

Geochemical influences on the quality of groundwater in eastern part of Semnan Province, Iran

Mohamad Sakizadeh¹ · Rouhollah Mirzaei² · Hadi Ghorbani³

Received: 12 February 2015 / Accepted: 7 May 2016 / Published online: 20 May 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract To identify the principal geochemical processes controlling groundwater quality in eastern part of Semnan Province, a total of 257 groundwater samples from wells and springs were collected. Since groundwater is the only source of water for various purposes, preservation of the quality of available groundwater resources in this region is of great importance. The order of abundance of anions was $\text{Cl}^- > \text{SO}_4^{2-} > \text{HCO}_3^- > \text{NO}_3^- > \text{F}^-$, and for them the average values were 509.41, 404.24, 120.48, 14.51, 0.59 mg/l, respectively. On the other hand, for cations, Na^+ and Ca^{2+} with average values of 442.10 and 135.09 mg/l were the most abundant parameters. Geochemical characteristics based on bivariate diagrams indicated that Na release from silicate weathering is an important process whereas the dissolution of dolomite and the effect of silicate materials on the aquifer were other determinative processes influencing the quality of groundwater in the area. In order to predict the overall quality of groundwater with respect to a calculated water quality index (WQI), support vector machine (SVM) and artificial neural network (ANN) were utilized. The correlation coefficients between the predicted and observed WQI for the training and test data set related to SVM were 0.97 and 0.96, respectively. On the contrary, the results of ANNs with early stopping showed roughly the same performance resulting in correlation coefficients of 0.98 and 0.96 for the training and test data set,

respectively. As a whole, the WQI for 8 % of the sampling stations was <50 and for 17 %, the calculated index was <60 which are important concerns.

Keywords Geochemical processes · Groundwater · Support vector machine · Artificial neural network

Introduction

More than 90 % of the Iran's area is located in arid and semiarid regions. Overpopulation in recent decade has diverted the groundwater consumption from agricultural use to industrial and drinking purposes. At present, about 55 % of the water consumed in Iran is provided from groundwater resources (Zehtabian et al. 2010). High rate of evaporation in addition to low rain fall in arid and semiarid area can lead to groundwater salinity (Umar and Absar 2003), a phenomenon which has been intensified in recent years in Iran. In the absence of major anthropogenic sources, water-rock interaction is the main process that affects the groundwater chemistry in an area (Corteel et al. 2005). These geochemical processes are responsible for the spatial-temporal variations in groundwater chemistry. There have been multiple studies on the impacts of geochemical processes on the quality of groundwater in different parts of Iran (e.g., Taheri Tizro and Voudouris 2007; Jalali 2009; Aghazadeh and Mogaddam 2011). In this respect, Panno et al. (2002) compared geochemical and isotopic techniques for identifying the natural and anthropogenic sources of Na and Cl contamination in groundwater and surface water resources in the Midwestern USA. Graphing techniques were used to discriminate among multiple sources and between unaffected and affected surface and groundwater. These graphical techniques were

✉ Mohamad Sakizadeh
msakizadeh@gmail.com

¹ Department of Environmental Sciences, Faculty of Sciences, Shahid Rajaei Teacher Training University, Tehran, Iran

² Department of Environmental Sciences, University of Kashan, Kashan, Iran

³ Department of Water and soil, University of Shahrood, Shahrood, Iran

useful for distinguishing the sources of Na^+ and Cl^- contamination in groundwater in Illinois and can be more widely applicable in other places. Hydrogeochemical characteristics based on bivariate diagrams of major and minor ions were also utilized by Kim et al. (2003) and showed that changes in the chemical composition of groundwater are mainly controlled by the salinization process followed by cation exchange reactions.

Semnan Province is one of the provinces in central parts of Iran that stretches along the Alborz mountain range and borders to Dasht-e Kavir desert in its southern parts (Mirhosseini et al. 2011). The main lithologic units of this area are ophiolitic complex accompanied by Eocene–Oligocene volcanoclastic and basic rock units. The ophiolitic complex is mostly dominated by serpentinized harzburgite and dunite, which is considered as the main body of ultramafic rocks of ophiolitic zones (Hajizadeh et al. 2011). Moreover, cretaceous carbonates are other dominant formations in the region (Bazargani-Guilani et al. 2010).

The main cities in the study area are Shahrood, Damghan and Byar. Considering the chemical composition and ion chemistry of groundwater in the study area, high Ca, Mg, and pH and higher-than-recommended hardness values were detected in the groundwater of Shahrood City and its nearby area reflecting the influence of carbonate rock formations on the groundwater composition (Kazemi 2004). In this respect, the limestone/dolomite has mainly influenced the hardness of the water whereas the samples for which the pH was measured are supersaturated with respect to calcite and were undersaturated with respect to all other mineral phases such as evaporates (Kazemi 2004). Moreover, in a study conducted by Rahimi et al. (2013) on the levels of fluoride in groundwater resources of Shahrood and Damghan, nine out of ninety five drinking water samples (9.5 %) had exceeded the standard level of 1.5 mg/l and the values ranged from 0.052 to 6.87 mg/l in that research.

Due to the arid climate of this province, and low recharge of the groundwater, preservation of the quality of available groundwater resources is of great importance. This study was initiated to identify the principal geochemical processes controlling groundwater quality in eastern part of Semnan Province. In addition, considering the large sample size collected in this research, prediction of groundwater quality using a water quality index was another considered goal in this study.

Materials and methods

Sample collection and field studies

A total of 257 groundwater samples from wells and springs were collected. Polyethylene bottles were used for

sampling following rinsing of each bottle with the sample to avoid any contamination during sampling. Electrical conductivity (EC) and pH were recorded in situ using pH meter and a portable EC meter. Samples were transported to the laboratory the same day and following with filtering with 0.45- μm Millipore filter paper and acidification with nitric acid, they were analyzed for cations. For anion analyses, these samples were stored below 4 °C. Major cations (Na^+ , Ca^{2+} , Mg^{2+} , K^+) and anions (Cl^- , SO_4^{2-} , HCO_3^- , F^- , NO_3^-) were determined using the procedures given in APHA (1995). A factor analysis, with respect to principal component analysis and varimax rotation method, was conducted to find the principal components responsible for the variation found in groundwater quality variables.

Characteristics of the area

The geology of the study area is mainly characterized by Shemshak Formation (including Dark gray shale and sandstone), Lar Formation (including bedded to massive limestone), Dorud Formation (including shale with subordinate sandy limestone), Mobarak Formation (including limestone with black shale), Ruteh limestone (including bedded to massive limestone) and Karaj Formation (including tuff and tuffaceous shale). In addition, there are outcrops of cretaceous rocks and marl with gypsiferous marl in the region as well. The main geological formations have been illustrated in Fig. 1. Regarding the industrial activities in the region, there are just two industrial complexes in the area including Shahrood and Damghan industrial complex located next to these cities; however, these industrial activities are less likely to have an impact of the levels of anions and cations analyzed in this study. The location of these industrial complexes has been given in Fig. 6.

Water quality index calculation

To calculate water quality index (WQI), the method proposed by Horton (Horton 1965) and followed by many researchers (e.g., Dwivedi and Pathak 2007) was applied. To sum up, a weight was assigned to each parameter with respect to its importance in the overall quality of water. In the next step, using a quality rating scale for each parameter the final WQI was calculated as the product of the weight and the associated rating scale for each parameter. The detail of this method can be found in other literatures (e.g., Rupal et al. 2012).

Support vector regression (SVR)

Support vector regression (SVM) was one of the modeling procedures used to predict the water quality index given 10 groundwater quality variables (pH, EC, Cl^- , NO_3^- , SO_4^{2-} ,

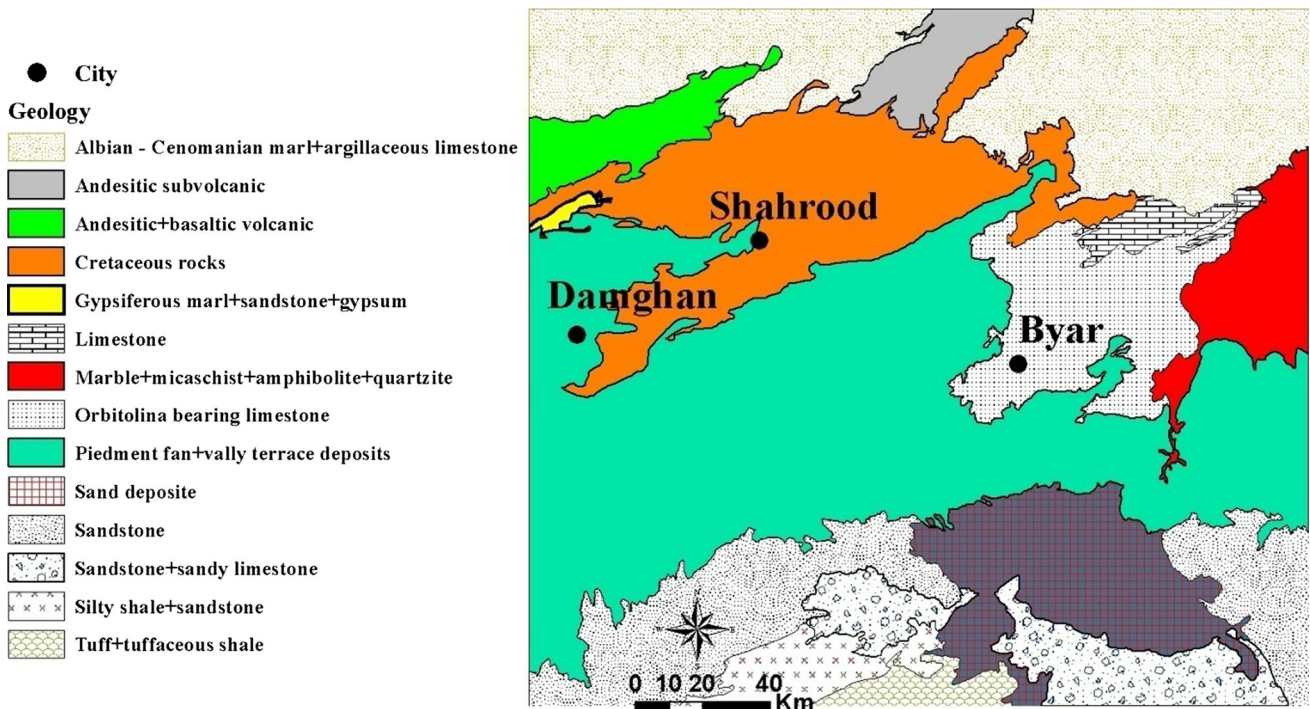


Fig. 1 Main geological formations in the study area

HCO_3^- , Na^+ , Ca^{2+} , Mg^{2+} , K^+) as the features. A popular regression version of SVM, ϵ -SVM, is used to find a function that has at most ϵ deviations from the actual obtained targets for all the training data, and is as flat as possible (Smola and Scholkopf 2004).

To enlighten this modeling method, consider a set of training points, $\{(\mathbf{x}_1, \mathbf{z}_1), \dots, (\mathbf{x}_l, \mathbf{z}_l)\}$, where $\mathbf{x}_i \in R^n$ is a feature vector and $\mathbf{z}_i \in R^1$ is the target output. Given the regulation parameter ($C > 0$) and insensitive loss function ($\epsilon > 0$), the standard form of support vector regression is as follows (Vapnik 1999):

$$\min_{\omega, b, \xi, \xi^*} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^*$$

subject to $\omega^T \phi(x_i) + b - z_i \leq \epsilon + \xi_i$,

$z_i - \omega^T \phi(x_i) - b \leq \epsilon + \xi_i^*$, $\xi_i, \xi_i^* \geq 0$, $i = 1, \dots, l$ (1)

where ξ_i and ξ_i^* are slack variables and ϵ is the accuracy demanded for the approximation. The constant $C > 0$ determines the trade-off between the flatness of linear functions and the amount up to which deviations larger than ϵ are tolerated (Smola and Scholkopf 2004). Transforming this quadratic programming problem to its corresponding dual optimization problem and introducing the kernel function in order to achieve the nonlinearity, yields the optimal regression function as (Singh et al. 2011):

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \tag{2}$$

where $C \geq \alpha_i, \alpha_i^* \geq 0, i = 1, \dots, N$. α_i and α_i^* (with $0 \leq \alpha_i, \alpha_i^* \leq C$) are the Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x})$ represents the kernel function.

For building SVR forecasting model, the LIBSVM package proposed by Chang and Lin (2001) was adapted in this study. In addition, RBF kernel was utilized for building the model.

The performance of SVM for regression depends on several parameters such as capacity parameter (C), ϵ -insensitive loss function and the variables associated with each kernel type (Aryafar et al. 2012). A trial and error procedure was followed for optimization of each parameter. The leave-one-out cross validation method was utilized using 80 % of the original data as the training set, and out-of-sample generalization error of the test data set was used for model selection.

Artificial neural network with early stopping

Artificial neural network (ANN) was utilized as another modeling method for prediction of WQI. The linear transfer function and the following transfer function were used for the output and hidden layers, respectively:

$$y_j = \tanh \left(\sum_{i=1}^d w_{ij} x_i + b_j \right) \tag{3}$$

where w_{ij} and b_j are the weight and bias parameters in which “ i ” and “ j ” subscripts refer to the input and neuron,

respectively. In addition, Levenberg–Marquardt algorithm was used to update the weight and bias of the network. To avoid the risk of over-fitting which is a common problem during neural network training, early stopping was used. To keep within the scope of this paper, we limited our survey of ANN models to the feed-forward neural network with one hidden layer.

Results and discussion

Principal geochemical processes controlling groundwater quality

Descriptive statistics associated with each groundwater quality variable are given in Table 1. The order of abundance of anions was $\text{Cl}^- > \text{SO}_4^{2-} > \text{HCO}_3^- > \text{NO}_3^- > \text{F}^-$ in which the average values were 509.41, 404.24, 120.48, 14.51, 0.59 mg/l, respectively. On the other hand, for cations Na^+ and Ca^{2+} with average values of 442.10 and 135.09 mg/l were the most abundant parameters. Moreover, the results of factor analysis for water quality parameters using principal component analysis and varimax rotation method are rendered in Tables 2 and 3. The Kaiser–Meyer–Olkin (KMO) test is a representative test of the sampling adequacy to conduct factor analysis. There is no cutoff point associated with this test; however, if the test result is smaller than 0.5, the factor analysis is not suitable (Wu and Kuo 2012). The result of KMO test was 0.788, indicating the suitability of factor analysis. In addition, Chi-square distribution (χ^2) of Bartlett's test of sphericity was high (1669.46), and highly significant, implying the existence of a common factor among the relevant matrices of the parent population (Wu and Kuo 2012).

There are many criteria for retaining the number of factors. For instance, according to Kaiser Criterion (Kaiser

1960), only factors with eigenvalues greater than 1 are retained. However, Jolliffe (1972) believed that Kaiser's criterion was too large and suggested using a cutoff of 0.7 on the eigenvalues instead. Therefore, based on the Jolliffe's criterion, eight components were kept accounting for 90.97 % of the total data variance. According to the results of initial and rotated sum of squares associated with each component (Table 3), ten initial water quality variables were reduced to five components accounting for 87.03 % of the total variance. It should be noted that due to missing values for bicarbonate in 17 stations, this parameter was excluded from the factor analysis. The first factor with an eigenvalue of 4.47 accounts for 44.73 % of the total variance (Table 2) and is the most important component. It has a high loading with chloride (0.932), sulfate (0.766), potassium (0.814), sodium (0.942), magnesium (0.911) and EC (0.778) as well (Table 3). As explained by Kumar et al. (2006), chloride levels are higher in area covered with sand dunes (like that of the study area). In addition, due to morphology of the area, high chloride weathering of ridge material can also contribute to the elevated levels of chloride (Kumar et al. 2006). The reaction of rainwater with evaporated deposits in the sand dunes can enhance the Cl^- , HCO_3^- and Na^+ as well (Subramanian and Saxena 1983). In urban area of the region, wastewater is the most likely source of extra chloride in the groundwater (Kazemi 2011); however, as explained by Jalali and Kolahchi (2008), the addition of salt to animal food and application of their manures in agricultural fields is a possible source of chloride in the study rural area next to geological origin of this element.

Sulfate levels in this research fluctuated between 9.63 and 4753 mg/l with an average level of 404.24 mg/l. Regarding the morphology of the region, the concentrations of sulfate have been proved to be higher in sand dunes (665–2531 mg/l) compared with that of quartzite

Table 1 Descriptive statistics for samples groundwater quality variables

Groundwater quality variables	Average (mg/l)	Min (mg/l)	Max (mg/l)	SD
pH	7.64	6.44	9.04	0.47
F^-	0.59	0.05	8.99	0.86
Cl^-	509.41	2.41	23,660.00	1616.97
NO_3^-	14.51	0.05	113.22	14.54
SO_4^{2-}	404.24	9.63	4753.00	534.47
HCO_3^-	120.48	0.43	385.51	93.99
Na^+	442.10	3.05	16,560.00	1189.37
K^+	3.35	0.02	39.00	3.50
Ca^{2+}	135.09	6.05	5342.00	401.83
Mg^{2+}	47.78	1.00	835.00	65.07
EC	1606.89	54.00	29,080.00	2703.44

SD stands for standard deviation

Table 2 Initial and rotated sum of squares associated with each components

Component	Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.636	46.363	46.363	4.473	44.734	44.734
2	1.283	12.828	59.191	1.123	11.227	55.961
3	1.035	10.349	69.540	1.042	10.417	66.378
4	0.937	9.369	78.909	1.035	10.347	76.725
5	0.812	8.124	87.033	1.031	10.308	87.033

Table 3 Rotated component matrix of groundwater quality variables

Water quality parameters	Component				
	1	2	3	4	5
pH	0.003	0.043	0.986	0.081	0.043
F ⁻	0.120	0.966	0.040	-0.036	0.034
Cl ⁻	0.932	0.097	0.065	-0.022	0.020
Sulfate	0.766	0.338	0.098	0.102	0.131
Na ⁺	0.942	0.093	0.120	0.027	0.026
K ⁺	0.814	-0.080	-0.145	-0.193	0.129
Ca ²⁺	0.142	0.041	0.044	0.022	0.982
Mg ²⁺	0.911	-0.120	-0.011	-0.001	0.166
EC	0.778	0.181	-0.097	0.077	-0.017
Nitrate	-0.002	-0.029	0.080	0.986	0.023

Loadings higher than 0.5 have been highlighted using bold fonts

(75–750 mg/l) and alluvium (200–831 mg/l) in a study conducted in Delhi, India (Kumar et al. 2006). In comparison with other groundwater resources in arid and semiarid area, the levels are in the same range as that of Ejina Basin in China in which average values as high as 565.95 mg/l have been recorded (Wen et al. 2005), whereas these are higher than the average sulfate levels of 88.64 mg/l in Hail Province in Saudi Arabia (Zaidi et al. 2014). In this respect, as explained by Doulati Ardejani et al. (2011), high levels of sulfate in some parts of the aquifer (as high as 4753 mg/l) may result from pyrite oxidation due to prevalent coal washing waste dump.

The sources of Mg in the aquifer are mainly from the dissolution of dolomite (Zaidi et al. 2014). The average concentrations of calcium and magnesium were 47.78 and 135.08 mg/l, respectively. As a whole, the average concentration of magnesium is higher than that of calcium. Since these two species have originated from geological formations and the solubility of CaCO₃ is much lower than that of MgCO₃, so, Ca²⁺ is precipitated as CaCO₃ resulting in the lower values of this cation in the groundwater.

The fixation of K⁺ by clay minerals and the greater resistance of this cation to weathering result in its low levels in natural groundwater (Subba Rao 2002). The average value of this cation was 3.35 mg/l which is far less than that of the other cations due to the foregoing reasons.

The second factor, with an eigenvalue of 1.12, accounts for 11.23 % of the total variance and has a high loading with fluoride (0.966). Fluoride concentration in most natural groundwater resources is <1 mg/l (Hem 1985). The main contributing factor for the concentrations of fluoride in groundwater is the solubility of fluorite (CaF₂) as the common fluoride mineral. In addition, through ion exchange reactions fluoride is adsorbed onto clay minerals including gibbsite, kaolinite and halloysite (Vengosh and Pankratov 1998). Another possible source for fluoride is evaporitic and crystalline rocks (both magmatic and metamorphic) (D'Alessandro et al. 2008) which are prevalent in the region. According to a recent study, Semnan Province is located in high-fluoride regions of Iran (Mesdaghinia et al. 2012). The mean of fluoride in this study area was 0.6 mg/l which is comparable with that of 0.64 mg/l reported from a study conducted on 78 sampling wells by Mesdaghinia et al. (2012).

The third factor encompassed 10.42 % of the total variance and has an eigenvalue of 1.04. It also has a high loading with pH (0.986). The fourth factor with an eigenvalue of 1.04 accounts for 10.35 % of the total variance. In addition, this factor is highly loaded with nitrate. The average value of nitrate in the groundwater of the study area was 14.51 mg/l which is less than the permissible level of 40 mg/l; however, values as high as 113.22 mg/l

have also been recorded in the region indicating gross pollution in some parts of the aquifer. The sources of this cation however are mainly due to the application of fertilizer and manure in agricultural fields (Berenji 1998) plus urban and rural absorbing wells (Kazemi 2011). For instance, the application rate of nitrogen fertilizers in two towns in the region (Mojen and Tash) is twofold and threefold higher than the common rate in other parts of Iran due to the tendency of farmers to use these fertilizers (Berenji 1998). On the contrary, in the urban area of the region, disposal wells and deep cesspits have been attributed as the main cause of higher-than-normal values of nitrate in the previous studies (e.g., Kazemi 2011).

The fifth factor has a high loading with calcium (0.982) and accounts for 10.31 % of the total variance as well. Because of the prevalence of limestone in the region, most of the calcium in groundwater has emanated from this mineral.

On the contrary, the correlation coefficients among groundwater quality variables are rendered in Table 4. In general, most ions are positively correlated with Cl, and Na, K, Br, Mg, and SO₄ show a strong correlation with Cl, indicating that such ions are derived from the same natural source which is geological formations of the study area. With respect to the correlation coefficient between Na⁺ and Sulfate ($r = 0.78$), it can be concluded that the excess of sodium in these samples mostly results from the dissolution of sodium sulfate minerals. The low correlation coefficient between Ca²⁺ and HCO₃⁻ ($r = 0.12$) shows that calcite may not be the source of Ca²⁺ in the study area. The Sharood Aquifer (the biggest city in the study area) is an unconfined alluvial aquifer bordered by limestone/dolomitic mountains in the north of the city to the marly-gypsiferous outcrops in the south (Kazemi 2011). Dolomite, is a sedimentary rock composed primarily of the mineral dolomite, CaMg(CO₃)₂. It is thought to be formed

by the postdepositional alteration of lime mud and limestone by magnesium-rich groundwater. On the contrary, gypsum is a soft sulfate mineral composed of calcium sulfatedihydrate, with the chemical formula CaSO₄·2H₂O (Klein et al. 1985). Therefore, the high levels of calcium and magnesium in the region can be attributed to existence of these geological formations in the study area.

High correlation between sulfate and Mg²⁺ ($r = 0.66$) may highlight the contribution of magnesium sulfate minerals to the levels of Mg²⁺ in the region. The sources of bicarbonate can be attributed to limestone and dolomite of Elica and Lar formations (Doulati Ardejani et al. 2010).

A Na/Cl molar ratio can be used to detect the source of Na⁺ in the groundwater. For instance, Na/Cl molar ratio greater than one indicates that the excess of sodium has probably originated from silicate-weathering reactions, whereas if halite dissolution is responsible for Na, the Na/Cl molar ratio is approximately equal to one (Meybeck 1987). If silicate weathering is a probable source of sodium, the water samples would have HCO₃⁻ as the most abundant anion (Rogers 1989). This is because of the reaction of the feldspar minerals with the carbonic acid in the presence of water, which releases HCO₃⁻ (Elango et al. 2003). Among groundwater samples, about 82 % have Na/Cl molar ratio greater than one indicating that Na release from silicate weathering is an important process in the study area. The plot of Na/Cl versus EC would give a horizontal line if evaporation is dominant in the area (Jankowski and Acworth 1997) meaning that no mineral species is precipitated.

With respect to the plot of Na/Cl versus EC (Fig. 2a), it can be concluded that most of the samples lie above equiline suggesting silicate weathering as the dominant process for the excess Na in the study area. In addition, if silicate weathering is the dominant process, the plot of HCO₃ versus Na should have samples falling above the

Table 4 Correlation coefficients among groundwater quality variables

	pH	F ⁻ (ppm)	Cl ⁻ (ppm)	NO ₃ ⁻ (ppm)	SO ₄ ²⁻ (ppm)	HCO ₃ ⁻ (ppm)	Na ⁺ (ppm)	K ⁺ (ppm)	Ca ²⁺ (ppm)	Mg ²⁺ (ppm)
pH	1.00									
F ⁻	0.10	1.00								
Cl ⁻	0.05	0.21	1.00							
NO ₃ ⁻	0.18	0.00	0.00	1.00						
SO ₄ ²⁻	0.09	0.34	0.69	0.11	1.00					
HCO ₃ ⁻	0.52	0.29	0.02	0.27	0.11	1.00				
Na ⁺	0.10	0.18	0.95	0.04	0.78	0.08	1.00			
K ⁺	-0.12	0.09	0.70	-0.13	0.51	-0.12	0.67	1.00		
Ca ²⁺	0.09	0.08	0.17	0.05	0.24	0.12	0.18	0.19	1.00	
Mg ²⁺	-0.00	0.04	0.81	-0.00	0.66	-0.03	0.81	0.79	0.27	1.00

Bold values indicate the significant correlation coefficients at 5 percents level

equiline (Elango and Kannan 2007). Regarding Fig. 2b, it is obvious that the majority of points lie above the equiline thus confirming the results of Na/Cl versus EC plot.

On the other hand, if dissolution of calcite dolomites and gypsum is responsible for the production of Ca^{2+} , Mg^{2+} , SO_4^{2-} and HCO_3^- , then a charge balance should exist between the cations and anions (Jalali 2006). In the absence of such a relationship, ion exchange process will shift the points to the right of the resultant 1:1 line of the plot of $\text{Ca} + \text{Mg}$ versus $\text{SO}_4 + \text{HCO}_3$ whereas the points will be shifted to the left side of this line in the case of reverse ion exchange process (Kumar et al. 2006; Fisher and Mulican 1997). The clustering of points around and below 1:1 line indicates the dominance of ion exchange process which is due to excess bicarbonate; however, a few points lie above this line proving the existence of reverse ion exchange in the study although to a lesser extent as well (Fig. 3).

On the other hand, since Ca and Mg do not increase with increasing salinity (Fig. 3b), it is an indication of reverse ion exchange in the clay/weathered layer (Kumar et al. 2006) confirming the above-mentioned results.

The $(\text{Ca}^{2+} + \text{Mg}^{2+})/\text{HCO}_3^-$ is a good indication for the sources of Ca and Mg in the groundwater. As explained by Sami (1992), if dissolution of carbonates in the aquifer and weathering of accessory pyroxene or amphibole minerals are responsible for the presence of Ca and Mg in the aquifer, this ratio would be about 0.5. As it is obvious from the plot of $(\text{Ca}^{2+} + \text{Mg}^{2+})/\text{HCO}_3^-$ versus salinity (e.g., Concentration of Cl), the Ca and Mg concentrations are added to the solution at a greater rate than the increase in the salinity of the system (Fig. 4a). These high ratios cannot be attributed to HCO_3 depletion; as under the existing alkaline conditions, HCO_3 does not form carbonic acid (H_2CO_3) (Spears 1986). High ratios, therefore,

indicate other sources for Ca and Mg, such as reverse ion exchange (Rajmohan and Elango 2004). A ratio <0.5 may be due to the exchange of calcium and magnesium in water by sodium bound in the clay. As a whole, the salinity level does not have a significant impact on this ratio.

Study on the Ca/Mg ratio contains valuable information about the source of these cations in the aquifer. That is if $\text{Ca}/\text{Mg} = 1$, dissolution of dolomite should occur, whereas a higher ratio is indicative of greater calcite contribution (Maya and Loucks 1995). In addition, higher Ca/Mg molar ratio (>2) indicates the dissolution of silicate minerals, which contribute calcium and magnesium to groundwater (Katz et al. 1998). Considering Fig. 3b, most of the samples lie around or slightly below $\text{Ca}/\text{Mg} = 1$ line indicating the dissolution of dolomite. There are some other samples having a ratio between 1 and 2 implying the dissolution of calcite. Those with greater ratios are indicative of the effect of silicate materials on the aquifer.

Prediction of WQI

The optimized values of regularization (C) and epsilon value (ϵ) parameters for modeling with SVMs were 270 and 5, respectively. Since the gamma parameter was the most sensitive parameter during SVM model building, the results for different values of this parameter are rendered in Table 5.

Regarding this table, the coefficient of determination (R^2) for the out-sample data has increased from 0.04 to 0.93 whereas that of mean squared error (MSE) for the test data set has reduced from 383.57 to 35.05. The predicted values of WQI for the training and test data have been compared, and the results have been given in Fig. 5 to show the performance of SVM. The correlation coefficients for the training and test data set were 0.97 and 0.96,

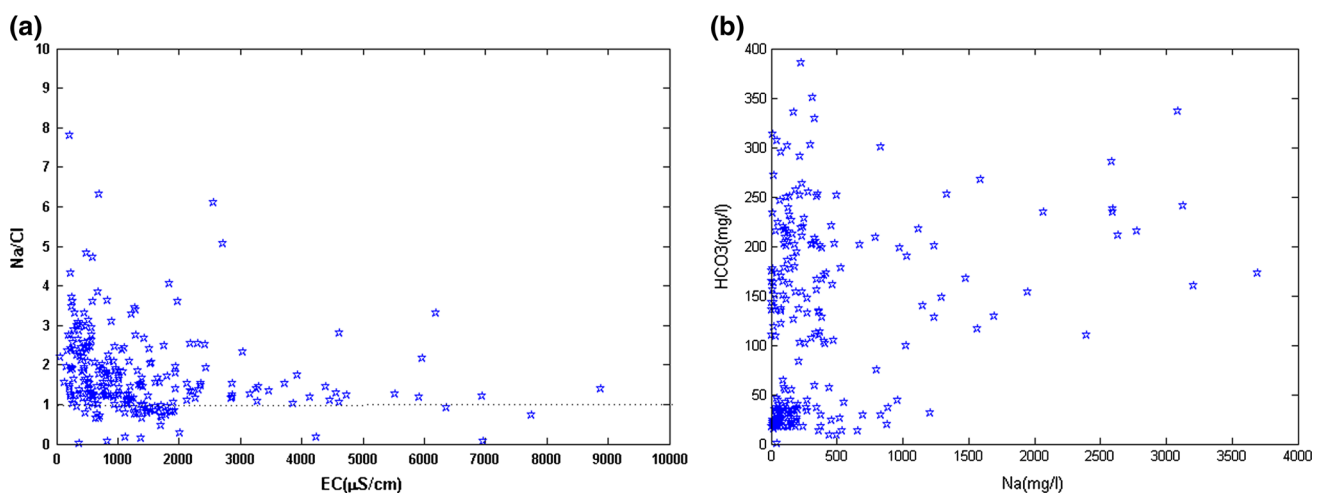


Fig. 2 Plot of Na/Cl versus EC (a) and plot of bicarbonate versus sodium (b) for the sampling stations

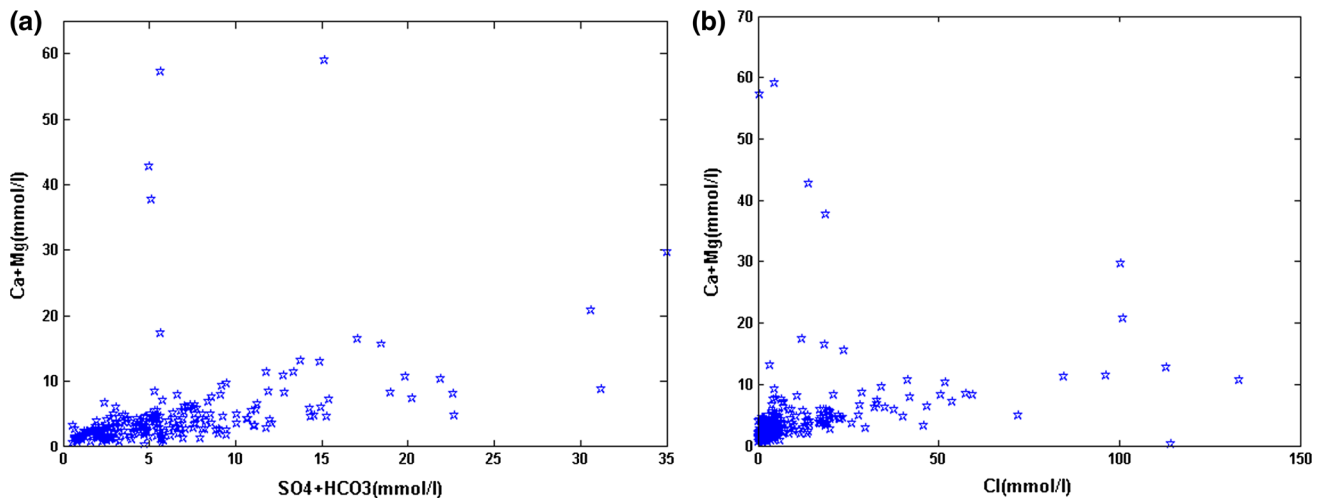


Fig. 3 Plot of Ca + Mg versus $SO_4 + HCO_3$ (a) and plot of Ca + Mg versus Cl (b) for the sampling stations

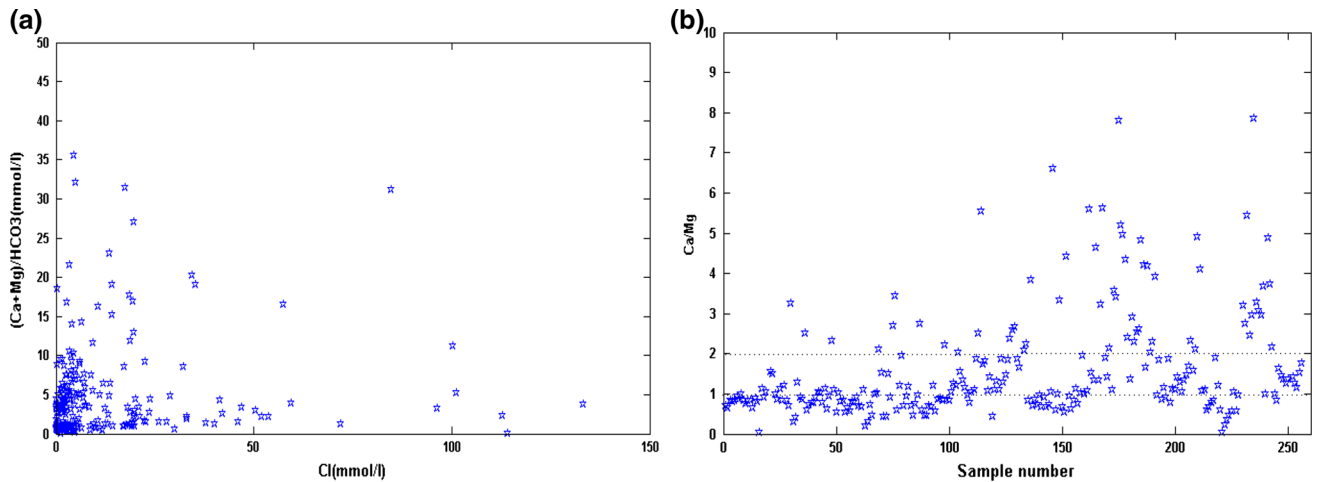


Fig. 4 Plot of $(Ca + Mg)/HCO_3$ versus Cl (a) and plot of Ca/Mg versus sample number (b) for the sampling stations

Table 5 Results of optimization of gamma parameter for RBF kernel

Gamma parameter	MSE training	R^2 testing	MSE testing	R^2 testing
0.015	22.26	0.97	383.57	0.04
0.0015	22.28	0.97	337.88	0.27
0.00015	20.37	0.97	201.36	0.60
0.000015	15.40	0.96	92.23	0.80
0.0000015	13.96	0.96	58.11	0.86
0.00000015	19.19	0.94	35.02	0.93

Bold values indicate the significant correlation coefficients at 5 percents level

respectively. On the contrary, the results of ANNs with early stopping have shown roughly the same performance with respect to Fig. 5 resulting in correlation coefficients of 0.98 and 0.96 for the training and test data set, respectively.

Although the danger of over-fitting is one of the problems during ANNs training in many of the previous researches (e.g., Sakizadeh et al. 2015; Piotrowski and

Napiorkowski 2013; Giustolisi and Simeone 2006) and SVMs have outperformed in many of the previous researches (Aryafar et al. 2012; Yoon et al. 2011; Behzad et al. 2009; Lamorski et al. 2008; Gill et al. 2006) however, this study indicated that in cases in which the number of samples is high enough compared with that of the number of features (e.g., water quality variables), using an

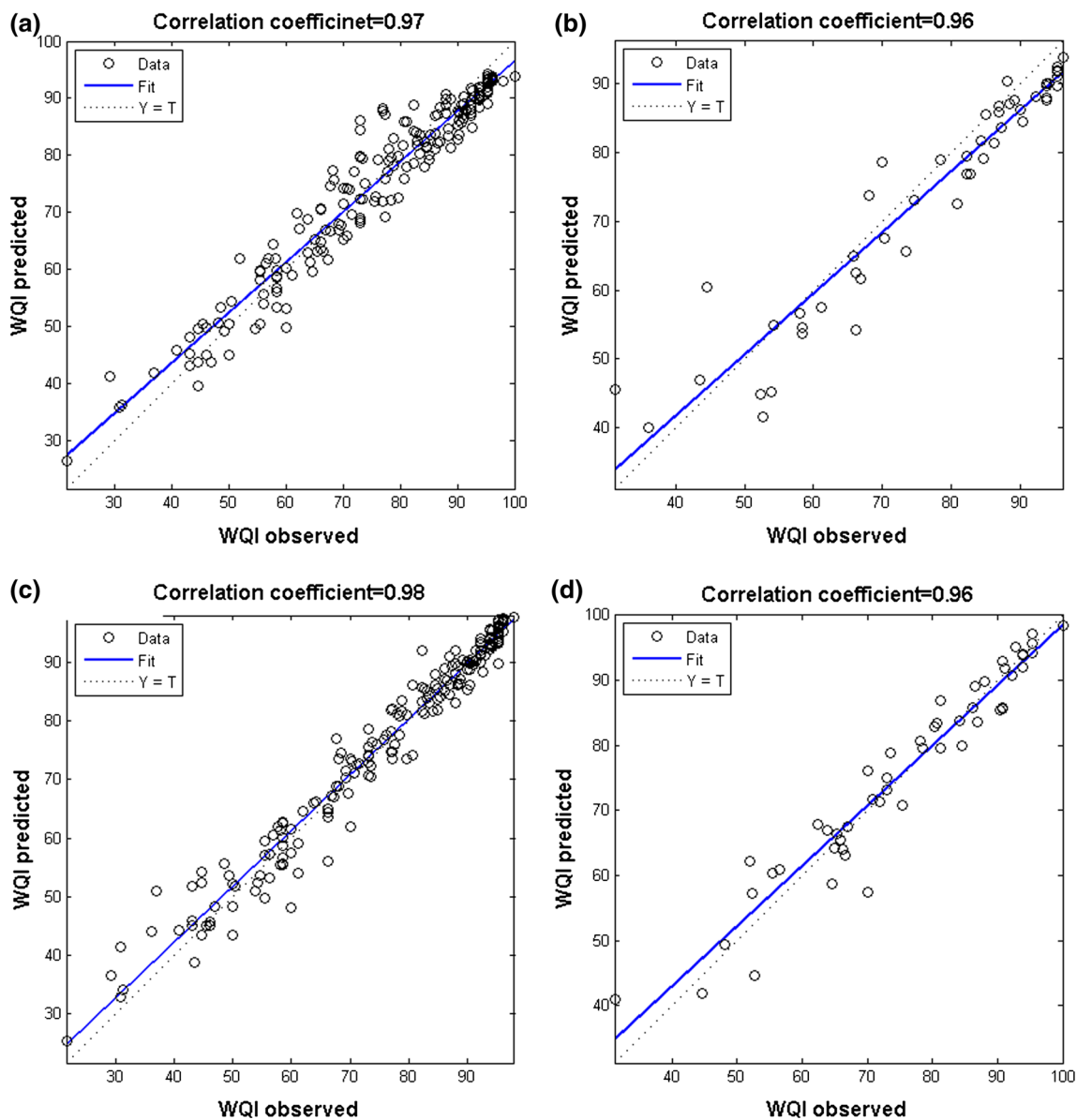


Fig. 5 Comparison between the observed and predicted values of WQI for training (a) and test (b) data of SVM and training (c) and test (d) data of ANNs

algorithm to avoid over-fitting, the same results can be obtained with ANNs modeling. Since the prediction of WQI was successful, it shows that the resultant WQI is a good representative of the available groundwater quality parameters. Therefore, the map of WQI for the study area was produced by ArcMap version 10.1, using IDW method (Gong et al. 2014).

Considering Fig. 6, there are some areas with WQI values as low as 21 in the vicinity of Byar City. The nitrate and fluoride concentrations in this station, for instance, were 49.39 and 3.93 mg/l, respectively. The area around Shahrood City, however, had the best quality compared with other parts of the study area. As a whole, with respect

to the measured parameters, the WQI for 8 % (22 stations) of the sampling wells and springs was <50 and for 17 % (44 stations), the calculated index was <60 which are important concerns that loom large.

Conclusion

A geochemical investigation was conducted in the groundwater of eastern part of Semnan Province to identify the geochemical characteristics controlling groundwater quality. This study demonstrated that the overall groundwater chemistry of the region is controlled

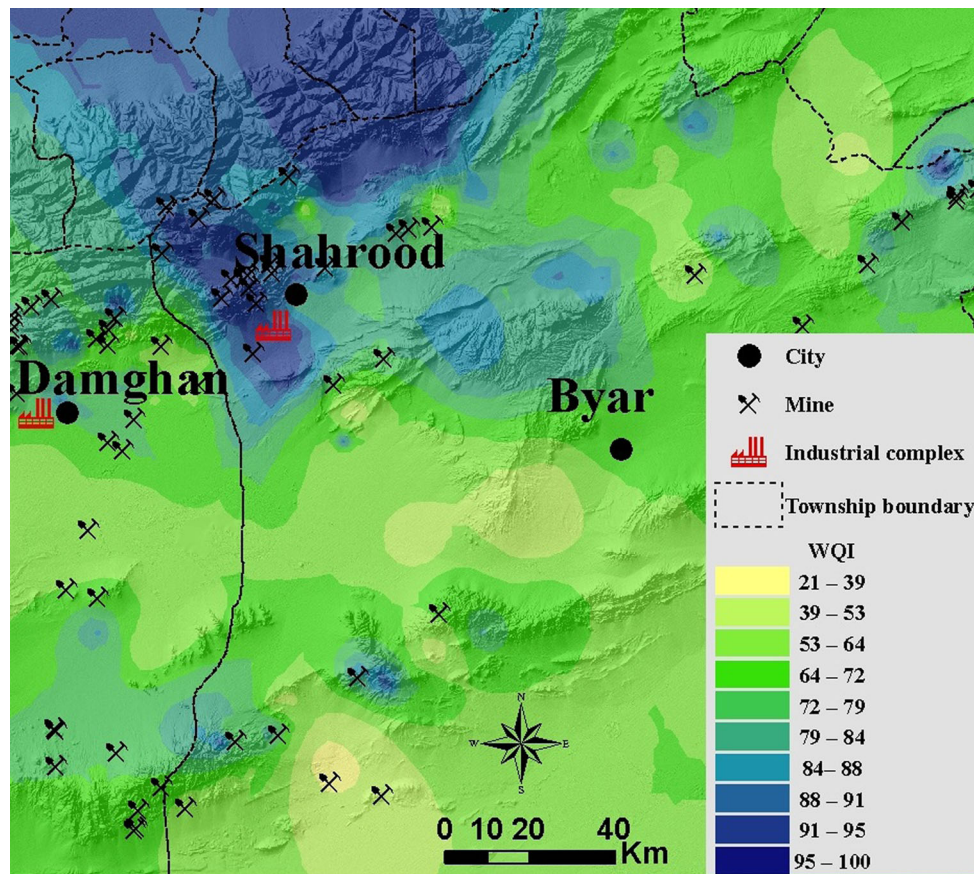


Fig. 6 Overall quality of groundwater with respect to WQI

by the rock–water interaction. There are some special conclusions drawn by studying the molar ratio of cations and anions in this research which deserve attention. Among groundwater samples, about 82 % of samples have Na/Cl molar ratio greater than one indicating that Na release from silicate weathering is an important process in the study area. The clustering of points around and below 1:1 line indicates the dominance of ion exchange process with respect to the plot of $\text{Ca} + \text{Mg}$ versus $\text{SO}_4 + \text{HCO}_3$, which is due to excess bicarbonate. Moreover, it can be concluded that most of the samples lie above equiline given the plot of Na/Cl versus EC, suggesting silicate weathering as the dominant process for the excess Na in the region. On the other hand, since Ca and Mg do not increase with increasing salinity, it is an indication of reverse ion exchange in the clay/weathered layer. Higher Ca/Mg molar ratio (>2) shows the dissolution of silicate minerals, which contribute calcium and magnesium to groundwater. In this study, most of the samples lie around or slightly below $\text{Ca/Mg} = 1$ line indicating the dissolution of dolomite. Factor analysis reduced the original water quality variables to five components accounting for 87.03 % of the total variance. The molar ratio between some groundwater quality parameters highlighted

geological sources of the studied cations and anions in the region. With respect to the results of WQI calculation, the area around Shahrood City had the best quality compared with other parts of the study area. The prediction of WQI was implemented through SVM and ANN methods. The finding of this study in this regard was that in cases in which the number of samples is high enough compared with that of the number of features (e.g., water quality variables), the same results for both of SVM and ANN can be obtained.

Acknowledgments The authors are grateful to Geological Survey of Iran for the help in analysis of heavy metals. The financial support of this project has been provided by the Grant number 100-2164 offered by Geological Survey of Iran.

References

- Aghazadeh N, Mogaddam AA (2011) Investigation of hydrochemical characteristics of groundwater in the Harzandat aquifer, Northwest of Iran. *Environ Monit Assess* 176:183–195
- APHA (1995) Standard methods for the examination of water and wastewater, 17th edn. APHA, Washington, DC
- Aryafar A, Gholami R, Rooki R, Doulati Ardejani F (2012) Heavy metal pollution assessment using support vector machine in the

- Shur River, Sarcheshmeh copper mine, Iran. *Environ Earth Sci* 67:1191–1199
- Bazargani-Guilani K, Faramarzi M, Nekouvaght Tak MA (2010) Multistage dolomitization in the cretaceous carbonates of the east Shahmirzad area, north Semnan, central Alborz, Iran. *Carbonates Evaporites* 25:177–191
- Behzad M, Asghari K, Eazi M, Palhang M (2009) Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Syst Appl* 36:7624–7629
- Berenji A (1998) Evaluation of agricultural development in the Shahrood District [in Persian]. In: Proceedings of Shahrood and development symposium, November 1998, Shahrood, Iran, pp 25–39
- Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. Software <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Corteel C, Dini A, Deyhle A (2005) Element and isotope mobility during water–rock interaction processes. *Phys Chem Earth* 30:993–996
- D'Alessandro W, Bellomo S, Parello F, Brusca L, Longo M (2008) Survey on fluoride, bromide and chloride contents in public drinking water supplies in Sicily (Italy). *Environ Monit Assess* 145:303–313
- Doulati Ardejani F, Jodieri Shokri B, Bagheri M, Soleimani E (2010) Investigation of pyrite oxidation and acid mine drainage characterization associated with Razi active coal mine and coal washing waste dumps in the Azad Shahr-Ramian region, northeast Iran. *Environ Earth Sci* 61:1547–1560
- Doulati Ardejani F, Jodieri Shokri B, Moradzadeh A, Ziadin Shafaei S, Kakaei R (2011) Geochemical characterisation of pyrite oxidation and environmental problems related to release and transport of metals from a coal washing low-grade waste dump, Shahrood, northeast Iran. *Environ Monit Assess* 183:41–55
- Dwivedi SL, Pathak V (2007) A preliminary assignment of water quality index to Mandakini River, Chitrakoot. *Indian J Environ Prot* 27:1036–1038
- Elango L, Kannan R (2007) Rock–water interaction and its control on chemical composition of groundwater. *Dev Environ Sci* 5:229–243
- Elango L, Kannan R, Senthil Kumar M (2003) Major ion chemistry and identification of hydrogeochemical processes of groundwater in a part of Kancheepuram district, Tamil Nadu, India. *J Environ Geosci* 10:157–166
- Fisher RS, Mullican WF (1997) Hydrochemical evolution of sodium sulphate and sodium-chloride groundwater beneath the northern Chihuahuan desert, Trans-Pecos, Texas, USA. *Hydrogeol J* 5:4–16
- Gill MK, Asefa T, Kemblowski MW, McKee M (2006) Soil moisture prediction using support vector machines. *J Am Water Resour As* 42:1033–1046
- Giustolisi O, Simeone V (2006) Optimal design of artificial neural networks by a multi-objective strategy: groundwater level predictions. *Hydrolog Sci J* 51:502–523
- Gong G, Mattevada S, O'Bryant SE (2014) Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environ Res* 130:59–69
- Hajizadeh H, Karami GH, Saadat S (2011) A study on chemical properties of groundwater and soil in ophiolitic rocks in Firuzabad, east of Shahrood, Iran: with emphasis to heavy metal contamination. *Environ Monit Assess* 174:573–583
- Hem JD (1985) Study and interpretation of the chemical characteristics of natural water, USGS Water-Supply Paper 2254
- Horton RK (1965) An index number system for rating water quality. *J Water Pollut Control Fed* 37:300–306
- Jalali M (2006) Chemical characteristics of groundwater in parts of mountainous region, Alvand, Hamadan, Iran. *Environ Geol* 51:433–446
- Jalali M (2009) Geochemistry characterization of groundwater in an agricultural area of Razan, Hamadan, Iran. *Environ Geol* 56:1479–1488
- Jalali M, Kolahchi Z (2008) Groundwater quality in an irrigated, agricultural area of northern Malayer, western Iran. *Nutr Cycl Agroecosyst* 80:95–105
- Jankowski J, Acworth RI (1997) Impact of debris-flow deposits on hydrogeochemical processes and the development of dryland salinity in the Yass River catchment, New South Wales, Australia. *Hydrogeol J* 5:71–88
- Jolliffe IT (1972) Discarding variables in principal component analysis. I: artificial data. *J Appl Stat* 21:160–173
- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
- Katz BG, Coplen TB, Bullen TD, Davis JH (1998) Use of chemical and isotopic tracers to characterize the interaction between groundwater and surface water in mantled Karst. *Groundwater* 35:1014–1028
- Kazemi GA (2004) Temporal changes in the physical properties and chemical composition of the municipal water supply of Shahrood, northeastern Iran. *Hydrogeol J* 12:723–734
- Kazemi GA (2011) Impacts of urbanization on the groundwater resources in Shahrood, Northeastern Iran: comparison with other Iranian and Asian cities. *Phys Chem Earth* 36:150–159
- Kim Y, Lee KS, Koh DC, Lee DH, Lee SG, Park WB, Koh GW, Woo NC (2003) Hydrogeochemical and isotopic evidence of groundwater salinization in a coastal aquifer: a case study in Jeju volcanic island, Korea. *J Hydrol* 270:282–294
- Klein C, Hurlbut CS Jr (1985) Manual of mineralogy, 20th edn. Wiley, New York, pp 352–353
- Kumar M, Ramanathan AL, Rao MS, Kumar B (2006) Identification and evaluation of hydrogeochemical processes in the groundwater environment of Delhi, India. *Environ Geol* 50:1025–1039
- Lamorski K, Pachepsky Y, Sławiński C, Walczak RT (2008) Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. *Soil Sci Soc Am J* 72:1243–1247
- Maya AL, Loucks MD (1995) Solute and isotopic geochemistry and groundwater flow in the Central Wasatch Range, Utah. *J Hydrol* 172:31–59
- Mesdaghinia A, Vaghefi KA, Montazeri A, Mohebbi MR, Saeedi R (2012) Monitoring of fluoride in groundwater resources of Iran. *Bull Environ Contam Toxicol* 84:432–437
- Meybeck M (1987) Global chemical weathering of surficial rocks estimated from river dissolved loads. *Am J Sci* 287:401–428
- Mirhosseini M, Sharif F, Sedaghat A (2011) Assessing the wind energy potential locations in province of Semnan in Iran. *Renew Sust Energ Rev* 15:449–459
- Panno SV, Hackley KC, Hwang HH, Greenberg S, Krapac IG, Landsberger S, O'Kelly DJ (2002) Source identification of sodium chloride contamination in natural waters: Preliminary results. In: Proceedings of the 12th Annual Conference of the Illinois Groundwater Consortium. <http://www.siu.edu/orda/igc/index.html>
- Piotrowski AP, Napiorkowski JJ (2013) A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modeling. *J Hydrol* 476:97–111
- Rahimi M, Ghorbani H, Naimi R, Rahimi S, Kafi S, Khosravi AA (2013) Study on the fluoride in groundwater resources of Shahrood and Damghan and its relation with dental fluorosis with respect to medical geology. *Iran J Geol* 88(4):3–13 (**in Persian**)
- Rajmohan N, Elango L (2004) Identification and evolution of hydrogeochemical processes in the groundwater environment in an area of the Palar and Cheyyar River Basins, Southern India. *Environ Geol* 46:47–61

- Rogers RJ (1989) Geochemical comparison of groundwater in areas of New England, New York, and Pennsylvania. *Groundwater* 27:690–712
- Rupal M, Tanushree B, Sukalyan C (2012) Quality characterization of groundwater using water quality index in Surat city, Gujarat, India. *Int Res J Environ Sci* 1:14–23
- Sakizadeh M, Malian A, Ahmadpour E (2015) Groundwater quality modeling with a small data set. *Ground Water* (in press)
- Sami K (1992) Recharge mechanisms and geochemical processes in a semi-arid sedimentary basin, Eastern Cape, South Africa. *J Hydrol* 139:27–48
- Singh KP, Basant N, Gupta S (2011) Support vector machines in water quality management. *Anal Chim Acta* 703:152–162
- Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222
- Spears DA (1986) Mineralogical control of the chemical evolution of groundwater. In: Trudgill ST (ed) *Solute processes*. Wiley, Chichester, p 512
- Subba Rao N (2002) Geochemistry of groundwater in parts of Guntur district, Andhra Pradesh, India. *Environ Geol* 41:552–562
- Subramanian V, Saxena K (1983) Hydrogeochemistry of groundwater in the Delhi region of India, relation of water quality and quantity. In: *Proceedings of the Hamberg symposium IAHS publication no. 146*: 307–316
- Taheri Tizro A, Voudouris KS (2007) Groundwater quality in the semi-arid region of the Chahardouly basin, West Iran. *Hydrol Process* 22:3066–3078
- Umar R, Absar A (2003) Chemical characteristics of groundwater in parts of the Gambhir River basin, Bharatpur District, Rajasthan, India. *Environ Geol* 44:535–544
- Vapnik V (1999) *The nature of statistical learning theory*, 2nd edn. Springer, New York
- Vengosh A, Pankratov I (1998) Chloride/bromide and chloride/fluoride ratios of domestic sewage effluents and associated contaminated ground water. *Ground Water* 36:815–824
- Wen X, WuY SuJ, Zhang Y, Liu F (2005) Hydrochemical characteristics and salinity of groundwater in the Ejina Basin, Northwestern China. *Environ Geol* 48:665–675
- Wu EMY, Kuo SL (2012) Applying a multivariate statistical analysis model to evaluate the water quality of a watershed. *Water Environ Res* 84:2075–2085
- Yoon H, Jun SH, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396:128–138
- Zaidi FK, Kassem OMK, Al-Bassam AM, Al-Humidan S (2014) Factors governing groundwater chemistry in paleozoic sedimentary aquifers in an arid environment: a case study from Hail Province in Saudi Arabia. *Arab J Sci Eng* (in press)
- Zehtabian G, Khosravi H, Ghodsi M (2010) High demand in a land of water scarcity: Iran. *Water and sustainability in arid regions*, Springer, pp 75–86