DESERT

Groundwater quality assessment using artificial neural network: A case study of Bahabad plain, Yazd, Iran

Z. Kheradpisheh[a*], A. Talebi[b], L. Rafati[a], M.T. Ghaneeian[a], M.H. Ehrampoush[a]

[a] *Environmental Health Faculty, Shahid Sadoughi University of Medical Sciences. Yazd, Iran*
[b] *Faculty of Natural Resources, Yazd University, Yazd, Iran*

Abstract

Groundwater quality management is the most important issue in many arid and semi-arid countries, including Iran. Artificial neural network (ANN) has an extensive range of applications in water resources management. In this study, artificial neural network was developed using MATLAB R2013 software package, and Cl, EC, SO_4 and NO_3 qualitative parameters were estimated and compared with the measured values, in order to evaluate the influence of key input parameters. The number of neurons in the hidden layer was obtained by the trial-and-error method. For this purpose, data from 260 water samples of 13 wells in Bahabad plain were collected during 2003- 2013. The results show that the performance of ANN model was more accurate for Cl (R=0.96), EC(R=0.98), and SO_4(R=0.95), using back-propagation algorithms according to the best chosen input parameters. It was observed that the use of ANN model for NO_3 was not very accurate, perhaps this was because of the different water sources or the impact of other parameters; thus, this result is in contrast with the study of Diamantopoulou *et al.* (2005). However, this study confirms that the number of neurons in the hidden layer cannot be found using a specific formula (double the number of inputs plus one) for all parameters but can be obtained using a trial-and-error method.

Keywords: Artificial neural networks; Modeling; Groundwater quality; Water resource

1. Introduction

The management and protection of groundwater resources is of great importance in countries such as Iran, located in arid and semiarid regions lacking surface water resources. Thus, preservation in addition to short and long-term planning of resources, is necessary to optimize their efficiency (Abu-Khalaf *et al.*, 2013). Bahabad is located in Iran's central plateau river basin, and groundwater is the main water source (Kheradpisheh *et al.*, 2014). In this region, the quality of water has been threatened in recent years by agricultural and industrial activities, and researchers have given more attention to quality parameters like EC, Cl, SO_4 and NO_3 (Kheradpisheh *et al.*, 2014). The utilization, management and protection of ground water should be a basic principle of Country planning. In this regard, the model can be used as an efficient tool for managers (Moosavi *et al.*, 2013). A neural network model is a conceptual model and in fact, is a simplified image of the mathematical model. The biggest problem faced by users and suppliers of mathematical models, is the need of these models to exact various input data. Artificial neural networks (ANN) which are driven by biological neural networks can help in solving such problems. These networks are a part of the intelligent systems, developed with various spread structures (Chitsazan *et al.*, 2013).

Artificial Intelligence is the hottest debate among experts in computer science and information technology as well as other scientists

---

* Corresponding author. Tel.: +98 912 6185023,
   Fax: +98 21 77043976.
   E-mail address: z.kheradpisheh@yahoo.com

and decision makers (Cuesta Cordoba, 2011). The ultimate goal of Artificial Intelligence is to build an artificial human, by making a software program that can think like humans, such machines can make the best decisions under critical conditions by combining advanced analytical techniques and the ability of analyst's engineers, politicians, and other scientists based on the huge volume of information resources available. Neuro Solution and MATLAB software are applications used in neural network modeling (Moosavi *et al.*, 2013). The first is the set of input and output layers, the number of this input and output are determined according to the conditions governing the issue, the existing database and

other factors. The second part is to determine the number of hidden layers; the hidden layer has a major role in the Network power. The third part of the neural network is the network training and testing. The meaning of network training in an artificial neural network is the adjustment of weight input parameters. The input database, related to a part of the network database, was presented during the training process. Output values and the target value were compared, and the weights were corrected according to the amount of error. After this step, the weight's values were stored and the network for the other part of the data not used during the training phase, was tested (Huang *et al.*, 2011).
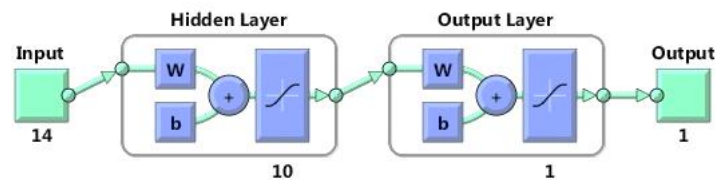

Fig .1. Example of Artificial Neural Network (ANN), MATLAB R2013 software

Artificial neural network models (ANN) are widely used as a black-box model for water in stream flow forecasting, groundwater, water quality, water management policy, precipitation forecasting, and reservoir operations, since the early 19th century (Nourani *et al.,* 2012). Seyam *et al.* (2011) and Li *et al.* (2012) recently showed that this model does not require complex processes, that occur in the environment. Cordoba (2011) analyzed the impact of physical parameters on water distribution using available historical data. Dutta *et al.* (2010) used the ANN model for adsorption and photo catalysis of reactive dye on the TiO_2 surface system.

Kulkarni *et al.* (2010) quantified the formation of trihalomethanes (THM_s), haloacetic acids (HAAs), and total organic halide (TOX) using the ANN model. Venkat Kumar *et al.* (2010) used the ANN model to predict water parameters from a few known parameters. Sadiq *et al.* (2004) and Chowdhury *et al.* (2009) developed an ANN model for predicting the disinfection of byproducts (DBP) formed in drinking waters. Diamantopoulou *et al.* (2005) and Venkat Kunwar *et al.* (2009) estimated dissolved oxygen (DO) and biochemical oxygen demand (BOD) levels in a river using the ANN model. Ming Kuo *et al.* (2004), Ying *et al.* (2007) and Palani *et al.* (2008)

used the ANN model to predict and forecast the quantitative characteristics of water bodies. Garcia *et al.* (2006) estimated the ability of the ANN to provide a data-driven approximation of the explicit relationship between transmissivity and hydraulic head, as described by the flow equation of ground water. Sahoo *et al*. (2005) applied the ANN for the prediction of pesticides in rural domestic wells. Karul *et al*. (2000) and Panda *et al.* (2004) estimated Lake Water Quality by the ANN model.

Fortunately, the importance of qualitative modeling of aquifer management in Iran, has been on the increase for the past few years. It is necessary to develop a system for monitoring groundwater quality. Sampling systems are the main sources of data for groundwater mathematical models and are the main cause of improper performance. Therefore, creating a system for monitoring and sampling groundwater is more important. Modeling of ground water seems necessary because of the status of water quality in each region. This study used an artificial neural network in estimating the quality parameters and compared them with the measured values, in order to evaluate the influence of key input parameters to the neural network.

2. Material and methods

2.1. Study area

Bahabad plain is one of the plains of Yazd Province in Central Iran. The geographical location of this plain is 56º20′ to 56º56′ east longitude and 31º40′ to 32º16′ north latitude (Figure 2). Its center is Bahabad city, located 85 km northeast of Bafq City. This city covers an area of 8 km² and has an altitude of 1398 m above sea level. The annual long term precipitation rate of the region is 55.1 mm; average temperatures are maximum (26ºC) and minimum (10ºC). Bahabad plane is located between the heights in the north west and east south, with remarkable ground tables which are used for agricultural development, animal breeding and drinking water. Due to the paucity of precipitation and over use of

water in the region, the annual loss of groundwater amounted to 25 cm withdrawal, and resulted in an annual regional ground water deficit of 8-12 million m³ (Kheradpisheh *et al.,* 2014, WRS, 2014).

2.2. Data collection and analysis

In this research, it was necessary to gather data for training purpose, therefore data gatherad from Yazd Regional Water Company and 260 samples of 13 wells were assessed in the Bahabad plain from 2003 to 2013. Model input parameters included: (Cl, EC, NO_3, SO_4, HCO_3, Anion, Cation, TDS, TH, pH, SAR,% Na, K, Mg, Ca, Evaporation, Water level, Q) and the model output parameters included: (Cl, EC, NO_3, SO_4). Artificial neural network modeling was developed using MATLAB R2013 software package.
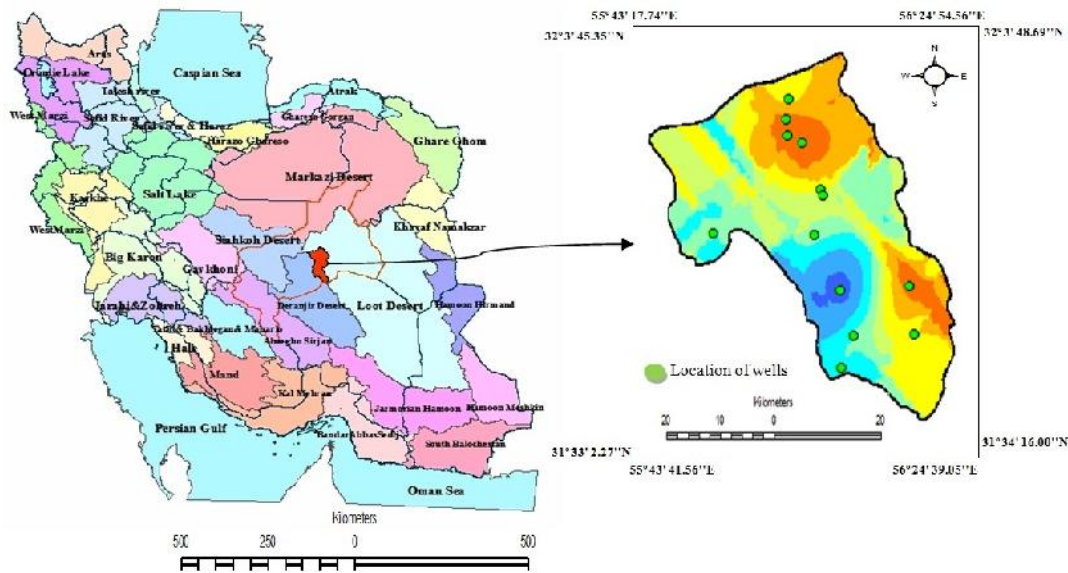

Fig. 2. Location of study region (Bahabad case study), (Kheradpisheh *et al.,* 2014)

In this research, 70% of inputs were used for training and the remaining 30% were used for test. At first, input data were normalized to the range 0–1, then the best number of epochs were estimated for each output parameter, and the root mean square error value used for the stopping criterion was set as 0.01. The best combination of inputs were estimated for Cl, EC, NO_3, and SO_4. The Coefficient of determination (R^2) (Equation 1), root mean squared error (RMSE) (Eq. 2) and Nash–Sutcliffe model efficiency coefficients were

used for modeling Cl, EC, NO_3, SO_4 for different ANN models with the best combination of inputs and best number of neurons in the hidden layer. The back-propagation algorithm was applied for modeling. This algorithm calculates the weights between the input, the hidden layers, and the output layer by modifying the number of hidden layers and the learning rate. Feed forward back propagation neural networks (FFN-BP) are relatively new tools in the earth sciences (Chitsazan *et al*., 2013, Huang *et al*., 2011). This

algorithm has been used in similar studies such as that of Moosavi *et al.* (2013), Nourani *et al.* (2012), Sahoo *et al.* (2005), Venkat Kunwar *et al.* (2009) and Ying *et al.* (2007).

3. Results and discussion

The results of the modeling were determined by input parameters, such as, Cl, EC(Electrical Conductivity), NO$_3$, SO$_4$, HCO$_3$, Anion, Cation, TDS(Total Dissolved Solids), TH (Total Hardness), pH, SAR (Sodium Absorption Ratio), %Na, K, Mg, Ca, Evaporation, Water level, Q, T(Temperature) using neural network and Coefficient of determination (R^2) (Eq. 1), root mean squared error (RMSE) (Eq. 2) and Nash–

Sutcliffe model efficiency coefficients were used to model Cl, EC, NO$_3$, and SO$_4$. Accordingly:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(o_i - \overline{o})(e_i - \overline{e})}{\sqrt{\sum_{i=1}^{n}(o_i - \overline{o})^2}\sqrt{\sum_{i=1}^{n}(e_i - \overline{e})^2}} \right)^2 \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(o_i - e_{i-}e)^2}{n}} \tag{2}$$

where o, e and n, represent the observed groundwater quality, estimated groundwater quality and number of data, respectively. In the Nash–Sutcliffe model efficiency coefficient, an efficiency of one corresponds to a perfect match of the simulated data to the observed data (Moosavi *et al.*, 2013).

Table 1. RMSE, COREL, N and R^2 for different ANN models with the best combination of inputs and number of neurons in hidden layer

	Network			
	RMSE	COREL	Nash	R^2
Cl				
	0.110968	0.894827	0.751953	0.800716
train SCG	0.040904	0.984817	0.966297	0.969865
train LM	0.093546	0.944696	0.823724	0.89245
EC				
train GDX	0.085031	0.952771	0.891993	0.907773
train SCG	0.033171	0.9925	0.983563	0.985057
train LM	0.080624	0.968828	0.902897	0.938628
NO$_3$				
train GDX	0.447729	-0.22482	-4.8985	0.050546
train SCG	0.327234	0.217898	-2.15086	0.04748
train LM	1.257356	0.311154	-2.11639	0.096817
SO$_4$				
train GDX	0.06598	0.954394	0.884451	0.910867
train SCG	0.057235	0.968582	0.91305	0.93815
train LM	0.043035	0.975265	0.950843	0.951142

The results of RMSE, COREL, N and R^2 for different ANN models with the best combination of inputs and number of neurons in hidden layer are shown in Table 1. The number of epochs was set as 200 for Cl, and the optimum combination of inputs was a combination of seven parameters: anionic, Cation, TDS, TH, SAR, Na, and Mg, the best number of neurons in the hidden layer was 19. The number of epochs was set as 100 for EC, and the optimal combination of inputs was a combination of nine parameters: Anion, Cation, TDS, TH, SAR, K, Na, Mg, and Ca, and the best number of neurons in the hidden layer was 15. The number of epochs was set as 400 for NO$_3$, and the best combination of inputs was a combination of two parameters: HCO$_3$ and Evaporation, and the best number of neurons in

the hidden layer was 7. The number of epochs was set as 400 for SO$_4$. The best combination of inputs was a combination of four parameters: Anion, Cation, TH, Na and the optimum number of neurons in the hidden layer was 10. The relationship between the estimated and observed values of the parameters Cl, EC, NO$_3$, and SO$_4$, are modeled as shown in Figure 3.

In this study, eighteen different parameters were used as input and four output parameters were examined, the results are shown in Table 1. Five training algorithms were examined, namely train GD, train GDA, train GDX, train SCG and train LM with different numbers of neurons in the hidden layer. As shown in Table 1, the optimal training algorithm for Cl was train SCG and the values of RMSE, COREL, N and R^2 were 0.04,

0.98, 0.96 and 0.96, respectively by the best combination of inputs. For EC, train SCG was the optimum training algorithms used and the values of RMSE, COREL, N and R^2 were 0.033, 0.99, 0.98 and 0.98, respectively. For NO_3, train LM was the desirable training algorithm and this result is similar to the research of Venkat Kumar *et al.* (2010), in which the values of RMSE, COREL N and R^2 were 1.25, 0.3, -2.1 and 0.096, respectively. The number of neurons in the hidden layer was equal to double the number of inputs plus one for all parameters and this result is in line with the research of Kavzoglu *et al.* (2003). It seems that the values should be obtained by the trial-and-error method, as pointed out by Venkat Kumar *et al.* (2010) and Moosavi *et al.* (2013). Figure 3 shows an increase in the parameter's value of groundwater in the standard range. The minimum and maximum values of parameters were Cl: (3.2, 77) mg / l, EC: (920, 11920) µmohs/cm, NO_3: (0, 15.4) mg / l, SO_4: (2, 82.5) mg / l for the last decade (2003-2013) and this increase could be a warning for the future. This

trend is an indication of drought and declining water quality, increasing number of industries, agricultural wastewater and population in this region. Industrial and municipal wastes and excess nitrate fertilizers increased these parameters in this area in recent years. Also, because of the increasing use of groundwater, resources have declined in the region in recent years, this trend will be exacerbated. As pointed out by Badeenezhad *et al.* (2012), Sharifi *et al*. (2012) and Abu-Khalaf *et al*. (2013), the increasing nitrate concentration in groundwater can be attributed to the transfer of surface and subsurface sewage by soil and fertilizer overuse. The maximum desired EC parameter is 1500 µmohs/cm, and the maximum allowable value is 2000 µmohs/cm, so the increase of this parameter, which has led to changes, has increased the salinity of water in recent decades. Therefore, underground aquifer changes can cause instability, and low groundwater level also affects water quality, and could be a subject for discussion at the Earth Summit.
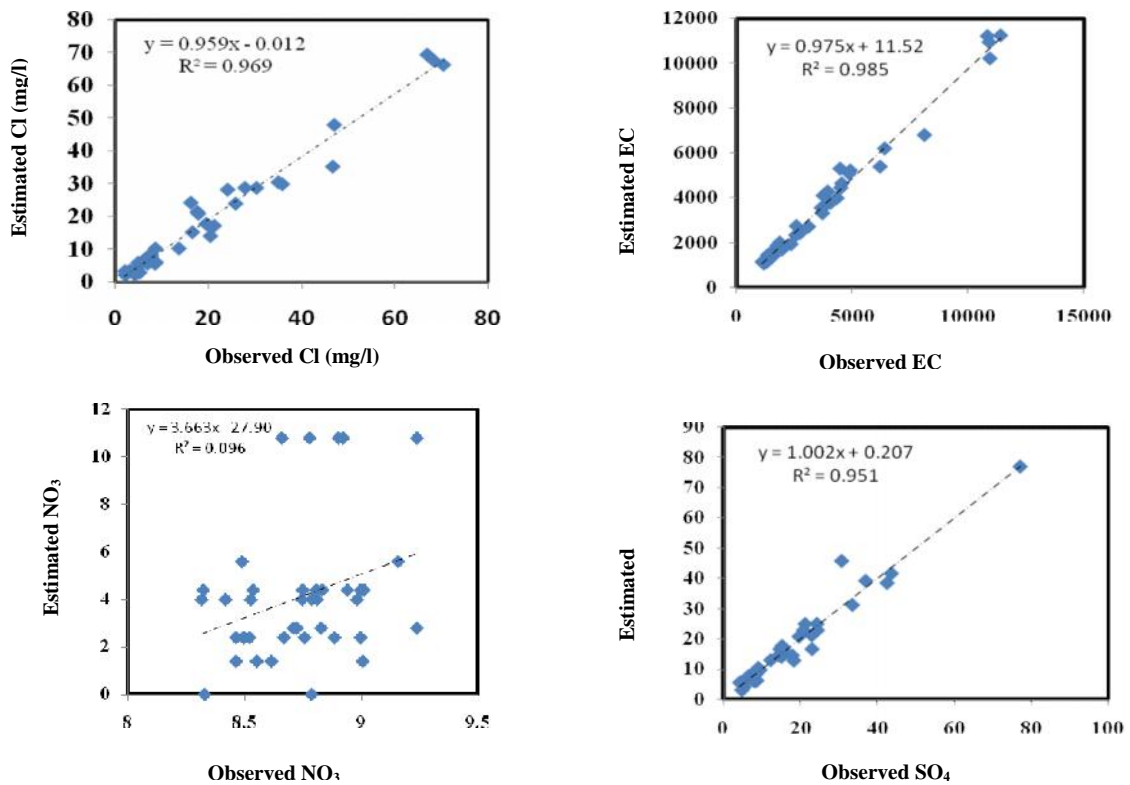
Fig. 3. The relationship between observed and simulated data for the best ANN model for CL, EC. SO_4, and NO_3 parameters

4. Conclusion

Based on this study's results, using artificial neural network with the back-propagation algorithms for modeling qualitative parameters of groundwater, such as Cl, EC, SO_4, was more accurate according to the chosen input parameters, and this finding is similar to those of Li *et al.* (2012) and Ying *et al.* (2007). However, the use of artificial neural networks for NO_3 modeling was not, accurate. This contradicts the finding of Diamantopoulou *et al.*'s (2005) research in Greece, that studied NO_3 modeling in surface water and used DO, EC, Na, Ca, Mg, HCO_3, Q, T, SO_4, and CL parameters as input. They also used back-propagation algorithm and obtained good results. Perhaps, it was because they used a different water source or experienced the impact of other parameters. According to the highly expensive and time-consuming tests, these parameters can be modeled to estimate their range. This is a quick and cost-effective method for management practices, especially in emergency situations.

Acknowledgment

References

Abu-Khalaf, N., S.Khayat, B. Natsheh, 2013. Multivariate Data Analysis to Identify the Groundwater Pollution Sources in Tulkarm Area/Palestine. Science and Technology, 3; 99-104.

Badeenezhad, A., M. Gholami, A. JonidiJafari, A. Ameri, 2012. Factors affectingnitrate Concentrationsin Shiraz Groundwater Using Geographical Information System (GIS). Tolooe Behdasht, 11; 47- 56.

Chitsazan, C., G.H. Rahmani, A .Neyamadpou, 2013. Groundwater level simulation using artificial neural network: a case Study from Aghili plain, urban area of Gotvand, south-west Iran. Geopersia Journal, 3; 35-46.

Chowdhury, SH., P. Champagne, P.J. McLellan, 2009. Models for predicting disinfection byproduct (DBP) formation in drinking waters: A chronological review. Science of the Total Environment, 407; 4189–4206.

Cuesta Cordoba, I. G. A., 2011. Using OF Artificial Neural Network for Evaluation and Prediction of Some Drinking Water Quality Parameters within a Water Distribution System. Water Management and Water Structures, 3; 1-11.

Diamantopoulou, M. J., V. Z. Antonopoulos, D. M. Papamichail, 2005. The Use of a Neural Network Technique for the Prediction of Water Quality Parameters of Axios River in Northern Greece. European Water, 11; 55-62.

Dutta, S., A.A. Parsons, C. Bhattacharjee, 2010. Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO2 surface. Expert Systems with Applications, 37; 8634–8638.

Garcia, L.A., A. Shigidi, 2006. Using neural networks for parameter estimation in ground water. Journal of Hydrology, 318; 215–231.

Huang, J., J. Xu, X. Liu, L. Wang, 2011. Spatial distribution pattern analysis of groundwater nitrate nitrogen pollution in Shandong intensive farming regions of China using neural network method. Mathematical and Computer Modelling, 54; 995–1004.

Karul, C., S. Soyupak, A.F.Cilesiz, 2000. Case studies on the use of neural networks in eutrophication modeling. Ecological Modelling, 134; 145–152.

Kheradpisheh, Z., S.A. Almodaresi, Y. Khaksar, L. Rafati, 2014. Zoning of groundwater Contaminated by Nitrate Using Geostatistics Methods (Case Study: Bahabad Plain, Yazd, Iran). Desert, 19; 83-90.

Kulkarni, P., SH. Chellam, 2010. Disinfection by-product formation following chlorination of drinking water: Artificial neural network models and changes in speciation with treatment. Science of the Total Environment, 408; 4202–4210.

Li, X., S. Lungcang, L. Liu, 2012. Sensitivity analysis of groundwater level in Jinci Spring Basin (China) based on artificial neural network modeling. Hydrogeology Journal, 20; 727–738.

Ming Kuo, Y., Ch. Wuing Liu, K. Hung Lin, 2004. Evaluation of the ability of an artificial neural networkmodel to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. Water Research, 38; 148–158.

Moosavi, V., M. Vafakhah, B. Shirmohamadi, 2013. A Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods. Water Resources Management, 27; 1301-1321.

Nourani, V., R. Goli Ejlali, 2012. Quantity and Quality Modeling of Groundwater by Conjugation of ANN and Co-Kriging Approaches. Water Resources Management and Modeling, Available from: http://www.intechopen.com/books/water-resources-managementand-modeling/quantity-and-quality-modeling-of-groundwater-by-conjugation-of-ann-and-co-krigingapproaches

Panda, S.S., V. Garg, I. Chaubey, 2004. Artificial Neural Networks Application in Lake Water Quality Estimation Using Satellite Imagery. Journal of Environmental Informatics, 4; 65-74

Palani, S., S.Y. Liong, P. Tkalich, 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin, 56; 1586–1597.

Sahoo, G.B., C. Ray, H.F. Wade, 2005. Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks. Ecological Modelling, 183; 29–46.

Sadiq, R., M. J. Rodriguez, 2004. Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. Science of the Total Environment, 321; 21-46.

Seyam, M., Y. Mogheir, 2011. Application of Artificial Neural Networks Model as Analytical Tool for Groundwater Salinity. Journal of Environmental Protection, 2; 56-71.

Sharifi, Z., A. ASafari, 2012. Assessment of Arsenic, Nitrate and Phosphorus Pollutions in Shallow Groundwater of the Rural Area in Kurdistan Province (Iran). Current World Environment, 7; 233-241.

Singh, K.P., A. Basant, A. Malik, 2009. Artificial neural network modeling of the riverwater quality-A case study. Ecological Modelling, 220; 888–895.

Venkat Kumar, N., S. Mathew, G. Swaminathan, 2010. Analysis of Groundwater for Potability from Tiruchirappalli City Using Backpropagation ANN Model and GIS. Journal of Environmental Protection, 1; 136-142.

WRS, 2014. Available From http://wrs.wrm.ir/m3/istgah-baranlist.asp?recperpage=ALL. Accessed 1st May 2014

Ying, Z., N. Jun, C. Fuyi, 2007. Water quality forecast through application of BP neural network at Yuqiao reservoir. Journal of Zhejiang University Science, 8; 1482-1487.