

Neural network prediction of nitrate in groundwater of Harran Plain, Turkey

M. Irfan Yesilnacar · Erkan Sahinkaya ·
Muhsin Naz · Bestamin Ozkaya

Received: 11 October 2007 / Accepted: 12 November 2007 / Published online: 24 November 2007
© Springer-Verlag 2007

Abstract Monitoring groundwater quality by cost-effective techniques is important as the aquifers are vulnerable to contamination from the uncontrolled discharge of sewage, agricultural and industrial activities. Faulty planning and mismanagement of irrigation schemes are the principle reasons of groundwater quality deterioration. This study presents an artificial neural network (ANN) model predicting concentration of nitrate, the most common pollutant in shallow aquifers, in groundwater of Harran Plain. The samples from 24 observation wells were monthly analysed for 1 year. Nitrate was found in almost all groundwater samples to be significantly above the maximum allowable concentration of 50 mg/L, probably due to the excessive use of artificial fertilizers in intensive agricultural activities. Easily measurable parameters such as temperature, electrical conductivity, groundwater level and pH were used as input parameters in the ANN-based nitrate prediction. The best back-propagation (BP) algorithm and neuron numbers were determined for optimization of the model architecture. The Levenberg–Marquardt algorithm was selected as the best of 12 BP algorithms and optimal neuron number was determined as 25. The model tracked the experimental data very closely ($R = 0.93$). Hence, it is possible to manage groundwater resources in a more cost-effective and easier way with the proposed model application.

Keywords Groundwater quality · Harran Plain · GAP project · Nitrate · Artificial neural network

Introduction

Harran Plain is undergoing large land use changes due to population growth and the accompanying industrial, commercial and agricultural development. This activity is producing multiple potential sources of contaminants from manure and artificial fertilizers, landfills, accidental spills and domestic or industrial effluent discharges. Among these sources, agriculture-related activities are well-known non-point source pollution. Agriculture activities may deteriorate the groundwater quality in small to large watersheds especially due to uncontrolled use of fertilizer and various carcinogenic pesticides (Almasri and Kaluarachchi 2005). Variation in groundwater quality is a function of physical and chemical parameters that are greatly influenced by geological formations and anthropogenic activities as well (Subramani et al. 2005).

Turkey is currently engaged in a large integrated water resources development project in its semi-arid southeastern region. Commonly referred to by its Turkish acronym GAP, the Southeastern Anatolia Project includes 22 dams in the upper Euphrates-Tigris Basin, and aims to provide irrigation for 1.7 million hectares of land by 2015 (Yesilnacar and Gulluoglu, in press). As in other semi-arid and arid parts of the world, water is a valuable resource in the GAP region. Despite large quantities of water currently available from the Euphrates and Tigris rivers, it is becoming increasingly important to improve management of these resources (Ozdogan et al. 2006) as the GAP region, and in particular in the Harran Plain, faces problems of salinity, excessive and uncontrolled irrigation, an

M. I. Yesilnacar (✉) · E. Sahinkaya · M. Naz
Department of Environmental Engineering, Harran University,
Osmanbey Campus, 63000 Sanliurfa, Turkey
e-mail: iyasilnacar@yahoo.com

B. Ozkaya
Department of Environmental Engineering,
Yildiz Technical University, 34349 Istanbul, Turkey

insufficient drainage system, and an increased groundwater level caused by irrigation (Kendirli et al. 2005).

Sources of nitrate, the most common pollutant found in shallow aquifers, may be point and non-point sources and it can be easily lost from soil, especially in sandy soils, by leaching processes due to its high mobility (Almasri and Kaluarachchi 2005; Freeze and Cherry 1979). Hence, it is very important to monitor or predict the nitrate concentration in such areas by means of cost-effective technologies. In this context, black-box models like artificial neural network (ANN) are very attractive to predict the nitrate concentration using easily measurable (by means of probes of portable in-situ instruments) water quality parameters such as temperature, electrical conductivity (EC), groundwater level, pH, etc. ANN does not require prior knowledge about the structure and relationships that exist between important variables. Moreover, their learning abilities make them adaptive to system changes (Strik et al. 2005). ANNs have already been used to simulate the effect of climate change on discharge and the export of dissolved organic carbon and nitrogen from river basins (Clair and Ehrman 1996), to forecast salinity (Maier and Dandy 1996), to simulate and forecast residual chlorine concentrations within urban water systems (Rodriguez and Sérodes 1998), to determine the relationship between sewage odour and BOD (Onkal-Engin et al. 2005), to

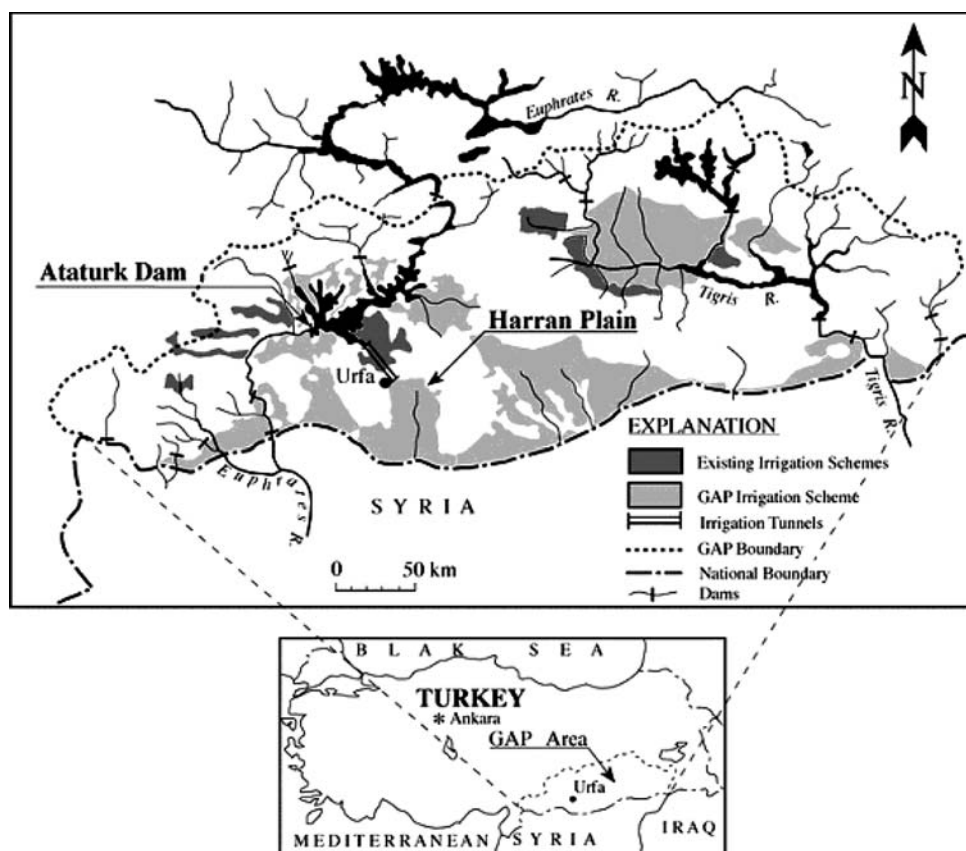
determine the performance of sulfidogenic bioreactor (Sahinkaya et al. 2007) and to determine the leachate amount from municipal solid waste landfill (Karaca and Özkaya 2006). There are also other similar applications of ANN in the field of environmental engineering.

To our knowledge, very few studies (Almasri and Kaluarachchi 2005) have been conducted on ANN-based prediction of nitrate in groundwater. In this sense, this study aimed at predicting the nitrate concentrations using ANN in 24 representative observation wells in Harran Plain.

Description of the study area

The Harran Plain is located in the south central part of the GAP project within the Sanliurfa-Harran irrigation district (Fig. 1), which is 30 km × 50 km and is located in a region of rolling hills and a broad plateau that extends south into Syria. The plain, the largest in the GAP region, has 141,500 ha of irrigable land, 3,700 km² of drainage area and 1,500 km² of plain area. Groundwater samples were taken monthly for 1 year from 24 representative observation wells (Fig. 2), which were drilled on the Pleistocene aged unit during the 2006 water year. The detailed information on geology and hydrogeology of

Fig. 1 Location map of the study area



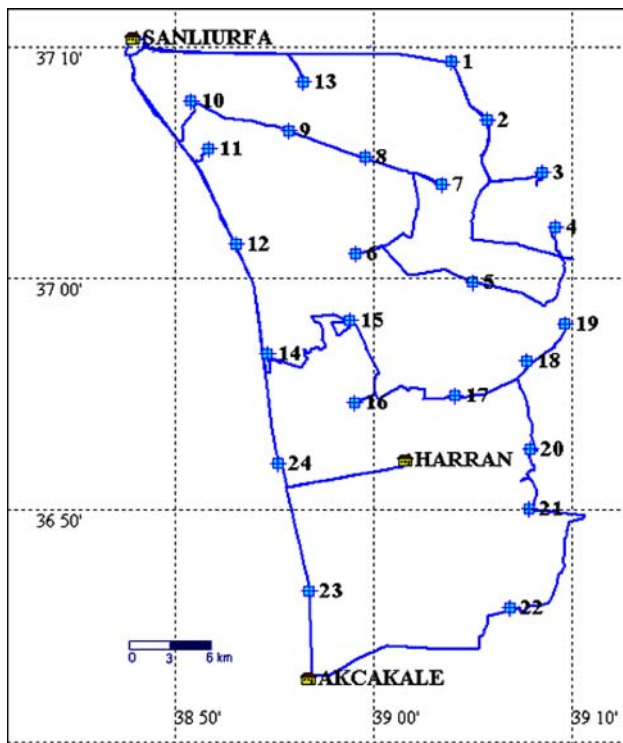


Fig. 2 Location of the sampling wells in the study area

the study area can be found elsewhere (Yesilnacar and Gulluoglu, in press).

Analytical methods

Electrical conductivity, temperature, pH and groundwater level were measured with a SevenGo pro-SG7 conductivity meter, YSI 6600 sonde, a portable pH meter and an electric contact meter immediately after sampling in the field.

Modelling

A neural network is defined as a system of simple processing elements, called neurons, which are connected to a network by a set of weights (Fig. 3). The network is determined by the architecture of the network, the magnitude of the weights and the processing element’s mode of operation. The neuron is a processing element that takes a number of inputs, weights them, sums them up, adds a bias and uses the result as the argument for a singular valued function, the transfer function, which results in the neuron’s output (Strik et al. 2005). At the start of training, the output of each node tends to be small. Consequently, the derivatives of the transfer function and changes in the connection weights are large with respect to the input. As learning

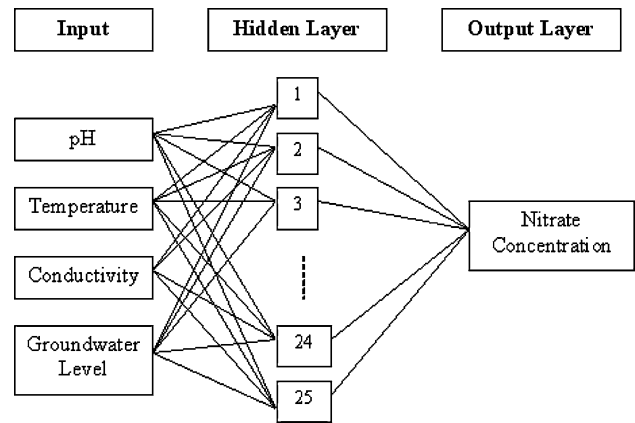


Fig. 3 The neural network structure for the prediction of nitrate concentration in monitoring wells

progresses and the network reaches a local minimum in error surface, the node outputs approach stable values. Consequently the derivatives of the transfer function with respect to input, as well as changes in the connection weights, are small (Maier and Dandy 1998).

Back-propagation (BP) algorithms use input vectors and corresponding target vectors to train ANN. The standard BP algorithm is a gradient descent algorithm, in which the network weights are changed along the negative of the gradient of the performance function (Abdi et al. 1996; Nguyen and Widrow 1990). There are a number of variations in the basic BP algorithm that is based on other optimization techniques such as conjugate gradient and Newton methods. For properly trained BP networks, a new input leads to an output similar to the correct output. This ANN property enables training of a network on a representative set of input/target pairs and achieves sound forecasting results.

Although some researchers suggest that one hidden layer is usually sufficient (El-Din and Smith 2002), the introduction of additional hidden layers allows the fit of a larger variety of target functions and enables approximations of complex functions with fewer connection weights (Toth et al. 2000). In this work, a two-layer ANN with a tan-sigmoid transfer function for the hidden layer and a linear transfer function for the output layer were used. Figure 3 shows the ANN structure used in the study. The

Table 1 Input and output parameters in ANN modelling

Input parameters (P)	
(P1)	PH
(P2)	Temperature (°C)
(P3)	Electrical conductivity (µS/cm)
(P4)	Groundwater level (m)
Output parameters (T)	
(T1)	Nitrate (mg/L)

input and output parameters used in the ANN modelling were as shown in Table 1. The data were divided into training, validation and test subsets. Half of the data were used for training and one fourth of the data was used for validation and tests, respectively.

Selection of back-propagation algorithm

Back-propagation neural networks have become a popular tool for modelling environmental systems (Maier and Dandy 1998). In this study, 12 BP algorithms were compared to select the best fitting one. For all algorithms, we used a two-layer network with a tan-sigmoid transfer function within the hidden layer and a linear transfer function within the output layer. In the selection of BP algorithm, the number of neurons was kept constant at 20. The learning rate parameter may also play an important role in the convergence of the network, depending on the application and network architecture. The learning rate can be used to increase the chance of preventing the training process being trapped in a local minimum instead of a global minimum (Hamed et al. 2004). A larger learning rate involves a bigger step. If the learning rate is too large, the algorithm becomes unstable. If the learning rate is set too small, the algorithm takes a long time to converge. In addition, the momentum allows a network to respond not only to the local gradient, but also to recent trends in the error surface. Without momentum, a network may get stuck in a shallow local minimum (Hagan et al. 1996). In this study, the learning rate and the momentum constant were 0.1 and 0.9, respectively. Training results are provided in Table 2. The performance of the BP algorithms was evaluated with the root mean square error (MSE) and

Table 2 Comparison of back-propagation algorithms for predicting nitrate concentration in monitoring wells (neuron number was 20)

Back-propagation algorithms	Nitrate		
	<i>R</i> -value	MSE	Iteration number
Trainlm	0.92	0.0794	18
Traincgp	0.91	0.1511	51
Traingd	0.80	0.3400	100
Traingda	0.84	0.2600	82
Traingdx	0.87	0.2156	100
Trainrp	0.88	0.1768	39
Trainscg	0.89	0.1754	31
Trainoss	0.86	0.2115	23
Traincgf	0.89	0.1768	29
Trainbfg	0.86	0.2076	18
Traingdm	0.17	18.6201	10
Traincgb	0.89	0.1804	26

Table 3 *R*-values, mean square errors and iteration numbers at different neuron numbers for predicting nitrate concentration in monitoring wells (Levenberg–Marquardt BP algorithm was used)

Neuron number	<i>R</i> -value	MSE	Iteration number
3	0.88	0.2179	51
5	0.91	0.1124	24
10	0.91	0.09361	12
15	0.91	0.09266	14
20	0.92	0.0794	18
25	0.93	0.0488	12
30	0.90	0.07108	11
40	0.91	0.06821	9

determination coefficient (*R*) between the modelled output and measured data set. The best BP algorithm with minimum training error and maximum *R* was the Levenberg–Marquardt (trainlm) algorithm.

Optimisation of neuron number

After selecting best BP algorithm, Levenberg–Marquardt (trainlm) algorithm, the number of neurons was optimised keeping all other parameters constant (Table 3). For the output variable (nitrate concentration), the MSE decreased for the training set with increasing neuron numbers (Table 3). However, after optimum neuron number (25 in our case), the MSEs did not change significantly. So, all the modelling was carried out using Levenberg–Marquardt (trainlm) algorithm with 25 neurons.

Model results and discussion

The applicability of ANN was investigated to predict the nitrate concentrations in 24 observation wells in Harran Plain. The groundwater quality in the observation wells was previously described in detail by Yesilnacar and Gulluoglu (2007). In this study, we have used the

Table 4 Influence of each variable on the performance of ANN prediction (Levenberg–Marquardt BP algorithm with 25 neurons was used)

Variable	<i>R</i> -value	MSE	Iteration number	Influence on <i>R</i> (%)
Base case	0.93	0.0488	12	–
Conductivity	0.64	0.2943	19	31
PH	0.87	0.1186	16	6.5
Groundwater level	0.88	0.1123	12	5.4
Temperature	0.91	0.1045	11	2.2

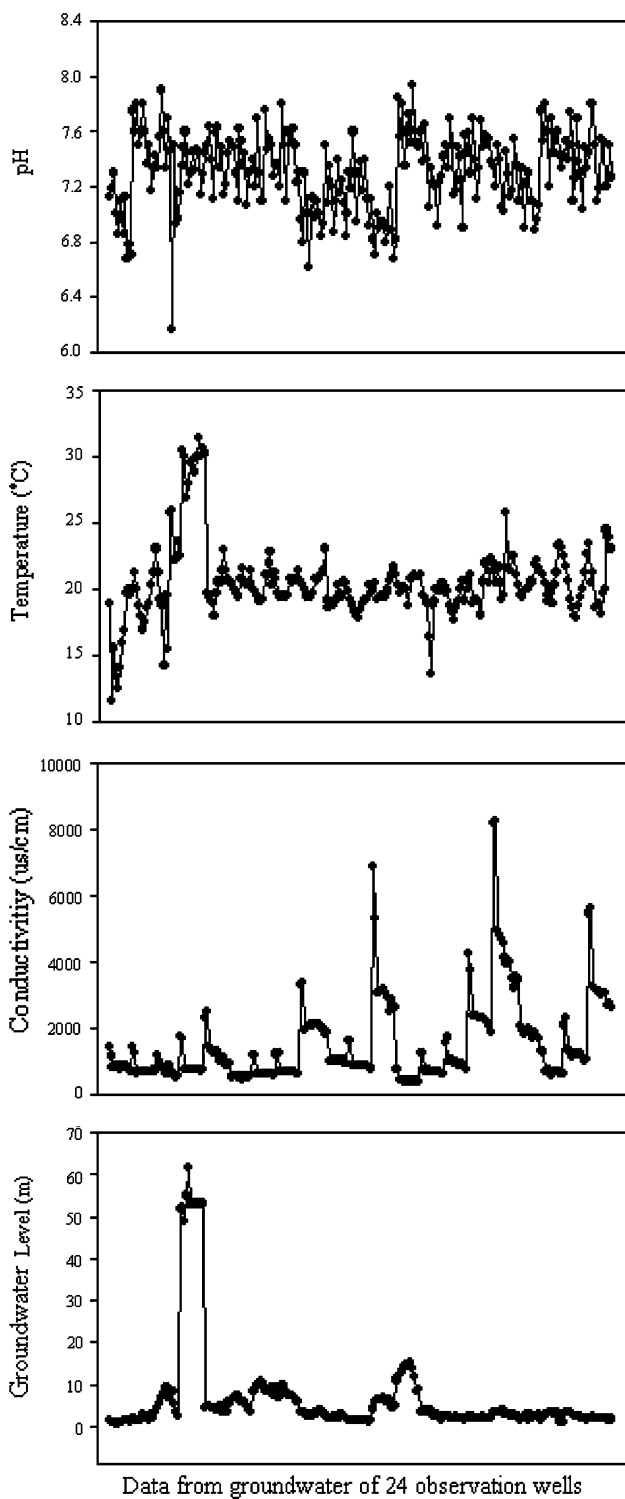


Fig. 4 The variation of input parameters in ANN modelling

aforementioned easily measurable water quality parameters (Table 1) in the prediction of nitrate concentrations.

The variation of input parameters is provided in Fig. 4. The groundwater of the study area is mainly alkaline in nature as the pH values of groundwater ranged from 7.0 to

7.5. The average temperature of the wells was around 20°C. The EC value in the wells varied between 400 and 8,235 µS/cm, with the average of 1,526 µS/cm. The maximum allowable conductivity value is 2,500 µS/cm in TS 266 (Turkish Standard Institution, standard of water intended for human consumption) and the European Union (EU) directives (Yesilnacar and Gulluoglu 2007). Hence, the groundwater conductivity in most of the observation wells exceeds the maximum allowable value. Previous studies have reported soil salinity of this area to be fairly high as well. Excessive amounts of dissolved ions in irrigation water affect plants and agricultural soil physically and chemically, thus reducing the productivity (Şahinci 1991). The groundwater levels in the wells were between 0.6 and 15 m except for well 4 in which groundwater level was around 50 m (Fig. 4).

The maximum allowable concentration of nitrate in drinking water is 50 mg/L according to the TS 266, the WHO guidelines and the EU directive (Yesilnacar and Gulluoglu 2007). Nitrate was well above the maximum admissible concentration and it ranged from 1.3 to 806 mg/L, with an average value of 153 mg/L. This is probably the result of the excessive use of artificial fertilizers in intensive agricultural activities, especially in the southeast part of Şanlıurfa and in the vicinity of Harran. The excessive nitrate concentration in the study area is triggered by excessive and uncontrolled irrigation.

Training, validation and test MSE for Levenberg–Marquard algorithm with 25 neurons were illustrated in Fig. 5. The training was stopped after 12 iterations as MSE did not

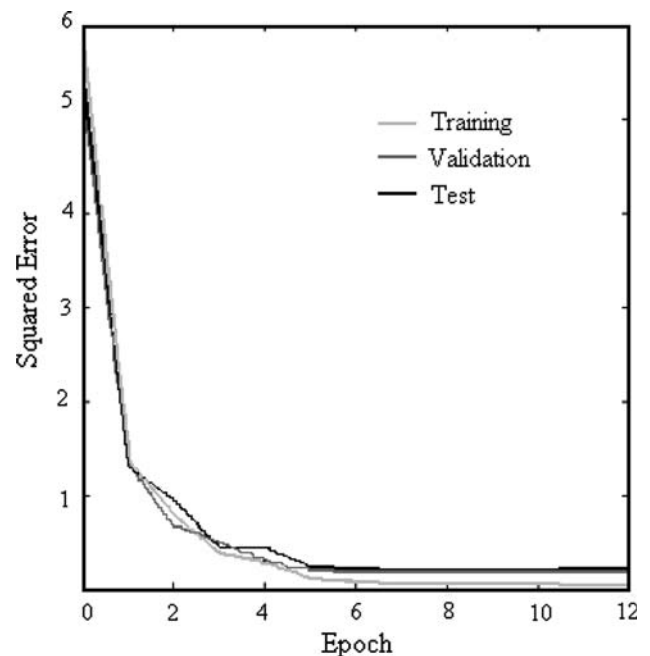


Fig. 5 Training, validation and test square mean errors for Levenberge–Marquardt algorithm with 25 neurons for nitrate prediction

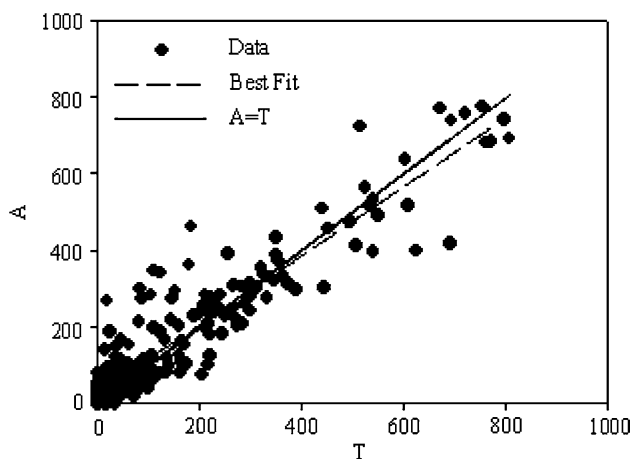


Fig. 6 Linear regression between the network outputs (A) and the corresponding targets (T) for output for Levenberge–Marquardt algorithm with 25 neurons

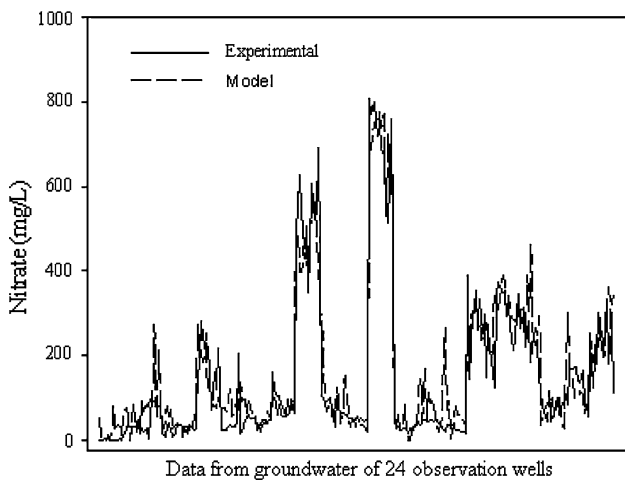


Fig. 7 Measured and the neural network prediction of nitrate concentration in groundwater of 24 observation wells

change significantly. We also performed a regression analysis between network output (A) and the corresponding target (T) (Fig. 6). The R and MSE values were observed as 0.93 and 0.0488, respectively (Table 3). The variations of measured and predicted data were also presented in Fig. 7 and the model data tracked the experimental data closely. Hence, with the use of easily measurable parameters as input data ANN can predict the nitrate concentration in groundwater.

The effect of eliminating each input variable on the ANN prediction performance was also analysed based on the correlation index (R) using the expression given below (Eq. 1) (Gontarski et al. 2000).

$$\text{Influence}(\%) = 100(1 - R_i/R_{CB}) \quad (1)$$

Where R_{CB} is the correlation index between predicted and observed values for the base case. R_i is the correlation

index for the case in which one input variable is eliminated. In the estimation of nitrate, conductivity is the most important parameter among the used input parameters as the elimination of conductivity decreased the R -value from 0.93 to 0.64 (Table 4).

Conclusions

The quality of groundwater of Harran Plain was monitored in 24 observation wells for 1 year. The concentrations of nitrate, the most common contaminant in groundwater, were well above the maximum admissible concentration of 50 mg/L for human consumption. The nitrate pollution is due to the intensive agricultural practices and the excessive use of artificial fertilizers. Nitrate can be easily predicted with the designed, trained, validated neural network model. EC was found to be the most significant parameter within input parameters used in the modelling. The developed model gave satisfactory fit to the experimentally obtained nitrate data in 24 observation wells in Harran Plain. Hence, with the proposed model applications it is possible to manage groundwater resources in a more cost-effective and easier way.

Acknowledgments This study was funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK project no: 104Y188) and the Scientific Research Projects Committee of Harran University (HÜBAK project no: 603). The authors would like to thank Yasemin Bayindir, Ozlem Demir, Atiye Atguden and Nuray Gok for their continuous help in the field and laboratory studies.

References

- Abdi H, Valentin D, Edelman B, O'Toole AJ (1996) A Widrow-Hoff learning rule for a generalization of the linear auto-associator. *J Math Psychol* 40(2):175–182
- Almasri MN, Kaluarachchi JJ (2005) Modular neural networks to predict the nitrate distribution in ground water using the on-ground nitrogen loading and recharge data. *Environ Model Softw* 20(7):851–871
- Clair TA, Ehrman JM (1996) Variations in discharge and dissolved organic carbon and nitrogen export from terrestrial basins with changes in climate: a neural network approach. *Limnol Oceanogr* 41(5):921–927
- El-Din AG, Smith DW (2002) A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Res* 36(5):1115–1126
- Freeze RA, Cherry JA (1979) *Groundwater*. Prentice-Hall, Inc., New Jersey, 604pp
- Gontarski CA, Rodrigues PR, Mori M, Prenem LF (2000) Simulation of an industrial wastewater treatment plant using artificial neural networks. *Comput Chem Eng* 24(2–7):1719–1723
- Hagan MT, Demuth HB, Beale MH (1996) *Neural network design*. PWS Publishing, Boston
- Hamed MM, Khalafallah MG, Hassanien EA (2004) Prediction of wastewater treatment plant performance using artificial neural networks. *Environ Model Softw* 19(10):919–928

- Karaca F, Özkaya B (2006) NN-LEAP: a neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site. *Environ Model Softw* 21(8):1190–1197
- Kendirli B, Cakmak B, Ucar Y (2005) Salinity in the Southeastern Anatolia Project (GAP), Turkey: issues and options. *Irrig Drain* 54(1):115–122
- Maier HR, Dandy GC (1996) The use of artificial neural networks for the prediction of water quality parameters. *Water Resour Res* 32(4):1013–1022
- Maier HR, Dandy GC (1998) Understanding the behaviour and optimizing the performance of back-propagation neural networks: an empirical study. *Environ Model Softw* 13(2):179–191
- Nguyen D, Widrow B (1990) Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In: *Proceedings of the international joint conference on neural networks*, vol 3. IEEE, San Diego, pp. 21–26
- Onkal-Engin G, Demir I, Engin SN (2005) Determination of the relationship between sewage odour and BOD by neural networks. *Environ Model Softw* 20(7):843–850
- Ozdogan M, Woodcock CE, Salvucci GD, Demir H (2006) Changes in summer irrigated crop area and water use in southeastern Turkey from 1993 to 2002: implications for current and future water resources. *Water Resour Manage* 20(3):467–488
- Rodriguez MJ, Sérodes JB (1998) Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems. *Environ Model Softw* 14(1):93–102
- Şahinci A (1991) *Geochemistry of natural waters*. Reform printing office, Izmir, 548pp (in Turkish)
- Sahinkaya E, Özkaya B, Kaksonen AH, Puhakka JA (2007) Neural network prediction of thermophilic (65°C) sulfidogenic fluidized-bed reactor performance for the treatment of metal-containing wastewater. *Biotechnol Bioeng* 97(4):780–787
- Strik DPBTB, Domnanovich AM, Zani L, Braun R, Holubar P (2005) Prediction of trace compounds in biogas from anaerobic digestion using the MATLAB neural network toolbox. *Environ Model Softw* 20(6):803–810
- Subramani T, Elango L, Damodarasamy SR (2005) Groundwater quality and its suitability for drinking and agricultural use in Chithar River Basin, Tamil Nadu, India. *Environ Geol* 47(8):1099–1110
- Toth E, Brath A, Montanari A (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. *J Hydrol* 239(1–4):132–147
- Yesilnacar MI, Gulluoglu MS (2007) Hydrochemical characteristics and the effects of irrigation on groundwater quality in Harran Plain, GAP Project, Turkey. *Environ Geol*. doi:[10.1007/s00254-007-0804-9](https://doi.org/10.1007/s00254-007-0804-9)