Self-organizing map clustering technique for ANN-based spatiotemporal modeling of groundwater quality parameters
Vahid Nourani, Mohammad Taghi Alami and Farnaz Daneshvar Vousoughi

ABSTRACT
The present study integrates co-kriging as spatial estimator and self-organizing map (SOM) as clustering technique to identify spatially homogeneous clusters of groundwater quality data and to choose the most effective input data for feed-forward neural network (FFNN) model to simulate electrical conductivity (EC) and total dissolved solids (TDS) of groundwater. The methodology is presented in three stages. In the first stage, a geostatistics approach of co-kriging is used to estimate groundwater quality parameters at locations where the groundwater levels are measured. In stage two, a SOM clustering technique is used to identify spatially homogeneous clusters of groundwater quality data. The dominant input data, selected by spatial clustering and mutual information are then imposed into the FFNN model for one-step-ahead predictions of groundwater quality parameters at stage three. The performance of the newly proposed model is compared to a conventional linear forecasting method of multiple linear regression (MLR). The results suggest that the proposed model decreases dimensionality of the input layer and consequently the complexity of the FFNN model with acceptable efficiency in spatiotemporal simulation of groundwater quality parameters. The application of FFNN for modeling EC and TDS parameters increases the accuracy of predictions respectively up to 84.5% and 17% on average with regard to the MLR model.

Key words | Ardabil plain, artificial neural network, co-kriging, groundwater quality, self-organizing map

INTRODUCTION
Groundwater quality is influenced by the geological formation and anthropogenic activities, e.g., changes in land use, urbanization, intensive irrigated agriculture, mining activities, disposal of untreated sewage in rivers, lack of rational management, etc. and clearly the groundwater contamination can cause various diseases and problems. In developing countries, as well as Iran, aquifers are experiencing an increasing threat of pollution from urbanization, industrial development, agricultural activities, and mining enterprises, because of their over-exploitation.

Groundwater quality management, as part of water resource management, requires adequate information about the chemical composition of groundwater which is controlled by many factors, including the composition of precipitation, mineralogy of watersheds, and aquifers' climate and topography. Such factors may combine to create diverse water types that change in composition spatially and temporally and cause the modeling of groundwater quality to become a complex issue.

Quantity and quality modeling of groundwater systems possesses features such as complexity, nonlinearity, being
multi-scale and random, all governed by natural and/or anthropogenic factors, which complicate the dynamic predictions. Therefore, several hydro-chemical models have been developed for simulation of groundwater complex systems. Models based on their involvement of physical characteristics generally fall into three main categories: black box models, conceptual models, and physical-based models (Nourani et al. 2014a). The conceptual and physical-based models are the main tools for predicting hydrological variables and understanding the physical processes that are taking place in the system. However, they have a number of practical limitations, including the need for large amounts of hydro-geological data, sophisticated programs for calibration using rigorous optimization techniques, and a detailed understanding of the underlying physical processes. Owing to the recognized limitations of these models, data-based methods such as artificial neural network (ANN) models (i.e., black box models) have been widely used for forecasting hydro-chemical time series.

Models based on their involvement of physical characteristics generally fall into three main categories: black box, conceptual, and physical-based models (Nourani & Mano 2007). The conceptual and physical based models are the main tools for predicting variables and understanding the physical processes involved in a system. However, they have a number of practical limitations, including the need for large amounts of field data and a detailed understanding of the underlying physical process. If sufficient field data are not available, and accurate predictions are more important than understanding the actual physics of the situation, black box models such as ANN remain a good alternative method and can provide useful predictions without the costly calibration time.

On the other hand, a number of mathematical techniques such as finite difference (FDM), finite volume (FVM), finite element (FEM), and boundary element (BEM) methods have been widely applied to solve the governing physical-based partial differential equation (PDE) of the process. Such numerical solutions of highly nonlinear and complex governing PDEs (such as Richards’ PDE for modeling flow through porous media and dispersion-diffusion PDE for simulating contamination concentration) in the lack of sufficient field data may be prone to some problems, instability, and weak convergence. Therefore, as an alternative, in this study the ANN model was employed for handling nonlinear time variability of the phenomenon to cope with the limitations of FEM (or FDM, etc.) method in temporal modeling of the process. Even so, to use such an ANN modeling approach, the user should have a good understanding of the physical processes involved in the aquifer system and such methods should be used with good engineering knowledge and judgments.

In recent years, the ANN has been used as an alternative approach for the estimation of aquifer water quality parameters (Maier & Dandy 1996). Lin & Chen (2005, 2006) proposed ANNs to determine aquifer parameters and examined the efficiency of back-propagation network (BPN) and radial basis function network (RBFN) models. Maier & Dandy (1996) presented a method using ANN to forecast salinity in water resources. Banerjee et al. (2011) evaluated the prospect of ANN simulation in estimating the safe pumping rate to maintain groundwater salinity in island aquifers. The proposed ANN model with the optimal number of spatial and temporal variables has surfaced as a simpler and more accurate alternative to the numerical method techniques. Cho et al. (2011) evaluated the predictive performance of four different models: multiple linear regressions (MLR), principal component regression (PCR), ANN, and the combination of principal components and an ANN (PC-ANN) for prediction of the arsenic contamination of groundwater for Southeast Asian countries. The modeling results showed that the prediction accuracy of PC-ANN is superior among the four different models. Nadiri et al. (2015) introduced a supervised committee machine based on artificial intelligence (AI) method to predict fluoride in groundwater of Maku, Iran. Four AI models, Sugeno fuzzy logic, Mamdani fuzzy logic, ANN, and neuro-fuzzy were employed to predict fluoride concentration. The results showed significant fitting improvement with regard to an individual AI model, especially for fluoride prediction in the mixing zone. Orzelewski et al. (2014) evaluated the quality of ground- and surface water in rural areas of the Silesian Lowlands under variable climatic conditions of Southwest Poland. The research revealed that statistical analysis employing ANN modeling can be a useful tool for estimation of water content in soils exposed to various climatic scenarios.

In spite of the reliable ability of the ANNs in temporal predictions, they found rare application for the spatial modeling of the environmental processes. Instead, powerful interpolating
Geostatistics tools have been widely used for unbiased estimation of the spatial variables at a given point. Geostatistics has made rapid advances in recent years since it was first developed by Matheron (1963). Recently, the term geostatistics has been used more generally to describe all applications of statistics in hydrogeology in which the attribute is a random field in space. The heterogeneity of the subsurface is often difficult to characterize adequately for use in deterministic models; therefore, geostatistical techniques are usually used to generate estimates of parameters in mathematical models where parameters are random variables in space.

For groundwater problems, attributes such as groundwater quality parameters are sampled at a limited number of sites whereas values at unsampled sites are sometimes needed for hydro-environmental management. Geostatistical techniques such as kriging and co-kriging can be applied to estimate the values of attributes in such unsampled sites. A comprehensive review of the applications of geostatistical tools to hydrogeology can be found in the ASCE Task Committee report (ASCE 1990). Also, a few papers have reported some applications of the geostatistics tools for groundwater quality predictions (e.g., Gaus et al. 2003; Barca & Passarella 2007).

To establish hydro-chemical classes of groundwater in a study, it is often necessary to use different approaches and methods for the graphical, spatial, and temporal representation of the relevant data. Multivariate data analysis has previously been used by researchers as part of the interpretation of groundwater quality (e.g., Voudouris et al. 2000; Lambrakis et al. 2004; Baalousha 2010). Cluster analysis is a statistical method of partitioning samples into homogeneous classes to produce an operational overview on data sets. The importance of cluster analysis techniques is evident in classifying multivariate data into sub-groups. The need for clustering techniques in hydro-environmental fields, especially groundwater-related researches, has been raised in recent decades (see Kim et al. 2003; Subyani & Al Ahmadi 2010; Chang et al. 2011; Aris et al. 2012).

A clustering technique may be employed as a spatial clustering method to improve the performance of the hydrological modeling. In the context of ANN-based hydrological modeling, clustering is usually performed for classification of the data, stations or zones into homogeneous classes (Nourani & Kalantari 2010) and/or for optimization of the model structure by selecting dominant and relevant inputs (Nourani & Parhizkar 2013; Karunasingha & Liong 2015). Clustering techniques identify structure in an unlabeled data set by objectively arranging data into homogeneous groups, where the within-group-object dissimilarity is minimized, and the between group-object dissimilarity is maximized. Therefore, the training of ANNs using such homogeneous data can lead to much better outcomes.

Among the various clustering methods, the self-organizing map (SOM) presented by Kohonen (1998) operates as an effective tool to convert the complex, nonlinear, statistical relationship between high-dimensional data items into a simple, geometric relationship on a low-dimensional display so as to allow the number of clusters to be determined by inspection. The SOM-based classification is attractive, due to its topology preserving properties for solving various problems that traditionally have been the domain of conventional statistical and operational research techniques. SOMs have been successfully accepted in science and engineering issues because they can yield unbiased and visualized results. Several studies have been carried out on SOM-based classification of groundwater quality parameters to develop regional monitoring network and chemical analyses (e.g., Hong & Rosen 2001; Sanchez-Martos et al. 2002; Peeters et al. 2007).

Modeling and understanding the quality of groundwater is as important as its quantity, because it is the main factor for determining the suitability of water for drinking, domestic, agricultural, and industrial usages. Two of the most important indicators of aquifer contamination are electrical conductivity (EC) and total dissolved solids (TDS). Most of the salts in water are present in the ionic form and are capable of conducting electric current, also EC is an important criterion in determining the suitability of water and wastewater for irrigation and is essential to assess water pollution water treatment processes. TDS is a measure of the amount of material dissolved in water and is a measure of the ‘freshness’ of water. Increasing levels of TDS and EC in an aquifer indicate that the aquifer is contaminated.

This study proposes, for the first time, a new method for time-space modeling of groundwater quality parameters (EC and TDS) based on the spatial estimator of co-kriging, spatial clustering technique of SOM, and temporal forecasting model of feed-forward neural network (FFNN). The proposed method is tested for single-step-ahead forecasting...
of groundwater EC and TDS of the Ardabil plain in Northwest Iran.

In the next sections, the concepts of the multivariate estimation method of co-kriging, SOM, FFNN, and mutual information (MI) are briefly reviewed. The following sections describe the study area and data sources, the proposed methodology, and the evaluation criteria. The results obtained using the proposed methodology are then presented and discussed, followed by some concluding remarks.

MATERIALS AND METHODS

Geostatistical modeling by co-kriging

Co-kriging is a spatial interpolation method derived from kriging, which is an optimal, in the sense of being a best linear unbiased estimator, interpolation method (Cressie 1993). In co-kriging, additional information from other variables, so-called co-variables (secondary variables), are incorporated into the analysis of the main variable. There are two situations when co-kriging usually provides better estimates than most other interpolation methods – the under sampled and the isotropic case (Myers 1982). In the first case, the primary variable is measured sparsely and additional, highly correlated, data are sampled also at additional locations. By including these stations the area is spatially better covered and more information is gained. Isotropic means that more than one variable is measured at every spatial location.

The aim of co-kriging is to estimate the primary variable of \( Z_1 \) (here, EC or TDS concentration) at some locations where no observations exist according to primary variable and secondary variable of \( Z_2 \) (here, groundwater level (GL)) in some sample points. For this, a weighted linear estimator is used as (Cressie 1995):

\[
Z(x_0) = \sum_{i=1}^{n} \lambda_{1i} Z_1(x_{1i}) + \sum_{j=1}^{m} \lambda_{2j} Z_2(x_{2j})
\]

(1)

\( x_{1i} \) and \( x_{2j} \) present the locations of primary and secondary variables. \( n \) and \( m \) denote the numbers of piezometers that measure primary and secondary variables, respectively. \( x_0 \) denotes a vector of geographical positions for the estimates.

For an unbiased estimator, the mean of the estimates must be equal to the true mean. Also, the best linear unbiased estimator must have minimum variance of estimation error (Isaaks & Srivastava 1989). The minimization of the estimation error variance under the constraint of unbiasedness leads to a set of simultaneous linear algebraic equations for co-kriging system with \((n + m + 2)\) unknowns and an equal number of equations, according to Equations (2) and (3), as follows:

\[
\begin{cases}
\sum_{i=1}^{n} \lambda_{1i} = 1, \sum_{j=1}^{m} \lambda_{2j} = 0 \\
2 \sum_{i=1}^{n} \lambda_{1i} R_{11}(x_{1i}, x_{1i}) + 2 \sum_{i=1}^{n} \lambda_{2j} R_{12}(x_{1i}, x_{2j}) - 2R_{12}(x_{1i}, x_0) + 2\mu_1 = 0 \quad i = 1, 2, \ldots, n \\
2 \sum_{i=1}^{n} \lambda_{2j} R_{22}(x_{2i}, x_{2i}) + 2 \sum_{i=1}^{n} \lambda_{1i} R_{12}(x_{1i}, x_{2j}) - 2R_{21}(x_{2j}, x_0) + 2\mu_2 = 0 \quad j = 1, 2, \ldots, m
\end{cases}
\]

(2)

(3)

\( R_{12} \) and \( R_{21} \) present cross-covariances that measure covariance between primary and secondary variables where \( \lambda_{1i} \) and \( \lambda_{2j} \) are weighting factors and can be determined by solving a nonlinear optimization problem using the Lagrange multipliers \( \mu_1 \) and \( \mu_2 \).

It is more common to write Equation (3) in terms of variogram. Cross-variogram functions describe the spatial correlation structure between different variables. The cross-variograms (and variograms when \( i = j \), i.e., the correlation structure within one variable) can be expressed as (Isaaks & Srivastava 1989):

\[
\gamma(x_i, x_j) = \sigma^2 - R(x_i, x_j)
\]

(4)

\( R(x_i, x_j), \sigma^2, \) and \( \gamma(x_i, x_j) \) present covariance, variance, and semi-variogram functions, respectively.
Self-organizing map

The SOM is an effective software tool for the visualization of high-dimensional data. It implements an orderly mapping of a high-dimensional distribution onto a regular low-dimensional grid. Thereby it is able to convert complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display while preserving the topology structure of the data (Kohonen 1998). The way SOMs go about reducing dimensions is by producing a map of usually one or two dimensions which plot the similarities of the data by grouping similar data items together. Thus SOMs accomplish two things: they reduce dimensions and display similarities. The SOM network generally consists of two layers, an input layer and a Kohonen layer. The input layer is fully connected to the Kohonen layer, which in most common applications is two-dimensional. A two-level SOM neural network is a promising approach to catch a preliminary overview on intricate data sets. It augments the conventional SOM network with an additional one-dimensional Kohonen layer in which each neuron is connected to neurons in the previous Kohonen layer. The schematic view of the two-level SOM network is shown in Figure 1.

The SOM is trained iteratively; initially the weights are randomly assigned. When the n-dimensional input vector \( x \) is sent through the network, the distance between the weight \( w \) neurons of SOM and the inputs is computed. The most common criterion to compute the distance is Euclidean distance (Kohonen 1998):

\[
||x - w|| = \sqrt{\sum_{i=1}^{n} (x_i - w_i)^2}
\]  

(5)

The weight with the closest match to the presented input pattern is called winner neuron or best matching unit (BMU). The BMU and its neighboring neurons are allowed to learn by changing the weights at each training iteration \( t \), in a manner to further reduce the distance between the weights and the input vector (Kohonen 1998):

\[
W(t + 1) = w(t) + \alpha(t)h_{lm}(x - w(t))
\]  

(6)

where \( \alpha(t) \) is the learning rate, ranging in \([0, 1]\), \( l \) and \( m \) are the positions of the winning neuron and its neighboring output nodes, and \( h_{lm} \) is the neighborhood function. The most commonly used neighborhood function is the Gaussian (Kohonen 1998):

\[
h_{lm} = \exp\left(-\frac{||l - m||^2}{2\sigma^2(t)}\right)
\]  

(7)

where \( h_{lm} \) is the neighborhood function of the best matching neuron \( l \) at iteration \( t \) and \( l-m \) is the distance between neurons and on the map grid; and \( \sigma \) is the width of the topological neighborhood. The training steps are repeated until convergence. After the SOM network is constructed, the homogeneous regions, i.e., clusters, are defined on the map.

Feed-forward neural network

The FFNN is widely applied in hydrology and water resource studies as a forecasting tool. It has already been demonstrated that an FFNN model trained by the backpropagation (BP) algorithm with three layers is satisfactory for forecasting and simulating hydrological problems.
(Hornik et al. 1989; ASCE 2000). Three-layered FFNNs, which have usually been used in forecasting hydrologic time series, provide a general framework for representing nonlinear functional mapping between a set of input and output variables (Figure 2).

They are based on a linear combination of the input variables, which are transformed by a nonlinear activation function. The term ‘feed-forward’ means that a neuron connection only exists from a neuron in the input layer to other neurons in the hidden layer or from a neuron in the hidden layer to neurons in the output layer. The neurons within a layer are not interconnected.

In Figure 2, \(i\), \(j\), and \(k\) denote the input layer, hidden layer, and output layer neurons, respectively, and \(w\) is the applied weight by the neuron. The explicit expression for an output value of a three-layered FFNN is given by (Nourani et al. 2009):

\[
\hat{y}_k = f_0 \left( \sum_{j=1}^{M} w_{kj} f_h \left( \sum_{i=1}^{N} w_{ij} x_i + w_{jo} \right) + w_{ko} \right) \]  

(8)

where \(w_{ij}\) is a weight in the hidden layer connecting the \(i\) th neuron in the input layer and the \(j\) th neuron in the hidden layer, \(w_{jo}\) is the bias for the \(j\) th hidden neuron, \(f_h\) is the activation function of the hidden neuron, \(w_{kj}\) is a weight in the output layer connecting the \(j\) th neuron in the hidden layer and the \(k\) th neuron in the output layer, \(w_{ko}\) is the bias for the \(k\) th output neuron, \(f_o\) is the activation function for the output neuron, \(x_i\) is \(i\) th input variable for input layer, and \(\hat{y}_k\) and \(y\) are computed and observed output variables, respectively. \(N_N\) and \(M_N\) are the number of the neurons in the input and hidden layers, respectively. The weights are different in the hidden and output layers, and their values can be changed during the network training process.

In any FFNN-based modeling, there are two important points to which attention must be paid: first, the architecture, i.e., the number of neurons in the input and hidden layers, and second, the training iteration (epoch) number. Appropriate selection of these two parameters improves model efficiency in both training and testing steps. Furthermore, a high epoch number and poor quality or quantity of data could cause the network to overfit during the training step. If this occurs, the model cannot adequately generalize new data outside of the training set.

**Mutual information**

Shannon mathematically formulated entropy \((H)\), now called the Shannon entropy or information content, for a discrete random variable of \(X\) with sample size of \(N\) (bin number), which takes values \(x_1; x_2; \ldots; x_N\) with probabilities

![Figure 2](https://iwaponline.com/jh/article-pdf/18/2/288/389451/jh0180288.pdf)
of \( p_1, p_2, \ldots, p_N \), respectively, as (Shannon 1948):

\[
H(X) = H(P) = - \sum_{i=1}^{N} p(x_i) \log[p(x_i)]
\]

Usually as a conventional method, linear correlation coefficients between candidate input and output are employed to determine the dominant inputs of ANN (Partal & Cigizoglu 2008), but as criticized by Nourani et al. (2014b), it is more appropriate to use a nonlinear measure such as MI in a nonlinear modeling framework such as ANN, since sometimes there may be a weak linear but strong nonlinear relationship between input(s) and output.

MI between two random variables \( X \) and \( Y \) is defined as (Yang et al. 2000):

\[
MI(A, B) = H(A) + H(B) - H(A, B)
\]

with \( H(A) \) and \( H(B) \) being the entropy of \( A \) and \( B \), respectively, and \( H(A, B) \) their joint entropy as:

\[
H(A, B) = - \sum_{a \in A} \sum_{b \in B} p_{AB}(a, b) \log p_{AB}(a, b)
\]

**Study area and data set**

The Ardabil plain aquifer (38°03’–38°27’ N and 47°55’–48°20’ E) is located in the northwest of Iran in the province of Ardabil. The region experiences relatively long winters and this plain is well known as the coldest region of Iran with an average annual precipitation of about 300 mm. May and August are the wettest and driest months of the region, respectively. The mean temperature in Ardabil plain is about 9°C. Figure 3 shows the location of the Ardabil plain.

Agriculture is the main occupation of the people in this area. Since 1980, the rapid expansion of Ardabil city, intense agricultural and some industrial activities have imposed stress on the Ardabil plain aquifers. According to Ardabil Regional Watercorp reports, the average use of drinking, industrial, and agricultural water in the Ardabil plain is about 26.4 and 177 (million m³/year), respectively, which accounts for 89% of total water demand supplied by groundwater with the remaining 11% being obtained from surface water. In the Ardabil plain 2,622 active pumping wells, 36 qantas, and 77 springs operate. The rapid increase in water demand has led to large-scale groundwater developments in this plain, and intense extraction of groundwater has caused GL to decline as much as 12 m during the past 25 years in the plain.

There are two high mountains located on the Ardabil plain, Alborz and Sabalan. Hydro-chemical studies indicated that geological units have the greatest effect on the groundwater quality in this plain. Ardabil plain has two aquifers; one is located on top of the other. The upper aquifer is multi-layer with varying thickness (20 to 40 m) across the plain and is extracted for different kinds of water supply. The number of penetrating wells to the lower aquifer is very low and may not be extracted because of its poor water quality. The general direction of groundwater flow is from other directions to the north-west of the plain.

The rock units consisting of conglomerate with some sandstone, marl, fresh water limestone, pumice, tuff and lahars of Neogene age are located in the southwestern Ardabil plain. These formations have a very low effect on aquifer recharge. Other rock units including cherty limestone, cherty dolomite, sandstone and conglomerate with thin beds of gypsum in the northeast and megaporphyric trachy andesite, trachy basalt, volcanic breccias, olivine basalt, tuff and sandy tuff located in the south, east, and north of the plain have moderate potential from the aquifer recharge point of view (Kord et al. 2013). The geological map of the Ardabil area is shown in Figure 4.

In the southern and northern parts of the Ardabil plain, although they are areas of recharge, Na-Cl water type is observed. These parts of the aquifer, which have a high concentration of sodium and chloride (poor chemical quality), are in direct contact with evaporated deposits (Kord et al. 2013).

The geo-chemical parameters (EC and TDS) of 38 piezometers, from 1996 to 2013, measured twice a year, were gathered and used in this study. One measurement per year belongs to the wet season in which the GL reaches its highest level and the other in the dry season when the level is at its lowest level. Also for the study period, GL time series of 26 piezometers, rainfall time series observed
in five stations, and observed runoff time series at the outlet of the basin, all in monthly time scale, were also employed in this study. Figure 3 presents the locations of piezometers and stations throughout the plain. A statistical summary of
the groundwater quality, quantity and hydrological data is presented in Table 1.

Input and output variables for ANN modeling are usually normalized to eliminate their dimensions. The following
simple linear mapping of the variables is the most common method for this purpose (Nourani & Kalantari 2014):

\[ r_i = \frac{I_i - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \]  

(12)

where \( I_i \) is the actual value and \( r_i \) is the respective normalized value. \( I_{\text{min}} \) and \( I_{\text{max}} \) are the minimum and maximum of the values, respectively. The normalized data were divided into training and verification sets. About 12 years of data were used for the training and the remaining 5 years for the validation purpose.

Model precision evaluation

Efficiency criteria in spatial clustering

To evaluate the performance of clustering results produced by the SOM, the silhouette coefficient is used as
the measure of cluster validity (Hsu & Li 2010). The silhouette coefficient of a cluster can indicate the degree of similarity of piezometers within a cluster defined as (Hsu & Li 2010):

\[ S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \] (13)

where \( S(i) \) is the silhouette of piezometer \( i \), \( a(i) \) measured as a Euclidean distance, is the average dissimilarity of cluster \( i \) to all other piezometers in cluster \( A \); and \( b(i) \) is the least average dissimilarity of piezometer \( i \) to the piezometers within a cluster different from cluster \( A \). Thus, a smaller \( S(i) \) value indicates a better similarity among piezometers within the same cluster. The overall quality of a clustering distribution can then be measured using the average silhouette width for the entire data set.

**Efficiency criteria in prediction stage**

In this study, two different criteria are used to measure the efficiency of the forecasting method: the root mean square error (RMSE) and the determination coefficient (DC) (Nourani & Kalantari 2010). The RMSE and DC demonstrate discrepancies between predictions and observations as:

\[ DC = 1 - \frac{\sum_{i=1}^{N} (G_{\text{com,}i} - G_{\text{obs,}i})^2}{\sum_{i=1}^{N} (G_{\text{obs,}i} - \bar{G}_{\text{obs}})^2} \] (14)

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (G_{\text{com,}i} - G_{\text{obs,}i})^2}{N}} \] (15)

where \( DC \), \( \text{RMSE} \), \( N \), \( G_{\text{obs,}} \), \( G_{\text{com,}} \), and \( \bar{G}_{\text{obs}} \) are determination coefficient, root mean square error, number of observations, observed data, computed values, and mean of observed data. In a best model, \( DC \) and \( \text{RMSE} \) go to one and zero, respectively. Legates & McCabe (1999) indicated that a hydrological model can be sufficiently evaluated by these two statistics.

**Proposed methodology**

To optimize the input layer and improve the efficiency of the FFNN-based groundwater quality model of the plain, the groundwater quality data (EC and TDS) were spatially estimated and classified, using a newly proposed hybrid method involving the co-kriging and SOM techniques. For the modeling, the dominant input data selected by SOM and MI were imposed into the FFNN model for single-step-ahead groundwater quality parameters modeling.

The proposed methodology considers the coupling of three stages for prediction of groundwater quality parameters (see Figure 5). At the first stage, since most of the piezometers which monitor the GL and groundwater quality piezometers are not co-locations, the co-kriging method and GL data observed in piezometers (as secondary parameter) are used to estimate quality parameters (as primary parameter) in the locations of GL monitoring. In stage 2, the SOM spatial clustering method is employed to classify piezometers based on GL and quality parameters. Thereafter, the central piezometer of each cluster is identified. At stage 3, first, dominant inputs are selected among groundwater quality (EC and TDS), GL and hydrological parameters (rainfall (P) and runoff (R)) data of central piezometer for each cluster using MI. Finally, the dominant inputs are imposed into the FFNN as a nonlinear forecasting method for predicting groundwater quality parameters in each cluster. The details of the obtained results are presented in the following sections.

To verify the proposed methodology, the method applied to a synthetic aquifer. In this example, two piezometers were selected from 26 groundwater level piezometers from different parts of the plain which represent different groundwater level patterns of the plain (i.e., GL1 and GL3, see Figure 3). Thereafter, best fitted equations for both piezometers data were consequently determined as:

\[ P_1(t) = 1305 \exp\left(-\frac{(t - 47.97)^2}{2084}\right) \]
\[ + 20.6 \exp\left(-\frac{(t - 317.4)^2}{71.42}\right) + 4.671 \exp\left(-\frac{(t - 249)^2}{16.16}\right) \]
\[ + 3.863 \exp\left(-\frac{(t - 202.8)^2}{23.9}\right), \quad R^2 = 0.8103 \] (16)
\[ P_3(t) = 1313 \exp \left( -\frac{(t-199.9)}{803.7} \right)^2 \\
+ 52.19 \exp \left( -\frac{(t+26.91)}{46.07} \right)^2 + 48.59 \exp \left( -\frac{(t-26.33)}{77.12} \right)^2 \\
+ 8.919e4 \exp \left( -\frac{(t-842.9)}{186.8} \right)^2 + 4.072 \exp \left( -\frac{(t-120.5)}{35.52} \right)^2, \]

\[ R^2 = 0.8615 \quad (17) \]

Finally, distinct noises (with normal distribution) and shifts were added to \( P_1(t) \) and \( P_3(t) \) equations to create five more synthetic GL time series (\( P_1 \) and \( P''1 \) according to \( P_1; P_3, P''3 \) and \( P''3 \) according to \( P_3; \) see Figure 6(a) and 6(b)). In this way, seven piezometers with synthetic GL time series data were considered in a hypothetical plain to verify the proposed clustering and forecasting methodology. According to obtained neighbor weighted distances (Figure 7), darker hexagonals divided the Kohonen layer approximately into two separate parts in the two-dimensional SOM and seven piezometers were correctly classified by spatial clustering method of SOM into two groups identifying piezometers of \( P_1 \) and \( P_3 \) as central piezometers of two clusters.

After clustering, ANN and MLR models were used to predict the GL one time step ahead. For the proposed methodology, the dominant inputs selected by SOM were used in FFNN and MLR models. The results are presented in Tables 2 and 3, respectively, which show the merit of the proposed methodology. In this way, owing to the use of synthetic data both FFNN and MLR could lead to good results, but for the real world data it is expected that the ANN will have a bit better result than the MLR model.

RESULTS AND DISCUSSION

The results of the proposed FFNN modeling of EC and TDS using spatial estimator (co-kriging) and spatial clustering
(SOM) techniques are presented in the following sub-sections. The FFNN results are also compared with those of the conventional MLR model.

Results of spatial estimation by co-kriging (stage 1)

Estimation of groundwater quality parameters in the locations where the GL are measured (GL piezometers) is the goal of this stage because the GL piezometers (26 piezometers) are mostly not co-locations with groundwater quality piezometers (38 piezometers). For this purpose, the geostatistical method of Co-Kriging is applied to estimate EC and TDS values in the location of GL piezometers.

Unlike ordinary kriging which only deals with the primary variable alone, co-kriging utilizes not only the primary variable (e.g., EC and TDS) but also cross-correlated secondary variables (here GL) in the spatial estimations and thus, co-kriging operates as a linear interpolator of both primary and secondary parameters. If only a limited number of observations are available for the primary variable of concern, knowledge of secondary variables that are correlated with the primary variable can be used to reduce the estimation error. The estimation error is thereby reduced since more information is being used for the estimation of the primary parameter.

The variogram measures dissimilarity, or increasing variance between points (decreasing correlation) as a function of distance. In addition to helping us assess how values at different locations vary over distance, the variogram provides a way to study the influence of other factors which may affect whether the spatial correlation varies only with distance (the isotropic case) or with direction and distance (the anisotropic case). The variogram map provides a visual picture of semi-variance in every compass direction. The cross-variogram calculates experimental semi-variogram values for two input variables and cross-variogram values for the combination of both variables. As two variables are handled simultaneously, the cross-variogram operation can be seen as the multivariate form of the spatial correlation operation.

The geostatistical model, having the least error, was selected by comparing the observed GL, EC, and TDS salinity values with the values estimated by variogram models.

Based on the fitted variograms and cross-variograms (between EC or TDS and GL values) models to the experimental variograms, spatial estimations of EC and TDS values for the area were carried out using co-kriging.
method. For this purpose, different variogram models (i.e., spherical, exponential, Gaussian; Myers 1982) were examined for spatial estimations of EC and TDS values for the piezometers that only monitor the GL values.

The performance of applied co-kriging method was also verified via cross-validation technique. Cross-validation is a process for checking the compatibility between a set of data, the spatial model, and neighborhood design. In cross-validation, each point in the spatial model is individually removed from the model, and then its value is estimated by a covariance model.

Since groundwater quality parameters have been measured twice a year, co-kriging estimations of EC and TDS were performed 34 times (for 17 years of study period) to create time series of quality parameters co-located with the GL monitoring piezometers.

The smallest RMSE and highest DC values in cross-validation step indicate the most accurate predictions via evaluating different variogram models. In this way, the Gaussian model was obtained as the best fitted variogram and cross-variogram models on EC, TDS, and GL data.

The smallest RMSE and highest DC values in cross-validation step indicate the most accurate predictions via evaluating different variogram models. In this way, the Gaussian model was obtained as the best fitted variogram and cross-variogram models on EC, TDS, and GL data.

Table 2 | Results of GL modeling via FFNN

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>Input variables</th>
<th>Output variable</th>
<th>Network structure*</th>
<th>Epoch</th>
<th>DC Calibration</th>
<th>RMSE (Normalized) Calibration</th>
<th>RMSE (Normalized) Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>P2(t), P3(t)</td>
<td>P1(t+1)</td>
<td>2-4-1</td>
<td>70</td>
<td>0.989</td>
<td>0.877</td>
<td>0.030</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>P5(t), P6(t), P7(t)</td>
<td>P4(t+1)</td>
<td>3-7-1</td>
<td>120</td>
<td>0.998</td>
<td>0.983</td>
<td>0.011</td>
</tr>
</tbody>
</table>

*The result has been presented for the best structure. First, second, and third numbers denote input, hidden, and output neuron number, respectively.

Table 3 | Results of GL modeling via MLR

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>MLR model</th>
<th>DC Calibration</th>
<th>RMSE (Normalized) Calibration</th>
<th>RMSE (Normalized) Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>P1(t+1) = 0.74P2(t) + 0.291P3(t)-0.024</td>
<td>0.811</td>
<td>0.0331</td>
<td>0.0433</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>P4(t+1) = 0.691P5(t) + 0.307P6(t) + 0.193P7(t)-0.068</td>
<td>0.823</td>
<td>0.028</td>
<td>0.033</td>
</tr>
</tbody>
</table>

Figure 8 | Gaussian semi-variogram model.

\[ \gamma(h) = C_0 + A \left(1 - e^{-\left(\frac{h}{r}\right)^2}\right) \text{ for } h > 0 \]  \hspace{1cm} (18)

By fitting a variogram model (e.g., Equation (16), Figure 8) with lag distance of \( h \) to the experimental variogram, the model parameters (sill, \( A \), range, \( r \), and nugget, \( C_0 \)) can be determined.

Table 4 illustrates the parameters of variogram models fitted to the experimental variograms and Table 5 presents the cross-validation results. It should be noted that since the co-kriging approach was applied to the data 34 times, the range (maximum–minimum) of obtained parameters has been presented for each parameter of the variograms (see Table 4). Consequently, Table 5 shows the range of efficiency criteria obtained through the cross-validation process. Also, Figure 9 shows an example of variogram and cross-validation results of TDS values.
Results of SOM-based clustering (stage 2)

In groundwater quality modeling, researchers are often confronted with large, multidimensional data sets. An exploratory data analysis is often carried out, aiming at summarizing the available data, extracting useful information and formulating hypotheses for further research.

Implanted piezometers all over the plain can prepare informative data about the quality of groundwater. Application of clustering techniques (e.g., SOM) on available piezometers interspersed spatially over the plain can be an appropriate method to capture the adequate information of homogeneous piezometers.

Owing to the existence of various piezometers over the plain and the importance of managing groundwater resources, it is a necessity to unite the adequate information about groundwater quality in various regions of the plain and identify the dominant piezometers to predict groundwater quality parameters of the plain for the future. In order to accomplish the spatial clustering, SOM was utilized to identify similar and predominant piezometers.

For the clustering at first, a two-dimensional SOM was applied to the groundwater quality (EC and TDS) data as input to classify the piezometers into the clusters with similar groundwater quality patterns. The purpose of such a two-dimensional SOM clustering was to have an overview on homogeneous regions and the approximate number of clusters with regard to the plain topology. In order to apply the proposed two-step SOM, the size of the Kohonen layer was considered to be 6 × 6 for the EC parameter in the first step. Figure 10(a) shows the hits plan of the output layer size of 6×6 for the EC parameter. The hits plan is an illustration of a SOM output layer, with each neuron showing the number of classified input vectors. The relative number of vectors for each neuron is shown via the size of a colored patch. Figure 10(b) shows the neighbor weighted distances obtained by the two-dimensional SOM, where the small hexagons represent the neurons. The colors in the linear

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Cross-validation results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>RMSE</td>
</tr>
<tr>
<td>EC</td>
<td>555.9–692.4 (μZimens/cm)</td>
</tr>
<tr>
<td>TDS</td>
<td>209.8–688 (mg/lit)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>The parameters of semi-varogram and cross-varogram in co-kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-varogram</td>
<td>Model</td>
</tr>
<tr>
<td>EC</td>
<td>Gaussian</td>
</tr>
<tr>
<td>TDS</td>
<td>Gaussian</td>
</tr>
<tr>
<td>GL</td>
<td>Gaussian</td>
</tr>
<tr>
<td>Cross-varogram</td>
<td>Model</td>
</tr>
<tr>
<td>EC</td>
<td>Gaussian</td>
</tr>
<tr>
<td>TDS</td>
<td>Gaussian</td>
</tr>
<tr>
<td>GL</td>
<td>Gaussian</td>
</tr>
</tbody>
</table>
(large) hexagons indicate the distances between neurons, with the darker colors representing larger distances, and the lighter colors representing smaller distances. According to the obtained neighbor weighted distances for EC (Figure 10(b)), darker linear hexagons divide the Kohonen layer approximately into five parts, in the two-dimensional SOM. Second, in order to be certain of the highlighted clusters, a one-dimensional SOM was used to classify the piezometers into a specific number of clusters, determined in the first step (i.e., $1 \times 4$). At this one-dimensional SOM, the number of neurons in the Kohonen layer was set equal to the number of clusters determined by two-dimensional SOM (i.e., four clusters).

To evaluate the performance of the clustering results produced by SOM, the silhouette coefficient (Equation (9)) was utilized and the results of clustering are presented in Tables 6 and 7 for EC and TDS parameters, respectively.

The Euclidean distance criterion was then utilized to determine the centroid piezometer (prominent piezometer) of each cluster, which is the best representation of the groundwater quality pattern of the cluster. Thus, the specific measures should be devoted to protecting and maintaining...
such important piezometers in order to have precise and reliable data for prediction and management of the plain. The determined central piezometers of clusters are presented in the fourth column of Tables 6 and 7.

Tables 6 and 7 (and see Figure 11) show that SOM clusters the EC and TDS piezometers into almost similar clusters. Owing to the fact that most of the salts in water are presented in the ionic form and capable of conducting electric current, conductivity is a good and rapid measure of concentration of TDS and denotes the relationship between EC and TDS. According to the geology of the plain, the north and south-west areas with high concentrations of sodium and chloride are placed in two separated clusters of EC and TDS parameters. The other parts of the plain (the south-east, east and central areas) with good chemical quality are placed in the other clusters.

Results of FFNN-based prediction (stage 3)

The spatial clustering of piezometers by SOM divided the plain into four homogenous clusters (zones) according to EC and TDS parameters. Then, two FFNNs were trained for each cluster to predict EC and TDS values of the central piezometer (as representative of zone) one time step ahead. Since the piezometers that lie in the same cluster may have a similar quality pattern, the quality behavior of piezometers of the cluster in future may be guessed from the quality predictions of the central piezometer and the behavior of central piezometers can be considered as the representative of groundwater quality pattern in various regions of the plain. To take into account the contribution of surface water in the hydrologic budget of the plain and its impact on the groundwater quality and quantity parameters, both precipitation and runoff data were also considered as potential inputs of the FFNNs as well as GL and quality data at previous time steps to predict quality parameters of EC and TDS one time step ahead.

In this plain, the groundwater quality data (EC and TDS) sampling have been reported at 6-monthly (T) intervals for all of the piezometers but the monthly (t) rainfall, runoff, and GL data are available and used in the FFNN models.

In ANN modeling, it is important that particular attention is paid to the appropriate selection of inputs which can improve the model’s efficiency in both the steps of calibration (training) and verification. It also prevents the model from being over-trained. For this purpose, different input combinations (from EC, TDS, rainfall (P), runoff (R), and GL time
series for each cluster) were examined as input neurons of FFNNs for one-step-ahead EC and TDS predictions.

In this study, the MI measure (Equation (10)) was employed to determine the dominant inputs and lag times between hydrological processes (rainfall, runoff, and GL) and groundwater quality parameters among different input combinations for each cluster. The results of the sensitivity analysis via MI in selecting dominant inputs and lag times for groundwater quality modeling are presented in Tables 8–11 in the column Input parameters.

For each cluster, two FFNNs were trained (for modeling EC and TDS) by back-propagation algorithm and the

![Figure 11 SOM clustering: (a) EC, (b) TDS.](image)

**Table 8 | Results of FFNN modeling for EC**

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>Input variables</th>
<th>Output variable</th>
<th>Network structure*</th>
<th>Epoch</th>
<th>DC Calibration</th>
<th>Veri Verification</th>
<th>RMSE (Normalized) Calibration</th>
<th>RMSE (Normalized) Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>EC4(T-1), R(t-2), P6(t)</td>
<td>EC4(T + 1)</td>
<td>3-3-1</td>
<td>40</td>
<td>0.932</td>
<td>0.767</td>
<td>0.07</td>
<td>0.143</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>EC16(T-1), GL16(t), R(t-2), P1(t)</td>
<td>EC16(T + 1)</td>
<td>4-4-1</td>
<td>20</td>
<td>0.955</td>
<td>0.708</td>
<td>0.045</td>
<td>0.165</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>EC5(T-1), GL5(t), R(t-2), P4(t-3)</td>
<td>EC5(T + 1)</td>
<td>4-3-1</td>
<td>20</td>
<td>0.867</td>
<td>0.803</td>
<td>0.081</td>
<td>0.126</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>EC7(T-1), GL7(t), R(t-3)</td>
<td>EC7(T + 1)</td>
<td>3-5-1</td>
<td>20</td>
<td>0.968</td>
<td>0.848</td>
<td>0.036</td>
<td>0.077</td>
</tr>
</tbody>
</table>

*The result has been presented for the best structure. First, second, and third numbers denote input, hidden, and output neuron number, respectively (T presents 6-monthly period and t presents monthly period).
efficiency criteria for each model were used to determine the best model. The results of temporal modeling of EC and TDS are shown for four clusters in Tables 8 and 9, respectively. The performance efficiencies of FFNNs in both training and validation steps show reliable results for predicting EC and TDS concentrations. Overall, it can be seen that the FFNN model well reproduced the variations of the observed groundwater quality parameters in the Ardabil plain. The result indicates that the proposed methodology can be used for groundwater quality modeling in such areas using groundwater data (quality and quantity) and some easily available hydro-climatologic data as inputs of the model.

The results for EC and TDS parameters indicate that all of the models produced acceptable outcomes, and confirm the appropriate identification of the representative groundwater quality patterns over the plain. In most cases, the east and central parts of the plain provided better results (for both EC and TDS parameters) than other parts which is due to them being rich in chemical quality. As mentioned in geology descriptions, the EC and TDS clusters that are placed in the north and

### Table 9 | Results of FFNN modeling for TDS

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>Input variables</th>
<th>Output variable</th>
<th>Network structure*</th>
<th>Epoch</th>
<th>DC Calibration</th>
<th>Verification</th>
<th>RMSE (Normalized) Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>TDS21(T-1), GL21(t), R(t-3)</td>
<td>TDS21(T + 1)</td>
<td>3-5-1</td>
<td>20</td>
<td>0.898</td>
<td>0.771</td>
<td>0.096</td>
<td>0.149</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>TDS24(T-1), GL24(t), R(t-2), P5(t)</td>
<td>TDS24(T + 1)</td>
<td>4-3-1</td>
<td>40</td>
<td>0.956</td>
<td>0.9305</td>
<td>0.0655</td>
<td>0.083</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>TDS26(T-1), GL26(t), R(t-2), P6(t-1)</td>
<td>TDS26(T + 1)</td>
<td>4-4-1</td>
<td>30</td>
<td>0.969</td>
<td>0.732</td>
<td>0.054</td>
<td>0.151</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>TDS9(T-1), GL9(t), R(t), P4(t-1)</td>
<td>TDS9(T + 1)</td>
<td>4-4-1</td>
<td>50</td>
<td>0.97</td>
<td>0.884</td>
<td>0.054</td>
<td>0.107</td>
</tr>
</tbody>
</table>

*The result has been presented for the best structure. First, second, and third numbers denote input, hidden, and output neuron number, respectively (T presents 6-monthly period and t presents monthly period).

### Table 10 | Results of FFNN modeling for EC without clustering

<table>
<thead>
<tr>
<th>Input variables</th>
<th>Output variable</th>
<th>Network structure*</th>
<th>Epoch</th>
<th>DC Calibration</th>
<th>Verification</th>
<th>RMSE (Normalized) Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC10(T-1), EC18(T-1), EC26(T-1)</td>
<td>EC4(T + 1)</td>
<td>3-3-1</td>
<td>30</td>
<td>0.947</td>
<td>0.604</td>
<td>0.062</td>
<td>0.212</td>
</tr>
<tr>
<td>EC8(T-1), EC15(T-1), EC17(T-1)</td>
<td>EC16(T + 1)</td>
<td>3-4-1</td>
<td>20</td>
<td>0.793</td>
<td>0.451</td>
<td>0.090</td>
<td>0.171</td>
</tr>
<tr>
<td>EC10(T-1), EC11(T-1), EC20(T-1), EC22(T-1)</td>
<td>EC5(T + 1)</td>
<td>4-2-1</td>
<td>30</td>
<td>0.875</td>
<td>0.590</td>
<td>0.094</td>
<td>0.185</td>
</tr>
<tr>
<td>EC5(T-1), EC18(T-1), EC21(T-1), EC23(T-1)</td>
<td>EC7(T + 1)</td>
<td>4-3-1</td>
<td>30</td>
<td>0.885</td>
<td>0.664</td>
<td>0.060</td>
<td>0.099</td>
</tr>
</tbody>
</table>

*The result has been presented for the best structure. First, second, and third numbers denote input, hidden, and output neuron number, respectively (T presents 6-monthly period).

### Table 11 | Results of FFNN modeling for TDS without clustering

<table>
<thead>
<tr>
<th>Input variables</th>
<th>Output variable</th>
<th>Network structure*</th>
<th>Epoch</th>
<th>DC Calibration</th>
<th>Verification</th>
<th>RMSE (Normalized) Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDS7(T-1), TDS18(T-1), TDS23(T-1)</td>
<td>TDS21(T + 1)</td>
<td>3-3-1</td>
<td>50</td>
<td>0.834</td>
<td>0.533</td>
<td>0.114</td>
<td>0.126</td>
</tr>
<tr>
<td>TDS11(T-1), TDS15(T-1), TDS19(T-1), TDS20(T-1)</td>
<td>TDS24(T + 1)</td>
<td>4-4-1</td>
<td>40</td>
<td>0.976</td>
<td>0.599</td>
<td>0.045</td>
<td>0.218</td>
</tr>
<tr>
<td>TDS4(T-1), TDS10(T-1), TDS22(T-1)</td>
<td>TDS26(T + 1)</td>
<td>5-2-1</td>
<td>30</td>
<td>0.791</td>
<td>0.430</td>
<td>0.137</td>
<td>0.175</td>
</tr>
<tr>
<td>TDS6(T-1), TDS8(T-1), TDS12(T-1), TDS17(T-1)</td>
<td>TDS9(T + 1)</td>
<td>4-3-1</td>
<td>40</td>
<td>0.750</td>
<td>0.541</td>
<td>0.089</td>
<td>0.111</td>
</tr>
</tbody>
</table>

*The result has been presented for the best structure. First, second, and third numbers denote input, hidden, and output neuron number, respectively (T presents 6-monthly period).
south-west of the plain produced less accurate outcomes in comparison to the other clusters.

It is clear from Tables 8 and 9 that the quality of piezometers placed in areas with a higher distance from the plain outlet show higher lag with the values of runoff time series. The higher distances of central piezometers from the outlet demonstrate a late response to the overland water condition (rainfall and runoff). Therefore, the central piezometers close to the outlet show minimum lag with runoff parameter (see Tables 8 and 9). Also, lag time between the rainfall and central piezometer’s EC and TDS values is proportional to the spatial distance between the related rainfall station and that piezometer.

For instance, scatter plots shown in Figures 12 and 13 compare the observed and predicted concentrations of EC and TDS in both training and validation steps for piezometers 5 and 24, respectively.

In dealing with many alternative inputs in FFNN modeling, an efficient algorithm is needed to select dominant inputs. Therefore, to optimize the input layer of FFNN and improve the modeling efficiency, SOM clustering approach was used to select dominant inputs instead of using whole potential inputs in the FFNN. Tables 10 and 11 show the results of FFNNs without clustering for modeling EC and TDS, respectively. In this way, the input variables were selected from three or four neighbor piezometers of the coordinate piezometer. The comparison of FFNNs results without and with using clustering approach (Tables 8 and 9) shows that modeling by clustering approach has more accuracy than without clustering (using neighbor

![Figure 12](https://i.imgur.com/9Q5Q5Q.png) **Figure 12** | Comparison of observed and predicted EC concentrations in piezometer 5: (a) train step, (b) verification step.

![Figure 13](https://i.imgur.com/5Q5Q5Q.png) **Figure 13** | Comparison of observed and predicted TDS concentrations in piezometer 24: (a) train step, (b) verification step.
piezometers data) for groundwater quality predictions. This is because in modeling without clustering irrelevant data as inputs in the prediction stage enter redundant information and noise into the modeling process which, in turn, leads to inappropriate and over-fitted results for most cases.

For further evaluation of the trained FFNNs and to have a comparison with conventional prediction methods, the classic MLR model was also employed for modeling groundwater quality parameters of the plain. Although a linear and classic time series forecasting model like MLR is likely not to be able to model a complex nonlinear hydrological process, it is still a commonly used method in practice and, accordingly, is useful as a comparison model. In this research, rainfall, runoff, antecedents of GL, and groundwater quality data were used as exogenous inputs to predict future groundwater quality parameter as the output. In the MLR model, input data are selected using correlation coefficient between potential inputs and target. The MLR model, owing to its being linearly inherent, could not detect nonlinear relationship between the data. Thus, correlation coefficient that detects linear correlation between input parameters and target is usually used as criterion for dominant input selection. However, in a nonlinear complex hydrological process such as groundwater quality modeling, in spite of a weak linear relationship, a strong nonlinear relationship may exist between input and output variables. MI measures the statistical nonlinear dependency between potential inputs and the output of the model. Therefore, the implication of these two input selection criteria in the study led to different best input combinations used in the MLR and FFNN models.

MLR models with the best forecasting performance for clusters are shown in Tables 12 and 13 for EC and TDS, respectively. The results show that the MLR model can model the TDS parameter more reliably than the EC parameter which means that the TDS time series relatively follows a linear pattern in temporal scale. This could be the reason why the inputs of the MLR model in modeling the TDS parameter are almost the same as the inputs of the ANN model, in contrast to the inputs of the EC parameter which are quite different from the inputs of ANNs.

Also, tabulated results indicate that the MLR model, owing to its linear inherence is unable to completely handle the complex nonlinear groundwater quality predictions. Thus, the nonlinear FFNN model is more efficient than the MLR model.

Table 12 | Results of MLR modeling for EC parameter

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>MLR model</th>
<th>DC Calibration</th>
<th>RMSE (Normalized) Calibration</th>
<th>Verification Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>EC4(T + 1) = 0.337-0.254P6(t)-0.189R(t-2)-0.02GL4(t) + 0.6EC4(T-1)</td>
<td>0.526</td>
<td>0.196</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>EC16(T + 1) = 0.436-0.484R(t-2) + 0.16GL16(t) + 0.364EC16(T-1)</td>
<td>0.468</td>
<td>0.155</td>
<td>0.189</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>EC5(T + 1) = 0.257-0.111P4(t-3)-0.254GL5(t) + 0.719EC5(T-1)</td>
<td>0.745</td>
<td>0.14</td>
<td>0.174</td>
<td></td>
</tr>
<tr>
<td>Cluster 4</td>
<td>EC7(T + 1) = 0.356-0.369R(t-3) + 0.296EC7(T-1)</td>
<td>0.509</td>
<td>0.152</td>
<td>0.121</td>
<td></td>
</tr>
</tbody>
</table>

Table 13 | Results of MLR modeling for TDS parameter

<table>
<thead>
<tr>
<th>Cluster no.</th>
<th>MLR model</th>
<th>DC Calibration</th>
<th>RMSE (Normalized) Calibration</th>
<th>Verification Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>TDS21(T + 1) = 0.639 + 0.103R(t-3)-0.67GL21(t) + 0.3TDS21(T-1)</td>
<td>0.711</td>
<td>0.161</td>
<td>0.156</td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>TDS24(T + 1) = 0.275-0.091P5(t)-0.04R(t-2)-0.19GL24(t) + 0.813EC24(T-1)</td>
<td>0.835</td>
<td>0.127</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>TDS26(T + 1) = 0.459-0.223P6(t-1)-0.131R(t-2)-0.177GL26(t) + 0.53TDS26(T-1)</td>
<td>0.913</td>
<td>0.09</td>
<td>0.099</td>
<td></td>
</tr>
<tr>
<td>Cluster 4</td>
<td>TDS9(T + 1) = 0.211 + 0.286P4(t-1)-0.058R(t)-0.168GL9(t) + 0.765 TDS9(T-1)</td>
<td>0.781</td>
<td>0.149</td>
<td>0.135</td>
<td></td>
</tr>
</tbody>
</table>
CONCLUDING REMARKS

In this paper, FFNN models have been developed for groundwater quality forecasting over the plain of Ardabil, northwestern Iran. The inputs of the models were monthly rainfall, runoff, and GL and groundwater quality parameters (EC and TDS) at 6-monthly intervals over the study area. Co-kriging estimator was used to estimate groundwater quality parameters in the locations of GL. As a result, the entire study area was divided into four clusters via SOM clustering approach. ANN modeling was then applied separately for central piezometers of each cluster.

Spatial clustering via SOM was shown to be useful in improving FFNN-based modeling of groundwater quality parameters. The computed MI between different input candidates and output for selecting dominant inputs and lags of model showed that the concentrations of groundwater quality parameters (EC and TDS) are significantly affected by GL fluctuations. Furthermore, although the effect of hydrological parameters (rainfall and runoff) in the plain under study was not significant on plain recharge, it affected the changes in qualitative parameters of the plain.

To evaluate model performance, the proposed methodology was also compared with a conventional model (i.e., the MLR method as a linear model). The results show that the MLR model, owing to its linear inherence, was not able to detect a nonlinear relationship between the studied parameters. The FFNN yielded better performance for clusters, on average 84.5% and 17% for modeling EC and TDS parameters, respectively.

The examination of other artificial intelligence approaches to forecast groundwater quality parameters in hydrological processes is also suggested. For instance, owing to the uncertainty of the hydro-geological processes and the ability of the fuzzy concept to handle uncertainties, the conjunction of the ANN and fuzzy inference system (FIS) models as an adaptive neural-fuzzy inference system (ANFIS) model, could provide useful results. Owing to the lack of monthly data for quality parameters of EC and TDS for the study area, the modeling was performed using monthly data of rainfall, runoff, and GL but 6-monthly data of EC and TDS, and it would be useful to apply the proposed methodology on other heterogeneous groundwater systems with longer groundwater quality data sets in monthly scale.

REFERENCES


First received 25 February 2015; accepted in revised form 28 July 2015. Available online 9 September 2015