Prediction of heavy metals contamination in the groundwater of Arak region using artificial neural network and multiple linear regression

Feridon Ghadimi

Assistant of professor, Department of Mining Engineering, Arak University of Technology, Arak, Iran.

* Corresponding Author: drghadimi@yahoo.com

Abstract

Prediction of the heavy metals in the groundwater is important in developing any appropriate remediation strategy. This paper attempts to predict heavy metals (Pb, Zn and Cu) in the groundwater from Arak city, using artificial neural network (ANN) algorithm by taking major elements (HCO_3, SO_4) in the groundwater from Arak city. For this purpose, contamination sources in the groundwater were recorded based on 150 data samples and several models were trained and tested using collected data to determine the optimum model in which each model involved two inputs and three outputs. The results obtained (the comparison between the predicted and the measured data) indicate that Multilayer Perceptron Neural Networks model (ANN) has strong potential to estimation of the heavy metals in the groundwater with high degree of accuracy and robustness.

Keywords: Groundwater, Artificial neural network, Heavy metals, Major elements, Arak.

1- Introduction

Chemical composition of groundwater is controlled by many factors which include composition of precipitation, mineralogy of watersheds and geochemical processes within aquifer (Rajmohan and Elango, 2004; Andre *et al*., 2005; Singh *et al*., 2014; Zghibi *et al*., 2014). These processes occurring within groundwater and reactions with aquifer minerals have a profound effect on water quality and are responsible for variations in the groundwater composition (Helstrup *et al*., 2007; Ghadimi and Ghomi, 2013). Hence quality of water along its underground movement is therefore dependent not only on chemical and physical properties of surrounding rocks but also varies as a result of human activity (Helstrup *et al*., 2007; Matiatos *et al*., 2014; Zapata *et al*., 2014; Devic *et al*., 2014; Ghomi *et al*., 2013). Variation in groundwater chemistry is mainly a function of the interaction between groundwater and mineral composition of aquifer materials through which it moves (Monjerezi *et al*., 2011; Yidana *et al*., 2010; Anderson *et al*., 2014). Hydro-chemical processes, including dissolution, precipitation, weathering together with residence time occurring along flow path, control variation in chemical composition of groundwater (Oinam *et al*., 2012; Wang *et al*., 2013; Srinivasamoorthy *et al*., 2014; Masoud, 2014).

Moreover, over the years, the application of artificial neural network (ANN) in different fields of engineering has been developing. ANNs have a special capacity to estimate the dynamics of nonlinear systems in many applications in a black box manner (Dovins and Vogel, 1993; Chen, 2015). Given enough input-output data, ANN is able to estimate any continuous function to arbitrary accuracy (Braik *et al*., 2008). In addition, several different efforts have been proposed by various researchers to propitiate this training problem

(Sexton *et al*., 2004). Almasri and Kaluarachchi (2005) applied the modular neural networks to predict the nitrate distribution in the groundwater using the on-ground nitrogen loading and recharge data. Khandelwal and Singh (2005) predicted the mine water quality by the physical parameters using back propagation neural network and multiple linear regression. Erzin and Yukselen (2009) used the back propagation neural network for the prediction of zeta potential of kaolinite. Singh *et al*. (2009) modeled the back propagation neural network to predict water quality in the Gomti River (India). Rooki *et al*. (2011) predicted the heavy metals in acid mine drainage using ANN from the Shur River of the Sarcheshmeh porphyry copper mine, southeast Iran (Rooki *et al*., 2011). The main objective of this study is to predict the heavy metals contamination in the groundwater of Arak city using artificial neural network approach. For this purpose the values of Zinc, Lead and copper are predicted using multilayer perceptron (MLP) neural network. The traditional method is used to compare the predicted results obtained from the developed ANN model. As a first case study for predicting the heavy metals pollution in the groundwater quality of Arak region, it is expected that the developed approach may be used as a new tool to improve current situation of groundwater quality management practices in Arak region.

2- Material and methods

2.1- Area Descriptions

The study area is located in the center of Iran and characterized by a semi-arid climate and an average precipitation and temperature of about 280 mm/year and 11º C, respectively (Zamani, 1999). Most of its inhabitants are concentrated in town of Arak with more than 400000 inhabitants and work mainly in the industrial plants (Fig. 1). The aquifer of Arak is developed into the medium to fine phases of the Pliestocene sediments, which occupy a broad graben between mounts Arak and Ashtian (Ghadimi, 2014). The bedrock of these formations is composed of Cretaceous crystalline limestone rocks. The study area is situated in the alluvial plain and aquifer is directly fed by stream water coming from different reliefs surrounding the depression inter-mountainous of Mighan playa. The plain hosts a large number of water-wells with depths varying from 70 to 150 m. Most of these wells supply water for drinking. The direction of groundwater flow around Arak plain is from southwest to northeast and toward saline Mighan playa.

Arak is one of the regions that its groundwater affected by contamination of industrial origin. The Arak is one of the industrial regions in Iran where the impact of rapid population growth and industrialization on limited natural sources and agricultural lands is progressively high and as a result, the size of uncontaminated areas is getting diminished. Due to expanding industrialization and urbanization in Arak and the unrestrained disposal of factory wastes to groundwater, it is thought that heavy metal contents in this region are high. Therefore, monitoring of this change and determination of contamination in the groundwater has gained importance.

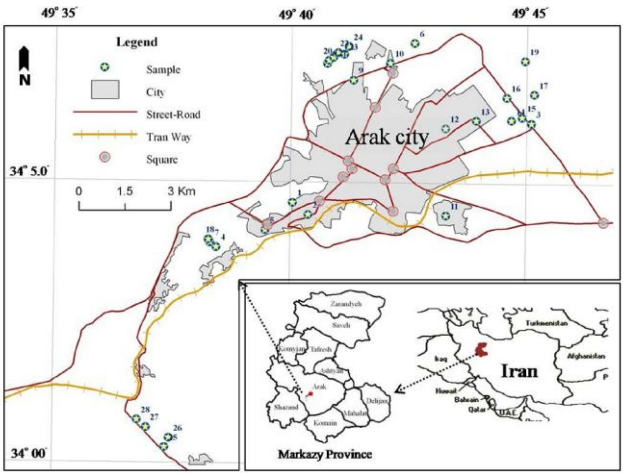

Figure 1) Map of the study area showing location and some of sample locations.

2.2- Groundwater sampling

Water samples were collected from shallow wells for urban water supply using standard sampling procedures during sampling campaigns in 2014. It can be predicted as one year ahead. The shallow wells were drilled to depths between 70 and 150 m. Total of 150 samples were taken for this study. Samples were collected in 250 ml sterilized polythene bottles. All samples were analyzed for main chemical descriptors using standard methods. Parameters analyzed include major ions of calcium (Ca), magnesium (Mg), potassium (K), sodium (Na), chloride (Cl), sulphate (SO_4) in milligram per liter using ion chromatograph (I.C). Bicarbonate ion concentration in water was determined by titration. Heavy metals were determined by Graphite Furnace Atomic Absorption Spectrophotometer (Perkin–Elmer Analyst 700) using multi element Perkin–Elmer standard solutions. Accuracy of chemical analysis was verified by calculating ion-balance errors where errors were generally within 10%.

2.3- ANN model

Artificial neural networks (ANN) are parallel information processing methods, which can express nonlinear relationship and complex use number of input–output training patterns from the experimental data. The ANN provides a nonlinear mapping between outputs and inputs by its intrinsic ability (Hornik *et al*., 1989). The success in obtaining a reliable and robust network depends on the appropriate data preprocessing, appropriate architecture selection, and appropriate network training choice robustly (Garcia *et al*., 2003). Multi-layer perceptron (MLP), the most famous type of ANN, consists of at least three layers: input, intermediate or hidden layers and output (Fig. 2). Difficulty level of the problem determines the number of the hidden layers and neurons (Simpson, 1990). The neurons are linked from a layer to the next one, but this connection is not within the same layer. Once a series of inputs presents to the network, the input values are transmitted through the links to the second layer. In every link, the transmitted value is multiplied to the weight of the link. The weighted values are come together at a node in the hidden layer and a bias is summed to the weighted values in that particular node. Consequently, the achieved value transfer to an activation function and a signal is created. Using the departing links of hidden nodes, the results are transmitted to the output layer. Similar to hidden nodes, the input values of the output nodes are weighted, biased, summed and transferred to the activation function. The created values of activation functions in output layers are the outputs of the network. Performance of an ANN is dependent on architecture of the network which is the pattern of the connections existing between the neurons. The network should be trained with sufficient input–output patterns that are known as the training data (Meulenkamp and Grima, 1999). As the error reached specified error goal, training is finished and the optimum model is determined. Then, this trained network can be used to test the model.

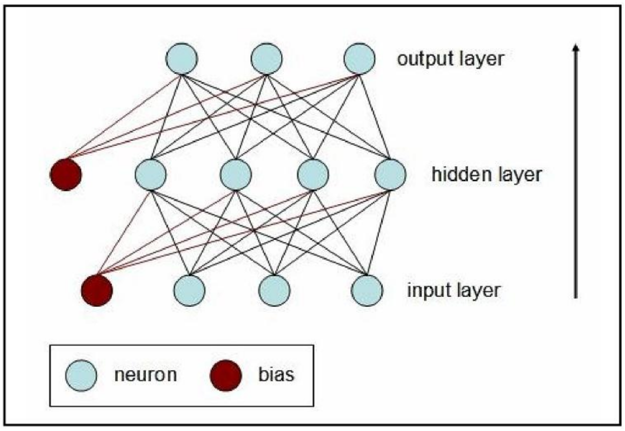

Figure 2) A schematic diagram of a fully connected MLP neural network with three inputs, four hidden units (neurons), and three outputs. Note that the hidden and output layers have a bias term. Bias is a neuron that emits a signal with strength 1.

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network (Collobert and Bengio, 2004). MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable. If a multilayer perceptron has a linear activation function in all neurons, that is, a linear function that maps the weighted inputs to the output of each neuron, then it is easily proved with linear algebra that any number of layers can be reduced to the standard two-layer input-output model. What makes a multilayer perceptron different is that some neurons use a nonlinear activation function which was developed to model the frequency of action potentials of biological neurons in the brain. This function is modeled in several ways. The main activation functions used in current applications is described by Eq.1:

$$y(v_i) = \tanh(v_i) \quad \text{and} \quad y(v_i) = (1+e^{-v_i})^{-1} \qquad \text{Eq.1}$$

Which the former function is a hyperbolic tangent which ranges from -1 to 1, and the latter, the logistic function, is similar in shape but ranges from 0 to 1. Here is the output of the node (neuron) and is the weighted sum of the input synapses. Alternative activation functions have been proposed, including the rectifier and soft plus functions. More specialized activation functions include radial basis functions which are used in another class of supervised neural network models. The multilayer perceptron consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes and is thus considered a deep neural network. Each node in one layer connects with a certain weight to every node in the following layer. For an MLP with two layers (MLP2) it is recommended that you use the tan h (hyperbolic) function although other types are also possible such as the logistic sigmoid and exponential functions. The output neuron activation functions are, for most cases, set to the identity but this may vary from task to task.

2.4- Multiple linear regression

Multiple linear regression (MLR) is an extension of the regression analysis that incorporates additional independent variables in the predictive equation (Rooki *et al*., 2011). Here, the model to be fitted is by Eq.2:

$$Y = B_1 + B_2 X_2 + ... + \mathrm{B}_n X_n \qquad \text{Eq.2}$$

Where Y is the dependent variable, X is the independent random variables The parameters B is, stand for the regression coefficients, are unknown and are to be estimated. However, there is usually substantial variation of the observed points around the fitted regression line. The deviation of a particular point from the regression line (its predicted value) is called the residual value. The smaller the variability of the residual values around the regression line, the better is model prediction.

3- Results and discussion

3.1- Hydrochemistry of groundwater

Table 1 presents the statistical summary of all the parameters analyzed for this study. The mean concentrations of the major ions in the groundwater of Arak city are within the Iran Standard Guidelines (TTPW, 2011) for drinking water. The maximum Cl and SO_4 concentrations of 242 mg/l and 320 mg/l respectively are, however higher than their respective Iran Standard Guidelines of 200 mg/l, and 250 mg/l. These are resulted from contamination of sources such as domestic sewage and agricultural activities.

Maximum concentrations of some of the major ions such as Na are higher than the Iran Standard. All other major parameters have concentrations lower than the standard guideline limits. The aquifers of the alluvial Arak, which are mostly sedimentary aquifers, therefore produce groundwater of acceptable quality for most uses. In the groundwater of Arak aquifers, concentrations of the Pb, Zn and Cu are higher than the recommended Iran Standard Guidelines. The value of Pb, Zn and Cu ranges from 3 to 9 mg/l, 4 to 50 mg/l and 2 to 52 mg/l in the groundwater and the recommended Iran Standard ranges 0.05mg/l, 5 mg/l and 0.05 mg/l, respectively.

Table 1) Statistical characteristics of hydrochemical variables in groundwater (All unites are mg/l).

Variables	Mean	Median	Minimum	Maximum
Cl	95	79	6.50	242
SO_4	213	245	23	320
HCO_3	166	156	55	410
Ca	242	250	80	400
Mg	22	20	7.50	43
Na	215	205	38	400
K	0.88	0.82	0.30	1.90
Fe	0.02	0.02	0.01	0.23
Mn	0.01	0.01	0.001	0.10
Pb	7.10	7	3	9
Zn	16	14	4	50
Cu	14	12	2	52

Hydrochemical data of groundwater were mapped on Piper Diagram (Fig. 3). The central diamond-shaped figure displays that most groundwater samples are concentrated in three groups: (1) Ca-HCO_3 type waters: (2) Ca-HCO_3-Cl type waters and (3) Ca-Na-HCO_3-Cl type waters. All of groundwater are rich in HCO_3 and Ca and these wells emerged from limestone rocks groundwater have Ca>Mg>Na and HCO_3>Cl>SO_4. Some of samples that have high Na and Cl belong to Ca-Na-HCO_3-Cl type water and show mixing of saline water with fresh water towards very low brackish.

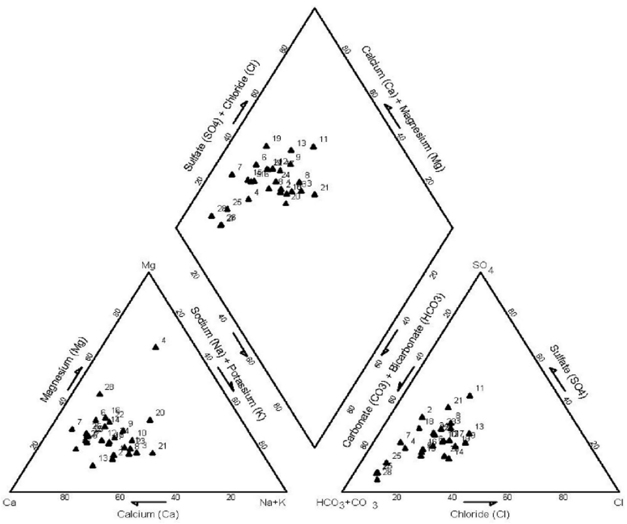

Figure 3) Piper classification diagram of groundwater samples.

3.2- Estimation of heavy metals using ANN model

To simulate heavy metals in groundwater using ANN model, all relevant parameters should be determined, due to the fact that ANN work based on given data and do not have previous knowledge about the subject of prediction. Following sections describe the input and output parameters and simulation of heavy metals in groundwater using ANN model. According to the correlation matrix HCO_3 and SO_4 that have most dependent on heavy metals (Pb, Zn and Cu) concentrations were selected as inputs of the network (Table 2). The outputs of network were heavy metals concentrations including Pb, Zn and Cu. In ANN modeling, any type of input can be used as long as they have effects on output results.

To train and verify the accuracy and ability of the ANN model, a total of 150 data samples records in groundwater from Arak city, were used in this research. In total, two input parameters including HCO_3 and SO_4 (major ions) and output including Pb, Zn and Cu (heavy metals) were used to estimation of heavy metals in groundwater from Arak city.

3.2.1- Pre-processing of data

In data-driven system modeling methods, some pre-processing steps are usually implemented prior to any calculations, to eliminate any outliers, missing values or bad data. This step confirms that the raw data retrieved from database is perfectly proper for modeling. In order to softening the training procedure and improving the accuracy of prediction, all data samples are normalized to adapt to the interval [0, 1] according to the following linear mapping function (Eq. 3) and selected 75% data as training and 25% as test data.

$$x_M = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \text{Eq.3}$$

Where *x* is the original value from the dataset, x_M is the mapped value, and x_{min} (x_{max}) denotes the minimum (maximum) raw input values, respectively. It is to be noted that model outputs will be remapped to their corresponding real values by the inverse mapping function ahead of calculating any performance criterion.

Table 2) Correlation matrix between heavy metals concentrations and independent variables.

	Cl	SO_4	HCO_3	Ca	Mg	Na	K	Fe	Mn	Pb	Zn	Cu
Cl	1.00											
SO_4	0.29	1.00										
HCO_3	-0.07	0.11	1.00									
Ca	0.35	0.56	0.36	1.00								
Mg	0.39	0.04	-0.01	0.19	1.00							
Na	0.49	0.68	0.42	0.39	0.00	1.00						
K	0.31	0.41	0.21	0.24	0.09	0.60	1.00					
Fe	0.05	-0.04	0.02	-0.11	0.03	0.13	0.05	1.00				
Mn	-0.05	0.05	0.15	0.06	0.08	0.04	0.09	0.07	1.00			
Pb	0.28	***0.85***	0.16	0.39	0.05	0.66	0.42	-0.07	0.07	1.00		
Zn	-0.03	0.16	***0.80***	0.36	0.11	0.35	0.19	0.01	0.19	0.19	1.00	
Cu	-0.08	0.15	***0.74***	0.35	0.23	0.28	0.14	-0.01	0.03	0.19	0.87	1.00

3.2.2- Network architecture

Architecture of the ANN model includes type of network, number of input and output neurons, transfer function, number of hidden layers as well as number of hidden neurons. Generally, the input neurons and output neurons are problem specific (Braik *et al*., 2008). In this paper, multi-input multi-output structure had been utilized. The architecture of the network is given in Table 3. Also, in this study, tansig was used as transfer function between input and hidden layer, as well as was used as transfer function between hidden and output layer, shown by Eq. 4:

$$\text{tansig} = \frac{2}{(1+\exp(-2x))} - 1 \quad \text{Eq.4}$$

To evaluate the performances of the ANN model, root mean squared error (RMSE), squared correlation coefficient (R^2) and variance account for (VAF) were chosen to be the measure of accuracy. Let y_k be the actual value and $\hat{y}_k$ be the predicted value of the k^{th} observation and *n* be the number of samples. The higher the R^2 and VAF the better is the model performance. For instance, a R^2 and VAF of 100% means that the measured output has been predicted exactly (perfect model). R^2 and VAF=0 means that the model performs as poorly as a predictor using simply the mean value of the data. Also, the lower RMSE indicates the better performance of the model. RMSE, R^2 and could be defined, respectively, as Eqs.5, 6, and 7:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2} \quad \text{Eq.5}$$

$$R^2 = \frac{(\sum_{k=1}^{n} y_k \hat{y}_k - n\mu_y \mu_{\hat{y}})^2}{(\sum_{k=1}^{n} \hat{y}_k^2 - n\mu_y{}^2)^2 (\sum_{k=1}^{n} \hat{y}_k^2 - n\mu_{\hat{y}}^2)^2} \quad \text{Eq.6}$$

$$\mathrm{VAF} = \left(1 - \frac{\mathrm{Var}(yk - \hat{y}_k)}{\mathrm{Var}(yk)}\right).100\% \quad \text{Eq.7}$$

Where μ_y ($\mu_{\hat{y}}$) denotes the mean value of the μ_k ($\mu_{\hat{k}}$), $k = 1,...,n,$ respectively and var. denotes the variance.

Table 3) The architecture of the network.

Parameter	Value
No. of input neurons	2
No. of output neurons	3
No. of hidden layers	2
No. of neurons in first hidden layer	3
No. of neurons in second hidden layer	9
No. of training data sets	120
No. of testing data sets	30

A comparison between predicted values of heavy metals in the groundwater by the ANN model and measured values for 150 data sets at training and testing phases is shown in Figures 4 and 5. As shown in figures 4 and 5 the results of the ANN model in comparison with actual data show a good precision of the ANN model (Table 4).

The performance indices obtained in Table 4 indicate the high performance of the ANN model that can be used successfully for the estimation of heavy metals in the groundwater. Furthermore, correlation between measured and predicted values of heavy metals in the groundwater for training and testing phases are shown in Figures 6 and 7.

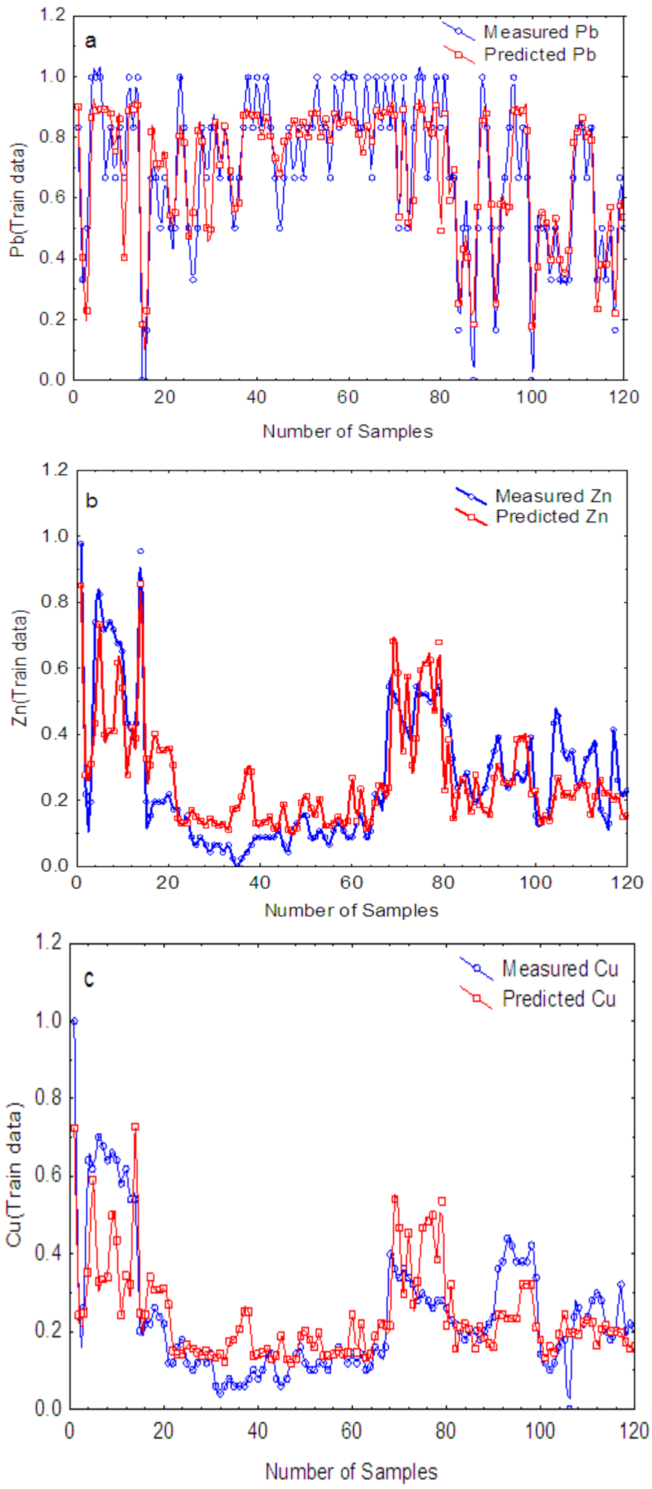

Figure 4) Comparison between measured and predicted heavy metals in the groundwater for training data sets: a) Pb; b) Zn; c) Cu.

In order to increase the accuracy and applicability of ANN for estimation of heavy metals in groundwater, ANN algorithm was used to weighting ANN. Several ANN models were trained and tested using obtained data from Arak city, to determine the optimum network. Performances of the selected ANN model using training and testing dataset are shown in figures 4, 5, 6, and 7 and Table 4.

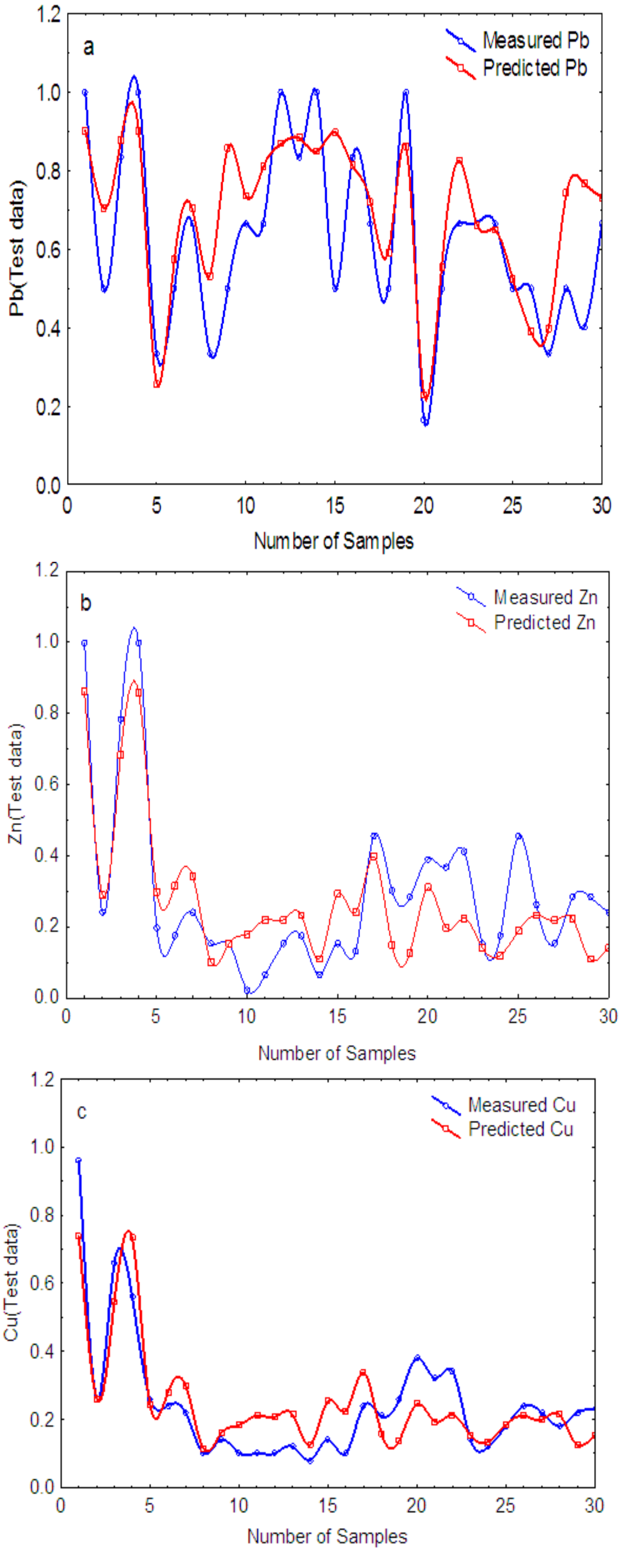

Figure 5) Comparison between measured and predicted heavy metals in the groundwater for testing data sets: a) Pb; b) Zn; c) Cu.

The predicted heavy metals fit the measured heavy metals almost perfectly for training datasets. Nevertheless, the predicted heavy metals denote fit perfectly to the measured heavy metals for testing datasets. This might be caused by a lack of training data in that range. In general, it can be said that the proposed ANN model is able to predict heavy metals with high degree of accuracy.

Table 4) Performance of the model for estimation of heavy metals in the groundwater.

Description		R^2	RMSE	VAF%
Pb	Training datasets	0.75	0.123	76
	Testing datasets	0.71	0.158	59
Zn	Training datasets	0.65	0.121	66
	Testing datasets	0.71	0.124	75
Cu	Training datasets	0.51	0.014	51
	Testing datasets	0.70	0.094	97

Table 5) Statistical characteristics of the multiple regression models.

	Pb		Zn		Cu	
	Train	Test	Train	Test	Train	Test
$SO_{4(B2)}$	0.866	0.737	0.073	0.062	0.09	0.005
$HCo_{3\ (B3)}$	0.036	0.209	0.782	0.822	0.697	0.832
$Intercept_{(B1)}$	0.201	0.16	-0.043	-0.03	0.0005	0.024
R^2	65	63	60	62	49	62

In this study, multiple linear regression (MLR) analysis was performed using the training and test data employed in neural network data. Heavy metal concentrations were considered as the dependent variables and HCO_3 and SO_4 were considered as the independent variables. A computer-based package called *Statistica* was used to carry out the regression analysis. The estimated regression relationships for heavy metals are given in Table 5. Heavy metal concentrations were estimated according to the Eq. 2. Figures 8 and 9 show the correlation between measured heavy metal concentrations and those predicted using MLR with two inputs.

Table 6 compares the correlation coefficient R associated with two methods for both training and test data. Low correlation values between the model predictions and measured data using MLR method describes its low capability in prediction heavy metals. As can be seen in

Table 6, the inappropriate predictions of the heavy metals shown by less values of R^2 is the most important disadvantage of the MLR method compared to ANN method.

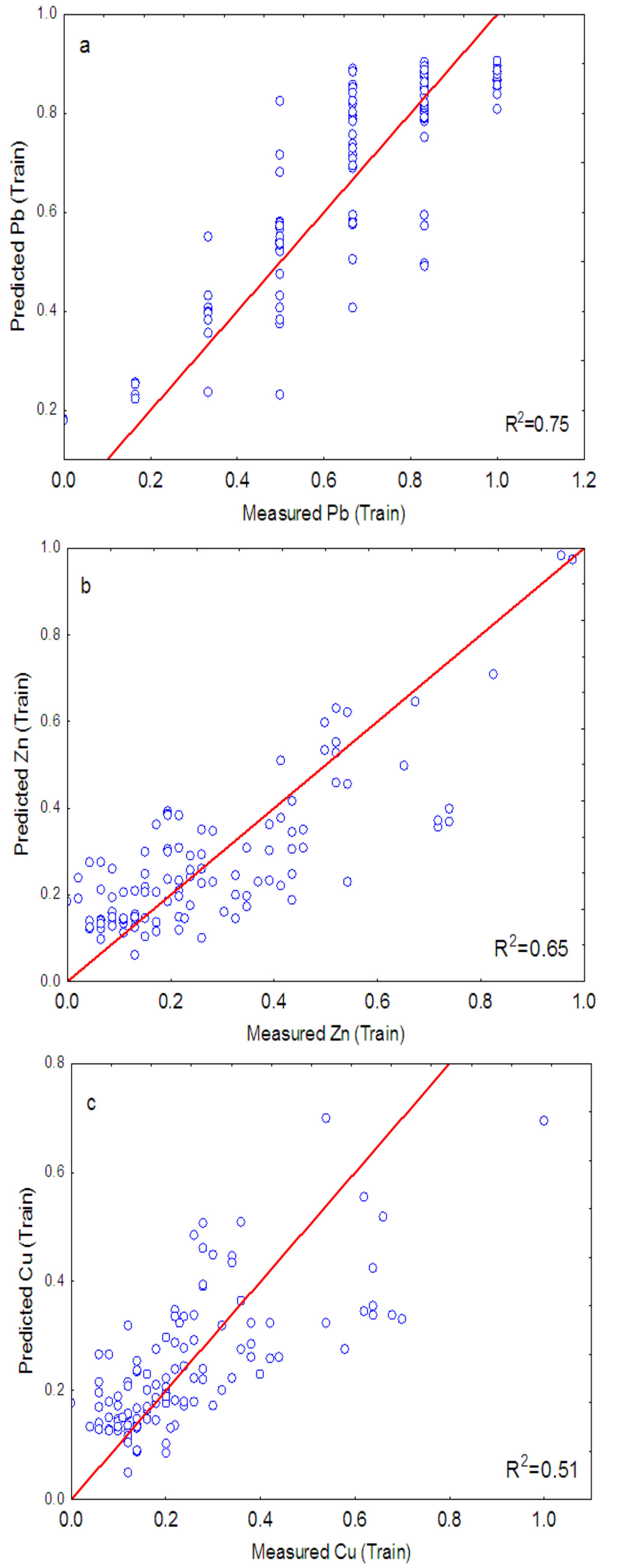

Figure 6) Correlation between measured and predicted values of heavy metals in the groundwater for training data sets: a) Pb; b) Zn; c) Cu.

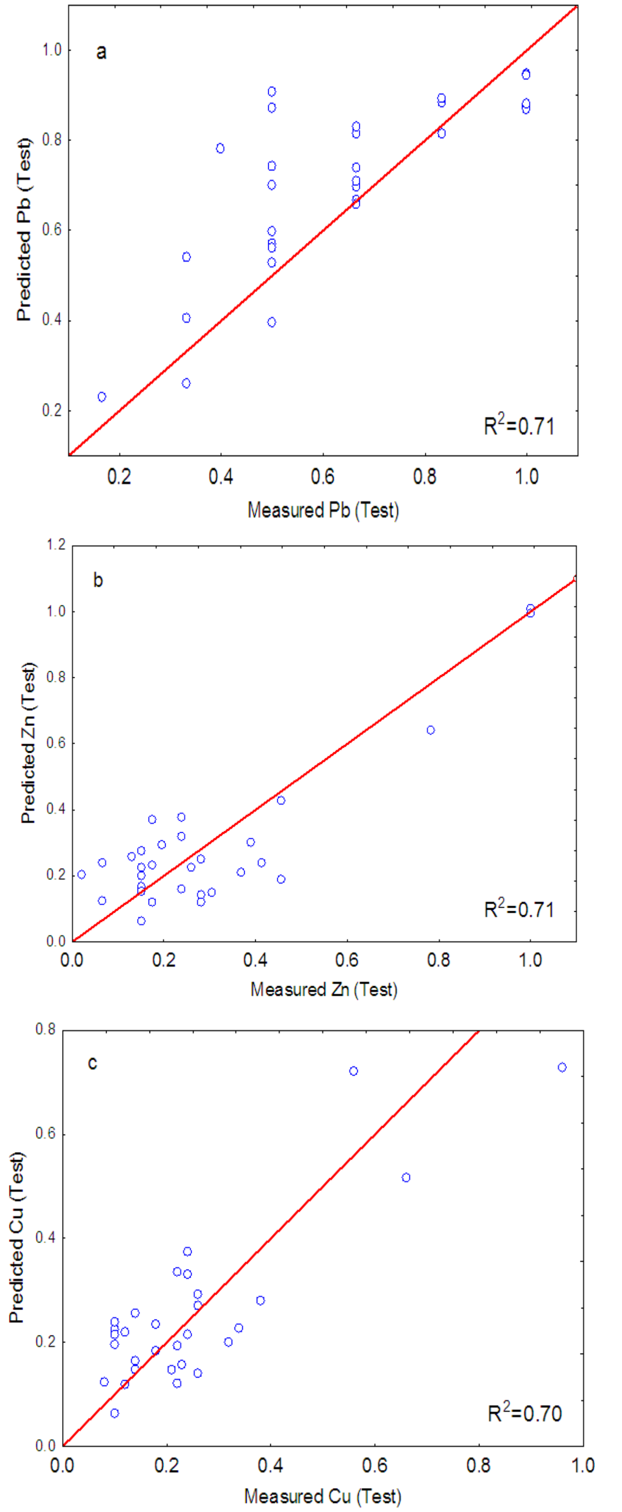

Figure 7) Correlation between measured and predicted values of heavy metals in the groundwater for testing data sets a) Pb, b) Zn c) Cu.

Table 6) Comparison of the results (R^2) of two methods in training and test data.

Method	Pb		Zn		Cu	
	Train R^2	Test R^2	Train R^2	Test R^2	Train R^2	Test R^2
ANN	0.75	0.71	0.65	0.71	0.51	0.70
MLR	0.65	0.63	0.60	0.62	0.49	0.62

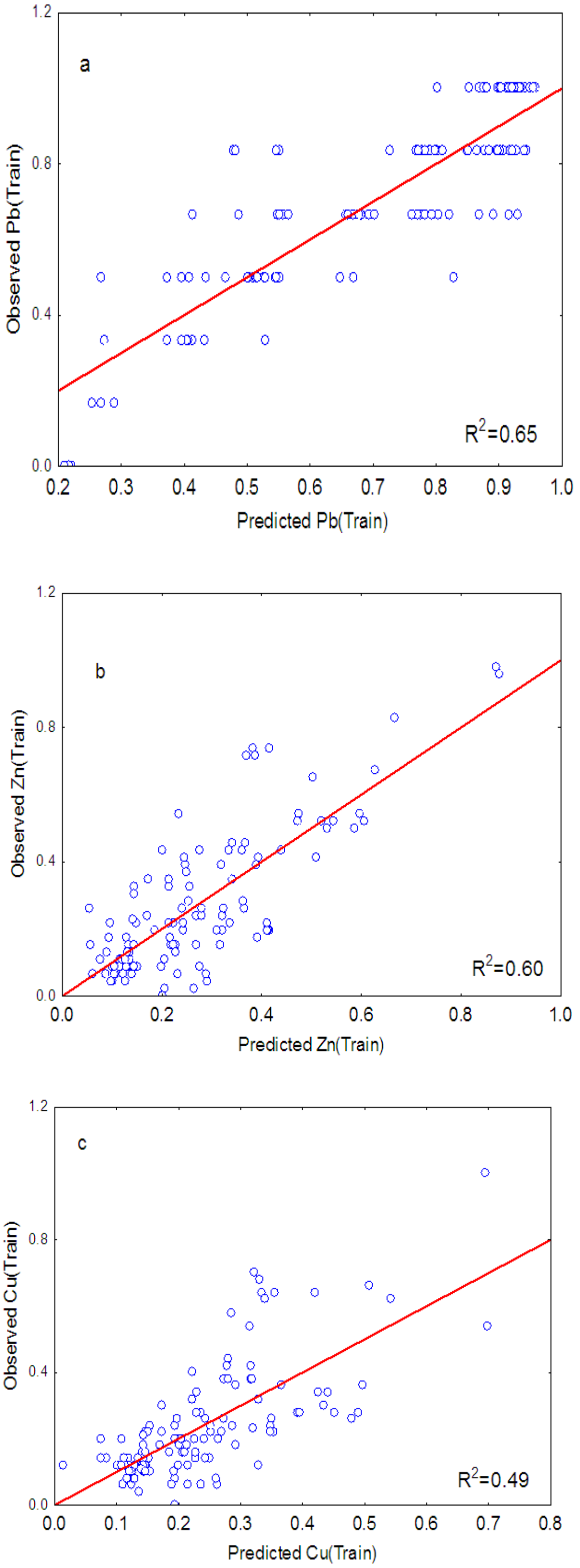

Figure 8) Comparison of the predicted concentrations using MLR and measured concentrations for training data set: a) Pb; b) Zn; c) Cu.

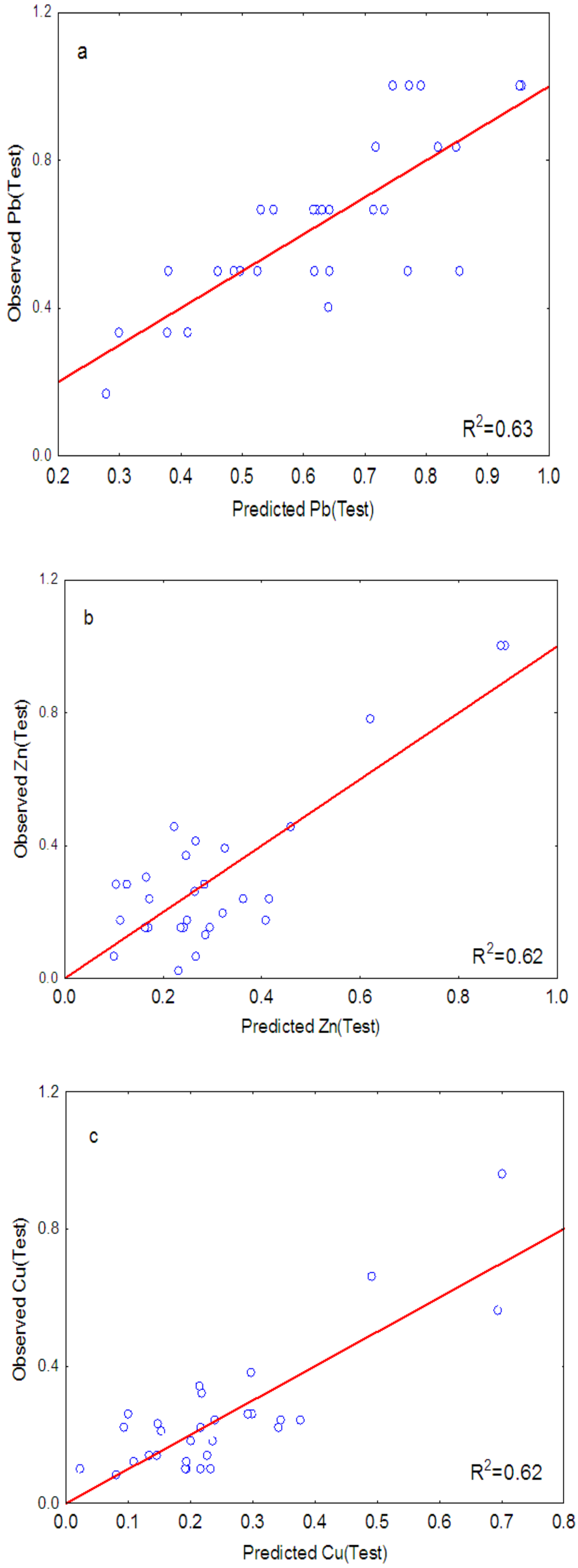

Figure 9) Comparison of the predicted concentrations using MLR and measured concentrations for testing data set, a) Pb, b) Zn, c) Cu.

4- Conclusion

High concentrations of Pb, Zn and Cu were found in the groundwater of Arak City. Heavy metals were emitted mainly by anthropogenic sources. In this paper, ANN model was developed to estimation of heavy metals in the groundwater from Arak city, Iran. To generate the proposed ANN model, a dataset consists of 150 samples was used. Two parameters including HCO_3, SO_4, and (major ions) were used as input parameters and Pb, Zn and Cu (heavy metals) were used as output parameters. Consequently, it may conclude that ANN is a reliable system modeling technique for estimation of heavy metals in the groundwater from Arak city with highly acceptable degree of accuracy and robustness. Low correlation values between the model predictions and measured data using MLR method describes its low capability in prediction heavy metals.

Acknowledgments:

Chemical analyses were conducted at the laboratory of the Department of Mining Engineering, Arak University of Technology. The authors thank Arak University of Technology for their continuous support.

References:

Almasri, M. N., Kaluarachchi, J. J. 2005. Modular neural networks to predict the nitrate distribution in groundwater using the on-ground nitrogen loading and recharge data. Environmental Modell Software: 20, 851–871.

Anderson, R. B., Naftz, D. L., Day-Lewis, F. D., Henderson, R. D., Rosenberry, D. O., Stolp, B. J., Jewell, P. 2014. Quantity and quality of groundwater discharge in a hypersaline lake environment. Journal of Hydrology: 512, 177–194.

Andre, L., Franceschi, M., Pouchan, P., Atteia, O. 2005. Using geochemical data and modeling to enhance the understanding of groundwater flow in a regional deep aquifer, Aquitaine Basin, south-west of France. Journal of Hydrology: 305, 40–62.

Braik, M., Sheta, A., Arieqa, A. A. 2008. comparison between GAs and PSO in training ANN to model the TE chemical process reactor. In: Proceedings of the AISB 2008 symposium on swarm intelligence algorithms and applications, 24–30.

Chen, S. T. 2015. Mining Informative Hydrologic Data by Using Support Vector Machines and Elucidating Mined Data according to Information Entropy. Entropy: 17, 1023–1041.

Collobert, R., Bengio, S. 2004. Links between Perceptrons, MLPs and SVMs. Proc. International Conference on Machine Learning (ICML).

Devic, G., Djordjevic, D., Sakan, S. 2014. Natural and anthropogenic factors affecting the groundwater quality in Serbia. Science of The Total Environment: 468–469, 933–942.

Downs, J. J., Vogel, E. F. 1993. A plant-wide industrial process control problem. Computers and Chemical Engineering: 17, 245–255.

García-Pedrajas, N., Hervás-Martínez, C., Muñoz-Pérez, J. 2003. COVNET: a cooperative coevolutionary model for evolving artificial neural networks. Neural Networks IEEE Transactions on: 14, 575–596.

Ghadimi, F. 2014. Assessment of heavy metals contamination in urban topsoil from Arak industrial City, Iran, Journal of Tethys: 2, 196–209.

Ghadimi, F., Ghomi, M. 2013. Assessment of the effects of municipal wastewater on the heavy metal pollution of water and sediment in Arak Mighan Lake, Iran. Journal of Tethys: 1, 205–214.

Ghomi, M., Ghadimi, F., Hajati, A. 2013. Evolution of soil contamination in the abandoned Lakan Lead and Zinc mine by heavy metals, Iran. Journal of Tethys: 1, 12–28.

Erzin, Y., Yukselen, Y. 2009. The use of neural networks for the prediction of zeta potential of kaolinite. Mathematical Geosciences: 41 779–797.

Helstrup, T., Jorgensen, N. O., Banoeng-Yakubo, B. 2007. Investigation of hydrochemical characteristics of groundwater from Cretaceous–Eocene limestone aquifers in southern Ghana and Togo using hierarchical cluster analysis. Hydrogeology: 15, 977–989.

Hornik, K., Stinchcombe, M., White, H. 1989. Multilayer feed forward networks are universal approximators. Neural networks, 2 , 359–366.

Khandelwal. M., Singh, T. 2005. Prediction of mine water quality by physical parameters. Journal of Scientific and Industrial Research: 64, 564–570.

Masoud, A. A. 2014. Groundwater quality assessment of the shallow aquifers west of the Nile Delta (Egypt) using multivariate statistical and geostatistical techniques. Journal of African Earth Sciences: 95, 123–137.

Matiatos, I., Alexopoulos, A., Godelitsas, A. 2014. Multivariate statistical analysis of the hydrogeochemical and isotopic composition of the groundwater resources in northeastern Peloponnesus (Greece). Science of The Total Environment, 476–477, 577–590.

Meulenkamp, F., Grima, M. A. 1999. Application of neural networks for the prediction of the unconfined compressive strength (UCS) from Equotip hardness. International Journal Rock Mechanic Minning Science: 36, 29–39.

Monjerezi, M., Vogt, M. R., Aagaard, P., Saka, J. D. K. 2011. Hydro-geochemical processes in an area with saline groundwater in lower Shire River valley, Malawi: An integrated application of hierarchical cluster and principal component analyses. Applied Geochemistry: 26, 1399–1413.

Oinam, J. D., Ramanathan, A. L., Singh, G. 2012. Geochemical and statistical evaluation of groundwater in Imphal and Thoubal district of Manipur, India. Journal of Asian Earth Sciences: 48, 136–149.

Rajmohan, N., Elango, E. L. 2004. Identification and evolution of hydrogeochemical processes in the groundwater environment in an area of the Palar and Cheyyar River Basins, Southern India. Environmental Geology: 46, 47–61.

Rooki, R., Ardejani, F. D., Aryafar, A., Asadi, A. B. 2011. Prediction of heavy metals in acid mine drainage using artificial neural network from the Shur River of the Sarcheshmeh porphyry copper mine, Southeast Iran. Environmental Earth Sciences: 64, 1303–1316.

Sexton, R.S., Dorsey, R.E., Sikander, N.A. 2004. Simultaneous optimization of neural network function and architecture algorithm. Decision Support Systems: 36, 283–296.

Simpson, P. K. 1990. Artificial neural system—foundation, paradigm, application and implementations. Pergamon Press, New York.

Singh, K. P., Basant, A., Malik, A., Jain. G. 2009. Artificial neural network modeling of the river water quality– a case study. Ecological Model: 220, 888–895.

Singh, K. P., Gupta, S., Mohan, D. 2014. Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. Journal of Hydrology: 511, 254–266.

Srinivasamoorthy, K., Chidambaram, S., Vasanthavigar, M., Sarma, V. S. 2014. Hydrochemical characterization and quality appraisal of groundwater from Pungar sub basin, Tamilnadu, India. Journal of King Saud University. Science: 26, 37–52.

TTPW. 2011. Tehran Province Water and Wastewater. Water standard of Iran.

Wang, P., Yu, J., Zhang, Y., Liu, C. 2013. Groundwater recharge and hydrogeochemical evolution in the Ejina Basin, northwest China. Journal of Hydrology: 476, 72–86.

Yidana, S. M. 2010. Groundwater classification using multivariate statistical methods: Southern Ghana. Journal of African Earth Sciences: 57, 455–469.

Zamani, F. 1999. Sedimentology of Arak Mighan lake. MSc Thesis in Shahid Beheshti University Tehran, Iran.

Zapata, E. P., Ruiz, R. L., Harter, T., Ramírez, A. I., Mahlknecht, J. 2014. Assessment of sources and fate of nitrate in shallow groundwater of an agricultural area by using a multi-tracer approach. Science of The Total Environment: 470–471, 855–864.

Zghibi, A., Merzougui, A., Zouhri, L., Tarhouni, J. 2014. Understanding groundwater chemistry using multivariate statistics techniques to the study of contamination in the Korba unconfined aquifer system of Cap-Bon (North-east of

Tunisia). Journal of African Earth Sciences: 89, 1–15.

Received: 18 April 2015 / Accepted: 12 July 2015 / Published online: 31 July 2015