriques verticaux par combinaison des méthodes statistique, géostatistique et d'inversion (in French). Thèse de doctorat. Univ Moulay Ismail-Univ Bordeaux 1, 2010.
22. World Health Organization, Guidelines for drinking water quality, recommendations, WHO, Geneva, 2008, vol. 1, p. 1308.
23. Gibbs, R. J., Mechanisms controlling world's water chemistry. *Science*, 1970, **170**, 1088–1090.
24. Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M. and Fernandez, L., Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.*, 2000, **34**(3), 807–816.
25. Güler, C., Alpaslan, M., Kurt, M. A. and Temel, A., Deciphering factors controlling trace element distribution in the soils of Karaduvar industrial–agricultural area (Mersin, SE Turkey). *Environ. Earth Sci.*, 2010, **60**(1), 203–218.
26. Nazzal, Y., Habes, G. and Marc, A. R., Application of multivariate geostatistics in the investigation of heavy metal contamination of roadside dusts from selected highways of the Greater Toronto Area, Canada. *Environ. Earth Sci.*, 2013, doi: 10.1007/s12665-013-2546-1.
27. Qin, R., Wu, Y., Xu, Z., Xie, D. and Zhang, C., Assessing the impact of natural and anthropogenic activities on groundwater quality in coastal alluvial aquifers of the lower Liaohe River Plain, NE China. *Appl. Geochem.*, 2013.
28. Al Bassam, A. A., Evaluation of groundwater quality in Al Qassim area, Saudi Arabia using cluster and factor analyses. *Kuwait J. Sci. Eng*., 2006, **32**(2), 101–121.
29. Nazzal, Y. *et al.*, A pragmatic approach to study the groundwater quality suitability for domestic and agricultural usage, Saq aquifer, northwest of Saudi Arabia. *Environ. Monit. Assess*, 2014; doi: 10.1007/s10661-014-3728-3.

ACKNOWLEDGEMENTS. This project was supported by NSTIP strategic technologies programmes (Project no. 12-WAT 2453-02) in the Kingdom of Saudi Arabia.

Received 11 July 2014; revised accepted 14 October 2014