Int. Agrophys., 2011, 25, 187-192

INTERNATIONAL
Agrophysics
www.international-agrophysics.org

Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models

A.H. Zare[1]*, *V.M. Bayat*[1], *and A.P. Daneshkare*[2]

[1]Agricultural Irrigation and Drainage Engineering Department, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran
[2]Agricultural Irrigation and Drainage Engineering Department, Faculty of Agriculture, Ghazvin University, Ghazvin, Iran

Received May 25, 2010; accepted October 27, 2010

A b s t r a c t. In this research, the ability of artificial neural network (ANN) to model groundwater nitrate of Arak Aquifer (Iran) is introduced. The ANN and linear regression (LR) methods were used to relate groundwater nitrate concentration to other water quality indices. Results showed that using the measured parameters is convenient to model nitrate concentration with acceptable and appropriate accuracy and ANN and LR methods were able to predict nitrate concentration at the desirable level of accuracy. Comparison of ANN analysis with LR model results showed that ANN requires fewer parameters with more accuracy in comparison to LR models. However, the ANN model with the highest correlation coefficient (r = 0.87), minimum root mean square error ($RMSE$=10.46 mg l^{-1}) and mean absolute error (MAE = 7.77 mg l^{-1}) provided the best results among the LR models.

K e y w o r d s: nitrate concentration, prediction, artificial neural network, regression

INTRODUCTION

In arid and semi-arid regions groundwater is the major source for domestic purposes and irrigation. Irrigation water quality has a significant role in crop production and has a profound impact on physical and chemical soil properties. Monitoring of water quality is one of the important tools for sustainable development and provides important information for water management (Jalali, 2009). Due to the correlations and interactions between water quality variables such as anions and cat-ions concentrations, it is interesting to investigate whether a domain-specific mechanism governing observed patterns exists to prove the predictability of these variables. The identification of such forecast models is particularly useful for ecologists and environmentalists, since they will be able to predict water pollution levels and take necessary precaution measures in advance (Palani *et al.,* 2008). Variation in groundwater quality is a function of physical and chemical parameters that are greatly influenced by geological formations and anthropogenic activities as well (Yesilnacar *et al.,* 2008).

Artificial neural networks (ANNs) are able to approximate accurately complicated non-linear input-output relationships. The ANN is used as an approximation tool rather than a complex mathematical calculation, which results in a ten percent deviation of predicted value from observed data (Lingireddy and Ormsbee, 1998). There are a number of studies in which neural networks are applied to water quality problems. Chau (2006) reviewed the development and current progress of the integration of artificial intelligence into water quality modelling. Hatzikos *et al.* (2005) utilized neural networks with active neurons as a modelling tool for the prediction of seawater quality indicators like water temperature, pH, dissolved oxygen (DO) and turbidity. Palani *et al.* (2008) demonstrated the application of ANNs to model the values of selected seawater quality variables, having the dynamic and complex processes hidden in the monitored data itself (Faruk, 2010).

The ANNs have been widely used in various studies on surface water pollution control for predicting stream nitrogen concentration (Lek *et al.,* 1999), forecasting raw water quality parameters (Zhang and Stephen, 1997), prediction of water quality parameters, water quality management (Wen and Lee, 1998), and identification of non-point sources of microbial contamination (Brion and Lingireddy, 1999; Zaheer and Cui, 2003). Due to increased agricultural activity which is necessary for enhanced food production and also due to industrial activity, there is an increasing evidence of nitrate pollution of groundwater (Prakasa Rao and Puttanna, 2000). Mueller *et al.* (1995) found that nitrate concentrations were generally twice as high in groundwater under agricultural

*Corresponding author's e-mail: Zareabyaneh@gmail.com

lands as under other areas. There are some models such as DRAINMOD-N, LEACHN, soil and water assessment tool (SWAT) to model nitrate concentration in groundwater. Very little work has been done so far in building stochastic models to predict nitrate concentration in groundwater using regression and neural networks. Ray and Klindworth (2000) utilized artificial neural networks to predict the pesticide and nitrate contamination in rural private wells. They used depth to aquifer materials from land surface, well depth and distance to cropland as input parameters and concentration of pesticides or nitrates were the outputs. Ramasamy *et al.* (2003) used regression and neural networks to model nitrate concentration in groundwater. In this study nitrate concentration (NO_3 measured in mg l^{-1}) was the dependent (response) variable and iron concentration (Fe measured in mg l^{-1}), season and distance of the well from a poultry house (measured in meters) were the independent (explanatory) variables. Those results showed the statistics from neural networks were better than the statistics from regression; neural networks underpredicted the log of the nitrate concentration in ground water. Almasri and Kaluarachchi (2005) used modular neural networks (MNN) to simulate the nitrate concentration in an agriculture-dominated aquifer. Those MNN simulations are further analysed and compared to obtain from a physically-based fate and transport model to evaluate the overall applicability of MNN. Kaluli *et al.* (1998) developed an ANN model that simulated nitrate leaching to ground water from a 4.2 ha site. ANN inputs included time of year, daily denitrification rate, cropping systems, water table depth, nitrogen fertilizer application rate, daily antecedent precipitation index, initial nitrate concentration, and daily drain flow. They used ANN to simulate the impacts of different management options on nitrate leaching.

Yesilnacar *et al.* (2008) predicted nitrate concentration in groundwater using four parameters of temperature, electrical conductivity, groundwater level and pH as input parameters in the ANN. The Levenberg-Marquardt algorithm was selected as the best one within 12 back propagation (BP) algorithms and optimal neuron number was determined as 25.

Recently, many studies have been conducted to investigate nitrate leaching hazards and groundwater pollution in Iran (Jalali, 2009). ANNs as a tool are still not widely used in the field of groundwater nitrate contamination prediction and forecasting.

The aim of this paper was to study of artificial neural network modelling to predict and forecast nitrate concentration in 53 representative observation wells and compared with linear regression models (LR) to select the best method.

MATERIAL AND METHODS

Industrial districts of Arak were considered because of large amount of nitrated wastewater production which is used to irrigate adjacent farmlands. Arak is located between 35°8'-35°23' N and 47°53'-48°37' E, at 1 752 m a.s.l. The climate of the study area is semi-arid, the annual average precipitation and temperature are approximately 340 mm and 5°C. Precipitation occurs mainly in winters. Because of high industrialization in the study area, industrial wastewaters as well as agricultural fertilizers are the main reason of groundwater contamination. Within all contaminants, nitrate is the most important pollutant during recent years. The nitrogen sources in the study area are application of agricultural fertilizers, irrigation with nitrogen-contaminated groundwater and wastewater of factories which discharge into surface water courses. Although, the study of Głąb *et al.* (2009) showed that long-term application of compost and nitrogen fertilizers did not have any significant influence on bulk density, water retention, pore-size distribution and the saturated hydraulic conductivity.

The methodology of research is divided to four separate parts: description of data sets; ANN and LR inputs; description of ANN characteristics of neurons, layers and in-out parameters; and method of sensitivity analysis.

The choice of the type of network depends on the nature of the problem to be solved (Goethals *et al.,* 2007). The number of input and output neurons is determined by the nature of the modelling problem, the input data representation and the form of the network output required. The number of hidden layers is related to the complexity of the system being modelled. Although some researchers suggest that one hidden layer is usually sufficient (El-Din and Smith, 2002). So in this study a three-layer ANN (input-output layers with a hidden layer) with Levenberg-Marquardt algorithm and a tan-sigmoid transfer function for the hidden layer and a linear transfer function for the output layer were used. In a study of Amiri-Chayjanl and Esna-Ashari (2009), the best result for prediction of sorption isotherm in rough rice by ANN was with Levenberg-Marquardt algorithm and tan-sigmoid transfer function.

The input layer included some water quality parameters and the output layer was nitrate concentration. In regression modelling, nitrate concentration (NO_3^- measured in mg l^{-1}) as dependent variable was related to some water quality parameters as independent variables. Fifty three groundwater wells were selected in Arak Aquifer and 818 groundwater samples were analyzed to model nitrate contamination change. Seventy percent of the samples were used to train the ANN and develop LR models, and the remaining 30% of data were used to evaluate the models.

Samples were collected after a pumping time of about 30 min from water wells. Samples were analysed in the laboratory for the major ions using standard methods. The analyzes were carried out within 48 h after sampling.

Parameters of pH, electrical conductivity (EC), magnesium (Mg^{2+}), chloride (Cl^-), sodium (Na^+), potassium (K^+), bicarbonate (HCO_3^-), sulphate (SO_4^{2-}) and calcium (Ca^{2+}) ionic concentrations, total dissolved solids (TDS) and total hardness (TH) were measured. Nitrates (NO_3^-) were measured

using colorimetric method with an UV – visible spectrophotometer. For localization of nitrate concentration changes in the study zone, the zoning map was drawn by kriging method in Surfer software.

First, 11 parameters of water quality were used as a primary input of artificial neural network. For the selection of the most important artificial neural network input parameters the periodic remove method was used. Therefore, by eliminating any input parameter, the structure of optimized artificial neural network was run. With comparing neural network output by eliminating any input parameter, the network sensitivity to any input parameter was calculated from the following equation:

$$PC = \frac{|X_1 X_2|}{X_1} 100, \quad (1)$$

where: PC – percent of change (%), X_1 – artificial neural network output with 11 input parameters, X_2 – output of artificial neural network with eliminating any input parameter.

To select the most appropriate input parameters, different combinations of parameters were used to predict nitrate concentration. The combined parameters are given in Table 1.

Three different forecast consistency measures of the root mean square error (*RMSE*), the mean absolute error (*MAE*) and the correlation coefficient (*r*) were used to evaluate models results and to compare ANN and LR:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(N_i(obs) - N_i(cal)\right)^2}, \quad (2)$$

$$MAE = \frac{1}{n}\sum\left|N(obs)N(cal)\right|, \quad (3)$$

$$r = \frac{\sum_{i=1}^{n}(N_i(obs) - \overline{N}(obs))(N_i(cal) - \overline{N}(cal))}{\sum_{i=1}^{n}(N_i(obs) - \overline{N}(obs))^2 (N_i(cal) - \overline{N}(cal))^2}, \quad (4)$$

where: $N_i(cal)$ and $N_i(obs)$ are the predicted and observed nitrate concentrations, respectively, $\overline{N}(cal)$ and $\overline{N}(obs)$ are the means of predicted and observed nitrate concentration, respectively, and n is the number of data.

Table 1. Different combinations of input parameters in ANN and LR models

No.	Input parameters
1	Na^+, Mg^{2+}, T, Ca^{2+}, HCO_3^-, $SO4_2^-$, Cl^-, TH, TDS, EC, pH
2	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, TH, TDS, EC, pH
3	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, TH, EC, pH
4	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, EC, pH
5	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, EC
6	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, EC

RESULTS AND DISCUSSION

The basic statistics for the analyzed groundwater samples are listed in Table 2. The minimum and maximum measured nitrate concentrations were 3 and 89 mg l^{-1}, respectively, for August and March. The average nitrate concentration was a little more than 32 mg l^{-1}. Nitrate concentration was more than the standards for drinking water in 28% of the wells.

Spatial variation of nitrate in groundwater of the study area is presented in Fig. 1. The highest nitrate concentrations occur in northwest and southeast parts of the aquifer.

As mentioned before, 11 indices shown in Table 2 were used to predict groundwater nitrate concentration using different ANN models. The best network structure for estimating nitrate concentration was determined. To reduce the input parameters and to determine parameters with less influence on nitrate concentration, sensitivity to each of the selected parameters was studied.

Percentage of variation of estimated nitrate concentration was determined for each index. The importance of each input parameter is shown in Fig. 2. Ca^{2+} has the highest influence on nitrate prediction. In contrast, TDS, T and SO_4^{2-} have the least effects. According to the results of Fig. 2, six different combinations of input parameters were used to evaluate the accuracy measures of Eqs 1 to 3. The results are presented in Table 3. The reduction of less effective parameters in definition of ANN structures was considered. ANN1 structure with 11 parameters had a greater error than those of other structures, which means that increasing the number of input parameters is not always effective. ANN3 structure with eight inputs was selected as the appropriate structure considering error and correlation coefficients. It is clearly noted in Table 3 that the proposed ANN3 model has impressively well learned the nonlinear relationship

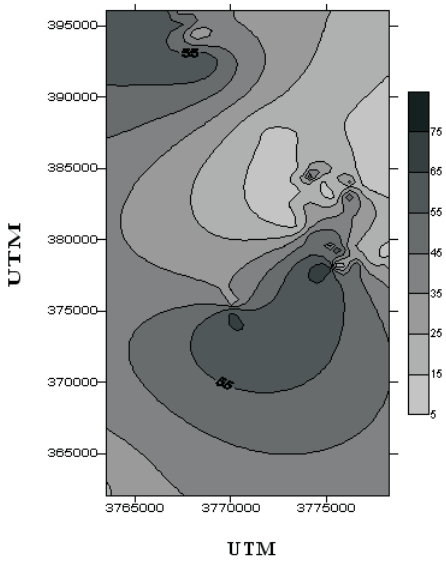

Fig. 1. Spatial distribution of nitrate concentration in groundwater.

T a b l e 2. Groundwater samples analysis

Parameter	Unit	Min	Max	Mean	S	CV
Na^+	mg l-1	7	540	69.0	56.6	0.82
T	°C	16	29	21.3	2.5	0.12
Mg^{2+}	mg l^{-1}	7.16	119.5	23.2	11.6	0.50
Ca^{2+}	mg l^{-1}	36	388	90.7	44.0	0.48
HCO_3^-	mg l^{-1}	97	383	183.1	57.4	0.31
SO_4^{2-}	mg l^{-1}	6.5	344	115.4	65.7	0.57
Cl^-	mg l^{-1}	5	1804	120.7	152.3	1.26
TH	mg l^{-1}	136	1460	322.2	152.9	0.47
TDS	mg l^{-1}	208	3920	649.1	333.1	0.51
EC	dS m^{-1}	298	5600	927.5	475.7	0.51
pH	–	7.04	8.17	7.5	0.2	0.02
NO_3^-	mg l^{-1}	3	89	32.2	22.2	0.63

T a b l e 3. Evaluation of the ANN structures

No.	Input parameters	*r*	*RMSE*	*MAE*
1	Na^+, Mg^{2+}, T, Ca^{2+}, HCO_3^-, SO_4^{2-}, Cl^-, TH, TDS, EC, pH	0.84	14.78	9.84
2	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, TH, TDS, EC, pH	0.84	12.68	8.02
3	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, TH, EC, pH	0.87	10.46	7.77
4	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, EC, pH	0.81	13.51	8.68
5	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, EC	0.79	13.56	9.28
6	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, EC	0.80	14.25	10.47

between the input and the output variables with *RMSE* = 10.46 mg l^{-1}, *MAE*=7.77 mg l^{-1} and *r* = 0.87. Therefore, Fig. 2 and Table 3 showed that eliminating three parameters of T, TDS and SO_4^{2-} improved the accuracy of the ANN.

The same input parameters were used in LR models as independent variables for modelling nitrate concentration. The results are presented in Table 4. The best regression model includes 11 independent variables (LR1). This structure is able to estimate groundwater nitrate concentration with an error of 14.45 mg l^{-1} and with correlation coefficient

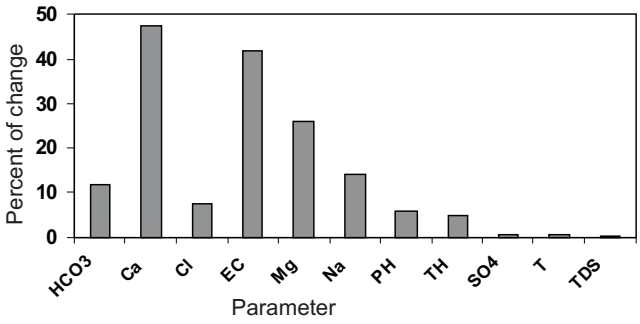

Fig. 2. Sensitivity of each input index for prediction of nitrate concentration.

of 0.73. Comparison of two selected models, ANN3 and LR1, shows that LR1 is less accurate. The reducing input variables of LR model decreases model accuracy, whereas for ANN it increases. However, the main limitations of statistical techniques are the rigid assumptions that are essential for justifying their applications, such as those of sample size, linearity, and continuity. The main advantage of reduction in ANN input parameters is computational economy, decrease in computation time and cost. This point is confirmed on estimation of free water evaporation and reference evapotranspiration (Wang *et al.*, 2008). In above mentioned studies using minimum parameters neural network to predict the unknown parameter (output) was investigated.

ANN3 structure, which was the most appropriate model to estimate nitrate concentration, was used for two different data sets of training and testing. The results of ANN3 for the two data sets are shown in Fig. 3. The comparison of nitrate concentration predicted by ANN3 and observed values at training and testing phases showed good agreement with *r* of 0.88 and 0.87, respectively.

Figure 4 shows the comparison between nitrate concentration values observed and those calculated by the selected ANN model (ANN3) for 53 sampling wells under testing of

T a b l e 4. Evaluation of LR models

No.	Input parameters	*r*	*RMSE*	*MAE*
1	Na^+, Mg^{2+}, T, Ca^{2+}, HCO_3^-, SO_4^{2-}, Cl^-, TH, TDS, EC, pH	0.73	14.45	10.36
2	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, TH, TDS, EC, pH	0.72	14.84	10.62
3	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, TH, EC, pH	0.63	16.72	11.94
4	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, EC, pH	0.63	16.72	12.04
5	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, Cl^-, EC	0.63	16.72	12.01
6	Na^+, Mg^{2+}, Ca^{2+}, HCO_3^-, EC	0.62	16.83	12.12

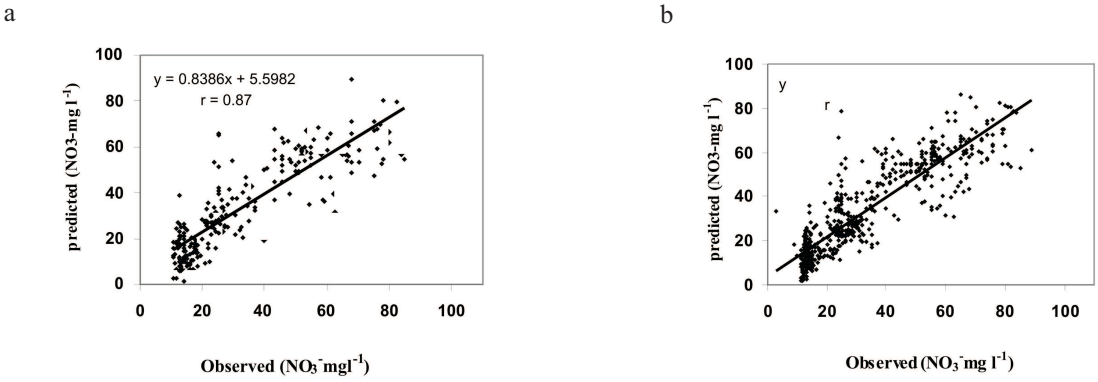

Fig. 3. Evaluation of ANN3 to model nitrate concentration for: a – testing-, and b – treining data sets.

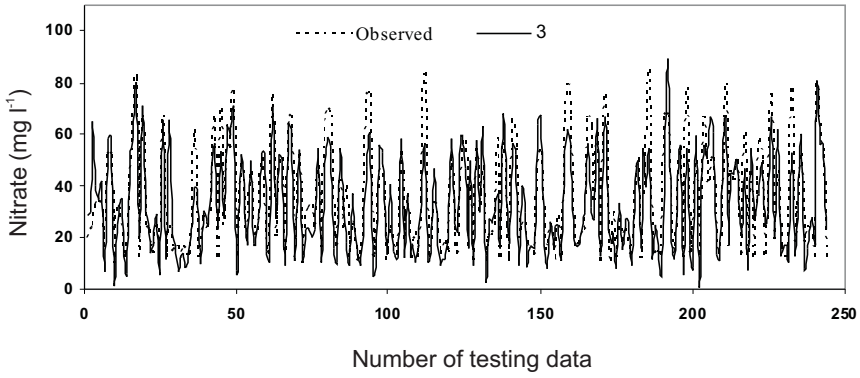

Fig. 4. ANN results for nitrate concentration prediction for sampling well No. 53.

the neural network. Evolution is similar and one line is practically superimposed over the other. This figure illustrates the close relationship between nitrate concentration observed and derived from the ANN model. The maximum difference between nitrate concentration values observed and those calculated by the ANN model is in the point of maximum nitrate concentration. But in most of the points, nitrate concentration values observed and estimations of the ANN model are the same. Figure 4 shows that ANN models lower concentrations of nitrate more accurately than higher concentrations. In other words, the difference between actual and predicted values of nitrate concentrations is higher for high nitrate concentrations. This can be due to the effects of other factors on nitrate concentration which is not considered in ANN. However, the results are reasonable for lower concentrations.

CONCLUSIONS

1. Artificial neural network (ANN) performed better than linear regression (LR) model. The results provided sufficient assessment of each model performance (r = 0.73, *RMSE*=14.45 mg l^{-1} and *MAE*=10.36 mg l^{-1} for LR and r = 0.87, *RMSE*=10.46 mg l^{-1} and *MAE*=7.77 mg l^{-1} for ANN model).

2. Artificial neural network (ANN) models need fewer parameters for prediction of nitrate concentrations compared to linear regression (LR) models. Moreover, because of all of their advantages, ANNs are easy and practical to apply from site to site. Their fast execution should also be helpful for simulation of nitrate concentrations on a large scale.

REFERENCES

Almasri M.N. and Kaluarachchi J.J., 2005. Modular neural networks to predict the nitrate distribution in ground water using the on-ground nitrogen loading and recharge data. Environ. Model. Software, 20, 851-871.

Amiri-Chayjan R. and Esna-Ashari M., 2010. Comparison between mathematical models and artificial neural networks for prediction of sorption isotherm in rough rice. Int. Agrophysics, 24, 1-7.

Brion G.M. and Lingireddy S., 2000. Artificial Neural Networks in Hydrology. Kluwer Academic Press, London, UK.

Chau K.W., 2006. A review on integration of artificial intelligence into water quality modeling. Marine Poll. Bull., 52, 726-733.

El-Din A.G. and Smith D.W., 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res., 36(5), 1115-1126.

Faruk D., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artificial Intell., 23(4), 586-594.

Głąb T., Zaleski T., Erhart E., and Hartl W., 2009. Effect of biowaste compost and nitrogen fertilization on water properties of Mollic-gleyic Fluvisol. Int. Agrophysics, 23, 123-128.

Goethals P., Dedecker A.P., Gabriels W., Lek S., and Pauw N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Ecol., 41, 491-508.

Hatzikos E., Anastasakis L., Bassiliades N., and Vlahavas I., 2005. Simultaneous prediction of multiple chemical parameters of river water quality with tide. Proc. 2nd Int. Sci. Conf. Computer Sci., IEEE Computer Society, May 11-13, Varna, Bulgaria.

Jalali M., 2009. Groundwater geochemistry in the Alisadr, Hamadan, western Iran. Environ. Monit. Assess., 166, 359-369.

Kaluli J.W., Madramootoo C.A., and Djebbar Y., 1998. Modeling nitrate leaching using neural networks. Water Sci. Technol., 38(7), 127-134.

Lek S., Maritxu G., and Giraudel J., 1999. Prediction of stream nitrogen concentration from watershed features using neural networks. Wat. Res., 33(16), 3469-78.

Lingireddy S. and Ormsbee L.E., 1998. Neural Networks in Optimal Calibration of Water Distribution Systems, Artificial Neural Networks for Civil Engineering: Advanced Features and Applications (Eds I. Flood, R. Kartam). ASCE Press, New York, USA.

Mueller D.K, Hamilton P.A., Helsel D.R., Hitt K.J., and Ruddy B.C., 1995. Nutrients in groundwater and surface water of the United States: an analysis of data through 1992. U.S. Geol. Survey Water Resour. Inv., Report, 95-4031, Dovenport, IO, USA.

Palani S., Liong S.Y., and Tkalich P., 2008. An ANN application for water quality forecasting. Marine Poll. Bull., 56, 1586-1597.

Prakasa Rao E.V.S. and Puttanna K., 2000. Nitrates, agriculture and environment. Current Sci., 79(9), 1163-1168.

Ramasamy N., Krishnan P., Bernard J.B., and Ritter W.F., 2003. Modeling Nitrate Concentration in Ground Water Using Regression and Neural Networks. ORES SP03-01, Food and Resource Economic, Delaware Univ. Papers, Newark, DE, USA.

Ray C. and Klindworth K.K., 2000. Neural networks for agrichemical vulnerability assessment of rural private wells. J. Hydrol. Eng., 4, 162-171.

Wang Y.M., Traore S., and Kerh T., 2008. Neural network approach for estimating reference evapotranspiration from limited climatic data in Burkina Faso. Transactions on Computers, 6(7), 704-712.

Wen C.G. and Lee C.S., 1998. A neural network approach to multi-objective optimization for water quality management in a river basin, Water Resour. Res., 34(3), 427-436.

Yesilnacar M.I., Sahinkaya E., Naz M., and Ozkaya B., 2008. Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. Environ Geol., 56, 19-25.

Zaheer I. and Cui G., 2003. Inverse modeling to identify nonpoint source pollution using a neural network, Taihu lake watershed, China. J. Environ. Hydrol., 11, 1-8.

Zhang Q. and Stephen J.S., 1997. Forecasting raw water quality parameters for North Saskatchewan River by neural network modeling. Water Res., 31(9), 2340-2351.