Journal of Groundwater Science and Engineering
ISSN 2305-7068
Indexed by ESCI CABI CSA Scopus GeoRef AJ CNKI

Spatial variation assessment of groundwater quality using multivariate statistical analysis(Case Study: Fasa Plain, Iran)

Mehdi Bahrami, Elmira Khaksar, Elahe Khaksar

Citation:
Mehdi Bahrami, Elmira Khaksar, Elahe Khaksar. Spatial variation assessment of groundwater quality using multivariate statistical analysis(Case Study: Fasa Plain, Iran)[J]. *Journal of Groundwater Science and Engineering*, 2020, 8(3): 230-243.

View online: https://doi.org/10.19637/j.cnki.2305-7068.2020.03.004

DOI: 10.19637/j.cnki.2305-7068.2020.03.004

Spatial variation assessment of groundwater quality using multivariate statistical analysis (Case Study: Fasa Plain, Iran)

Mehdi Bahrami[1*], Elmira Khaksar[1], Elahe Khaksar[2]

[1] Department of Water Engineering, Faculty of Agriculture, Fasa University, Fasa 74617-81189, Iran.
[2] Department of Biostatistics and Epidemiology, School of Health, Shahid Sadoughi University of Medical Sciences, Yazd 89169-78477, Iran.

Abstract: Groundwater is considered as one of the most important sources for water supply in Iran. The Fasa Plain in Fars Province, Southern Iran is one of the major areas of wheat production using groundwater for irrigation. A large population also uses local groundwater for drinking purposes. Therefore, in this study, this plain was selected to assess the spatial variability of groundwater quality and also to identify main parameters affecting the water quality using multivariate statistical techniques such as Cluster Analysis (CA), Discriminant Analysis (DA), and Principal Component Analysis (PCA). Water quality data was monitored at 22 different wells, for five years (2009-2014) with 10 water quality parameters. By using cluster analysis, the sampling wells were grouped into two clusters with distinct water qualities at different locations. The Lasso Discriminant Analysis (LDA) technique was used to assess the spatial variability of water quality. Based on the results, all of the variables except sodium absorption ratio (SAR) are effective in the LDA model with all variables affording 92.80% correct assignation to discriminate between the clusters from the primary 10 variables. Principal component (PC) analysis and factor analysis reduced the complex data matrix into two main components, accounting for more than 95.93% of the total variance. The first PC contained the parameters of TH, Ca^{2+}, and Mg^{2+}. Therefore, the first dominant factor was hardness. In the second PC, Cl^{-}, SAR, and Na^{+} were the dominant parameters, which may indicate salinity. The originally acquired factors illustrate natural (existence of geological formations) and anthropogenic (improper disposal of domestic and agricultural wastes) factors which affect the groundwater quality.

Keywords: Groundwater; Iran; Multivariate statistical methods; Pollution

Received: 16 Dec 2019/ Accepted: 22 Mar 2020

Introduction

Groundwater is a basic renewable resource, the conservation of which has been overlooked in many places. Groundwater is considered one of the most important sources for water supply for agriculture, industry, drinking, laboratory, and recreational uses. In arid regions, groundwater oftentimes shows significant spatiotemporal variability. Groundwater is regarded as a limited resource in most countries including Iran. Groundwater pollution has been increasing in recent years, coupled with concerns about the human health and environmental effects of pollutants. For example, chemicals like pesticides, herbicides, and fertilizers which have been used in agriculture may transport into groundwater with rain or irrigation water leaching to the underground (El Alfy and Faraj, 2016). Groundwater quality depends not only on natural variables such as the

*Corresponding author. E-mail: bahrami@fasau.ac.ir

lithology of the aquifer, the quality of recharged water and the type of interaction between surface water and aquifer but also on anthropogenic activities, which can vary these groundwater systems either by contaminating them or by changing the hydrological cycle (Mahmood *et al*. 2011; Matiatos, 2016). Pesticides and fertilizers are subsidized by the Iran government and therefore applied extensively in agriculture in Southern Iran (Zarei and Bahrami, 2016; Amiri *et al*. 2018; Bahrami *et al*. 2018). Irrigation water leaches these chemicals into groundwater aquifers and agricultural run-off. Agricultural wastewater is thus highly contaminated with chemicals that regularly contains heavy metals (Bahrami *et al*. 2013). Unfortunately, polluted water can cause many diseases, particularly in regions with high population density.

Groundwater quality monitoring has been one of the highest priorities in environmental protection policy. It is of great importance and necessity to develop a representative and credible groundwater quality monitoring plan.

Consequently, comprehensive monitoring plans that contain frequent water sampling at abundant sites and consist of a full analysis of a large number of physicochemical variables are to be designed for proper groundwater quality management. Real hydrological data are mostly noisy, which means they are not normally distributed, often co-linear or auto-correlated, and including outliers or errors, *etc*. In order to avoid these noises, multivariate techniques have been applied. The use of multivariable statistical techniques provides a better understanding of water quality when interpreting the complicated datasets (Mahmood *et al*. 2011). Different multi-variate statistical techniques have been applied to discover the relationship among factors and sampling sites, to identify the parameters and sources which influence groundwater quality and to propose practical tools for both water resources management and groundwater quality assessment (Nosrati and Van Den Eeckhaut, 2012; El Alfy *et al*. 2017). The most commonly used techniques are Cluster Analysis (CA), Discriminant Analysis (DA), and Principal Component Analysis (PCA), which were applied in the present study. CA was applied to investigate the spatial grouping of the sampling wells. This method is a usual technique to categorize variables into clusters (Hummel *et al*. 2017). CA and PCA are usually supported by DA as verification and commonly referred to as pattern reconnaissance methods (Azhar *et al*. 2015). Recently, many researchers have applied multivariate statistical techniques to identify the main parameters affecting the water quality of groundwater. For example, Ebrahimzadeh *et al*. (2011) used PCA to evaluate the main variables responsible for the concentration of dissolved ions in the groundwater of Zarghan plain, Iran. The results indicated a strong and positive loading related to Cl, Ca, Mg, Na, Zn and EC, and strong and negative loading related to As Mahmood *et al*. (2011) applied multivariate statistical methods including factor analysis (FA), CA, and DA to assess the spatial variations and the interpretation of the water quality dataset in Punjab, Pakistan. FA shows five factors of salinization, alkalinity, temperature, domestic waste, and chloride, which explained 74% of the total variance in the water quality dataset. Hierarchical cluster analysis grouped nine sampling stations into three clusters, *i.e.* less polluted (LP), moderately polluted (MP), and highly polluted (HP). DA recognized 10 significant factors that discriminate the groundwater quality with close to 100% correct assignment for spatial variations. Noshadi and Ghafourian (2016) investigated the groundwater quality in Fars Province, Iran, using multivariate statistical methods. Cluster analysis resulted in three quality groups in groundwater of the research area. The principal component analysis reduced the complex and voluminous data matrix into three main components, accounting for more than 80% of the total variance. Matiatos *et al*. (2016) used the multivariate statistical methods to identify the main parameters and mechanisms controlling the hydrogeochemistry of groundwater in the deltaic environment of River Pinios (Thessaly). Ghassemi Dehnavi (2018) evaluated the groundwater quality in Aliguodarz, Lorestan, west of Iran using statistical methods. The cluster diagram classified the parameters into six clusters. Interpretation of PCA results indicated that the nitrate in groundwater increased in the area which was caused by the fertilizer leaching.

The Fasa Plain, located in the east of Fars Province, Southern Iran is one of the major areas with wheat cultivation which uses groundwater resources as irrigation water supply. Hence, the principal goal of the present study was to assess the spatial variability of groundwater quality and to identify the main parameters affecting the groundwater quality using multivariate statistical techniques in this plain. Major parameters related to the water quality of the plain were measured and analyzed using the statistical methods.

1 Materials and methods

1.1 Study area

The study area in the present research is the Fasa Plain, a part of Fars Province, Iran, with a total area of 4 196.93 km^2. It is located between 53° 19' to 54° 15' E, and 28° 31' to 29° 24' N with the altitude of 1 370.2 m (Fig. 1). In the 2016 census, the county's population was 205 187 in 61 509 families. According to the De Martonne aridity index, the climate of the Fasa Plain is semi-arid (Bahrami *et al*. 2017). The mean annual temperature is about 20.1 degrees Celsius and the mean annual rainfall is 289 mm. Since the rainfall regime of the Fasa Plain is Mediterranean, rainfall is mostly concentrated in winter months and occasionally in summer months (under the influence of monsoon rains of the Indian Ocean).

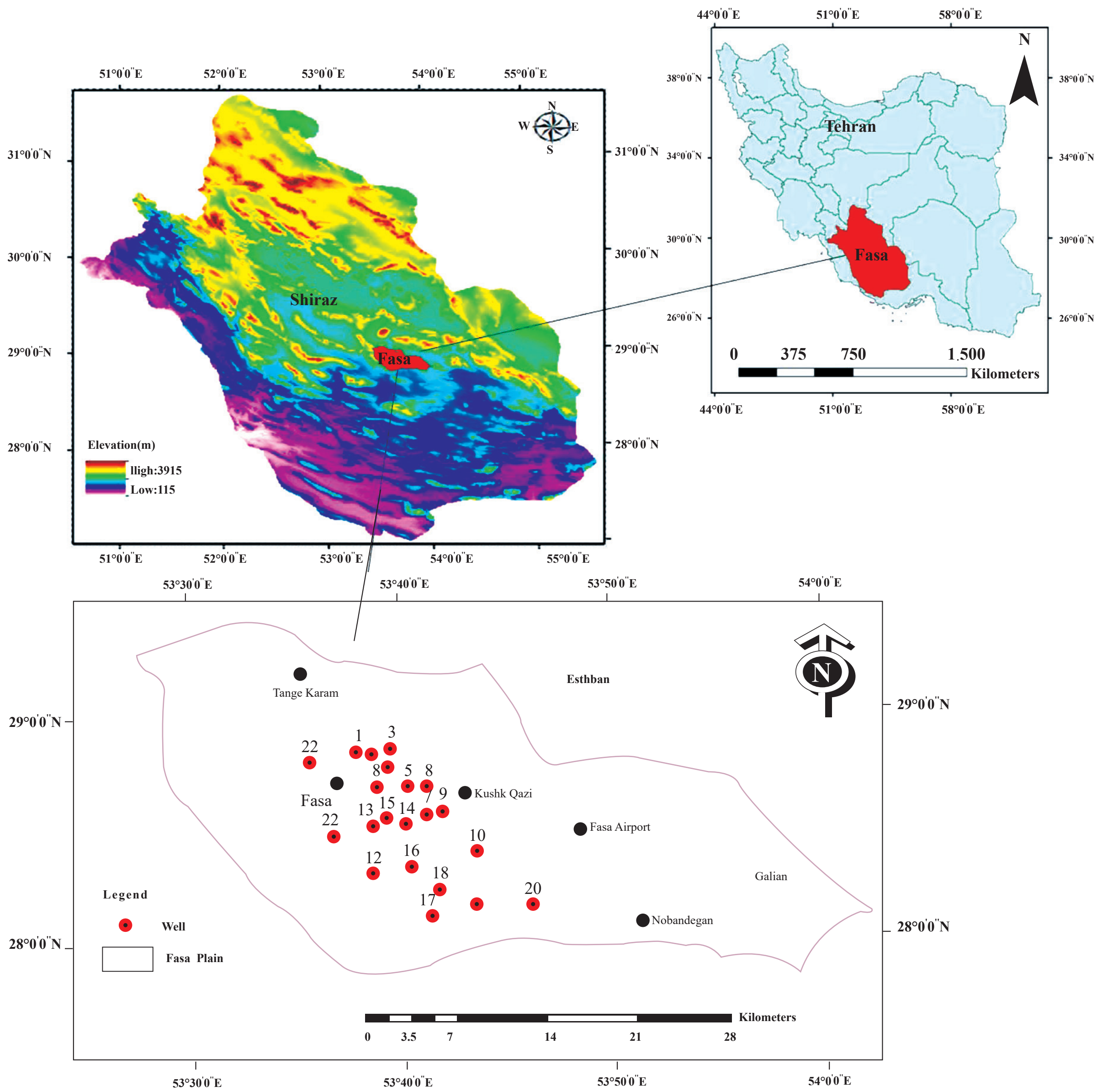


Fig. 1 Location map of the study area showing the sampling wells of the Fasa Plain

Table 1 The ID of sampling wells

ID	Sampling site	ID	Sampling site
1	Jangalkari	12	Qanatno
2	Tonbakan	13	Kamal Abad
3	Toureh	14	Kheir Abad
4	Baniyan	15	Firuzemard
5	Kahnekoyeh	16	Dastjeh
6	Shomal Fasa	17	Baghe Jafari
7	Kushk Qazi 1	18	Sahraroud
8	Kushk Qazi 2	19	Saad Abad
9	Rahmat Abad	20	Ghiyas Abad
10	Harom	21	Soghad
11	Chaghad	22	Cheshme Abnarak

Geologically, Fasa Plain is located in the Zagros Mountains Range that consists of a series of sub-parallel, NW-SE trending anticlines and synclines (Alavi, 2004). The exposed geological formations, in descending order of age, are the Hormuz salt formation (Palaeozoic); the Sarvak limestone formation (Cretaceous); the Pabdeh-Gurpi shales and gypsiferous marl formation (Paleocene Oligocene); the Sachun gypsum formation (Paleocene-Eocene); the Asmari-Jahrom limestone and dolomite formation (Oligocene-Miocene); the Razak evaporite formation (Miocene); the Gachsaran gypsum and marl formation; the Aghajari sandstone formation (late Miocene to Pliocene); the Bakhtiari conglomerate formation (late Pliocene-Pleistocene); and recent alluvium (Fig. 2). The research area is located in the Quaternary alluvial plain. The deposits in the center of the area are mainly sandy loam and silt, while the sediments near the edges are gravel and sand. The aquifer system in highly permeable karstified carbonate rocks often discharges groundwater through springs. The alluvial aquifer in the area is recharged mainly by subsurface groundwater inflows from the adjacent carbonate rocks and by rainfall. In this region, the groundwater generally flows from the north to the south of the plain (Fig. 3).

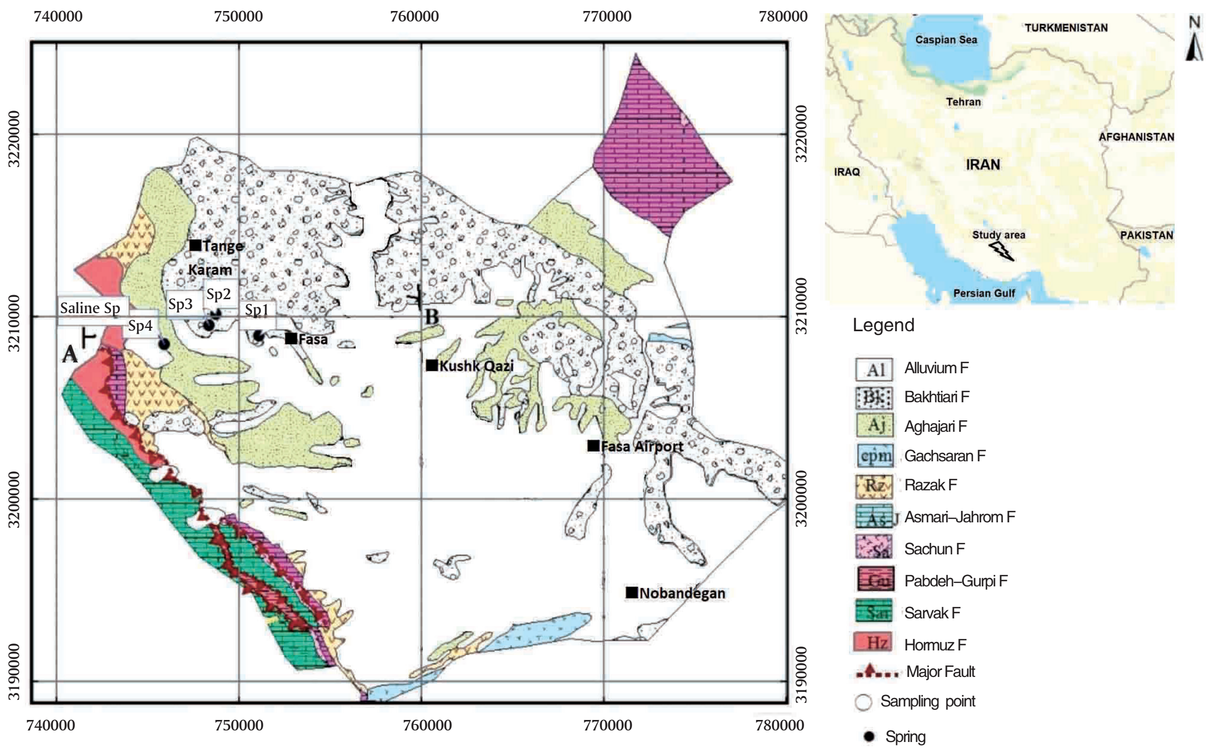


Fig. 2 Geological map of the study area

1.2 Data collection and treatment

The water quality data in the present research was gathered from 22 monitoring wells among the Fasa Plain between October 2009 and September 2014, by the laboratory of water engineering, Fasa University. Fig. 1 represents the map of the study area with the corresponding sampling locations.

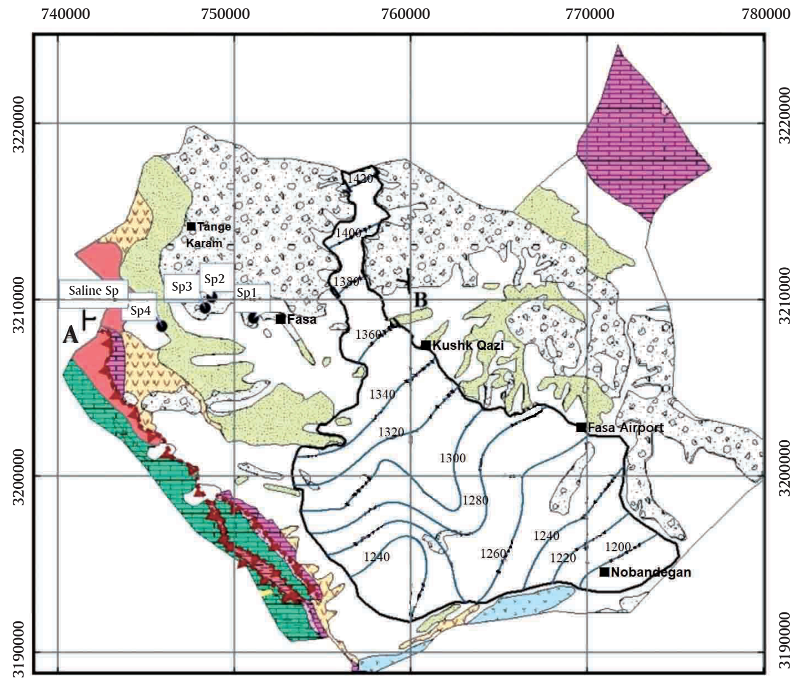


Fig. 3 Iso-Potential map of the study area

For the assessment of groundwater quality, the samples were taken below the water table. Prior to each sampling, the water was pumped for about 15 minutes. Each sample was collected in a 1.5-liter polythene bottle. In the field, each bottle was filled and emptied twice with the water to be sampled before sampling. The sample bottle should be submerged and completely fill without air to mix with the sample until the cap is firmly in place. All the glassware and plastic containers were cleaned with 1 M HNO_3 and rinsed with double distilled water prior to use in order to prevent the contamination of the sample. Sampling, preservation, and transportation of the water samples to the laboratory were as per standard methods (Mahmood *et al*. 2011). All samples were collected in the same fashion and under the specified conditions in a standardized operating procedure. In each of the selected monitoring wells, 10 groundwater parameters including electrical conductivity (EC), total hardness (TH), sodium absorption ratio (SAR), chlorine (Cl^-), sulfate (SO_4^{2-}), calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+), and cations were measured monthly during the entire sampling period. Electrical conductivity was measured in the field immediately after sampling utilizing a digital portable water analyzer (ZX44XL). The electrical conductivity meter was calibrated by immersing the probe in a standard KCl solution (0.1 N). The remaining parameters were determined in the laboratory within 24 h using standard methodologies. Calcium, sodium, and potassium were determined by flame emission photometry (Jenway, UK, PFP7/C), magnesium by atomic absorption spectrophotometer (Analytik-Jena, Germany, Vario 6), chloride by argentometric titration, sulfate by turbidimetric method, and total hardness by EDTA titrations method. Sodium absorption ratio (SAR), described as the relative concentration of sodium to calcium and magnesium (U.S. Salinity Laboratory Staff, 1954), was estimated to ascertain the sodium hazard for irrigation water because of the salinity content.

1.3 Analytical methods

To provide insight into the relationships between the factors, we used multivariate statistical techniques, *i.e*. Cluster Analysis (CA), Discriminant Analysis (DA), and Principal Component Analysis (PCA), to analyze the data. All analysis was performed using Minitab 16 and R 3.3.1 softwares and P-value less than 0.05 was considered significant.

1.3.1 Cluster analysis (CA)

The aim of cluster analysis is to classify objects into two or more groups based on the similarity between objects with respect to a set of special characteristics (Matiatos and Evelpidou, 2013; Rogerson, 2001; McKenna, 2003; Hardle and Simar, 2007; Lokhande *et al*. 2008).

In the present research, to classify the sampling wells based on the proximity of variables, hierarchical cluster analysis was carried out by applying Ward's technique (Pandit and Gupta, 2011; Singh *et al*. 2013). Independent T-test was utilized to compare the clusters and to determine which one is more prominent.

1.3.2 Discriminant analysis (DA)

Discriminant analysis (DA) is utilized to categorize instances into categorical-dependent values, commonly a dichotomy. If the discriminant analysis is effective for a set of data, the categorization table of correct and incorrect approximations will result in a high correct percentage. In this study, the Lasso discriminant analysis (LDA) was used to verify the CA results. This method manipulates raw data and builds up a discriminant function for each category (Singh *et al*. 2005; Witten and Tibshirani, 2011).

1.3.3 Principal component analysis (PCA)

Principal component analysis (PCA) is extensively used for data reduction in hydrochemical and hydrogeological researches (Matiatos, 2016). This method is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear compositions of the primary variables. In this study, the PCA method was used to categorize the wells based on the correlation of the variables (Sarbu and Pop, 2005). PC prepares information on the most significant parameters that explain a whole data set affording data decrease with minimum loss of primary information (Helena *et al*. 2000).

2 Results

2.1 Descriptive statistics

Basic statistics were performed to provide basic information on water quality data. The descriptive statistics on the water quality variables sampled in five years are shown in Table 2. Results indicated that mean values of EC and TH (1 441.43 µmhos/cm and 570.79 ppm as $CaCO_3$, respectively) in the study area are a little higher than the permissible limits (1 400 µmhos/cm and 500 ppm as $CaCO_3$, respectively) suggested by the Institute of Standards and Industrial Research of Iran (ISIRI, 2009). On the other hand, the mean concentrations of Cl^-, K^+, Na^+, Mg^{2+}, Ca^{2+}, and SO_4^{2-} (4.89, 0.07, 3.77, 5.30, 6.11, and 5.47 ppm, respectively) are much lower than the permissible limits (400, 12, 200, 30, 300, and 400 ppm, respectively). Large standard deviations of most of the variables indicated their randomly fluctuating concentration levels in the groundwater.

Table 2 Descriptive statistics of groundwater quality parameters ranges and their comparison with the Iranian standard for drinking water

Variable	Iranian permissible limit	N	Minimum	Maximum	Mean	Std. Deviation
EC (µmhos/cm)	1 400	110	333	52 00	1 441.43	843.99
Cl^- (ppm)	400	110	0.15	22.50	4.89	3.73
TH (ppm as $CaCO_3$)	500	110	166	2 000	570.79	367.80
SAR (-)	-	110	0.11	4.17	1.56	0.92
K^+ (ppm)	12	110	0.01	0.32	0.07	0.05
Na^+ (ppm)	200	110	0.15	17.39	3.77	2.80
Mg^{2+} (ppm)	30	110	0.20	22.50	5.30	3.85
Ca^{2+} (ppm)	300	110	2.00	27.00	6.11	4.44
Cations (ppm)	-	110	3.72	57.66	15.26	9.40
SO_4^{2-} (ppm)	400	110	0.19	33.92	5.47	6.14
Valid N (listwise)		110				

2.2 Cluster analysis

CA was performed on the water quality dataset to estimate the spatial variability amongst the monitoring wells of the Fasa Plain. The sampling wells were categorized into two clusters as represented in Fig. 4. The dendrogram shows the clustering of monitoring wells according to groundwater quality specifications of the Fasa Plain.

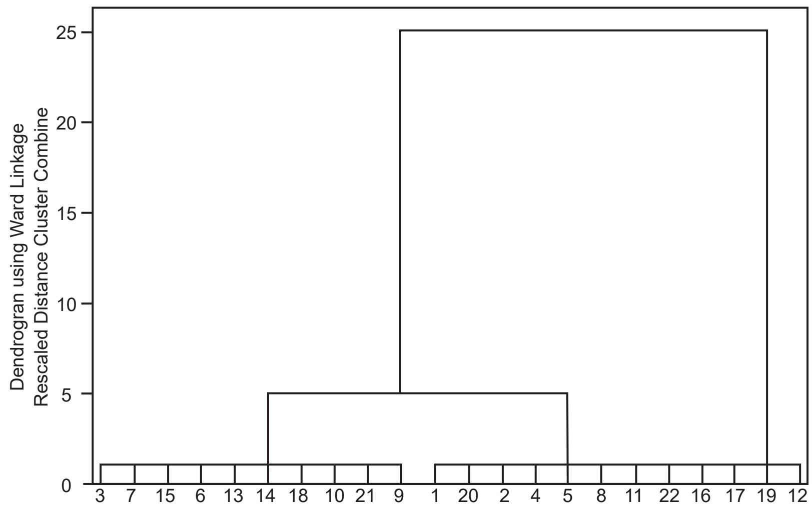


Fig. 4 A Dendrogram representing the two groups of the cluster

Cluster 1 contains eighteen wells (Toureh, Kushk Qazi 1, Shomal Fasa, Kamal Abad, Kheir Abad, Firuzemard, Sahraroud, Harom, Soghad, Rahmat Abad, Jangalkari, Ghiyas Abad, Tonbakan, Baniyan, Kahnekoyeh, Kushk Qazi 2, Chaghad, and Cheshme Abnarak), classified as less polluted (LP) wells due to the lowest average value. Cluster 2 contains four wells (Dastjeh, Baghe Jafari, Saad Abad, and Qanatno), classified as highly polluted (HP) wells. As indicated in the results, the wells of Cluster 2 (No. 12, 16, 17, and 19) are located in the southern part of the plain, while the wells of Sahraroud and Ghiyas Abad (No. 18 and 20) are in the same region but in Cluster 1. This is due to different well depths, which may cause the recharge water of the wells from different layers of geological formations.

The mean values of water quality parameters confirm the division of wells into two groups (Table 3). The values of most parameters were increased from the first to the second group and indicated the degradation of groundwater quality (Table 3). According to Table 3, the values of all studied parameters are increased from Group 1 to 2.

According to Table 4, the means of all variables of the HP cluster except SAR are significantly higher than those of the LP cluster.

Table 3 Spatial clustering of sampling wells

Groups	EC	Cl^-	TH	SAR	K^+	Na^+	Mg^{2+}	Ca^{2+}	Cations	SO_4^{2-}	No. of wells
1	1 154.7	4.0	445.4	1.4	0.06	3.1	4.1	4.8	12.1	3.4	18
2	2 731.8	9.1	1 135	2.1	0.14	6.8	10.7	12.0	29.7	14.6	4

Table 4 Independent sample test

Levene's test for equality of variances				t-test for equality of means				
	Variable	F	sig.	t	df	Sig. (2-tailed)	Mean difference (LP-HP)	Std. error difference
EC	Equal variances assumed	4.215	0.053	-8.462	20	0.000	-1 577.222	186.374
Cl^-	Equal variances assumed	1.047	0.318	-4.777	20	0.000	-5.122	1.072
TH	Equal variances assumed	0.547	0.468	-8.305	20	0.000	-689.589	83.031

Table 4 (Continued)

	Levene's test for equality of variances			t-test for equality of means				
	Variable	F	sig.	t	df	Sig. (2-tailed)	Mean difference (LP-HP)	Std. error difference
SAR	Equal variances not assumed	7.445	0.013	-1.219	3.380	0.150	-0.626	0.514
K^+	Equal variances assumed	0.079	0.782	-7.410	20	0.000	-0.077	-0.010
Na^+	Equal variances not assumed	9.556	0.006	-2.570	3.293	0.037	-3.750	1.460
Mg^{2+}	Equal variances assumed	0.862	0.364	-6.679	20	0.000	-6.595	0.987
Ca^{2+}	Equal variances not assumed	19.526	0.000	-2.545	3.073	0.041	-7.193	2.826
Cations	Equal variances assumed	2.731	0.114	-9.022	20	0.000	-17.613	1.952
SO_4^{2-}	Equal variances not assumed	10.417	0.004	-3.891	3.138	0.014	-11.186	2.875

The alluvial aquifer in this area is near the Salloo diapir (Fig. 5). The saline water of the salt dome, which is at a shallower level than the groundwater below, probably flows through faults in the southwest of the area to increase the salinity in the study area. Besides, the dominant groundwater flow can transport these properties from north to south of the Plain.

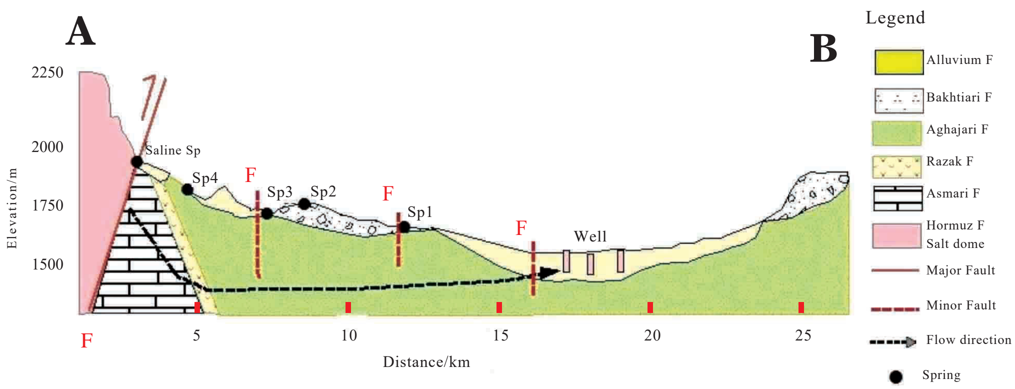
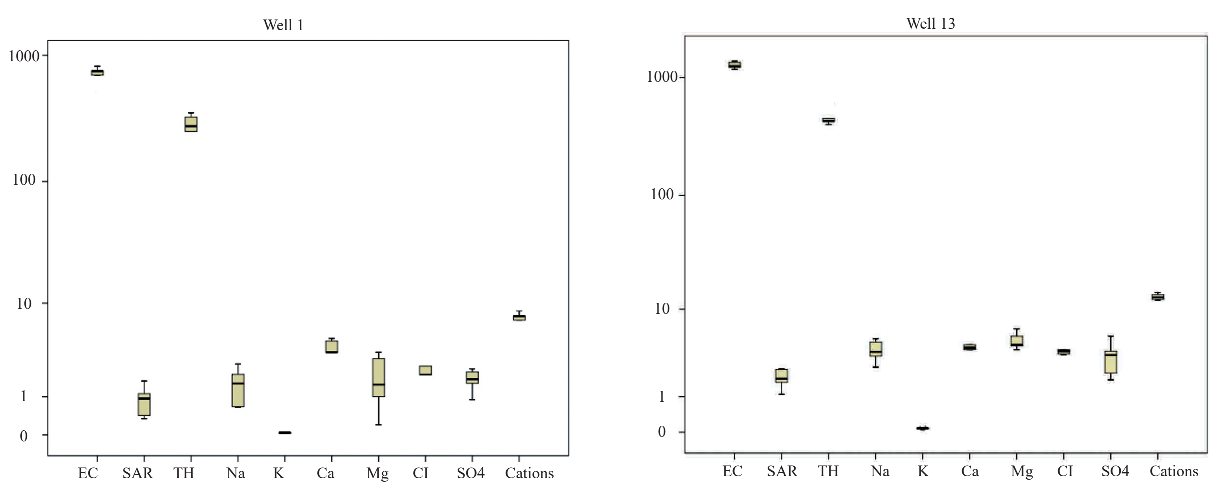


Fig. 5 Schematic hydrogeological cross section of the study area (Bagheri R *et al*. 2017)

To further investigate the groundwater quality in studied wells, the variability of all parameters for two wells out of each Cluster (Wells 1 and 13 for Cluster 1 and Wells 12 and 16 for Cluster 2) was presented in Fig. 6. The results showed higher values of the quality parameters in wells of Cluster 1 (northern and central areas of the Plain) than in Cluster 2 (southern areas of the Plain).



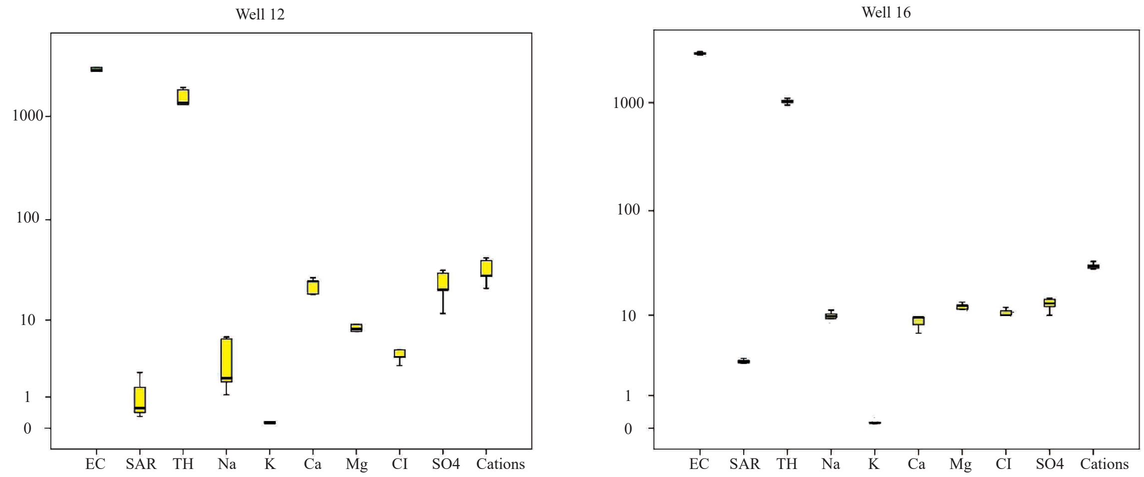


Fig. 6 Box plots of groundwater quality parameters for typical wells of two clusters

2.3 Discriminant analysis

One of the main assumptions of the DA method is the absence of a linear relationship between independent variables. Due to a high correlation between the independent variables and the linear relationship establishment, the DA method cannot be used. Therefore, another technique called Lasso Discriminant Analysis (LDA) was used, which has solved the linear problem as well as appropriate for the available data. In this method, the value of Lambda with the lowest cross validation error is 0.237 and the coefficient values were calculated with the contribution of this Lambda value. The resulted coefficients are tabulated in Table 5.

Table 5 Resulted coefficients of LDA method

Variable	EC	Cl^-	TH	SAR	K^+	Na^+	Mg^{2+}	Ca^{2+}	Cations	SO_4^{2-}
LDA coefficient	-0.423	-0.215	-0.388	0.000	-0.297	-0.144	-0.351	-0.298	-0.430	-0.345

After the classification into two major clusters by utilizing CA, the LDA method was applied based on the original data of 10 parameters in order to assess the spatial variation of groundwater quality amongst studied wells. Lasso discriminant functions and classification matrices acquired from the standard mode of LDA was 92.80%. The LDA results were presented in Table 6. Based on these results, 95.6% and 90% of groundwater samples were classified correctly in the two predetermined groups of Cluster 1 and Cluster 2, respectively.

Table 6 Classification matrix obtained from LDA of spatial variation of the groundwater in the Fasa Plain

	Predicted cluster determined by LDA	
Actual cluster	Cluster 1	Cluster 2
Cluster 1	95.60	4.40
Cluster 2	10.00	90.00
Total accuracy	92.80	

2.4 Principal component analysis (PCA)/ factor analysis (FA)

The major goal of factor analysis is to decrease the quota of less significant parameters by further simplifying the data structure eventuating from the PCA method. PCA is planned to transform the primary parameters into new, uncorrelated parameters, called the principal ingredient, which is derived from linear combinations of the primary parameters. Factor loadings are the simple correlations between the groundwater quality parameters and each factor.

At first, the significance of data in the matrix and the adequacy of the data for performing factor analysis were tested by using Kaiser-Meyer-Olkin (KMO) and Bartlett's tests, respectively. The KMO index more than 0.6 is indicant of the adequacy of data and the significance of Chi-Square of Bartlett's test is the minimum required condition for factor analysis. The results of KMO and Bartlett's tests were tabulated in Table 7.

Table 7 Results of KMO and Bartlett's tests

Kaiser-Meyer-Olkin measure of sampling adequacy		0.623
Bartlett's Test of Sphericity	Approx. Chi-Square	919.709
	df	45
	Sig.	0.000

According to Table 7, the resulted KMO index of 0.623 (>0.6) indicated the sampling adequacy for factor analysis. In addition, the significance of the data in a matrix is determined through the Bartlett Chi-Square test. The null hypothesis rejection implies that the correlation matrix has significant information and there is the minimum required condition for factor analysis. With respect to P-Value of 0.000, it can be concluded that there is the minimum required condition for factor analysis.

Table 8 represents the percentage of total variance explained with the components. Based on the Kaiser criterion, two main factors were extracted for varimax rotation (Srivastava and Ramanathan, 2008). As can be seen in Table 8, the first and second components accounted for 79.60% and 16.33 % of the total variance, respectively, the combination of which is 95.93% of the total variance. But the important point is to realize which variables have high loadings in each component.

In Table 9, variables with high loadings are marked in bold. In the first principal component (PC1), Ca^{2+}, SO_4^{2-}, TH, K^+, cations, EC, and Mg^{2+} parameters have high positive loadings, respectively. PC2 is originally participated by trace elements of SAR, Na^+, and Cl^-, respectively.

To evaluate the relationship between the variables further, the correlation matrix between the factors was formed (Table 10).

Table 8 Total variance explained with two principal components

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
1	7.960	79.600	79.600	7.960	79.600	79.600	5.839	58.393	58.393
2	1.633	16.327	95.928	1.633	16.327	95.928	3.753	37.534	95.928
3	0.284	2.837	98.765						
4	0.084	0.842	99.607						
5	0.010	0.253	99.860						
6	0.004	0.098	99.958						
7	0.017	0.035	99.993						
8	0.001	0.007	100.000						
9	2.774E-6	2.774E-5	100.000						
10	3.456E-9	3.456E-8	100.000						

Extraction method: Principal component analysis

Table 9 Rotated component matrixa

	Component	
Variable	1	2
EC (µmhos/cm)	0.828	0.559

Table 9 (Continued)

	Component	
Variable	1	2
Cl^- (ppm)	0.516	0.799
TH (ppm as $CaCO_3$)	0.943	0.329
SAR	0.013	0.954
K^+ (ppm)	0.883	0.445
Na^+ (ppm)	0.339	0.932
Mg^{2+} (ppm)	0.697	0.638
Ca^{2+} (ppm)	0.976	0.008
Cations (ppm)	0.851	0.525
SO_4^{2-} (ppm)	0.970	0.187

Extraction method: Principal component analysis

Rotation method: Varimax with Kaiser Normalizationa

a. Rotation converged in 3 iterations.

Table 10 Correlation matrixa of studied variables

Variable	EC	Cl^-	TH	SAR	K^+	Na^+	Mg^{2+}	Ca^{2+}	Cations	SO_4^{2-}
EC	1.000									
Cl^-	0.878	1.000								
TH	0.966	0.757	1.000							
SAR	0.535	0.694	0.317	1.000						
K^+	0.978	0.787	0.977	0.462	1.000					
Na^+	0.798	0.903	0.622	0.908	0.713	1.000				
Mg^{2+}	0.946	0.908	0.879	0.542	0.883	0.804	1.000			
Ca^{2+}	0.804	0.490	0.917	0.068	0.875	0.354	0.616	1.000		
Cations	0.998	0.860	0.975	0.509	0.982	0.779	0.930	0.833	1.000	
SO_4^{2-}	0.904	0.609	0.972	0.226	0.944	0.514	0.771	0.961	0.923	1.000

a. Determinant = 1.87E-024

3 Discussion and conclusions

Two clusters obtained from the cluster analysis indicate higher values of water quality parameters in the southern areas of the plain, while the wells located in the northern and center of the plain had lower values of these parameters. Such variation may be attributed to the natural hydrogeological environment and the multipurpose nature of land use in the research area. There are various types of water in the groundwater which is natural because the integration of the chemical composition with groundwater is due to the interaction of groundwater with geology (Noshadi and Ghafourian, 2016). Matiatos and Evelpidou (2013) reported that high SAR values indicate the saline water which promotes soil dispersion and hardening and reduces infiltration rates. Also, Freeze and Cherry (1979) indicated that when surface water is charged with atmospheric and biogenic CO_2 and infiltrates into the soil, CO_2 attacks the aluminosilicates including micas and feldspars, causing liberation of Ca^{2+} and Mg^{2+} cations into the water. Usman *et al*. (2014) proposed that the multipurpose nature of land use and their effects on groundwater quality prevent the exact spatial categorization of monitoring sampling wells. The result showed that for quick assessment of groundwater quality, only one well in each cluster is required to illustrate a rational and correct spatial evaluation of the water quality for the entire cluster. In other words, the CA separation eliminates the requirement for sampling from multiple areas and makes it possible to monitor two areas for assessment of water quality and pollution.

The Lasso Discriminant Analysis that tests the CA capability in clustering the sampling sites, identified EC, Cl^-, TH, K^+, Na^+, Mg^{2+}, Ca^{2+}, cations, and SO_4^{2-} as the most important parameters discriminating between two clusters. Generally, the total accuracy of the LDA method in the prediction of included wells in designated clusters was 92.8%. A similar result was reported by Matiatos *et al*. (2014) in the multivariate statistical analysis of the hydrogeochemical and isotopic composition of the groundwater resources in northeastern Peloponnesus (Greece). Therefore, LDA is a technique, which can verify the classification into predetermined groups.

Identification of the root of each water quality variable using the principal component analysis method resulted in two varimax factors (VFs) accounting for 95.93% of the total variance in the data set. According to high loadings for TH, Ca^{2+}, and Mg^{2+} in PC1, the first factor would be hardness. Also, based on the high loadings for Cl^-, SAR, and Na^+ in PC2, the second factor would be salinity. Bencer *et al*. (2016) suggested that when these parameters are accumulated in one factor, it reflects the impact of natural factors like the dissolution of carbonate and dolomitic. The evaluation of the correlation matrix between the factors revealed that: Positive correlation of Cl^- with Na^+ and Mg^{2+} indicates the dominance of these soluble salts in these samples. A strong correlation between Cl^- and Na^+ ($r = 0.90$) ions showed that the existence of Na^+ ions in water is due to the dissolution of halite by water (Yidana, 2010). However, Na^+ and Cl^- ions in the groundwater can originate from several sources. The high positive correlation of EC with ions of K^+, Mg^{2+}, and SO_4^{2-} showed the high mobility of ions. Moreover, the strong correlation between EC and TH ($r = 0.97$) and cations ($r = 0.998$) can be due to ions in TH and especially cations that conduct electricity. The strong correlation between Mg^{2+} and Cl^- ($r = 0.91$) can be as a result of domestic wastewater entering groundwater in this area since municipal wastes of salts and soaps contain $MgCl_2$. The strong correlation of SO_4^{2-} and Ca^{2+} is indicative of the existence of gypsum in the groundwater that is related to the formations of the plain. The dominant formations in the plain are Bakhtiari and Aghajari, which are Zagros geological formations. The Bakhtiari formation features alluvial and foothill sediments derived from altitude erosion, including conglomerates and calcareous sandstones. Aghajari formation is composed of brown-grey limestone sandstones, gypsum veins, red marl, and siltstone. Also, a strong correlation of TH with K^+, Ca^{2+}, and SO_4^{2-} is a demonstrator of the permanent hardness of the water. The strong and positive correlation between ions of K^+ and SO_4^{2-} ($r = 0.94$) may result from the application of chemical fertilizers in agricultural activities in the study area.

These results are consistent with the ones reported by several other studies. Mozafarizadeh and Sajadi (2013) studied the reasons for high salinity and intrusion of saltwater of Dalaki and Helleh rivers into Borazjan aquifer, southern Iran. Also, the results of Dehghani *et al*. (2015) indicated that overexploitation, dissolution of dolomite, and gypsum deposits are some of the main causes of high salinity (TDS above 1 000 mg/L) of groundwater for the Seyed Gholi Region, Saveh, Iran. Noshadi and Ghafourian (2016) showed that three major variables influencing the quality of groundwater in Fars Province within a 10-year period were salinity, weathering of silicate minerals, and improper disposal of domestic wastes or the use of chemical fertilizers in agriculture. Finally, the results of this research can be attributed to the type of geological formations in the study area as well as the agricultural and domestic wastewaters entering groundwater.

Acknowledgements

The authors would like to thank the Laboratory of Water Engineering, Fasa University for providing the facilities to perform this research.

References

Alavi M. 2004. Regional stratigraphy of the Zagros fold-thrust belt of Iran and its pro-foreland evolution. American Journal of Science, 304: 1-20.

Amiri MJ, Bahrami M, Beigzadeh B, *et al*. 2018. A response surface methodology for optimization of 2, 4-dichlorophenoxyacetic acid removal from synthetic and drainage water: A comparative study. Environmental Science and Pollution Research, 25(34): 34277-34293.

Azhar SC, Aris AZ, Yusoff MK, *et al*. 2015. Classification of river water quality using multivariate analysis. Procedia Environmental Sciences, 30: 79-84.

Bagheri R, Bagheri F, Eggenkamp HGM. 2017. Origin of groundwater salinity in the Fasa Plain, southern Iran, hydrogeochemical and isotopic approaches. Environmental Earth Sciences, 76: 662.

Bahrami M, Amiri MJ, Beigzadeh B. 2018. Adsorption of 2, 4-dichlorophenoxyacetic acid using rice husk biochar, granular activated carbon, and multi-walled carbon nanotubes in a fixed bed column system. Water Science and Technology, 78(8): 1812-1821.

Bahrami M, Brumand-Nasab S, Kashkooli HA, *et al*. 2013. Cadmium removal from aqueous solutions using modified magnetite nanoparticles. Iranian Journal of Health and Environment, 6(2): 221-232.

Bahrami M, Zarei AR, Chakav S. 2017. Analysis of drought transitions using log-linear models in Iran. International Journal of Water, 11(3): 266-278.

Bencer S, Boudoukha A, Mouni L. 2016. Multivariate statistical analysis of the groundwater of Ain Djacer area (Eastern of Algeria). Arabian Journal of Geosciences, 9(4): 1-10.

Dehghani R, Mahvi AH, Rabani D, *et al*. 2015. Evaluation of chemical quality and salinity origin of groundwater in a semi aried area; Seyed Gholi region Saveh, Iran. Archives of Hygiene Sciences, 4(2): 100-108.

Ebrahimzadeh S, Boustani F, Shakeri A. 2011. Groundwater quality assessment of the Zarghan Plain, Shiraz, Iran. 2nd International Conference on Environmental Science and Technology IPCBEE, 6. Singapore: IACSIT Press.

El Alfy M, Faraj T. 2016. Spatial distribution and health risk assessment for groundwater contamination from intensive pesticide use in arid areas. Environmental Geochemistry and Health, 39: 231-253.

El Alfy M, Lashin A, Abdalla F, *et al*. 2017. Assessing the hydrogeochemical processes affecting groundwater pollution in arid areas using an integration of geochemical equilibrium and multivariate statistical techniques. Environmental Pollution, 229: 760-770.

Freeze RA, Cherry JA. 1979. Groundwater. Newjersey: Prentice-Hall, inc: 604.

Ghassemi Dehnavi A. 2018. Hydrochemical assessment of groundwater using statistical methods and ionic ratios in Aliguodarz, Lorestan, west of Iran. Journal of Advances in Environmental Health Research, 6: 193-201.

Hardle W, Simar L. 2007. Applied multivariate statistical analysis, 2nd edn. Berlin: Springer.

Helena B, Pardo R, Vega M, *et al*. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Research, 34(3): 807-816.

Hummel M, Edelmann D, Kopp-Schneider A. 2017. Clustering of samples and variables with mixed-type data. PLOS ONE, 12(11):e0188274. https://doi.org/10.1371/journal.pone.0188274

Institute of Standards and Industrial Research of Iran. 2009. Drinking water-physical and chemical specifications. ISIRI No. 1053, the 5th Revision.
http://www.isiri.org/Portal/Home/Default.aspx?CategoryID=5f6bbf1b-ac23-4362-a309-9ee95a439628

Lokhande PB, Patil VV, Mujawar HA. 2008. Multivariate statistical analysis of groundwater in the vicinity of Mahad industrial area of Konkan Region, India. International Journal of Applied Environmental Sciences, 3(2): 149-163.

Mahmood A, Muqbool W, Mumtaz MW, *et al*. 2011. Application of multivariate statistical techniques for the characterization of groundwater quality of Lahore, Gujranwala and Sialkot (Pakistan). Pakistan Journal of Analytical & Environmental Chemistry, 12(1): 102-112.

Matiatos I, Alexopoulos A, Godelitsas A. 2014. Multivariate statistical analysis of the hydrogeochemical and isotopic composition of the groundwater resources in northeastern Peloponnesus (Greece). Science of the Total

Environment: 476-477, 577-590.

Matiatos I. 2016. Nitrate source identification in groundwater of multiple land-use areas by combining isotopes and multivariate statistical analysis: A case study of Asopos basin (Central Greece). Science of the Total Environment, 541: 802-814.

Matiatos I, Evelpidou N. 2013. Assessment of groundwater quality contamination by nitrate leaching using multivariate statistics and Geographic Information Systems. Understanding freshwater quality problems in a changing world, 361: 183-190.

Matiatos I, Paraskevopoulou V, Botsou F, *et al*. 2016. Hydrogeochemical assessment of groundwater quality in a river delta using multivariate statistical techniques. EGU General Assembly Conference Abstracts, 18: 14568.

McKenna Jr JE. 2003. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling and Software, 18(3): 205-220.

Mozafarizadeh J, Sajadi Z. 2013. Investigation of saline water intrusion in the Borazjan freshwater aquifer from the Dalaki and Helleh rivers. Journal of Electromagnetic and Application, 6(16): 69-78.

Noshadi M, Ghafourian A. 2016. Groundwater quality analysis using multivariate statistical techniques (Case study: Fars Province, Iran). Environmental Monitoring and Assessment, 188: 1-13.

Nosrati K, Van Den Eeckhaut M. 2012. Assessment of groundwater quality using multivariate statistical techniques in the Hashtgerd Plain, Iran. Environmental Earth Sciences, 65(1): 331-344.

Pandit S, Gupta S. 2011. A comparative study on distance measuring approaches for clustering. International Journal of Computer Science, 2(1): 29-31.

Rogerson PA. 2001. Statistical methods for geography. London: Sage Publications Ltd.

Sarbu C, Pop HF. 2005. Principal component analysis versus fuzzy principal component analysis. A case study: The quality of Danube water (1985-1996). Talanta, 65: 1215-1220.

Singh KP, 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(1-2): 355-374.

Singh A, Yadav A, Rana A. 2013. K-means with three different distance metrics. International Journal of Computer Applications, 67(10): 13-17.

Srivastava SK, Ramanathan AL. 2008. Geochemical assessment of groundwater quality in vicinity of Bhalswa landfill, Delhi, India, using graphical and multivariate statistical methods. Environmental Geology (Berlin), 53: 1509-1528.

US Salinity Laboratory Staff. 1954. Diagnosis and improvement of saline and alkali soils. Washington DC: US Department of Agriculture Handbook, 60: 160.

Usman UN, Toriman ME, Juahir H, *et al*. 2014. Assessment of groundwater quality using multivariate statistical techniques in Terengganu. Science and Technology, 4(3):42-49. DOI: 10.5923/j.scit.20140403.02

Witten DM, Tibshirani R. 2011. Penalized classification using Fisher's linear discriminant. Journal of the Royal Statistical Society, 73(5): 753-772.

Yidana SM. 2010. Groundwater classification using multivariate statistical methods: Birimian Basin, Ghana. Journal of Environmental Engineering, 136(12): 1379-1388.

Zarei AR, Bahrami M. 2016. Evaluation of quality and quantity changes of undergroundwater in the Fasa Plain, Fars (2006-2013). Iranian Journal of Irrigation & Water Engineering, 6(24): 103-113.