



Evaluation of Groundwater Quality Variations Using Geographical Information System (GIS) and Multivariate Statistical Techniques: A Case Study from Qa'Jahran Basin, Dhamar, Yemen

Nabil M. AL-Areeq, Hadi A. Quriaa, Ahmed M. AL-Areeq, and Amer M. Al-Sabri

Abstract

The current study investigates the water quality of the water resources in Qa'Jahran basin, Dhamar, Yemen. Thirty water wells with a depth ranging from 100 to 600 m below ground level were used to collect water samples. A total of seventeen water quality parameters were measured during 2019, allowing investigation of spatial variation and indication of potential pollution sources. It is essential to have adequate understanding on the interaction effect, trends, and the correlation structure of these parameters to control water quality. Geographical Information System (GIS) and multivariate statistics techniques were used to understand the water resource quality and generate a water quality index map of Qa'Jahran basin. Based on multivariate statistical analysis, strong correlations were observed between total dissolved solids (TDS), conductivity, chloride, total alkalinity, bicarbonate, sodium, and chloride. Sodium has strong correlation with total alkalinity, chloride, and bicarbonate, while it has moderate correlation with fluoride. On the other side, principal component analysis shows that 85.22% of the overall variances were explained by four major principal components. Accordingly, 52% of the data were required to explain 85.22% of total variances as shown by the factor analysis indicating significant data reduction was resulted in this study. According to the water quality index (WQI), most of the water samples were classified as excellent to good for drinking purpose.

Keywords

Water quality parameters • Multivariate statistical analysis • GIS • Water quality index • Qa'Jahran basin • Yemen

1 Introduction

A great increase in the demand in Yemen for the freshwater in the last few decades due to the rapid growth of the population leads to imbalance in available water resource and demand. Hence, the gap between water supply and demand is increasing with time. The groundwater resources are considered as the main source of freshwater in Yemen. This resource is depleted at high rates, and many of groundwater basins in Yemen such as Sana'a are showing decline in the groundwater levels and increase in the salinity which is a strong sign of depletion. The expansions of urbanization, irrigation activities, and industrialization have a significant effect on the groundwater resources. Accordingly, conserving and monitoring this important resource is essential. The aquifer in Qa'Jahran basin consists of two layers (sedimentary unit and volcanic fractures unit), and as a result of this weakness, the probability of water contamination by human or nature is very high. There are many sources of water contamination in the study area which disposed randomly such as wastewater, oils, structures waste, and municipal solid waste. Large quantities of human and industrial waste disposals pose a serious threat to this valuable resource. The sources of these contaminants are the study area itself. These contamination sources will affect the quality of water resources which are used by about 137,804 persons. According to the report of WHO, water is the main reason for 80% of all human diseases. Regular monitoring of the groundwater is required to protect it where it is impossible in most of the cases to stop the pollutants from the source (Al-hadithi, 2012).

N. M. AL-Areeq (✉) · H. A. Quriaa · A. M. Al-Sabri
National Water Resource Authority (NWRA), Sanaa, Yemen

N. M. AL-Areeq · A. M. Al-Sabri
Department of Geology and Environment, Faculty of Applied Sciences, Thamar University, Dhamar, Yemen

A. M. AL-Areeq
Civil and Environmental Engineering Department King, Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia

Manual, tube wells, and springs constitute the main resources for domestic, industrial, and irrigation uses in many areas. But, few of them are maintained in their original condition due to intensive human activities (Zhao et al., 2011). Today, the shortage of freshwater sources in many countries is a great concern (Alberto et al., 2001; Singh et al., 2004; Qadir et al., 2008). A constant polluting source due to municipal and industrial wastewater discharge and a seasonal phenomenon due to surface runoff can be most affected by climate (Singh et al., 2004). Decision-makers and water quality specialists often face significant challenges to control water contamination due to these complexities (Elhatip et al., 2008). To keep water sources useable, protecting water quality is essential to minimize water treatment cost and to satisfy regulatory limitations (Chowdhury et al., 2007; WHO, 2017; Canada, 2017).

Statistical analyses of water chemistry data are widely used in water analysis interpretation. Complex data matrices of water quality can be better interpreted and understood by applying different multivariate statistical techniques, such as factor analysis (FA) and principal component analysis (PCA). Rapid solution to pollution problems and the reliable management of water resources can be done using such statistical techniques which facilitate the identification of possible factors that influence water quality (Lee et al., 2001; Reghunath et al., 2002). Voza et al. (2015) proved the utility of multivariate statistical methods, factor analysis (FA), principal component analysis (PCA), and cluster analysis (CA) for the analysis and interpretation of the complex water quality data sets. Many studies demonstrated that multivariate statistical analysis is useful to assess the spatial water quality variations (Simeonov et al., 2003; Zeilhofer et al., 2006; Zhang et al., 2009a, 2009b; Olsen et al., 2012; Marinović & Ružđjak, 2015; Phung et al., 2015; Muangthong & Shrestha, 2015). Freshwater quality can be characterized and evaluated using multivariate statistical techniques which are also useful in verifying temporal and spatial variations (Singh et al., 2005; Chowdhury & Al-Zahrani, 2014).

Environmental data management is one of the important applications of Geographic Information System (GIS). To avoid possible environmental crisis, managers and scientists can be provided with different scenarios using GIS for predicting future trends and spatial distribution of the data. GIS is widely used in pollution and water resource management due to its importance in playing a critical role in these fields (Zeilhofer et al., 2007; Igboekwe & Akankpo, 2011; Yan et al., 2015). Nasher and El-Sagheer (2012) studied bacteriological, chemical, and physical parameters of water resources in Bura'a natural protected area, Hodeidah, Yemen, to evaluate the water quality for agricultural and drinking purpose. They found that the water is suitable for irrigation purposes, but it is unsuitable for drinking. Saleh

et al. (2018) assessed the suitability of groundwater quality of rural areas (Zabid, Hodeidah) for drinking purpose using water quality index (WQI). They found that the groundwater quality is generally good–poor water quality status. Aqeel et al. (2017) used GIS to create a geospatial distribution map of fluoride concentrations and identified three risky zones in the Al-Howban Basin, Taiz-Yemen. Mohamed and Hassane (2016) studied the impact of urbanization on groundwater resources in Al Ain city, UAE, by performing hydrochemistry assessment of the quality of groundwater. Nasher et al. (2013) identified and evaluated the hydrogeochemical processes for the management of the groundwater resources of the lower part of Wadi Siham, Tihama plain, Yemen. Nazzal et al. (2015) investigated Saq aquifer, northwestern part of KSA by conducting a combined hydrochemical–statistical analysis to assess groundwater resources. Toumi et al. (2015) studied the water quality and the hydrochemical characteristics in Al-Ula Region, Saudi Arabia, and they found that the lithological influences are the guide of water chemistry. Gebrehiwot et al. (2011) evaluated the suitability of groundwater quality in El Khairat aquifer, Enfidha, Tunisia, for drinking purposes. They conclude that GIS is an effective tool for creating maps showing the spatial distribution of water quality parameters. Eslami et al. (2013) used GIS interpolation techniques to evaluate the groundwater quality of Mianab plain, Iran, and they found that groundwater quality in their study area is not suitable for drink and agricultural uses. They also confirm the importance of selecting an appropriate interpolation technique and illustrated the effective of geostatistical technique in groundwater management system. Al-Omran et al. (2018) used different techniques such as water quality index (WQI) and Durov diagrams to assess the groundwater quality of Al-Hasa region, Saudi Arabia, for drinking and irrigation purposes, and they found that the water is excellent for irrigation and poor to unsuitable for drinking purposes.

The main objective of this study is for evaluation of groundwater quality variations in Qa'Jahran basin in Yemen. This study focuses on using the Geographical Information System (GIS) and multivariate statistics to understand the water resource quality and generate a water quality index map for the first time in this area.

2 Study Area and Data Acquisition

The Qa'Jahran basin is located in the middle western part of Yemen, at the Central Highland Plains in Jahran district. It represents the main catchments of Jahran plain (Fig. 1). The geographic extent of the basin area lies between longitudes 44° 12' 20" and 44° 22' 30" E and latitudes 14° 38' 11" and 14° 57' 30" N with total area extent 406 km², and it is about 25 kma distance toward the north of Dhamar city. The

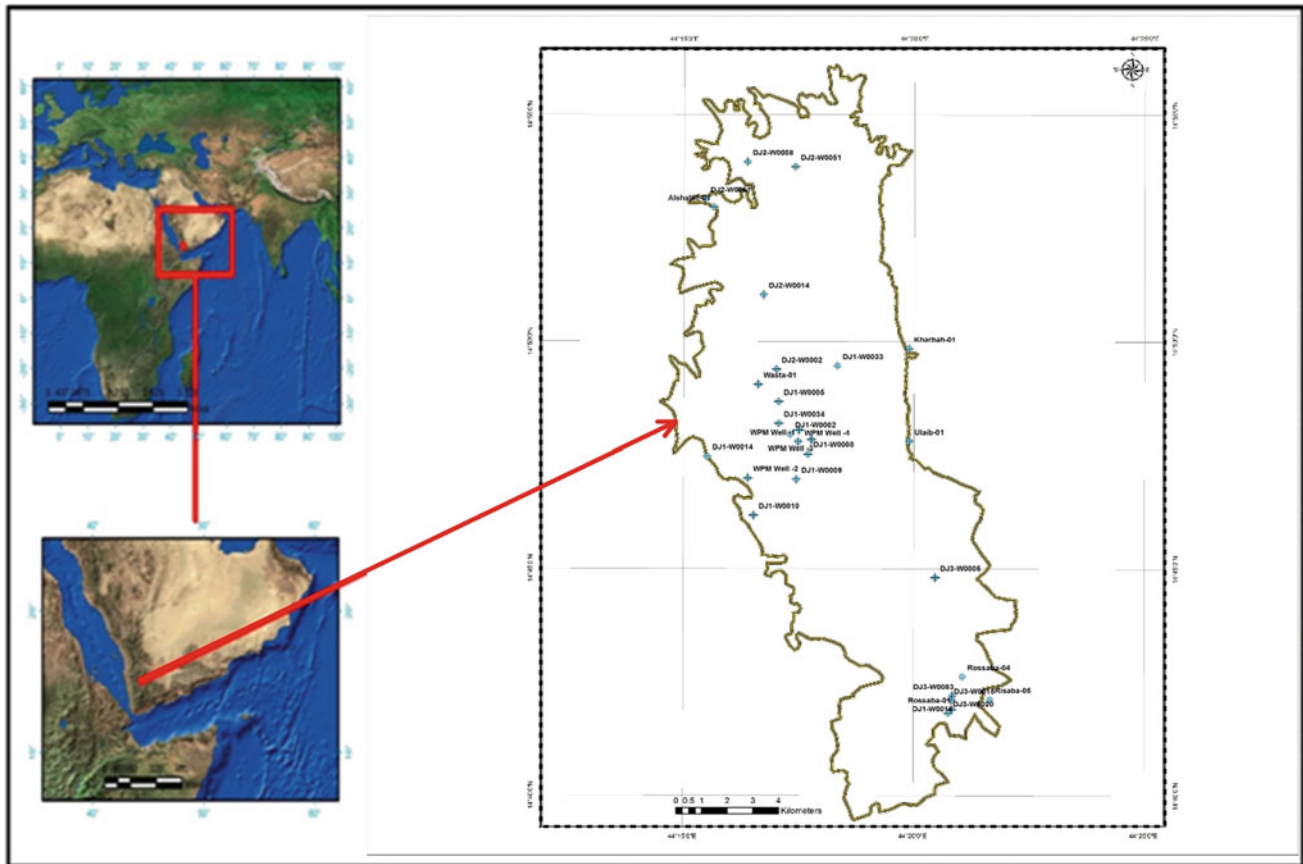


Fig. 1 Location map of the groundwater samples from the Qa'Jahran basin, Dhamar, Yemen

Qa'Jahran basin is one of the largest and most completely flat areas of its size in the Highland Plains of Yemen. The catchment area of the basin is affected by soil erosion. The intensive drilling activity for water quest in the catchment and water scarcity related to human interface with water storage has resulted in groundwater to sink down and the basin threatened by drought due to groundwater depletion. The manual and tube wells are the main resource of the water in Qa'Jahran basin. The people in the area depend mainly on these resources for their water requirements. Thirty water wells were selected from Qa'Jahran basin from different locations for the study taking into consideration the hydrologic, social, and economic variation. All statistical and mathematical computations were made using JMP, MINITAB™, SPSS Statistics, XLSAT, and Microsoft Office Excel 2016.

2.1 Scatterplots and Pairwise Correlations

The indication of linear or nonlinear relationships and possible correlations were represented by scatterplots and pairwise correlations. Scatterplots and pairwise correlations of

Qa'Jahran basin data were investigated by this study using JMP and the MINITAB™ statistical package.

2.2 Principle Component Analysis (PCA)

PCA is designed to convert the original variables (correlated variables) into new, linearly uncorrelated variables, called principal components. The majority of the data set can be described by the most meaningful parameter information which is provided by PCA (Helena et al., 2000). As a result of the reduction in the dimensionality of data by PCA, a smaller number of transformed variables can be used to represent the bulk of the variation in the data (Jackson & John Wiley & Sons, 1991; Wold et al., 1987). The following equation can represent the principal component (PC):

$$Z_{ji} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \dots + a_{im}x_{mj}$$

where *Z* is the component score, *i* is the component number, *j* is the sample number, *a* is the component loading, *x* is the measured value of variable, and *m* is the total number of variables.

The transformed data coming from the PCA can be simplified even more with the benefit of the factor analysis (FA) which construct new variables, called varifactors (VFs) by reducing the contribution of less significant variables. VF includes hypothetical, unobservable, latent variables, while PC is a linear combination of observed water quality variables (Helena et al., 2000; Vega et al., 1998). Normalized variables are used in PCA to extract PCs with major significance and reduce that have minor significance (Zhang et al., 2009a,2009b; Bu et al. 2010). As a result, the amount of information obtained from a less number of variables will be similar to that resulted from the much larger set of original variables. The following equation can represent the factor analysis (FA):

$$Z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \dots + a_{fm}f_{mi} + e_{ji}$$

where Z is the measured variable, i is the sample number, a is the factor loading, f is the factor score, m is the total number of factors, and e is the residual term accounting for errors or other source of variation. In this study, fifteen of Qa'Jahran water quality data set was subjected to FA/PCA.

2.3 Water Quality Index (WQI)

It is a rating that can be used to reflect the composite influence of different water quality parameters on the overall water quality. The main objective of computing water quality index (WQI) is to obtain usable and understandable information from the complex water quality data. A parameter is selected based on its impact on health and the overall quality of water. The WQI can be computed by the following three steps.

First step: Assign weight (w_i) to the selected water parameters (e.g., TDS, pH, TH, HCO_3 , SO_4 , Cl, NO_3 , Fe, etc.) according to their relative importance in the overall water quality for drinking purposes (weight may be from 1 to 5).

Second step: Compute a relative weight (W_i) of the chemical parameter using the following equation:

$$W_i = w_i / \sum_{i=1}^n w_i$$

where W_i is the relative weight, ' n ' is the number of parameters, and w_i is the weight of each parameter.

Third step: Assign a quality rating scale (q_i) for each parameter, as below:

$$q_i = (C_i / S_i) \times 100$$

where q_i is the quality rating, S_i is the guideline value/desirable limit as given in Yemeni drinking water

standard, and C_i is the concentration of each chemical parameter in mg/l. For the computation of WQI, firstly the subindex (SI) is determined for each water quality parameter, as given below:

$$\begin{aligned} \text{SI} &= W_i \times q_i \\ \text{WQI} &= \sum \text{SI}_i \end{aligned}$$

where W_i is relative weight of the parameter, SI_i is the subindex of the parameter, and q_i is the rating based on concentration of the parameter.

2.4 Classification of Water

The water can be classified into five types based on computed WQI, as given below:

Excellent water (<50); good water (50–100); poor water (100–200); very poor water (200–300); and water unsuitable for drinking (>300).

3 Results and Discussion

The statistical analysis of water quality parameters (WQPs) is calculated for 30 (WQP) in Qa'Jahran wells as shown in Table 1. Distribution fitting was examined using maximum likelihood as an estimation method and tested using Kolmogorov–Smirnov and chi-square method. The gamma (2) distribution is the distribution that fits best the data as shown in Table 2. The water quality data sets of Qa'Jahran basin were subjected to multivariate statistical techniques: FA and PCA.

3.1 Correlation and Scatterplots

The correlations between different water quality parameters are carried out as shown in Table 3. The pH has strong correlation with carbonate ($r = 0.857$) and negative strong correlation with hardness ($r = -0.632$) and calcium (-0.639). Total dissolved solids is perfectly correlated with conductivity ($r = 1$). Total dissolved solids and conductivity show strong correlation with chloride and bicarbonate ($r = 0.881$), total alkalinity ($r = 0.868$), sodium ($r = 0.803$), and sulfate ($r = 0.743$), and moderate correlations with magnesium ($r = 0.597$) and potassium ($r = 0.475$). Sodium has strong correlation with total alkalinity, bicarbonate, and chloride ($r = 0.845$, 0.822 , and 0.628 , respectively), and moderately correlated with fluoride ($r = 0.484$). The strong positive correlation of sodium with chloride and between them with conductivity reflects a common source of these

Table 1 Statistical analysis of water quality parameters (WQPs) of Qa'Jahran wells

Statistic	Minimum	Maximum	Median	Mean	Variance	Standard deviation	WHO	G.Y.L
Turbidity	0.040	4.800	0.290	1.073	2.162	1.470		
PH (9–6.5)	7.960	9.260	8.385	8.473	0.112	0.335	6.5–8.5	6.5–8.5
EC	117.800	1006.000	435.250	462.461	41,484.80	203.678	400–1500	450–2500
TDS	76.570	653.900	282.913	300.600	17,527.33	132.391	1000.00	650–1000
TH	0.000	192.200	72.300	77.737	3287.23	57.334	100–500	100–500
TA	38.800	272.570	146.470	148.238	3256.08	57.062		
HCO ₃	42.600	332.500	178.100	177.038	4799.40	69.278	150–500	150–500
CO ₃	0.000	17.500	0.000	3.128	28.01	5.293		
Cl	7.800	110.600	25.500	32.472	555.70	23.573	250.00	200–600
SO ₄	1.000	102.000	21.500	30.111	734.45	27.101	25–400	200–400
F	0.220	3.700	0.580	0.973	0.97	0.986	0.5–1.5	0.5–1.5
Ca	0.000	54.600	18.650	21.622	296.80	17.228	75–200	75–200
Mg	0.000	20.600	3.300	5.728	28.00	5.292	30–50	30–150
Na	7.500	190.600	71.500	82.200	2560.48	50.601	20–175	200–400
K	0.400	5.400	0.950	1.611	2.06	1.436	8–12	8–12
NO ₃	2.200	34.500	7.700	9.244	49.25	7.018	25–50	10–50
Fe	0.010	0.340	0.035	0.053	0.006	0.076	0.3–1	0.3–1

Table 2 Distribution fitting of the data analysis

Distribution	<i>p</i> -value	Distribution	<i>p</i> -value	Distribution	<i>p</i> -value
Beta4	0.181	Gamma (1)	0.044	Normal	0.096
Chi-square	0.733	Gamma (2)	0.745	Normal (Standard)	<0.0001
Erlang	0.002	GEV	0.014	Student	<0.0001
Exponential	0.063	Gumbel	0.007	Weibull (1)	0.318
Fisher–Tippett (1)	<0.0001	Log-normal	0.683	Weibull (2)	0.496
Fisher–Tippett (2)	<0.0001	Logistic	0.147	Weibull (3)	0.366

ions. Potassium is moderately correlated with chloride ($r = 0.52$). Hardness shows strong correlation with calcium ($r = 0.945$) and magnesium ($r = 0.76$) and moderate correlation with chloride, sulfate, and nitrate ($r = 0.474$, 0.533 , and 0.494 , respectively), while it has negative moderate correlation with carbonate ($r = -0.534$). Magnesium has strong correlations with chloride ($r = 0.705$) and moderate correlations with sulfate and calcium ($r = 0.544$ and 0.51 , respectively). Total alkalinity is strongly correlated with bicarbonate ($r = 0.996$) and shows moderate correlation with chloride ($r = 0.598$). Bicarbonate has strong correlation with chloride ($r = 0.624$) and moderate correlation with sulfate ($r = 0.473$). Carbonate shows moderate correlation with calcium ($r = 0.575$) and same correlation between calcium and nitrate ($r = 0.503$). Strong correlations were

shown between chloride and sulfate ($r = 0.632$) and between nitrate and fluoride ($r = 0.851$).

3.2 Principal Component Analysis (PCA)

PCA was performed for Qa'Jahran basin data. Based on the eigenvalue–one and scree plot (Fig. 2 and Table 4), the number of significant principal components (PCs) was determined. PCs are regarded as significant if eigenvalues are greater than one. Four significant PCs with eigenvalues > 1 were extracted in this study and explained 40.137%, 26.277%, 11.943%, and 6.863% of the variance, respectively. The findings indicate that 85.22% of total variance associated with WQP in Qa'Jahran basin can be explained by the first four PCs.

Table 3 Correlation of different parameters of chemical analysis for water samples in Qa'Jahran basin

Variables	Turbidity NTU (5)	PH (9– 6.5)	EC µS/cm (2500)	TDS mg/L (1500)	TH mg/L (500)	TA mg/L (–)	HCO ₃ mg/L (500)	CO ₃ mg/L (–)	Cl mg/L (600)	SO ₄ mg/L (400)	F mg/L (1.5)	Ca mg/L (200)	Mg mg/L (150)	Na mg/L (400)	K mg/L (12)	NO ₃ mg/L (50)	Fe mg/L (1)		
Turbidity NTU (5)	1																		
PH (9– 6.5)	0.030	1																	
EC µS/cm (2500)	-0.322	-0.242	1																
TDS mg/L (1500)	-0.322	-0.242	1.000	1															
TH mg/L (500)	0.100	-0.632	0.443	0.443	1														
TA mg/L (–)	-0.558	0.008	0.868	0.868	0.205	1													
HCO ₃ mg/L (500)	-0.556	-0.072	0.881	0.881	0.256	0.996	1												
CO ₃ mg/L (–)	-0.050	0.857	-0.086	-0.086	-0.534	0.099	0.006	1											
Cl mg/L (600)	-0.089	-0.354	0.881	0.881	0.474	0.598	0.624	-0.243	1										
SO ₄ mg/L (400)	-0.057	-0.377	0.743	0.743	0.533	0.455	0.473	-0.173	0.632	1									
F mg/L (1.5)	-0.326	0.117	0.355	0.355	-0.160	0.318	0.302	0.187	0.141	0.557	1								
Ca mg/L (200)	0.048	-0.639	0.288	0.288	0.945	0.097	0.151	-0.575	0.274	0.433	-0.147	1							
Mg mg/L (150)	0.166	-0.398	0.597	0.597	0.764	0.348	0.375	-0.268	0.705	0.544	-0.131	0.510	1						
Na mg/L (400)	-0.406	0.180	0.803	0.803	-0.160	0.845	0.822	0.292	0.628	0.454	0.484	-0.312	0.195	1					
K mg/L (12)	-0.334	-0.242	0.475	0.475	0.183	0.365	0.393	-0.284	0.521	0.245	0.000	0.146	0.196	0.327	1				
NO ₃ mg/L (50)	0.630	-0.341	0.033	0.033	0.494	-0.209	-0.180	-0.316	0.161	0.125	-0.413	0.503	0.304	-0.302	-0.189	1			
Fe mg/L (1)	0.796	-0.113	0.003	0.003	0.332	-0.317	-0.311	-0.070	0.166	0.246	-0.090	0.292	0.294	-0.218	-0.340	0.851	1		

Values in bold are different from 0 with a significance level alpha = 0.05

Fig. 2 Scree plot of water samples in Qa'Jahran basin

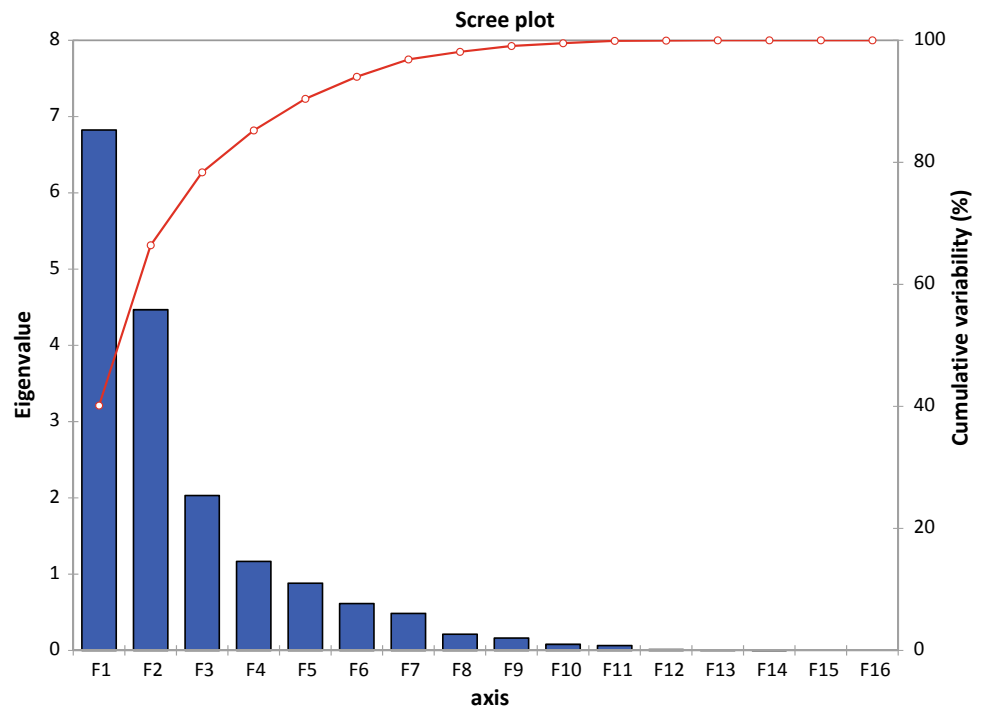


Table 4 Total variance explained of water samples in Qa'Jahran wells

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	6.823	40.137	40.137	6.823	40.137	40.137
2	4.467	26.277	66.414	4.467	26.277	66.414
3	2.03	11.943	78.357	2.03	11.943	78.357
4	1.167	6.863	85.22	1.167	6.863	85.22
5	0.882	5.186	90.406			
6	0.614	3.615	94.021			
7	0.485	2.853	96.874			
8	0.211	1.242	98.116			
9	0.161	0.945	99.061			
10	0.08	0.469	99.53			
11	0.063	0.373	99.903			
12	0.009	0.054	99.957			
13	0.007	0.039	99.996			
14	0.001	0.004	100			
15	7.61E-07	4.48E-06	100			
16	2.96E-08	1.74E-07	100			
17	1.01E-16	5.91E-16	100			

3.3 Loading and Scatterplots

Understanding the contribution of original WQP to the major PCs is essential to decide which parameter is important and which is not. Figures 3 and 4 show the scatterplots and loading on the major PCs for Qa’Jahran basin. The WQPs, EC, TDS, TA, HCO₃, Cl, SO₄, Mg, and Na exceeded the threshold limit of 0.55 on PC1 (Fig. 3). Therefore, they had significant loadings in PC1 and PC2 and have significant loadings from turbidity, pH, TH, CO₃, Ca, Na, NO₃, and Fe. In PC3, significant loadings were assigned from iron turbidity, CO₃, and Fe. The PC4 obtained significant loading from fluoride (F) (Table 5). Figure 4 shows the scatterplots of the loadings. The plot of PC1 × PC2 shows that nitrate and iron were significantly associated with PC2 only, while conductivity, TDS, alkalinity, chloride, bicarbonate, magnesium, sulfate, and sodium were significantly associated with PC1. To represent water quality, it may be required one or two parameters from each cluster. In general, the four PCs were mainly associated with conductivity, TDS, alkalinity, chloride, bicarbonate, magnesium, sulfate, sodium, nitrate, and iron.

The corresponding VFs were obtained for Qa’Jahran basin (Table 5). The factor loadings were classified

according to absolute loading values as strong > 0.75, moderate, 0.75–0.50, and weak 0.50–0.30 (Liu et al., 2003). The percent (40.137%) of total variance was explained by VF1 and had strong positive loadings from alkalinity, bicarbonate, sulfate, chloride, total dissolved solids, and conductivity, while hardness, sodium, manganese, and potassium had a medium loading on VF1. The percent (26.277%) of total variance was explained by VF2, had strong positive loadings from nitrate, and had medium loading from turbidity, hardness, and calcium. The percent (11.943%) of total variance was explained by VF3 which had medium loadings from iron, turbidity, and carbonate. The VF4 explained 6.863% of the total variance and had weak negative loading from sulfate.

3.4 Data Reduction

The original parameters on the major VFs were assessed to investigate the possibility of reduction of data. It was reported by the past studies that the original parameters have the greatest contribution to temporal variation if they have loadings greater than 0.7 or 0.75 in the major VFs (Liu et al., 2003; Varol et al., 2012). The value (85.22%) of variance is

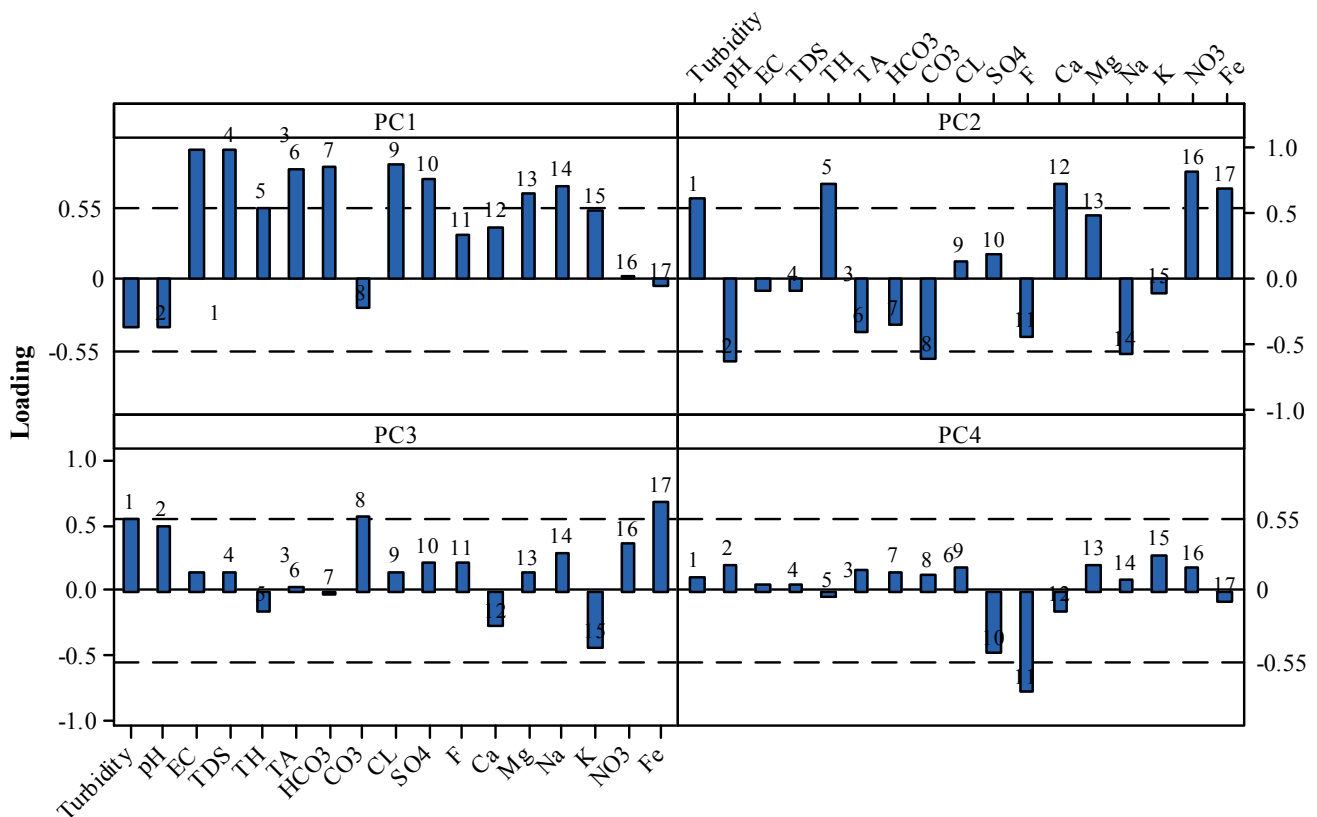


Fig. 3 Scatterplots and loading on the major PCs for Qa’Jahran basin

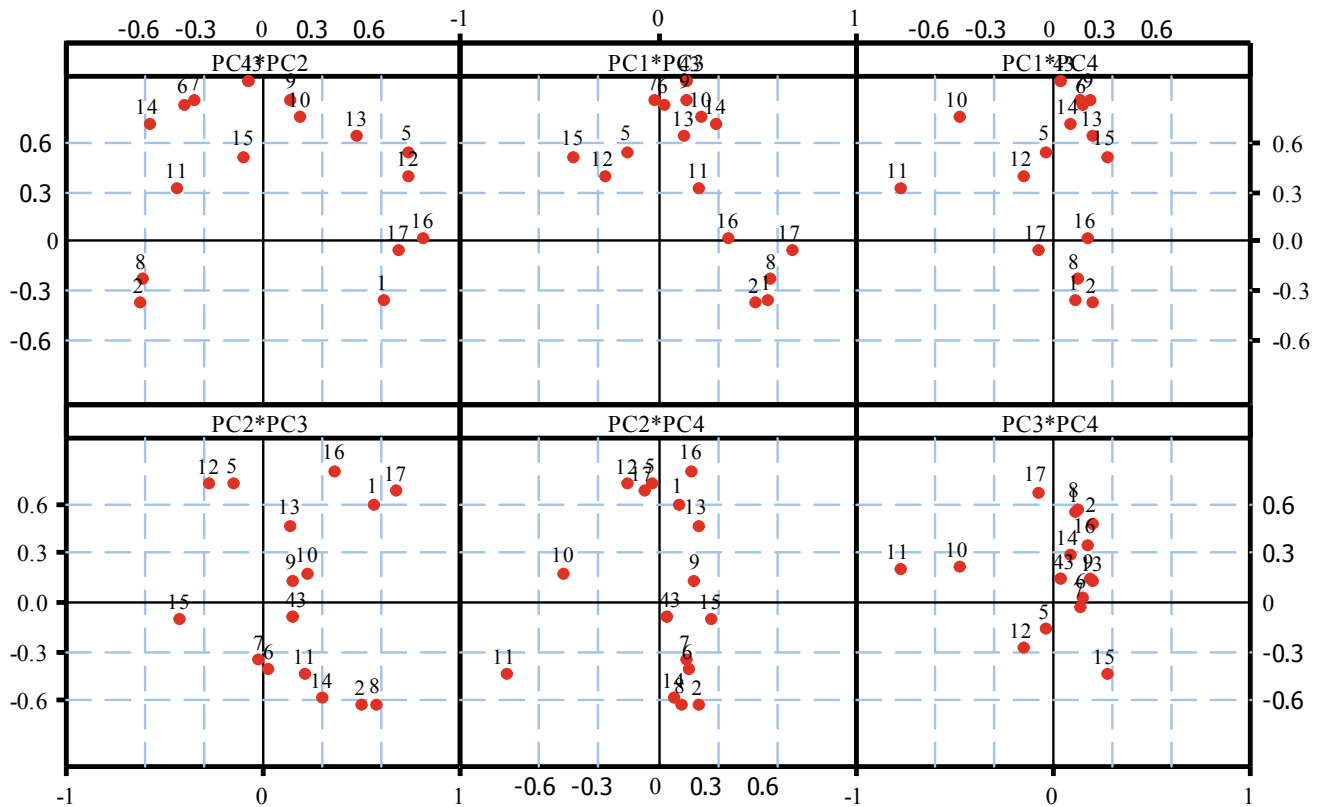


Fig. 4 Scatterplots of the (1) turbidity, (2) pH, (3) conductivity, (4) TDS, (5) hardness, (6) alkalinity, (7) bicarbonate, (8) carbonate, (9) chloride, (10) sulfate, (11) fluoride, (12) calcium, (13) magnesium, (14) sodium, (15) potassium, (16) nitrate, and (17) iron

explained by the major four VFs which had strong loadings from sulfate, sodium, chloride, manganese, calcium, total dissolved solids, hardness, bicarbonate, conductivity, iron, total alkalinity, and nitrate as shown in Table 7. As TDS, conductivity, chloride, hardness, magnesium, sodium, and sulfate are strongly correlated as shown in Table 5, TDS can be selected to be significant. Accordingly, a total of twelve parameters (sulfate, chloride, sodium, manganese, calcium, hardness, total dissolved solids, conductivity, bicarbonate, total alkalinity, nitrate, and iron) are needed (approximately 80% of the 15 parameters) to explain 85.22% of the data variance (Table 5).

3.5 Water Quality Parameters

The minimum and maximum values of each parameter were identified, and the spatial distributions for sixteen parameters were determined for the study area. The value of iron (Fe) varies from 0.01 to 1.08 mg/l, and the highest concentration is in the Al-ulaib well east part of the study area (Fig. 5a). The spatial distribution of nitrate (NO_3) is shown in Fig. 5b which shows the concentration variation range from 0.8 to 101 with highest concentration only in DJ2-W-9

well. The concentration of potassium is low in the whole study area with values between 0.4 and 5.4 mg/l (Fig. 5c). The spatial distribution of sodium varies from 7.5 to 190.6, and the highest concentration is in the DJ1-W8 (Fig. 5d). Figure 5e shows that spatial distribution of magnesium which varies from 0.0 to 37 mg/l. The spatial distribution of calcium varies from 0.0 to 150, and the highest concentration of calcium is in the DJ2-W-9 well. Fluoride varies from 0.2 to 3.7 with high concentration in the northern part and low concentration only in the central part of the study area (Fig. 5f). The value of sulfate (SO_4) varies from 1 to 102 mg/l, and the low concentration is in the northern and southern parts of the study area (Fig. 5g). The spatial distribution of chloride is shown in Fig. 5h which shows the concentration variation is from 7.8 to 110.6. The concentration of bicarbonate varies from 42.6 to 332.5 and is low in the whole study area except DJ1-W33 within the permissible range (Fig. 5i). The value of pH range between 7.9 and 9.2 mg/l is shown in Fig. 5k, which shows that the water is alkaline in most of the study area. TDS varies from 76.5 to 653.9 mg/l which proved that the TDS in groundwater is low (Fig. 5l). That spatial distribution of total hardness of water varies from 0 to 192, and most of the study area has low concentration of TH as shown in Fig. 5n. Figure 5o

Table 5 Factor loadings on the major PCs for Qa'Jahran basin

Variable	Factor				Communalities
	VF1	VF2	VF3	VF4	
Turbidity	-0.364	0.61	0.559	0.111	0.83
pH	-0.369	-0.627	0.494	0.206	0.817
EC	0.982	-0.08	0.148	0.043	0.995
TDS	0.982	-0.08	0.148	0.043	0.995
TH	0.549	0.733	-0.156	-0.035	0.865
TA	0.842	-0.4	0.03	0.153	0.892
HCO ₃	0.867	-0.344	-0.023	0.143	0.89
CO ₃	-0.224	-0.614	0.575	0.119	0.772
CL	0.87	0.142	0.145	0.188	0.833
SO ₄	0.763	0.185	0.22	-0.479	0.895
F	0.329	-0.438	0.21	-0.777	0.948
Ca	0.403	0.733	-0.276	-0.151	0.799
Mg	0.649	0.479	0.135	0.206	0.711
Na	0.714	-0.58	0.298	0.085	0.941
K	0.52	-0.103	-0.429	0.272	0.539
NO ₃	0.02	0.812	0.359	0.172	0.819
Fe	-0.049	0.69	0.68	-0.074	0.947
% variance	40.137	26.277	11.943	6.863	
Cumulative (%)	40.137	66.414	78.357	85.22	

shows that spatial distribution of total alkalinity varies from 38 to 278. Figure 5p shows that spatial distribution of turbidity which varies from 0.04 to 4.8.

3.6 Water Quality Index (WQI)

WQI is widely used tool to assess the water quality. According to the relative importance of each water quality parameter, the relative weight was determined as shown in Table 6. Water quality index was calculated for each parameter as shown in Table 7 and Fig. 6. It can be noticed that 72% of the samples fall in excellent water category and 28% of the samples fall in good water category that reflect good-quality water in Qa'Jahran basin.

4 Conclusions

Spatial variability of water quality parameter (WQP) in Qa'Jahran basin was evaluated in this study. Several techniques, including pairwise correlations, FA, and PCA, were performed on the data collected in 2019. Data reduction was also investigated based on assessing their loadings and correlation structure on the major PCs and VFs. The possibility of data reduction was investigated in this study. About 80% of data in Qa'Jahran basin can be reduced in expense of 20% of overall variance. This research points out that the data reduction can be possible in the monitoring program. Approximately 48% can be reduced in expense of 14.78% of the overall variance. A total of nine parameters out of 17

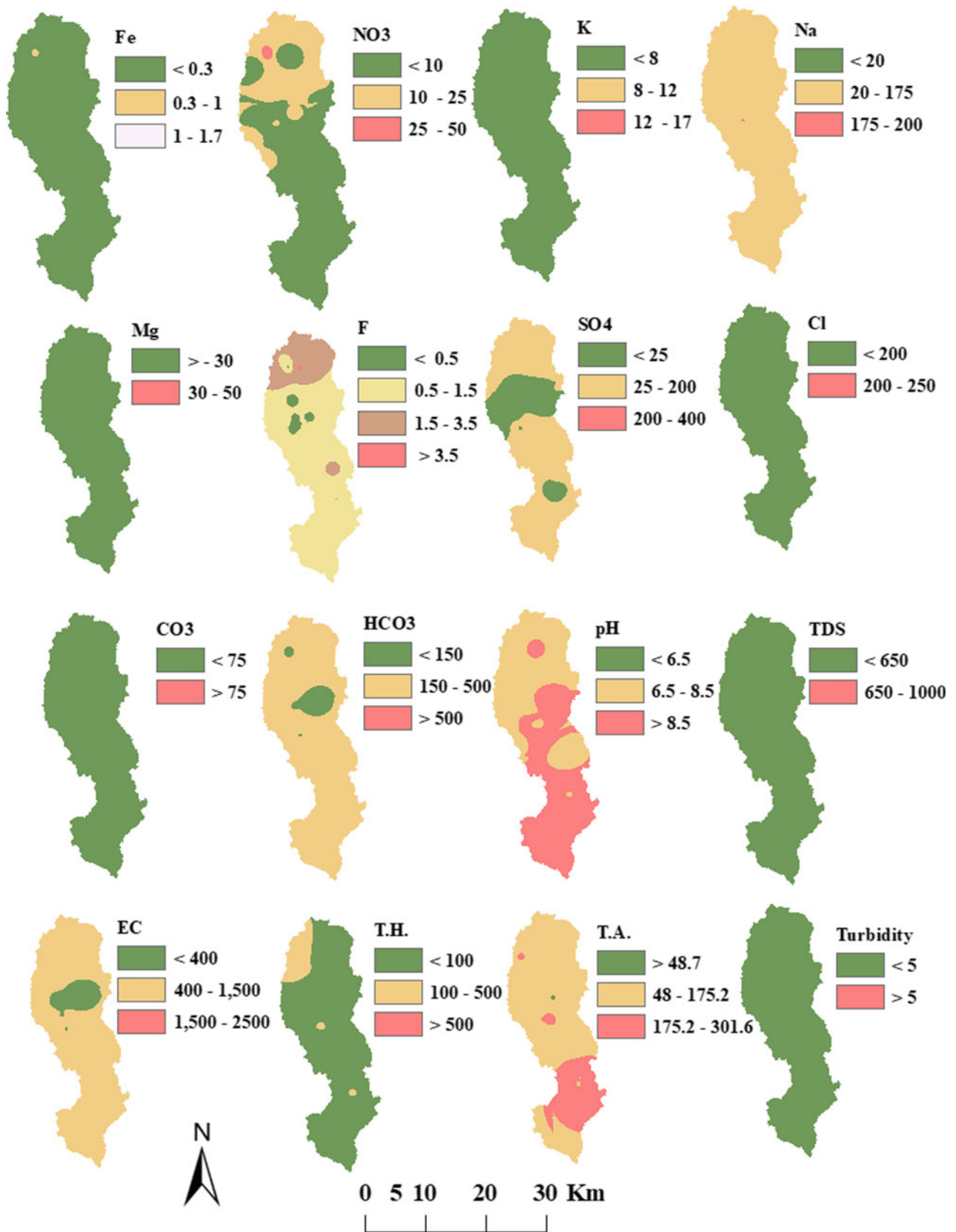


Fig. 5 Water quality parameters (spatial distribution maps) of Qa'Jahran basin

Table 6 Relative weight for each water quality parameter (Gebrehiwot et al., 2011)

Chemical parameters	G.Y.L	WHO (2011)	Weight (w_i)	Relative weight $W_i = w_i / \sum n w_i$
Turbidity		5	2	0.03922
pH	6.5–8.5	8.5	4	0.07843
EC ($\mu\text{S/cm}$)	450–2500	500	4	0.07843
TDS (mg/L)	650–1000	500	5	0.09804
TH (mg/L)	100–500	–	3	0.05882
TA		100	2	0.03922
HCO ₃ [–] (mg/L)	150–500	500	3	0.05882
Cl [–] (mg/L)	200–600	250	3	0.05882
SO ₄ ^{2–} (mg/L)	200–400	250	4	0.07843
F [–] (mg/L)	0.5–1.5	1	4	0.07843
Ca ²⁺ (mg/L)	75–200	75	2	0.03922
Mg ²⁺ (mg/L)	30–150	50	2	0.03922
Na ⁺ (mg/L)	200–400	200	2	0.03922
K ⁺ (mg/L)	8–12	12	2	0.03922
NO ₃ [–] N (mg/L)	10–50	45	5	0.09804
Fe ²⁺ (mg/L)	0.3–1	0.3	4	0.07843

Table 7 Water quality index for the Qa'Jahran wells

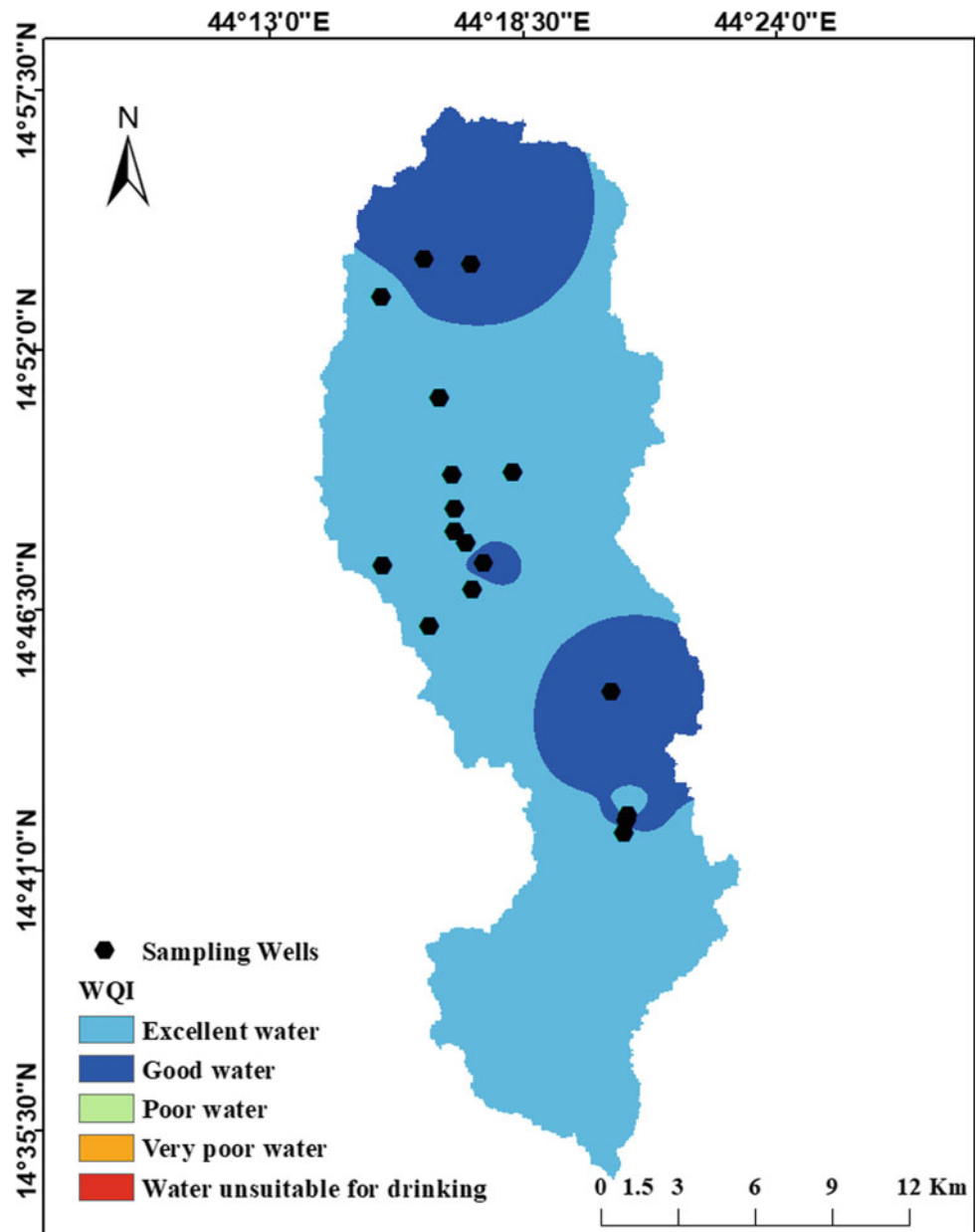
ID	WQI	Water classification
DJ1-W0005	38.243843	Excellent water
DJ1-W0015	64.904057	Good water
DJ1-W0002	39.601739	Excellent water
DJ1-W0008	73.76476	Good water
DJ1-W0014	43.2662	Excellent water
DJ1-W0034	35.64832	Excellent water
DJ1-W0010	43.719578	Excellent water
DJ2-W0051	64.260741	Good water
DJ3-W0020	42.006772	Excellent water
DJ2-W0002	34.78738	Excellent water
DJ3-W0005	61.615663	Good water
DJ3-W0083	41.377107	Excellent water
DJ2-W0014	37.560144	Excellent water
DJ2-W0067	45.272093	Excellent water
DJ2-W0008	58.925567	Good water
DJ1-W0009	34.209919	Excellent water
DJ1-W0033	23.970572	Excellent water
DJ3-W0015	19.979173	Excellent water

(conductivity, TDS, hardness, total alkalinity, bicarbonate, chloride, sulfate, calcium, nitrate) explained 85.22% of the data variance.

According to WHO and G.Y.L., most of the chemical elements in the study area fall in the permissible limits. WQI and GIS approach were integrated to evaluate water quality

in the study area and identify the potential risky zones. This water quality index map shows that the water in the study area is mostly excellent to good water, and based on that, this water is recommended for domestic use and human drinking.

Fig. 6 Spatial distribution of water quality index in Qa'Jahran basin



Acknowledgements The authors are grateful to the National Water Resources Authority (NWRA) of Yemen for providing the facilities to analysis water samples for this research.

References

- Al-hadithi, M. (2012). Application of water quality index to assess suitability of groundwater quality for drinking purposes in Ratmao-Pathri Rao watershed, Haridwar District, India. *American Journal of Scientific and Industrial Research*, 3(6), 396–402. <https://doi.org/10.5251/ajsir.2012.3.6.395.402>.
- Al-Omran, A. M., Mousa, M. A., AlHarbi, M. M., & Nadeem, M. E. A. (2018). Hydrogeochemical characterization and groundwater quality assessment in Al-Hasa, Saudi Arabia. *Arabian Journal of Geosciences*, 11(4). <https://doi.org/10.1007/s12517-018-3420-y>
- Alberto, W. D., María del Pilar, D., María Valeria, A., Fabiana, P. S., Cecilia, H. A., & María de los Ángeles, B. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Córdoba-Argentina). *Water Research*, 35(12), 2881–2894. [https://doi.org/10.1016/S0043-1354\(00\)00592-3](https://doi.org/10.1016/S0043-1354(00)00592-3)
- Aqeel, A., Al-Amry, A., & Alharbi, O. (2017). Assessment and geospatial distribution mapping of fluoride concentrations in the groundwater of Al-Howban Basin, Taiz-Yemen. *Arabian Journal of Geosciences*, 10(14). <https://doi.org/10.1007/s12517-017-3069-y>
- Bu, H., Tan, X., Li, S., & Safety, Q. Z.-E. (2010). Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. *Elsevier*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0147651309002735>
- Canada, H. (2017). Guidelines for Canadian drinking water quality summary table. Retrieved from <https://www.canada.ca/content/dam/>

- [hc-sc/migration/hc-sc/ewh-semt/alt_formats/pdf/pubs/water-eau/sum_guide-res_recom/sum_guide-res_recom-eng.pdf](https://doi.org/10.1007/s10661-014-3844-0)
- Chowdhury, S., & Al-Zahrani, M. (2014). Water quality change in dam reservoir and shallow aquifer: Analysis on trend, seasonal variability, and data reduction. *Environmental Monitoring and Assessment*, 186(10), 6127–6143. <https://doi.org/10.1007/s10661-014-3844-0>.
- Chowdhury, S., Champagne, P., & Husain, T. (2007). Fuzzy risk-based decision-making approach for selection of drinking water disinfectants. *Water Supply*. <https://doi.org/10.2166/aqua.2007.090>.
- Elhatip, H., Himis, M. A., & Gülbahar, N. (2008). Evaluation of the water quality at Tahtali dam watershed in Izmir-Turkey by means of statistical methodology. *Stochastic Environmental Research and Risk Assessment*, 22(3), 391–400. <https://doi.org/10.1007/s00477-007-0127-0>.
- Eslami, H., Dastorani, J., Javadi, M. R., & Chamheidar, H. (2013). Geostatistical evaluation of ground water quality distribution with GIS (case study: Mianab-Shoushtar plain. *BEPLS Bulletin Environment, Pharmacology and Life Science*, 3(31), 78–82. Retrieved from <https://www.bepls.com/dec2013/15.pdf>
- Gebrehiwot, A. B., Tadesse, N., & Jigar, E. (2011). Application of water quality index to assess suitability of groundwater quality for drinking purposes in Hantebet watershed, Tigray, Northern Ethiopia. *ISABB Journal of Food and Agriculture Science*, 1(1), 22–30. Retrieved from <https://www.isabb.academicjournals.org/JFAS>
- Helena, B., Pardo, R., Vega, M., Barrado, E., & Research, J. F.-W. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Elsevier*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0043135499002250>
- Igboekwe, M. U., & Akankpo, A. O. (2011). Application of Geographic Information System (GIS) in Mapping Groundwater Quality in Uyo, Nigeria. *International Journal of Geosciences*, 02(04), 394–397. <https://doi.org/10.4236/ijg.2011.24042>.
- Jackson, J. E., & John Wiley & Sons. (1991). *A user's guide to principal components*. Wiley. Retrieved from https://books.google.com.sa/books?hl=ar&lr=&id=f9s6g6cmUTUC&oi=fnd&pg=PR7&dq=A+user+guide+to+principal+components&ots=Lcl7kJWoP4&sig=PBRTmtL0IFqY-_xluYIrCLLLKw4&redir_esc=y#v=onepage&q=Auserguidetoprincipalcomponents&f=false
- Lee, J. Y., Cheon, J. Y., Lee, K. K., Lee, S. Y., & Lee, M. H. (2001). Statistical evaluation of geochemical parameter distribution in a groundwater system contaminated with petroleum hydrocarbons. *Journal of Environmental Quality*, 30(5), 1548–1563. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11577860>
- Liu, C.-W., Kao-Hung Lin, Y.-M. K. (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *The Science of the Total Environment*, 313(77–89). Retrieved from https://ac.els-cdn.com/S0048969702006836/1-s2.0-S0048969702006836-main.pdf?_tid=1aed8f8e-e69b-11e7-a470-00000aabb0f02&acdnat=1513894139_fbe8c3f1459c9905ba53e96ed9587261
- Marinović Ruždjak, A., & Ruždjak, D. (2015). Evaluation of river water quality variations using multivariate statistical techniques. *Environmental Monitoring and Assessment*, 187(4), 215. <https://doi.org/10.1007/s10661-015-4393-x>.
- Mohamed, M. M. A., & Hassane, A. B. (2016). Hydrochemistry assessment of groundwater quality in Al-Ain city, UAE. *Environmental Earth Sciences*, 75(4), 1–17. <https://doi.org/10.1007/s12665-015-4942-1>.
- Muangthong, S., & Shrestha, S. (2015). Assessment of surface water quality using multivariate statistical techniques: Case study of the Nampong River and Songkhram River. *Thailand. Environmental Monitoring and Assessment*, 187(9), 548. <https://doi.org/10.1007/s10661-015-4774-1>.
- Nair, H. C., Padmalal, D., Joseph, A., & Gopinthan, V. P. (2018). Hydrogeochemistry and water quality assessment of shallow aquifers in the western flanks of Southern Western Ghats, SW India.
- Nasher, G., Al-Sayyaghi, A., & Al-Matary, A. (2013). Identification and evaluation of the hydrogeochemical processes of the lower part of Wadi Siham catchment area, Tihama plain, Yemen. *Arabian Journal of Geosciences*, 6(6), 2131–2146. <https://doi.org/10.1007/s12517-011-0471-8>.
- Nazzal, Y., Ahmed, I., Al-Arifi, N. S. N., Ghrefat, H., Batayneh, A., Abuamarah, B. A., & Zaidi, F. K. (2015). A combined hydrochemical-statistical analysis of Saq aquifer, northwestern part of the Kingdom of Saudi Arabia. *Geosciences Journal*, 19(1), 145–155. <https://doi.org/10.1007/s12303-014-0016-8>.
- Olsen, R. L., Chappell, R. W., & Loftis, J. C. (2012). Water quality sample collection, data treatment and results presentation for principal components analysis—literature review and Illinois River watershed case study. *Water Research*, 46(9), 3110–3122. <https://doi.org/10.1016/j.watres.2012.03.028>.
- Phung, D., Huang, C., Rutherford, S., Dwirahmadi, F., Chu, C., Wang, X., ... Dinh, T. A. D. (2015). Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam. *Environmental Monitoring and Assessment*, 187(5), 229. <https://doi.org/10.1007/s10661-015-4474-x>
- Qadir, A., Malik, R., Assessment, S. H.-E. M. (2008). Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. *Springer*. Retrieved from <https://www.springerlink.com/index/E13326127137RKR4.pdf>
- Reghunath, R., Murthy, T. R. S., & Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research*, 36, 2437–2442. Retrieved from https://ac.els-cdn.com/S0043135401004900/1-s2.0-S0043135401004900-main.pdf?_tid=9fa18c4c-e8f5-11e7-ae81-00000aabb0f26&acdnat=1514152910_f7dce303e1c01b0339b9b99040034200
- Rice, E. W., Baird, R. B., & Eaton, A.D. (Eds.) (2017). *Standard methods for the examination of water and wastewater* (23rd ed). Retrieved May 6, 2018, from <https://www.awwa.org/store/productdetail.aspx?productid=65266295>
- Saleh, S. M. K., Al-alaiy, S. H. G., & Abdul-razzak, B. (2018). Application of water quality index to assessment of groundwater quality Application of water quality index to assessment of groundwater quality (April 2017).
- Sh Nasher, G. A., & A El-Sagheer, A. A. (2012). Water quality evaluation of Bura'a natural protected area, Hodeidah, Yemen. *Academic & Scholarly Research Journal*, 4. Retrieved from www.aasrc.org/aasrj
- Simeonov, V., Stratis, J., Samara, C., Rresearch, G. Z.-W. (2003). Assessment of the surface water quality in Northern Greece. *Elsevier*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0043135403003981>
- Singh, A. K., Tewary, B. K., & Sinha, A. (2011). Hydrochemistry and quality assessment of groundwater in part of NOIDA metropolitan city, Uttar Pradesh. *Journal of the Geological Society of India*, 78(6), 523–540. <https://doi.org/10.1007/s12594-011-0124-2>.
- Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research*, 38(18), 3980–3992. <https://doi.org/10.1016/j.watres.2004.06.011>.
- Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques— A case study. *Analytica Chimica Acta*, 538, 355–374. <https://doi.org/10.1016/j.aca.2005.02.006>.

- Toumi, N., Hussein, B. H. M., Rafrafi, S., & El kassas, N. (2015). Groundwater quality and hydrochemical properties of Al-Ula Region, Saudi Arabia. *Environmental Monitoring and Assessment*, 187(3), 1–16. <https://doi.org/10.1007/s10661-014-4241-4>
- Varol, M., Gökot, B., Bekleyen, A., & Şen, B. (2012). Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *CATENA*, 92, 11–21. <https://doi.org/10.1016/J.CATENA.2011.11.013>.
- Vega, M., Pardo, R., Barrado, E., & Research, L. D.-W. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Elsevier*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0043135498001389>
- Voza1*, D., Ljiljana Takic2, M. V., Nikolic1, D., & Mladenovic-Ranisavljevic2. I. (2015). Application of multivariate statistical techniques in the water quality assessment of Danube river, Serbia. *Archives of Environmental Protection*, 41, 96–103. <https://doi.org/10.1515/aep-2015-0044>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Tutorial n Chemometrics and Intelligent Laboratory Systems Elsevier Science Publishers B.V*, 2, 37–52. Retrieved from https://ac.els-cdn.com/0169743987800849/1-s2.0-0169743987800849-main.pdf?_tid=5be82b5a-dd10-11e7-a4d8-00000aacb362&acdnat=1512844979_52f5a071600a5e73564aabe03448d36f
- WHO. (2017). *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*. World Health Organization. Geneva, Switzerland. License: CC BY-NC-SA 3.0 IGO.
- Yan, C. A., Zhang, W., Zhang, Z., Liu, Y., Deng, C., & Nie, N. (2015). Assessment of water quality and identification of polluted risky regions based on field observations & GIS in the Honghe River Watershed. *China. PLoS ONE*, 10(3), 1–13. <https://doi.org/10.1371/journal.pone.0119130>.
- Zeilhofer, P., Lima, E. B. N. R., & Lima, G. A. R. (2006). Spatial Patterns of Water Quality in the Cuiabá River Basin. *Central Brazil. Environmental Monitoring and Assessment*, 123(1–3), 41–62. <https://doi.org/10.1007/s10661-005-9114-4>.
- Zeilhofer, P., Zeilhofer, L. V. A. C., Haroim, E. L., de Lima, Z. M., & Oliveira, C. S. (2007). GIS applications for mapping and spatial modeling of urban-use water quality: A case study in District of Cuiabá, Mato Grosso. *Brazil. Cadernos De Saúde Pública*, 23(4), 875–884. <https://doi.org/10.1590/S0102-311X2007000400015>.
- Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q., ... Ye, F. (2009a). Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment*, 152(1–4), 123–131. <https://doi.org/https://doi.org/10.1007/s10661-008-0301-y>
- Zhang, Y., Guo, F., Meng, W., & Wang, X.-Q. (2009). Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. *Environmental Monitoring and Assessment*, 152(1–4), 105–121. <https://doi.org/10.1007/s10661-008-0300-z>.
- Zhao, J., Fu, G., Lei, K., Li, Y., Li, X., Li, J., & Zheng, Y. (2011). Assessment of spatial-temporal patterns of surface water quality in the Min River (China) and implications for management. In *Proceedings—International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, CDCIEM 2011*, 1983–1989. <https://doi.org/10.1109/CDCIEM.2011.499>