Desalination and Water Treatment
www.deswater.com

11 (2009) 258–263

Determination of relationship between hardness and groundwater quality parameters by neural networks

N. Balkaya[a], H. Kurtulus Ozcan[a,*], O. Nuri Ucan[b]

[a]*Environmental Engineering Department, Istanbul University, Engineering Faculty. 34 320, Avcilar, Istanbul, Turkey*
Tel. +90-212-4737070/17726; Fax. +90-212-4737180; email: hkozcan@istanbul.edu.tr
[b]*Electrical-Electronics Engineering Department, Istanbul University, Engineering Faculty. 34 320, Avcilar, Istanbul, Turkey*

Received 27 March 2009; Accepted 26 August 2009

ABSTRACT

All life forms on the earth contain water and water is crucial for any life form on the earth. Apart from being the essential ingredient if living organisms, water has numerous other uses and benefits. Groundwaters form a circle of the natural hydrologic chain like surface waters and the other water in the atmosphere. Hydrologic, hydraulic and geologic processes play important roles during underground water's formation, storage, underground flow and coming up to the surface of the earth. In this study, groundwater hardness quality at Samsun Incesu-Derekoy region was modeled by the use of Artificial Neural Network (ANN) structure. In the data set arrangement effective input variables are the five different water quality parameters (pH, chlorine, calcium, magnesium and total hardness) concentrations in the time *"t"*, and the output variable (total hardness) is the concentrations in the time *"t + 1"*. For the model 10,000 epochs were performed and the learning rate is equal to 0.1, and correlation coefficient (r) that achieved in this study was found 0.591. As a result, we conclude that ANN is the effective modeling technique on estimation of environmental complex water quality problems.

Keywords: Artificial neural network (ANN); Ground water quality; Hardness

1. Introduction

Water is an inevitable component for all the creatures on earth. People meet their need for water from groundwater which moves and accumulates in the pores of the masses within the earth's crust along with the surface waters such as rivers and lakes. In spite of the abundant natural waters existing on earth, obtaining water requires time and effort. The quality and the quantity of the water resources are affected by factors such as urbanization, expanding population and pollution seen in available water resources [1, 2]

Groundwaters form a circle of the natural hydrologic chain like surface waters and the other water in the atmosphere. The essence of groundwaters is the water circulating in various ways in nature. Hydrologic, hydraulic and geologic processes play important roles during groundwater formation, storage, underground flow and coming up to the surface of the earth [3].

The waters taking place in different depths of the crust of earth are connected with the masses of various compositions. According to the dissolution levels of these masses, more or less dissolved substances flow into groundwater. The quantity of the dissolved substances alters according to the connection period of the masses with groundwater, the speed of the water, the temperature, the type of the mass, the pH and the pressure of the environment [3]. Carbon dioxide which is a gas dissolved from the air by the water, inclines the acidity of the water and plays an important role for the dissolution of the minerals in the soil by the water. Hence

*Corresponding author.

doi: 10.5004/dwt.2009[illegible]

groundwaters involve more dissolved substances once it is compared with surface waters. The major mineral salts which are dissolved by water are calcium, magnesium, sodium, potassium and iron salts. These substances are generally dissolved up to a level that does not harm the drinkability of the water. Moreover the existence of these substances in the water in certain amounts makes positive contributions to the drinkability feature of the water. However if these substances exceed a certain amount, the usage areas of groundwater are restricted.

Statistical models are based on semi-empirical statistical relations among available data and measurements. They do not necessarily reveal any relation between cause and effect. They attempt to determine the underlying relationship between sets of input data (predictors) and targets (predictands) [4]. ANNs are well suited to environmental modeling as they are nonlinear [5], relatively insensitive to data noise, and perform reasonably well when limited data are available [6]. When ANNs are used for the prediction of environmental variables, the modeling philosophy employed is similar to that used in the development of more conventional statistical models.

In recent years, artificial neural networks (ANNs) have become a popular and useful tool for modeling environmental systems. For example, they have already been successfully used to simulate the export of nutrients from river basins [7], to forecast salinity [8] and ozone levels [9,10], to predict air pollution [11,12], the functional characteristics of ecosystems [13], to model algal growth, and transport in rivers [14]. Over the last few years, recurrent networks (i.e., networks with feedback connections) have been proposed as an alternative to feed forward networks, particularly in time series applications [15,16,17]. They have the ability to model time structure implicitly with the aid of their feedback connections, although problems with "remembering" longer term dependencies have been encountered [18,19,20].

2. Experimental details

2.1. Investigated area

Research area is situated between Incesu Creek and Derekoy in the provincial border of Samsun in Middle Black Sea Region, Turkey. The area is restricted with Incesu Creek in the east, Derekoy Channel in the west, Samsun-Sinop Highway (D010) in the south and Black Sea in the north (Fig. 1). The topography of the area is a seaside plain the altitude of which ranges from 0.5 to 1. Derekoy Stream and Incesu Creek are the important rivers in the region. There are three important soil groups in research area: grey-brown podzolic soil, brown forest soil and colluvial soil [1].

Fig. 1. Investigated area.

Seawater input is expected due to the closeness of research area to the sea. It is an important area in terms of excessive taking for water use because there are septic tanks since there is no sewer; the ground is permeable and the area is used as countryside in summer.

2.2. Analytical procedure

While delivering the samples, the water was drained by making use of pumps until it was assured that the bottom water that maintains water for those areas came up. The sample water was put into glass and pyrex bottles and they were covered in order to prevent connection with air. During the study period, pH, chlorine, calcium, magnesium and total hardness experiments were conducted according to the Standard Methods [20].

2.3. Artificial neural networks

The neural models are basically based on the perceived work of the human brain. The artificial model of the brain is known as Artificial Neural Network (ANN) or simply Neural Networks (NN). Rapid, efficient propagation of electrical and chemical impulses is the distinctive characteristic of neurons and the nervous system in general. The neurons operate collectively and simultaneously on most for all data and inputs, which performs as summing and non-linear mapping junctions. In some cases they can be considered as threshold units that fire when total input exceeds certain bias level. Neurons usually operate in parallel and are configured in regular architectures. They are often organized in layers, and feedback connections both within the layer and toward adjacent layers are allowed. Strength of each connection is expressed by a numerical value called a weight that can be updated. Also they are characterized by their time domain behavior, which is often referred as dynamics. In general, the neuron could be modeled as a non-linear activated function of which the total potential inputs into synaptic weights are applied. It is assumed that synapses can impose excitation or inhibition but not both on the receptive neuron. In present Artificial Neural Network (ANN) model, input (i), hidden (k) and output (1) neurons are used in model were shown in Fig. 2.

Where, ($x_1, x_2, x_3, \ldots, x_i$) are input and y output parameters. In this model ($w_{1(1,1)}, \ldots, w_{1(i,k)}$) and ($w_{2(1,1)}, \ldots, w_{2(k,1)}$) are input and output weight coefficients, re spectively, which will be trained to find an optimum solution.

The artificial model of neuron consists of three elements. These are; A set of synapses or connection links, each of which is characterized by a weight or strength of its own. Specially, a signal xj at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj}. Unlike a synapse in the brain, the synaptic weight of an artificial neuron may lie in a range that includes negative as well as positive values. An adder for summing the input signals, weighted by the respective synapses of the neuron. An activation function or transfer functions for limiting the amplitude of the output of a neuron.

In present model, neural network is trained and tested using MATLAB 7.0. A three-layer neural network that consists of an input layer, output layer and one hidden layer is used and structure of this network is presented in Fig 2. The monitoring data (pH, Ca^{+2}, Cl^{-}, Mg^{+2}, total hardness) belonging to study period, was designed to meet the requirements of training and testing the neural network. Groundwater samples have conducted for five months. This database is divided into training and testing sets taking the odd numbered patterns as training data and even numbered ones as test data. Size of training and test matrices are 55×5 and 54×5 respectively. In training and testing set, y presents total hardness in Fig. 2. In simulations, various ANN models are tested changing the number of neurons in the hidden layer between 2 and 30. All the data

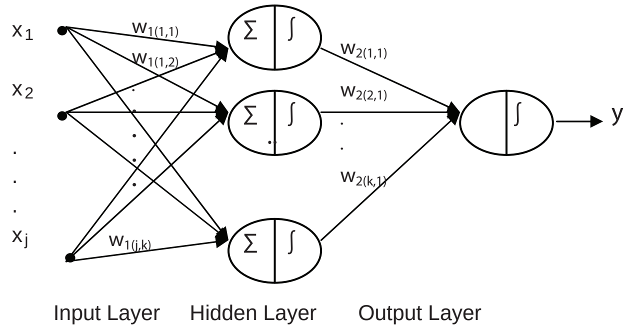

Fig. 2. Structure of the three layers ANN.

are normalized into the range {–1.0, 1.0}. This is carried out by determining the maximum and minimum values of each variable over the whole data period and calculating normalized variables using the equation 1.

$$Xnorm = 2 * \left[\frac{(X - X\min)}{(X\max - X\min)} \right] - 1.0. \quad (1)$$

3. Results and discussion

In this study, groundwater hardness quality at Samsun Incesu-Derekoy region was modeled by the use of Artifcial Neural Network (ANN) structure. In the data set arrangement effective input variables are the five different water quality parameters (pH, chlorine, calcium, magnesium and total hardness) concentrations in the time *"t"*, and the output variable (total hardness) is the concentrations in the time *"t + 1"*. Matrix structure which used in ANN modeling was shown at Fig. 3.

As it is seen in Fig. 3, arranged input and output data sets were used in training and test phases. Active input variations are five different water quality parameters in *"t"* period and output variation is hardness parameter in *"t + 1"* period. Training coefficient was chosen as (*lr*) 0.1 in used model structure and algorithm was operated. Training algorithm learning rate and sum-squared error values were presented in Fig. 4 and training algorithm outputs were presented in Fig. 5.

The neuron number of hidden layer is an important factor where else the number cascade connected hidden layer is not so effective. In our study, we have compared various input, hidden and output layers and for our problem. We have altered number of neurons for hidden layer and the best-suited model for ANN is found as (6,12,1) corresponding number of neurons in input, hidden and output layers respectively. For the model 10,000 epochs were performed and the learning rate is equal to 0.1. After the training phase of the model, coefficients were calculated and test phase was started. Model estimations acquired from test data set were given in Fig. 6.

Evaluation can also be undertaken by considering measures of agreement, such as the Pearson product moment correlation coefficient (*r*) values. The index of agreement, bounded, relative measure that is capable to measure the degree of which predictions are error-free. The denominator accounts for the model's deviation from the mean of the observations as well as to the observations deviation from their mean. In a good model, *r* should approach to 1. Correlation coefficient (*r*) that achieved in this study was found 0.591 (Fig. 7).

Correlations between model variables and hardness were calculated and results were given in Table 1. These correlation values indicate that how to model input

P=[	$Hardness_{(t1)}$	$Hardness_{(t2)}$	$Hardness_{(t3)}$	.	.	$Hardness_{(t54)}$
	$Cl_{(t1)}$	$Cl_{(t2)}$	$Cl_{(t3)}$	.	.	$Cl_{(t54)}$
	$Ca_{(t1)}$	$Ca_{(t2)}$	$Ca_{(t3)}$	.	.	$Ca_{(t54)}$
	$Mg_{(t1)}$	$Mg_{(t2)}$	$Mg_{(t3)}$	.	.	$Mg_{(t54)}$
	$pH_{(t1)}$	$pH_{(t2)}$	$pH_{(t3)}$	.	.	$pH_{(t54)}$]
T=[	$Hardness_{(t1+1)}$	$Hardness_{(t2+1)}$	$Hardness_{(t3+1)}$	.	.	$Hardness_{(t54+1)}$]

Fig. 3. Input (P) and output (T) structure of ANN matrix.

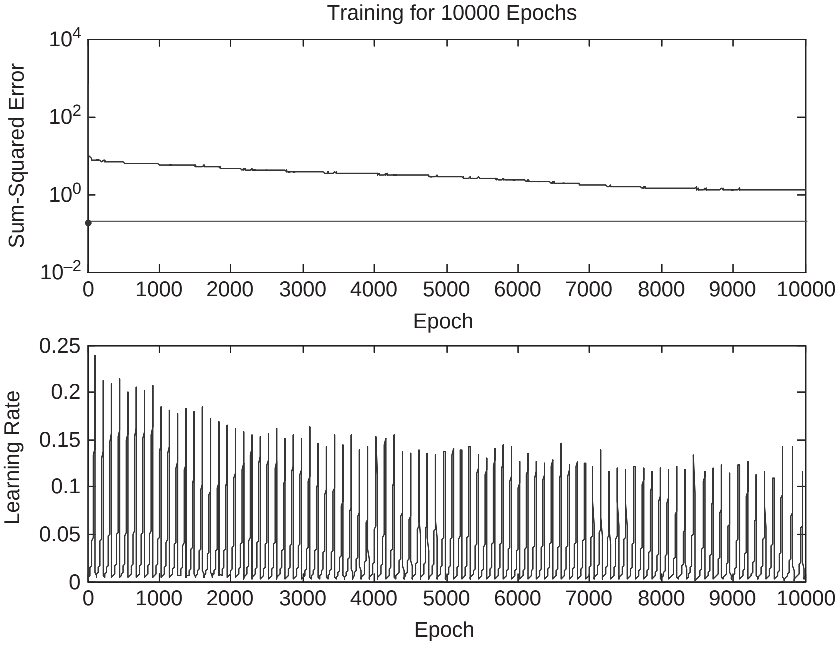

Fig. 4. Sum-squared error and learning rate of the ANN model.

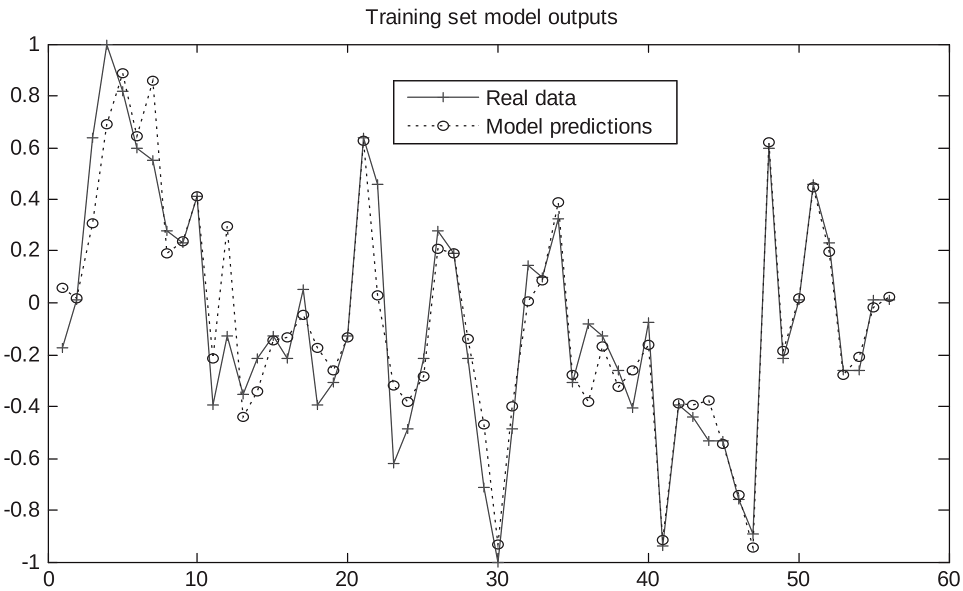

Fig. 5. Training outputs of ANN model.

Table 1
Correlations between model parameters and hardness.

	Hardness	*Cl*	*Ca*	*Mg*	*pH*
Hardness	1	0.336	0.670	0.170	0.294
Cl	0.336	1	0.217	0.237	0.056
Ca	0.670	0.217	1	0.052	0.220
Mg	0.170	0.237	0.052	1	0.272
pH	0.294	0.056	0.220	0.272	1

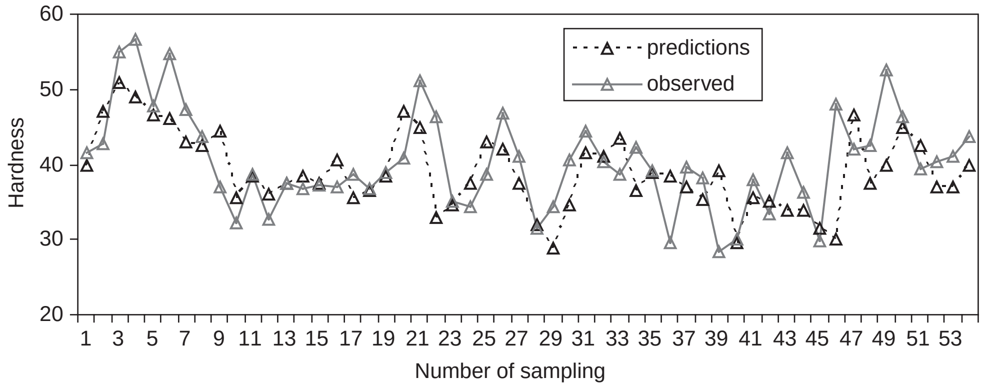

Fig. 6. Outputs for ANN model structure.

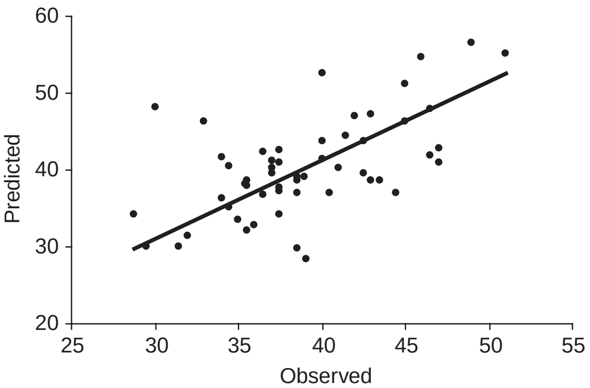

Fig. 7. Scatter plots of observed hardness concentration against predictions for Artificial Neural Network (ANN) model.

parameters affect on model results. As seen that in Table 1, inputs which correlations lower than the others can be eleminated and varius model structure for different conditions can be tested. As it is shown in Table 1 the highest correlation is between hardness and calcium.

4. Conclusion

In this study, various quality parameters of the groundwaters in Cakırlar, Yakamoz and Incesu Regions taking place in the seaside plain in the west of Samsun City were examined and the results were modeled via ANN modeling technique. Once the data sets acquired from the examined groundwater and used in the model during the study have been analyzed; it is seen that pH value ranges from 6.81 to 7.83. The chlorine values acquired in this study are between 29.78 mg/L and 58.93 mg/L. Also it is seen that calcium values are between 86 mg/L and 131.6 mg/L. While the magnesium values vary between 21.30 and 43.68 mg/L, the total hardness of the samples is calculated between 32.90 and 49.10 FS°. It is known that in groundwaters the values such as calcium, magnesium and chloride increase after rain. The acidic feature of the rain in research area facilitates the dissolution of the minerals in soil and enhances the alteration of these values in time.

During the modeling phase of the study, the hardness of groundwater was estimated via ANN modeling technique and the results were compared with the real values statistically (Fig. 7). The correlation coefficient between ANN model estimations and observed values is $r = 0.591$. Once the Figs. 5, 6 and 7 are analyzed, it is seen that ANN modeling technique can be used to determine the quality of the groundwater effectively.

The proposed ANN model can be used to predict other groundwater quality parameters almost without further adaptation. It also can be implemented in the form of water quality forecast models to provide predictions of pollutant trends within different time series, e.g., 1 week, 1 month, or even 1 year in advance, etc. One of the advantages of the artificial neural network (ANN) lies in the fact that the deterministic model needs a lot of information, whereas the neural network acts like a black-box model [21]. The drawback of the neural approach is that no deep understanding on the physical phenomena is gained by using a neural network, since it resembles the behavior of a black-box method. Moreover, the ANN, once trained, is fast at predicting the desired values. The last and the most convincing advantage is that the accuracy of the neural prediction is generally higher than the other kind of models. Another advantage is that the ANN network may also be applied to other investigated areas. More studies are needed to apply the developed system to other groundwater quality monitoring areas and to assess the changing of groundwater quality predictions due to inherent errors in the measured database.

References

[1] N. Balkaya, Groundwater pollution investigation between Incesu Derekoy plain at Samsun Region, Master Thesis-Ondokuz Mayis University, Natural Science Institute, Samsun, Turkey, 1989.

[2] A. Coynel, J. Schafer, A. Dabrin, N. Girardot and G. Blanc, Groundwater contributions to metal transport in a small river affected by mining and smelting waste, Water Res., 41 (2007) 3420–8.

[3] A. Furman and O. Yenigun, The environmental dimension, Bogazici University press, 2nd edn, Istanbul, 1998.

[4] S.A. Wahab-Abdul and S.M. Al-Alawi, Assesment and prediction of trospheric ozone concentration levels using artifical neural networks, Environ. Model. Softw., 17 (2002) 219–28.

[5] K. Chakraborty, K. Mehrotra, C.K. Mohan and S. Ranka, Forecasting the behavior of multivariate time series using neural networks, Neural Netw., 5 (1992) 961–70.

[6] Z. Tang, C. De Almeida and P.A. Fishwick, Time series forecasting using neural networks vs. Box-Jenkins methodology, Simulation, 57(5) (1991) 303–10.

[7] T.A. Clair and J.M. Ehrman, Variations in discharge and dissolved organic carbon and nitrogen export from terrestrial basins with changes in climate: A neural network approach, Limnol. Uceanogr., 41(5) (1996) 921–7.

[8] L. Desilets, B. Golden, Q. Wang and R. Kumar, Predicting salinity in the Chesapeake Bay using backpropagation, Comput. Oper. Res., 19(3/4) (1992) 227–85.

[9] A.C. Comrie, Comparing neural networks and regression models for ozone forecasting, J. Air Waste Manage. Assoc., 47(6) (1997) 653–63.

[10] H.K. Ozcan, U. Sahin, O.N. Ucan and C. Bayat, Modelling Of Trophospheric Ozone (O_3) Concentrations Using Cellular Neural Networks, J. Fac. Eng. Arch. Gazi Univ., 21(2) 2006 239–45.

[11] M. Boznar, M. Lesjak and P. Mlakar, A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain, Atmos. Environ., 27 B(2) (1993) 221–30.

[12] H.K. Ozcan, O.N. Ucan, U. Sahin, M. Borat and C. Bayat,-Artificial neural network modeling of methane emissions at Istanbul Kemerburgaz-Odayeri Landfill Site, J. Sci. Ind. Res., 65(2) (2006) 128–34.

[13] J.M. Paruelo and F. Tomasel, Prediction of functional characteristics of ecosystem, A comparison of artificial neural networks and regression models, Ecol. Model., 98(2/3) (1997) 173–86.

[14] P.G. Whitehead, A. Howard and C. Arulmani, Modeling algal growth and transport in rivers-A comparison of time series analysis, dynamic mass balance and neural network techniques, Hydrobiologia, 349 (1997) 39–46.

[15] R. Gencay and T. Liu, Nonlinear modelling and prediction with feedforward and recurrent networks, Physica, D 108(1/2) (1997) 119–34.

[16] V.G. Krishnapura and A. Jutan, ARMA neuron networks for modeling nonlinear dynamical systems, Can. J. Chem. Eng., 76(3) (1997) 574–82.

[17] K. Udono and R. Site, Modeling seawater desalination powered by waste incineration using a dynamic systems approach, Desalination, 229 (2008) 302–17.

[18] T. Lin, B.G. Horne, P. Tino and C.L. Giles, Learning long-term dependencies in NARX recurrent neural networks, IEEE Trans. Neural Netw., 7(6) (1996) 1329–38.

[19] H.T. Siegehmann, B.G. Horne and C.L. Giles, Computational capabilities of recurrent NARX neural networks, IEEE Trans. Syst. Man Cyb., B 27(2) (1997) 208–15.

[20] H.R. Maier and G.C. Dandy, Neural Network based modeling of environmental variables: A systematic approach, Math. Computer Modeling, 33 (2001) 669–82.

[21] APHA-AWWA-WPCF. Standard Methods for the Examination of Water and Wastewater, 17th edition, New York, 1989.

[22] M. Anitha and H. Singh, Artificial neural network simulation of rare earths solvent extraction equilibrium data, Desalination, 232 (2008) 59–70.

[23] P. Viotti, G. Liuti and P. Di Genova, Atmospheric urban pollution: Applications of an artificial neural network (ANN) to the city of Perugia, Ecol Model., 148 (2002) 27–46.