Journal of Civil Engineering and Urbanism
Volume 4, Issue 1: 08-14 (2014) ISSN-2252-0430

Artificial Neural Network Modeling of Total Dissolved Solid in the Simineh River, Iran

Samira Nemati[1], Leila Naghipour[2*] and Mohammad Hasan Fazeli Fard[1]

[1]*Department of Water Engineering, University of Tabriz, Tabriz, Iran*
[2]*Department of Civil Engineering, University of Tabriz, Tabriz, Iran*

*Corresponding author's Email address: naghipour.l@gmail.com

ABSTRACT: This research aims to model Total Dissolved Solid (TDS) values at the Simineh River in northwest Iran by application of Artificial Neural Networks (ANNs) to evaluate existing water quality conditions and also to predict future conditions in this river. The input parameters of the ANNs model are Calcium (Ca), Chloride (Cl), Magnesium (Mg), Sodium (Na), Bicarbonate (HCO_3), Sulfate (SO_4), and water discharge (Q) from 1993 to 2011. The performance of the ANNs model was assessed in accordance with Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) between the measured and predicted values. The study also includes an estimation of the relative importance of these variables to determine appropriate input combinations. A method is used in this paper to calculate the relative importance of each input parameters, showing that magnesium and calcium concentrations are the most and least influential parameters, with approximate values of 18 and 12 %, respectively. The ANNs with different numbers of neurons in the hidden layer were constructed, and the model with 14 hidden neurons was selected as the best. Comparisons between the measured and predicted values show that the ANNs model could be successfully applied and provide high accuracy and reliability for water quality parameters forecasting.
Keywords: Artificial Neural Networks, Total Dissolved Solid, Simineh River, Relative Importance, Water Quality

ORIGINAL ARTICLE
Received 15 Aug. 2013
Accepted 28 Nov. 2013

INTRODUCTION

Contamination of surface water has become a major concern of local organizations involved to the management of water quality and quantity for human needs. However, assessment of surface water is difficult, because contamination is a function of numerous, complex interacting parameters. Nevertheless, hydrological modeling is a powerful technique of hydrologic system investigation for both research hydrologists and water resource engineers involved in the planning and development of integrated approaches for management of water resources. Since management decisions are usually made under conditions entailing considerable predictive uncertainties, realistic estimates of the possible errors contained in predictions are necessary. Limited water quality data and the high cost of water quality monitoring often pose serious problems for process-based modeling approaches. Ideally, these should be achieved by continuous updating of the parameters and predictive models. Higher precision can be achieved most efficiently using estimation methods.

Water quality in the Simineh River which is used mainly for agriculture is degraded by diffuse (non-point) sources. Among water quality parameters, Total Dissolved Solid (TDS) is defined as the quantity of dissolved material in water, and it is one of the vital water quality parameters and continuously used to determine the water quality of rivers. For this reason this parameter was assessed to evaluate the water quality condition in this basin.

Uncertainty is inherent in all methods of assessing surface water vulnerability to contamination and arises from errors in obtaining data, the natural spatial and temporal variability of the hydrogeology parameters in the field, and in numerical approximation. Traditional methods are poor at addressing the non-linearity, subjectivity, and complexity of the cause-effect relationships between water quality variables and water quality status; yet, they are the currently accepted methods. New approaches such as Artificial Intelligence (AI) techniques have proven their ability and applicability for simulating and modeling various physical phenomena in the water engineering field. Artificial Neural Networks (ANNs), which is another AI approach, could provide a framework from which real-time or simulated assessment of non-point source (NPS) pollution could be made in the Simineh River Basin. ANNs are parallel information processing system and emulate the working processes in the brain. A neural network is an adaptable system that learns relationships from the input and output data sets and then is able to predict a previously unseen data set of similar characteristics to the input set (Haykin, 1999).

Different aspects of this problem area are reported, among others, by Maier and Dandy (1996), Diamantopoulou et al. (2005), Kuo et al. (2006), Palani et al. (2008), Anctil et al. (2009), Singh et al. (2009), Chang et al. (2010), He et al. (2011), Gazzaz et al. (2012), Najad et al. (2013). ANNs also have been used successfully for predicting TDS (Kanani et al., 2008; Memon et al., 2009; Abudu et al., 2011; Asadollahfard et al., 2012; Mehrdadi et al., 2012; Abbasi et al., 2013; Heydari et al., 2013; Moasheri et al., 2013).

The present work describes the development and training of the ANNs model for the purpose of estimating TDS. Calcium (Ca), Chloride (Cl), Magnesium (Mg), Sodium (Na), Bicarbonate (HCO_3), Sulfate (SO_4), and water discharge (Q) for a set of recorded data from 1993 to 2011 in the Simineh River were used as input parameters to predict TDS. The paper also estimates the relative importance of these input variables. The remainder of this paper is outlined as follows: in Section 2, overview of the ANNs model, relative importance index, the model performance evaluation, case study and data specification are presented; in Section 3, the results achieved with the ANNs model are presented and discussed; and finally, in Section 4, is leaved out some conclusions.

MATERIALS AND METHODS

Artificial Neural Networks (ANNs)

McCulloch and Pitts (1943) are recognized as the first designers of Artificial Neural Networks (ANNs), which are generally inspired by the operation of the brain and nerve systems in biological organisms with a capability for self-learning and automatic abstracting. Neural networks consist of a set of neurons or nodes arranged in layers, and in the case weighted inputs are used, these nodes provide suitable inputs by conversion functions. Each neuron in a layer is connected to all the neurons of the next layer but without any interconnection among the neurons in the same layer. The weight learned for each neuron in ANNs model remains internal, and therefore, their associations with physical systems are often overlooked.

The ANNs modeling strategy is implemented by Neural Network Toolbox in *MATLAB* (MATLAB® software, 2013), which is a Feed Forward Neural Networks (FFNNs), or the Multi-Layer Perceptron (MLP). The neural architecture involves three different layers: (i) an input layer, (ii) a hidden layer, and (iii) an output layer. The number of neurons in the input and output layers is defined by the number of input and output variables respectively, while the number of neurons in the hidden layer(s) is usually determined by trial-and-error, and the neurons of each layer are connected to the neurons of the next layer by weights. In the hidden layer, each neuron computes w_{ij}, a weighted sum of its p input signals, x_i for $i = 1, 2, \ldots, n$, and applies a nonlinear activation function to produce an output signal, u_j. A neuron j is described mathematically by the following pair of equations:

$$u_j = \sum_{i=1}^{p} w_{ij}.x_i \qquad (1)$$

And

$$x_j = \varphi(u_j - \theta_j) \qquad (2)$$

Where θ is a threshold function and its use has the effect to apply an affine transformation to the output of the linear combiner in the model of Figure 1 (see Haykin (1999)). In this study, the logistic sigmoid nonlinear function is used for this purpose, expressed as

$$\varphi_x = \frac{1}{1 + e^{-x}} \qquad (3)$$

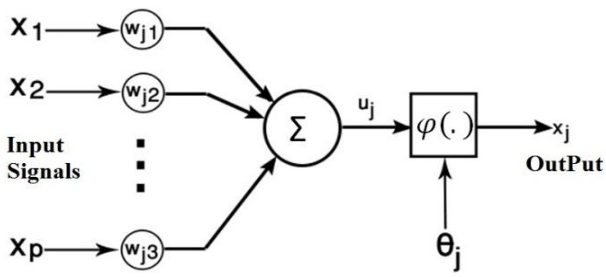


Figure 1. Nonlinear model of a neuron- see Haykin (1999).

Relative importance index

In order to assess the relative importance of the input variables, the Garson equation is used, which is based on the neural net weight matrix. Garson (1991) proposed the following equation based on the partitioning of connection weights:

$$I_j = \frac{\sum_{m=1}^{m=Nh} \left(\left(\frac{|W_{jm}^{ih}|}{\sum_{k=1}^{k=Ni} |W_{km}^{ih}|} \right) \times |W_{mn}^{ho}| \right)}{\sum_{k=1}^{k=Ni} \left\{ \sum_{m=1}^{m=Nh} \left(|W_{km}^{ih}| / \sum_{k=1}^{k=Ni} |W_{km}^{ih}| \right) \times |W_{mn}^{ho}| \right\}} \qquad (4)$$

Where I_j is the relative importance of the j^{th} input variable on the output variable, N_i and N_h are the number of input and hidden neurons, respectively, and W is connection weight, the superscripts i, h, and o refer to input, hidden, and output layers, respectively, and subscripts k, m, and n refer to input, hidden, and output neurons, respectively. For further details, see Weckman et al. (2009).

Model performance evaluation

In this study, performance of the models is assessed in accordance with the root mean square error (*RMSE*), mean absolute error (*MAE*), and coefficient of determination (R^2) between the observed and predicted values as shown in the tables. The *RMSE*, and *MAE* measure the errors while the R^2 indicates the goodness of fit. Scatter plots and time series plots are used for visual comparison of the observed and predicted values.

These performance measures and information criteria are calculated by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_i - \hat{T}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|T_i - \hat{T}_i| \quad (6)$$

$$R = \frac{\sum_{i=0}^{n}(T_i - \overline{T})(\hat{T}_i - \widetilde{T})}{\sqrt{\sum_{i=0}^{n}(T_i - \overline{T})^2}\sqrt{\sum_{i=0}^{n}(\hat{T}_i - \widetilde{T})^2}} \quad (7)$$

Where T_i and $\overline{T}$ are the observed values and their mean, respectively; $\hat{T}$ and $\widetilde{T}$ are the predicted values and their mean, respectively; n is the total number of records.

Case study and data specification

The Simineh River in northwest Iran is one of the largest and important rivers. The length of this river is about 200 km with a catchment of 2090 km^2. The geographical coordinates of Simineh basin lies between 45° 35′ to 46° 25′ East longitudes and 36° 1′ to 37° 56′ North latitudes (Figure 2).

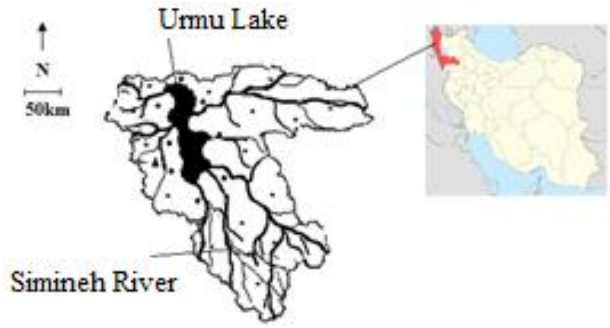


Figure 2. Location of the Simineh River

In this study, the water quality data at Dashband gauging station in the Simineh River during 1993-2011 are used. Concentrations of the parameters have been measured in 531 streams, and each record consists of 7 parameters including: Calcium (Ca), Chloride (Cl), Magnesium (Mg), Sodium (Na), Bicarbonate (HCO_3), Sulfate (SO_4), and water discharge (Q). In order to develop ANN model for prediction TDS, the data are divided into two sets: (i) the data used for training the models and these make up approximately 80 percent of data (425 sample); (ii) the data used for testing the models and these make up the remaining 20 percent of water quality data (106 sample). The mean variations of TDS and the other parameters of the gauging site used in this study are shown in Figure 3(a)–(h).

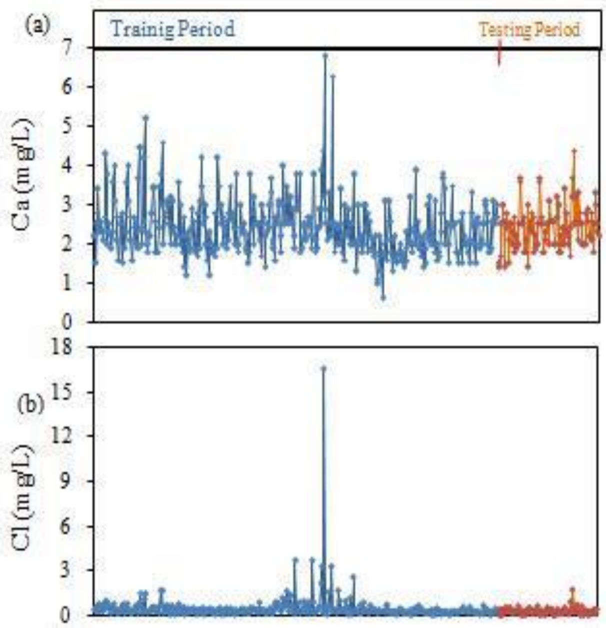
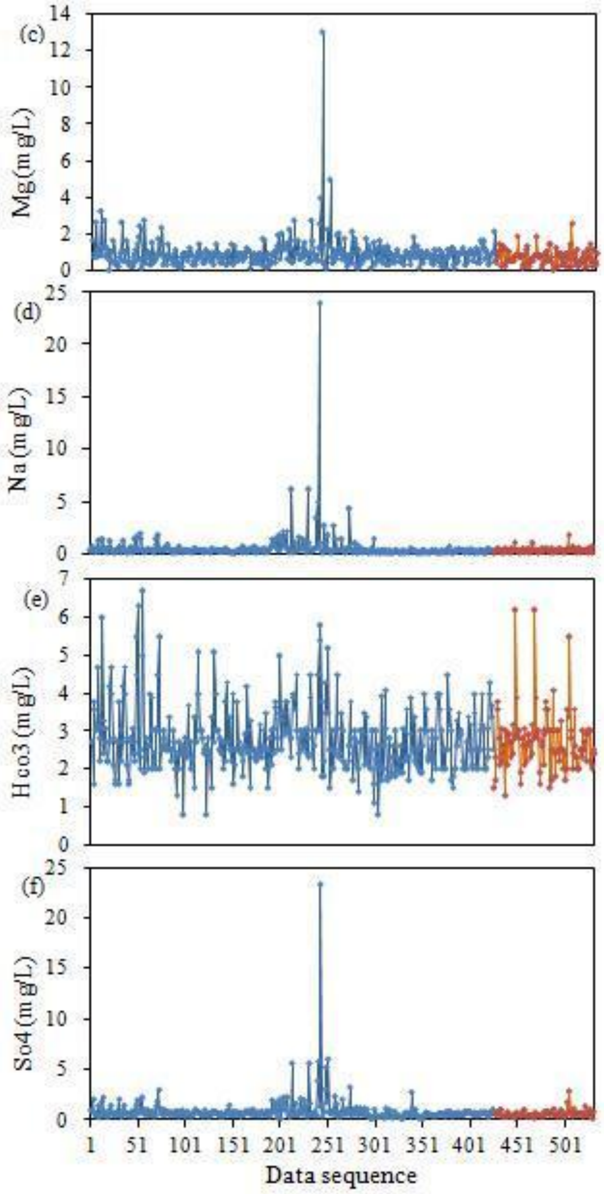


Figure 3. (a & b).



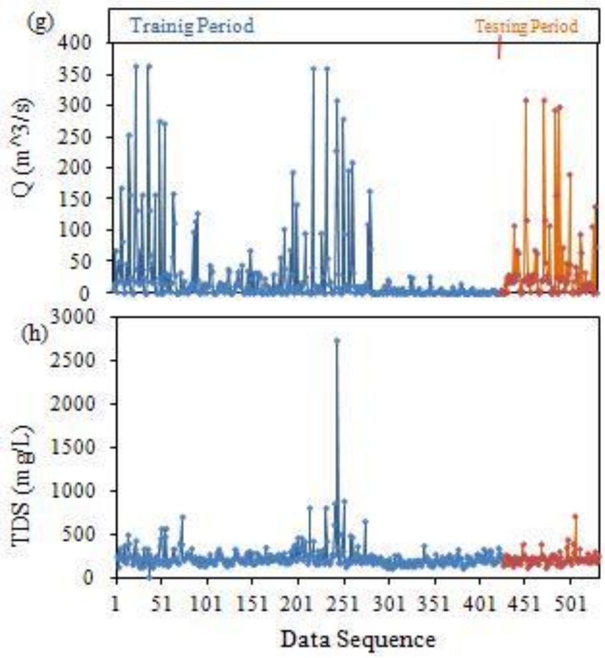


Figure 3. Measured time series of the water quality parameters at the Simineh River: a) Calcium, b) Chloride, c) Magnesium, d) Sodium, e) Bicarbonate, f) Sulfate, g) Water discharge, h) TDS

Table 1. Shows the statistical values of the used water quality data in this study.

Table 1. Statistics for water quality parameters of Simineh River at the Dashband station, period 1993-2011.

Data		Unit		Mean	Min.	Max.	Std. Dev.	Skew	Kurtosis
Input	Ca		Total	2.43	0.60	6.80	0.69	1.45	4.93
		mg/L	Training	2.43	0.60	6.80	0.72	1.51	5.04
			Testing	2.43	1.40	4.40	0.56	0.71	0.87
	Cl		Total	0.46	0.05	16.50	0.80	15.83	308.06
		mg/L	Training	0.48	0.05	16.50	0.89	14.48	254.33
			Testing	0.36	0.10	1.80	0.22	3.17	18.20
	Mg		Total	0.94	0.10	13.0	0.75	8.67	128.42
		mg/L	Training	0.98	0.10	13.0	0.81	8.55	117.88
			Testing	0.78	0.10	2.60	0.42	1.20	2.82
	Na		Total	0.50	0.10	24.0	1.18	15.83	304.52
		mg/L	Training	0.54	0.10	24.0	1.31	14.31	247.29
			Testing	0.35	0.10	1.80	0.21	4.07	23.35
	HCO_3		Total	0.76	0.80	6.70	0.84	1.32	3.02
		mg/L	Training	2.78	0.80	6.70	0.85	1.20	2.45
			Testing	2.72	1.30	6.20	0.80	1.89	6.38
	SO_4		Total	0.76	0.08	23.40	1.19	13.91	252.16
		mg/L	Training	0.80	0.08	23.40	1.31	12.76	209.25

			Testing	0.61	0.10	2.94	0.35	3.42	19.23
	Q		Total	26.41	0.0	364.31	57.20	3.78	15.51
		m^3/s	Training	23.29	0.0	364.31	55.17	4.09	18.36
			Testing	38.94	0.14	308.71	63.46	3.02	9.53
Output	TDS		Total	225.70	0.0	2730	145.63	10.51	167.16
		mg/L	Training	229.32	0.0	2730	158.43	10.08	148.15
			Testing	211.19	110.50	695.50	73.70	3.29	17.74

RESULTS AND DISCUSSION

The ANNs model architecture refers to the layout of neurons and the number of hidden layers; in this study, a typical ANNs model (Figure 1) with a back-propagation algorithm is constructed to predict TDS values. The back-propagation training algorithm is a supervised training mechanism and is normally adopted in most of the engineering application. The primary goal is to minimize the error at the output layer by searching for a set of connection strengths that cause the ANNs to produce outputs that are equal to or closer to the targets. Neurons in the input layer have no transfer function.

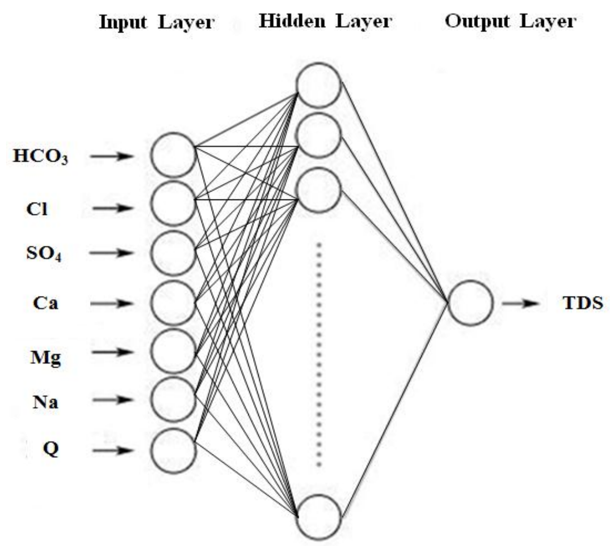


Figure 4. Implementation of the ANN Model.

Logarithmic linear transfer function was used in the hidden layer and linear transfer function was employed from the hidden layer to the output layer as an activation function, because the linear function is known to be robust for a continuous output variable.

The network was trained in 1000 epochs using the Levenberg–Marquardt learning algorithm with a learning rate of 0.001 and a momentum coefficient of 0.9. Preliminary model runs were tested with Ca, Cl, Mg, Na, HCO_3, SO_4 and Q; these led to the identification of the number of the hidden layer neurons (Figure 4).

In this study, in order to adopt the most appropriate network geometry, trial-and-error procedure was used and different numbers of hidden neurons were investigated for finding the optimum. A three-layer network was selected, and the study tested the model with 2I+1 as recommended by Lippmann (1987).

where I is the number of input variables. The number of hidden layers is one for all runs.

Table 2 presents the effect of changing the number of the hidden layer neurons on the RMSE, MAE, and R^2 statistics in training and testing periods.

The results indicated that the network geometry with fourteen hidden neurons is required for a relatively better performance of RMSE, MAE, and R^2 (30.119 mg/L, 16.986 mg/L, and 0.841, respectively).

In ANNs modeling, the contributions of the individual input variables to the output variable are normally unknown. This study employs the Garson Eq. (4) to assess the relative importance of these input variables, which uses the network weights produced by the ANNs model.

Table 2. The results of ANNs model for the training and testing periods to the identification of the number of the hidden layer neurons.

Hidden Layer Neurons	Training			Testing		
	RMSE (mg/L)	MAE (mg/L)	R^2	RMSE (mg/L)	MAE (mg/L)	R^2
1	35.970	22.455	0.950	32.149	18.803	0.840
2	40.060	19.138	0.950	29.515	17.451	0.840
3	33.840	21.523	0.955	31.616	18.652	0.831
4	29.495	17.458	0.965	29.847	16.580	0.837
5	30.532	17.779	0.963	31.108	18.043	0.829
6	30.046	17.777	0.964	31.343	18.441	0.833
7	30.636	17.623	0.963	31.587	17.481	0.828
8	30.498	18.649	0.963	31.355	17.760	0.837
9	33.948	21.733	0.958	34.160	19.309	0.809
10	32.438	20.337	0.958	33.455	19.905	0.805
11	30.032	18.023	0.964	33.310	19.667	0.799
12	34.992	20.243	0.955	32.786	19.096	0.800
13	26.828	17.536	0.971	30.744	17.488	0.826
14	29.853	17.868	0.964	30.119	16.986	0.841
15	29.515	16.642	0.965	31.726	17.577	0.816

Figure 5 shows the scatter plots of the results obtained from the optimum ANNs model for testing dataset. The model gave close approximations of the actual observations, suggesting that these approaches are applicable for modeling the TDS dataset.

Figure 6 presents relative importance of the input variables on TDS, and indicates that magnesium and calcium concentrations are the most and least influential parameters, with approximate values of 17.56 and 11.78 %, respectively.

Table 3 gives the values of matrices of weights between input and hidden layers and weights between hidden and output layers.

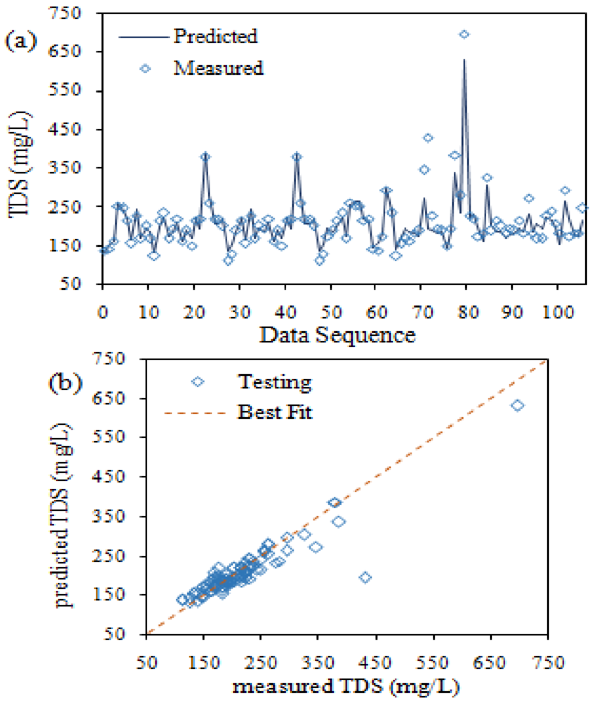


Figure 5. Comparison of predicted ANNs time series with observed values for testing dataset: (a) Sequence plot, (b) Scatter plot.

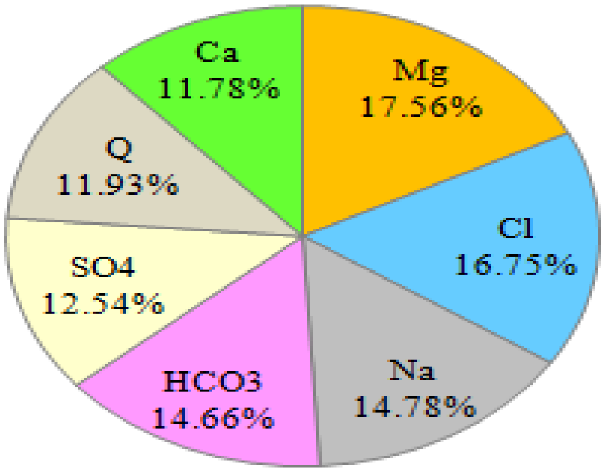


Figure 6. Relative importance of input variables on TDS

Table 3. Matrices of weights- w1 weights between input and hidden layers, w2 weights between hidden and output layers.

W1								W2	
Neuron	Variable							Neuron	Variable
	Ca (mg/L)	Cl (mg/L)	Mg (mg/L)	Na (mg/L)	HCO_3 (mg/L)	SO_4 (mg/L)	Q (m^3/s)		TDS (mg/L)
1	-0.0359	-0.7709	1.0689	2.2827	0.1683	1.1615	-0.1600	1	0.8078
2	-0.7143	-1.0290	-0.3720	-1.1589	1.0623	-1.2420	0.3313	2	-0.5754
3	0.2375	0.4220	-0.4355	1.1874	0.9715	-2.4002	0.4365	3	-0.4464
4	-0.0974	1.7889	-0.4964	0.4307	0.2117	-0.7557	-0.6488	4	0.8517
5	3.4288	-0.7376	-1.1994	0.8234	-2.8677	2.0686	-1.0979	5	-0.1644
6	0.5613	-1.3707	1.4443	2.7392	-3.5131	-1.4783	-1.6958	6	-0.0952
7	-1.4813	0.0735	-1.1773	-0.2775	1.6875	-0.1507	-0.3742	7	-0.7445
8	-0.5992	2.2376	2.0662	-0.1764	0.3731	0.6566	-1.8344	8	0.1144
9	-1.0443	0.6192	-0.5616	1.2724	-0.7222	0.6219	-0.9965	9	-0.2183
10	-0.14063	0.9416	1.1090	0.4562	0.3200	-0.0154	0.0335	10	0.1979
11	0.3532	1.5300	0.6438	-0.4494	-0.2224	1.0119	1.4426	11	-0.1320
12	0.7681	0.5201	-1.1019	0.0737	-1.0853	-0.0376	-1.0260	12	0.2854
13	-2.3626	0.0514	1.9098	-2.6700	0.5379	0.7493	-1.5977	13	-0.1089
14	0.3468	-1.2038	-1.3103	0.3519	0.8189	0.0153	-0.0908	14	-0.1238

CONCLUSION

A study of Total Dissolved Solid (TDS) time series is reported in this paper using local water quality parameters of Calcium (Ca), Chloride (Cl), Magnesium (Mg), Sodium (Na), Bicarbonate (HCO_3), Sulfate (SO_4), and water discharge (Q) for a set of recorded data in Simineh River at Dashband gauging station during 1993-2011. The paper also employs the Garson equation to assess the relative importance of the variables to determine appropriate input combinations. The general objective of this study is to investigate the performance of Artificial Neural Networks (ANNs) for the estimation of the TDS amounts without assuming or applying significant knowledge of the physics of the process.

The ANNs with different numbers of neurons in the hidden layer were constructed, and model performance has been estimated by means of several indicators, including data sequence, scatter diagrams, and quantitative measures of *RMSE*, *MAE,* and R^2. The modeling results indicated that reasonable prediction accuracy was achieved for the ANNs model.

REFERENCES

1. Abbasi Maedeh P, Mehrdadi N, Nabi Bidhendi GR and Zare Abyaneh H. (2013). Application of artificial neural Network to predict total Dissolved Solids variations in groundwater of Tehran Plain, Iran. International Journal of Environment and Sustainability, 2(1):10-20.
2. Abudu S, King JP, Sheng Z. (2011). Comparison of the performance of statistical models in forecasting monthly Total Dissolved Solids in the Rio Grande. Journal of the American Water Resources Association, 48(1):10-23.
3. Anctil, F, Filion M, Tournebize J. (2009). A neural network experiment on the simulation of daily nitrate-nitrogen and suspended sediment fluxes from a small agricultural catchment. Ecological Modelling, 220:879–887.
4. Asadollahfard G, Taklify A, Ghanbari A. (2012). Application of artificial neural network to predict TDS in Talkheh Rud River. Journal of Irrigation and Drainage Engineering, 138(4):363–370.
5. Chang FJ, Kao LS, Kuo YM, Liu CW. (2010). Artificial neural networks for estimating regional arsenic concentrations in a black foot disease area in Taiwan. Journal of Hydrology, 388:65–76.
6. Diamantopoulou MJ, Papamichail DM, Antonopoulos VZ. (2005). The use of a Neural Network technique for the prediction of water quality parameters. Operation Research, 5(1):115-125.
7. Garson GD. (1991). Interpreting neural-network connection weights. Artificial. Intelligent Expert, 6:47-51.
8. Gazzaz NM, Yusoff MK, Aris AZ, Juahir H, Ramli MF. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Marine Pollution Bulletin, 64(11): 2409-2420.
9. Haykin S. (1999). Neural networks: A comprehensive foundation. New York: Macmillan Publishing.
10. He B, Oki T, Sun F, Komori D, Kanae S, Wang Y, Kim H, Yamazaki D. (2011). Estimating monthly total nitrogen concentration in streams by using artificial neural network. Journal of Environmental Management, 92:172–177.
11. Heydari M, Olyaie E, Mohebzadeh H, Kisi Ö. (2013). Development of a neural network technique for prediction of water quality parameters in the Delaware River, Pennsylvania. Middle East Journal of Scientific Research, 13(10):1367-1376.
12. Kanani S, Asadollahfardi G, Ghanbari A. (2008). Application of artificial neural network to predict Total Dissolved Solid in Achechay River Basin. World Applied Sciences Journal, 4 (5):646-654.
13. Kuo JT, Wang YY, Lung WS. (2006). A hybrid neural–genetic algorithm for reservoir water quality management. Water Research, 40(7):1367–1376.
14. Lippmann RP. (1987). An introduction to computing with neural nets. ASSP Magazine IEEE., 4(2):4-22.
15. Maier HR, Dandy GC. (1996). The use of artificial neural networks for the prediction of water quality parameters. Water Resource Research, 32(4):1013-1022.
16. McCulloch WS, Pitts W. (1943). A logical calculus of the ideas immanent in neurons activity. Bulletin of Mathematical Biophysics, 5:115-133.
17. Mehrdadi N, Hasanlou H, Jafarzadeh MT, Hasanlou H, Abdolabadi H. (2012). Simulation of low TDS and biological units of Fajr industrial waste water treatment plant using artificial neural network and principal component analysis hybrid method. Journal of Water Resource and Protection, 4(6):370-376.
18. Memon NA, Unar MA, Mastorakis NE, Khaskheli GB. (2009). Total Dissolved Solids (TDS) modeling by artificial neural networks in the distribution system of drinking water of Hyderabad city. Proceedings of the 13th WSEAS International Conference on Computers, 607-611.
19. Moasheri SA, Khammar GA, Poornoori Z, Beyranvand Z, Soleimani M. (2013). Estimate the spatial distribution TDS the fusion method Geostatistics and artificial neural networks. International Journal of Agriculture and Crop Sciences, 6:410-420.
20. Najad A, El-Shafie A, Kari, OA, El-Shafie AH. (2013). Application of artificial neural networks for water quality prediction. Neural Computing & Application, 22(1):187-201.
21. Palani S, Liong SY, Tkalich P. (2008). An ANN application for water quality forecasting. Marine Pollution Bulletin, 56(9):1586-1597.
22. Singh KP, Basant A, Malik A, Jain, G. (2009). Artificial neural network modeling of the river water quality – a case study. Ecological Modelling, 220(6):888-895.
23. Weckman GR, Millie DF, Ganduri C, Rangwala M, Young W, Rinder M, Fahnenstiel GL. (2009). Knowledge extraction from the neural 'Black Box' in ecological monitoring. Journal of Industrial and Systems Engineering, 3(1):38-55.