Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2014.995713

57 (2016) 4888–4897
March

Taylor & Francis
Taylor & Francis Group

Application of artificial neural networks to estimating DO and salinity in San Joaquin River basin

E. Salami Shahid, M. Ehteshami*

Environmental Engineering Department, KN Toosi University of Technology, P.O. Box 1587-544-16, Tehran, Iran, Tel. +98 21 88770006; email: Esisalami@Yahoo.com (E. Salami Shahid), Tel./Fax: +98 21 88779476; email: Maehtesh@gmail.com (M. Ehteshami)

Received 10 March 2014; Accepted 1 December 2014

ABSTRACT

In the current research, which is based on recorded and collected data from one of the San Joaquin River data recording stations, an artificial neural network (ANN) model was developed to simulate water quality parameters. Then, the results were compared with traditional salinity formula to optimize parameters. We also chose the best parameters for estimation of dissolved oxygen (DO) through an ANN model. In both models, we used feed-forward perceptron training algorithm along with Levenberg–Marquardt as the learning algorithm and tansign(x) as transfer function algorithm. To simulate the salinity, we used more than 5,000 water quality data-sets. Also, we used two groups of data-sets, with 16,000 related data for simulating DO. We developed a highly precise model and verified the results with the most recommended formula. Mean squared error is 12.5 for the presented model and 9061 for the traditional formula (Salinity = 0.64*EC). We also recommend a formula whose result is very close to the pilot-recorded data. It showed a large disagreement between traditional formula and the proposed model. We used MATLAB software to optimize the design parameters of the model.

Keywords: Water quality; ANN model; Simulation; San Joaquin River; Electrical conductivity

1. Introduction

Water scarcity in the world is turning into a more serious challenge every year. During the last century, water consumption grew at twice the rate of population increase [1], and the complexity of managing natural resources generally increases as human population grows [2]. Therefore, the assessment of the properties and processes of running water is a major issue in aquatic environmental modeling [3]. Accurate determination of the concentration of nutrients and other substances in water bodies is an essential requirement for supporting effective management and legislation [4]. Water management decisions are increasingly based on model studies, [5] while modeling tools are becoming progressively more sophisticated [6]. Making models for the study of water quality parameters also has the benefit that the value of data modeling can be demonstrated as overall saving in maintenance or development costs. Viewed broadly across an organization's entire budget, this can be truly significant. The value of data modeling can also be seen at a more detailed level by the saving it will provide for development tasks on a specific project. Additionally, its value

*Corresponding author.

can be determined by identifying specific benefits that data modeling provides and then quantifying those benefits per project. Finally, data models can be reapplied in whole or in part to multiple projects which can result in significant saving for any organization [7]. Models can also generate the missing data [8].

The dissolved oxygen (DO) is an important quality index of water resources. However, it is difficult to simulate the DO concentration by traditional mathematical methods due to the effects of different factors on different waters [9]. Salinity is also a significant parameter. Many researchers, including McNeil and Cox [10], Granlund et al. [11], and CWT 2004 [12], have tried to estimate salinity using EC measurement data. They have confirmed that the median ratio of salinity/conductivity alters for different salinity ranges and different temperatures. In the current research as is presented, we recommend the following definition:

$$\text{Salinity (ppm)} = 0.5 \times \text{EC}(\mu\text{s/cm}) \quad (1)$$

Moreover, the above equation is only true for the current case (the water from San Joaquin) and for other saline waters, the relationship should be reedited. We have also compared the recommended equation with most recommended formula which is:

$$\text{Salinity (ppm)} = 0.64 \times \text{EC}(\mu\text{s/cm})[13] \quad (2)$$

Using artificial neural network (ANN) as a model for estimating water salinity parameters is increasing due to its exceptional capabilities for solving and modeling complicated problems. ANN focuses on completion of all requested parameters. Its use is also increasing in optimization problems [14]. Many studies have been accomplished in modeling and data analysis with ANN. For example: Patki et al. [15] worked more on finding the optimum number of neurons and hidden layers by finding the optimum status when changing these numbers. They included such parameters as pH, alkalinity, hardness, TS, and MPN as the input variables, which can be used to forecast water quality index in various zones of municipal distribution system. Rounds [16] have worked on the DO concentration in Tualatin River (in northwest Oregon, Oswego Dam). At first, he explained the importance of research on DO before showing an ANN model using a feed-forward algorithm. The information about air temperature, solar radiation, rainfall, and stream flow are considered as input, while DO concentration is the output of the model. The whole data were collected between 1991 and 2001. Cordoba [17] carried out a research on a four-year-old database with a set of inputs collected in the city of Našiměřice, Czech Republic. He used a feed-forward network to make the model and Multi-Layer Perceptron. Temperature, pH, flow, pipe material, diameter, and age of pipes were considered as input of the model while free chlorine was the output obtained using 1965–2002 data-set.

Using ANN for groundwater calculations was carried out for the first time by Aziz and Wong [18]. Paulin [19] calibrated three types of ANN (PNN, GRBF, RNN) using groundwater and hydrometeorology data to simulate the fluctuation of groundwater in Gondo. Chitsazan et al. [20] presented an ANN model to predict the groundwater depth according to such values as rain condition, mean monthly temperature, relative humidity, discharge of irrigation canal, and groundwater recharge from the plain boundary. Using MATLAB, they changed some network parameters and found the optimum groundwater network. Rak [21] presented an ANN model to predict turbidity of treated water in a newly-operating water treatment system for surface and retention water at Sosnówka reservoir, Poland. The input data are psychochemical parameters of interim water. The work demonstrates parameters effective on the NTU.

Pandan et al. [22] reviewed a model to predict the water level of a river branch analyzed in this study, the Kushabhadra, which originates at Balianta gauging site, flows downstream for a distance of about 90 km and finally drains into the Bay of Bengal. The presented model is a feed-forward network with a learning function of Levenberg–Marquardt (LM) and back propagation training algorithm. Simulated water levels by MIKE 11 HD were compared with the corresponding water levels predicted by the ANN model. The results obtained from the ANN model were found to be much better matched than MIKE 11 HD results as indicated by the values of the goodness of fit indices used in the study.

Chu et al. [23] represented an ANN model that could estimate the quality of the surface water parameters using some given parameters. The results showed that the factor analysis technique was introduced to identify important water quality parameters. Results revealed that biochemical oxygen demand, permanganate index, ammonia nitrogen, nitrogen, Cu, Zn, and Pb were the most important parameters in assessing water quality variations in the study area. This project was based on GB3838-2002 "Environmental quality standard for surface water." The model is a one-layer network using the algorithm of Hopfield Neural Network and created by the MATLAB. Ghazi Zade and Noori [24] presented an ANN model with a feed-forward network in order to predict generated solid waste in Mashhad. The utilized data is based on

weekly sampling of solid waste in Mashhad from 2004 to 2007. They showed that their model is performing better than the traditional methods. Nejadkoorki and Baroutian [25] presented an ANN model using a three-layer feed-forward back propagation neural network that was trained with (LM) training algorithm. It can estimate maximum PM10 concentration (that is smaller than 10 μm) in Tehran 24 h in advance. They used data collected between 2001 and 2009. Diamantopoulou et al. [8] presented a model using the data from one of Axios river quality monitoring stations as input data for training a feed-forward neural network. The data was collected from 1980 to 1994. In the presented model, such parameters as temperature, flow, EC, HCO_3, SO_4, Na, Cl, Ca, and DO were input data and nitrate was the model's output. Steyl [26] reviewed the application of ANN algorithms in geohydrology. The function of ANN model (standard neural network) trained by LM algorithm to predict fluctuation of groundwater depth was examined in Maheshwaram basin in India's Heidar Abad by Sreckanth et al. [27]. The model efficiency and accuracy were measured based on the root mean square error and regression coefficient (R). They implied that the ANN appears to be a promising tool for precise and accurate groundwater level forecasting. Nadiri [28] had dealt with evaluation of ANN's (FFN-LM) ability in modeling complex aquifer of Tabriz. The main purpose of this article also is to use ANNs, especially feed-forward back propagation neural networks to simulate and predict groundwater level. Zealand et al. [29] compared the performance of a stochastic–deterministic model with ANN for short-term river flow forecasting and concluded that the ANN model performed better than the stochastic-deterministic model during the testing phase. A similar study by Demirel et al. [30] revealed that the ANN model is better than the physically based model in predicting peak flow values. In the current paper, we trained two separate models that are used to predict the Salinity and DO. The models are based on the Grant Line Canal near Clifton Court Fore bay. The data was collected continuously in 2010 and was also verified. The simulation tests and results are presented here. Despite large and increasing number of studies that have been carried out in the field of modeling the properties of waters, different behavior of various waters necessitate making specific model(s) for particular water(s).

2. Methods and materials

Multi-layer feed-forward networks trained with back propagation algorithm are among the most popular kinds of networks [14,31]. Neuron is the smallest data processor unit that is the base of neural networks operations. You can see a single input neuron in Fig. 1. The input and the output signals are scalars (vectors). The effect of P on A is shown by the w scalar (matrix) [32]. The Product of this summarization is n, which will be the pure input for transfer (activation) function (F) [20,26] and, therefore, the output of neuron will be calculated as:

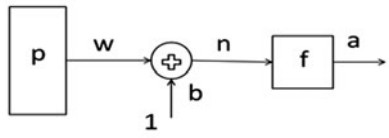

Fig. 1. A simple neuron model.

$$a = f(w_p + b) \tag{3}$$

The parameters b and w are adjustable and the activation function can be also chosen by designer of the network. Training means that b and w will change many times in a direction to get closer to a desired relation between input and output. The activation function (f) can be linear or non-linear. Function (f) would be chosen according to defined problem. A few sample functions are shown in Fig. 2.

Fig. 3 shows a neuron that has a number of R inputs. All the elements of p vector multiply in the related element of w matrix to form the bios (b) [33,34].

The input (n) calculated as:

$$n = \sum_{i=1}^{R} p_i \cdot w_{1,i} + b = [w] \cdot p + b \tag{4}$$

$$p = [p_1, p_2, \ldots, p_R]^T, \quad w = [w_{1,1}, w_{1,2}, \ldots, w_{1,R}] \tag{5}$$

Then the formula for the output will be like:

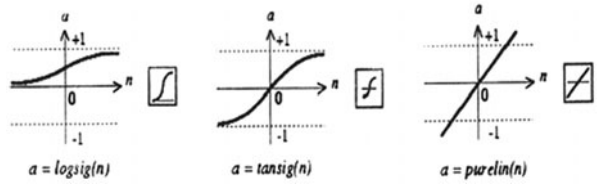

Fig. 2. Transfer functions.

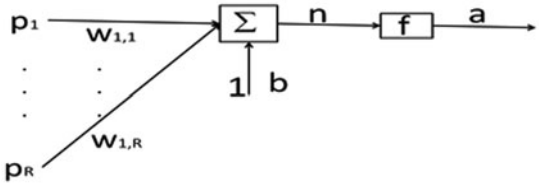

Fig. 3. A neuron with R inputs.

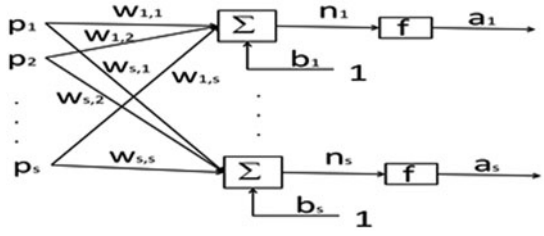

Fig. 4. S neuron with R inputs.

$$a = f(w \cdot p + b) \tag{6}$$

A single-layer network with S neurons and R inputs is shown in Fig. 4.

3. Learning rules

We define learning rule as a process for correcting or improving weights and biases. We have two kinds of learning rules (functions): supervised and unsupervised rules. In the supervised mode, e.g. Perceptron, we compare the network output with learning examples (which is related to the input data). Unsupervised learning method is being used mainly for division problems.

4. Feed-forward networks

Architecture of network: The basic architecture contains three types of layers: input layer, hidden layer, and output layer. The input layer is responsible for introducing the input data and hidden layer(s) is the place where they are processed. The output layer produces the results [16, 35–39]. In the Fig. 5, a five-layer feed-forward network with three hidden layers is demonstrated. Each layer can contain different numbers of neurons.

In feed-forward networks, data stream signal always goes straight forward from the input side to the output site. This process can take place in many units (i.e. layers) [20]. We don't have a return data

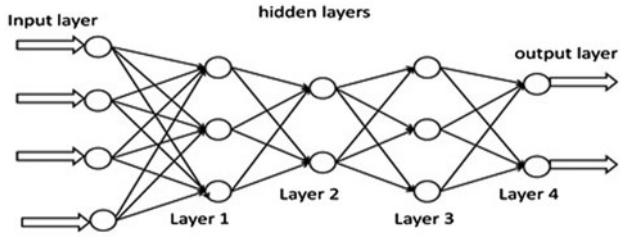

Fig. 5. A network with five layers.

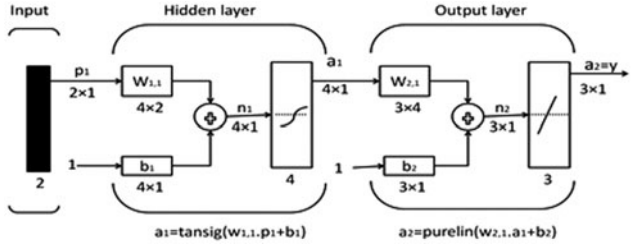

Fig. 6. A three-layer feed-forward network.

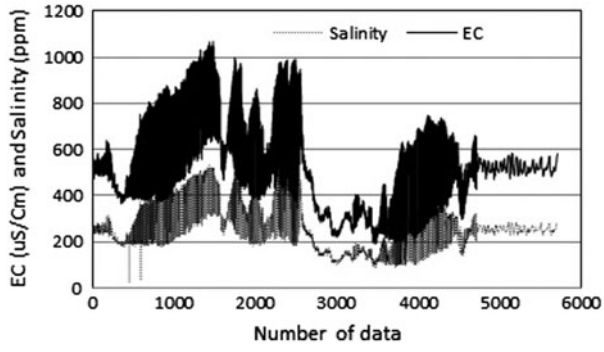

Fig. 7. Simulated EC (μs/cm) and salinity (ppm).

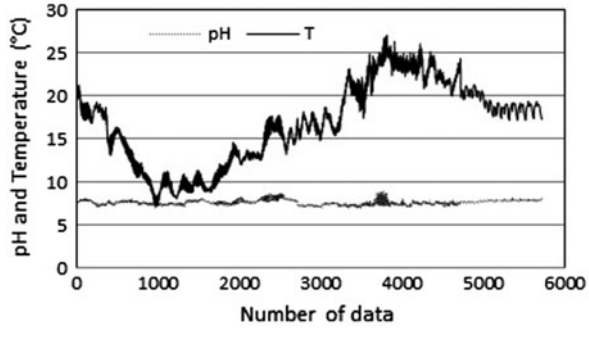

Fig. 8. Simulated pH and temperature (C).

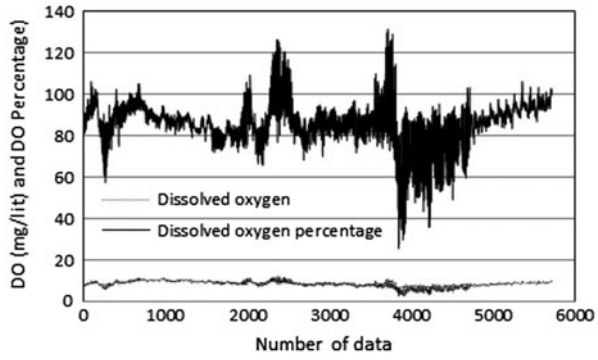

Fig. 9. Simulated DO (mg/l) and DO percentage.

stream here. When being trained by constant stream of data, the network changes weights and biases in each step and compares the output layer with answers. The supervised mode is the next step in training weights and biases to minimize the error.

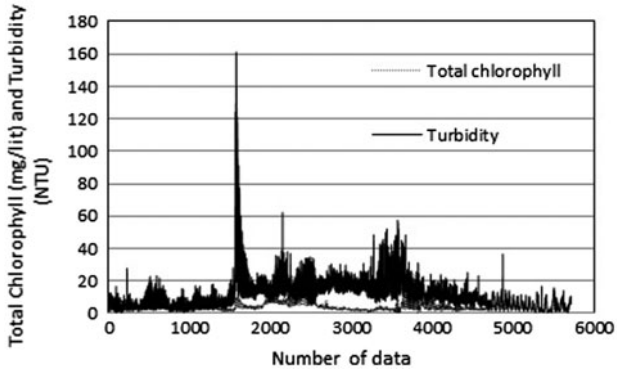

Fig. 10. Simulated total chlorophyll (mg/l) and turbidity (NTU).

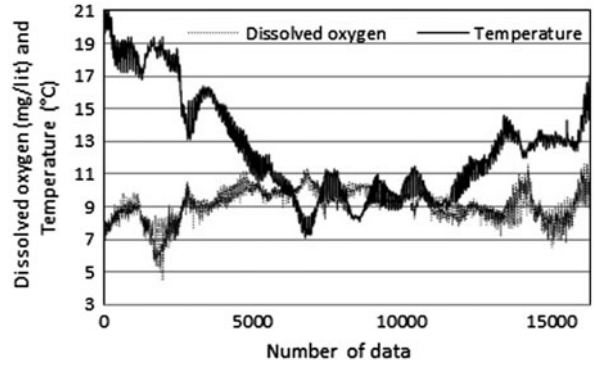

Fig. 11. Simulated DO (mg/l) and temperature (C).

You see a three-layer (tansign/pureline) network in Fig. 6. This network can be used to estimate any function with any number of rupture points (Figs. 7–11).

5. Calculation method

To start the calculation, we can use a two-layer model. Then, we calibrate weight and bias matrixes (by guessing) as shown below (according to: [31,33]) to reduce the error and get closer to the required precision. Here, α is the learning rate ($\alpha > 0$) and l is the number of each step:

$$w_{i,j}^{(l+1)} = w_{i,j}^{(l)} - \alpha \frac{\partial e(w,b)}{\partial w_{i,j}^{(l)}} \tag{7}$$

$$b_{i,j}^{(l+1)} = b_{i,j}^{(l)} - \alpha \frac{\partial e(w,b)}{\partial b_{i,j}^{(l)}} \tag{8}$$

$$b_{i,j}^{(l+1)} = b_{i,j}^{(l)} - \alpha \frac{\partial e(w,b)}{\partial b_{i,j}^{(l)}} \tag{9}$$

$$\frac{\partial e(w,b)}{\partial w_{i,j}^{(l)}} = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial w_{i,j}^{(l)}} e(w,b;x^{(i)},y^{(i)})\right] + \alpha w_{i,j}^{(l)} \tag{10}$$

$$\frac{\partial e(w,b)}{\partial b_{i,j}^{(l)}} = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b_{i,j}^{(l)}} e(w,b;x^{(i)},y^{(i)})\right] + \alpha b_{i,j}^{(l)} \tag{11}$$

The scale of precision (in supervised mode) will be also the quantity of mean square error (MSE), t_i is the answer (matrix) and a_i (matrix) is the output of the network:

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m} e^2 = \frac{1}{m}\sum_{i=1} (t_i - a_i)^2 \tag{12}$$

MATLAB is used for calculations [20,23]. Tables 1 and 2 show the design parameters of the network. The network contains three layers of neurons. The number of the layers and the neurons must be chosen in a way that would get us closer to the target. By increasing the number of input data, we can reach a higher precision, (false data decreases the precision of the network more than the increase caused by true data). However, it is not possible to get a higher precision by increasing the number of layers or neurons in each layer.

Data come from resources provided by California Department of Water Resources, the State of California.

Table 1
Some of the best networks that we tested

Number	Target	NL[a]	NNL1[b]	NNL2	NNL3	ALF[c]	PF[d]	TF[e]	Gradient	Performance
1	Salinity	4	15	8	8	LearrnGDM[f]	MSE[g]	tansig	0.239	0.209
2	Salinity	4	15	15	15	LearrnGDM	MSE	tansig	1.81	0.2
3	Salinity	4	15	8	8	LearrnGD	MSE	tansig	0.903	0.211
4	Salinity	4	15	8	8	LearnGDM	SSE[h]	tansig	2,400	1,700
5	Salinity	4	15	8	8	LearnGDM	MSE	logsig	1,450	299
6	Oxygen	4	15	8	8	LearnGDM	MSE	tansig	0.0188	0.436
7	Oxygen	4	15	15	15	LearnGDM	MSE	tansig	0.0784	0.441
8	Oxygen	3	10	10		LearnGDM	MSE	tansig	0.629	0.433

[a]Number of layers.
[b]Number of neurons in layer1.
[c]Adapting learning function.
[d]Performance function.
[e]Transfer function.
[f]Gradient descent with momentum weight and bias learning function.
[g]Mean squared normalized error performance function.
[h]Sum squared error performance function.

Table 2
Design parameters of optimal networks

Network type	Feed—forward back propagation	A	0.001
Training function	Trainlm[a]	α-dec	0.1
Adaption learning function	Learngdm	α-Inc.	10
Performance function	MSE	α-max	10^{10}
Transfer function	Tansign	epoches	1,000
Minimum gradient of error	1.00E-10	Goal	0

[a]Levenberg–Marquardt back propagation.

They use the data to manage and predict water specifications and levels [40]. The data is also used for water resources management as well as environmental and industrial controls. This project was based on San Joaquin River data related to 2009 and 2010, collected in Grant Line Canal near Clifton Court Forebay station. The data, which were registered every 15 min, cover the following parameters: 1- the time and date of each test; 2- EC; 3- DO; 4- DO percentage; 5- Chlorophyll concentration; 6- Temperature; 7- pH; 9- Turbidity; and 10- Salinity.

The parameters were chosen based on their relation with total salinity to achieve a higher degree of precision. Therefore, we developed an ANN model to calculate the salinity as the output when the other parameters were assumed as the input. To generate the model, all data related to 2010 according to B9529500 document of the site [40] were downloaded and prepared for modeling. More than 35,000 groups of data-sets were recorded for each year and we needed at least 1,000 data groups to reach the required precision. We decided to use more than 5,000 groups of data to achieve a higher precision. Therefore, we chose almost one parameter from seven data groups. In the following charts you can see 5,000 groups of 2010 data that we used to develop the ANN model. We found that we could not achieve a precision lower than 2 mg/l by using 5,000 groups of data. As a result, we used 16,300 data groups. In order to develop a model for predicting the DO, only temperature data were assumed as input of the model and DO was the output parameter. During the current study, the number of input data was tested to achieve higher algorithm efficiency (Table 1). Figs. 7 through 11 show the data that we used to develop the current model.

6. Results and discussion

In fact, in order to get the best results, the number of layers and neurons must be chosen properly according to existing complications problems and the

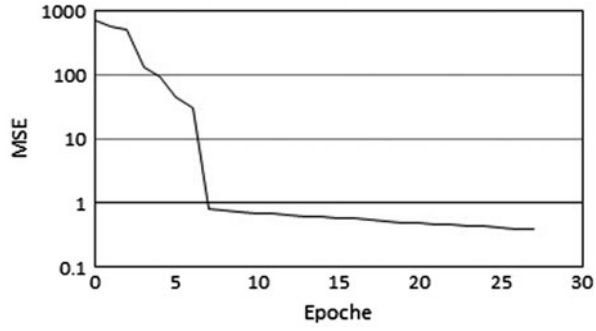

Fig. 12. Performance plot.

number of input and output parameters [34]. Our experience in this field shows that problem complications are related to the number of layers and the number of problem parameters related to the number of neurons. One of the best ways to optimize the model is Monte Carlo method. We found out the most proper number for each design parameter. So we tested hundreds of times with different numbers for each parameter. We saw that with 3 or more than 5 layers, the results were not acceptable. However, with 4 (and sometimes 3) layers, we were able to have a low number of errors. Some of the best tested networks are shown in Table 1.

In salinity model, the other seven parameters are considered input and, in an optimal state, the number of layers is four while the number of neurons in each layer will be: 15, 8, 8. As mentioned before, the number of the last layer's neurons is chosen by the program.

In the case of designing the DO predictor model, we found out that we could reach a higher precision (an error lower than 10 ppm) by choosing just the temperature as the input data. Therefore, for the proposed model, DO will be the output and the temperature will be the input. This model contains four layers and the number of neurons in each layer is: 15, 10, and 10. Design parameters of optimal networks are shown in Table 2.

Then, we could find the best models according to performance charts (represented by MATLAB, see Fig. 12). We saved the models and tested them by randomly chosen data from year 2009 that have been given in Table 3. Fig. 12 shows a performance plot generated by MATLAB, which indicates MSE value changes in the training process.

We have shown the results in Table 4. According to the current model, additional analyses were performed on model results and the results of traditional formulas. Many researchers have shown that

Table 3
Selected data for network testing

Date and time	EC (μs/cm)	DO (mg/l)	DO (%)	pH	Clph (mg/l)	Tu (NTU)	T (°C)	Salinity (ppm)
10/1/2008 15:59	653	7.69	89.9	7.63	3.4	12.3	23	320
10/20/2008 10:00	483	9.32	96.3	7.52	3.6	7	16.9	230
11/3/2008 8:59	515	8.21	84.7	7.63	2.4	4.5	16.8	250
11/12/2008 0:29	743	11.51	113.1	8.09	11.7	17.1	14.5	370
12/21/2008 1:00	844	11.12	92	7.91	4.5	5.6	7.1	420
12/24/2008 12:15	923	11.07	92.3	7.94	3.5	6.1	7.4	460
1/1/2009 16:30	969	10.43	90.1	7.82	3.7	7.2	8.8	480
1/16/2009 5:30	1,061	11.76	102.2	7.84	10.2	9.5	9.1	530
2/11/2009 23:16	739	11.07	99.7	7.78	4	12.2	10.6	360
2/17/2009 12:01	654	11.03	97.4	7.8	4.3	6.1	9.8	320
3/13/2009 15:30	863	11.01	108.2	8.04	10	21.5	14.5	430
3/24/2009 16:01	432	10.12	99.9	7.92	7.8	9.5	14.7	210
4/2/2009 15:16	364	9.63	97.6	7.95	7.1	15.5	16	180
4/10/2009 4:46	401	8.98	90.6	7.74	9.9	14.1	15.7	190
5/8/2009 10:00	424	8.67	95.2	7.61	3.1	17.8	19.9	200
5/18/2009 13:45	371	6.25	73.7	7.23	2.2	13.7	23.6	180
6/10/2009 1:45	464	7.4	82	7.55	8.5	30.3	20.3	220
6/14/2009 1:30	343	8.4	93.9	7.66	5.4	14.2	20.8	160
7/12/2009 15:00	631	5.25	61.7	7.33	7.9	34.3	23.3	310
7/19/2009 3:45	226	7.37	89	7.4	3.4	12.8	24.8	110
8/12/2009 16:45	698	6.33	77.4	7.46	10.7	27.5	25.4	340
8/14/2009 12:45	607	5.88	69.6	7.4	2.6	9.7	23.7	290
9/6/2009 18:14	473	7.89	92.6	7.73	2.6	5.9	23.3	230
9/25/2009 5:00	541	7.47	88.4	7.77	5.2	10.9	23.7	260

Table 4
Error analysis and comparison between networks, formula and pilot results

Date &Time	PSD[a]	EC (μs/cm)	0.5.EC	F1E[b]	0.64.EC	F2E[c]	CD[d]	CDE[e]	POD[f]	COD[g]
10/1/2008 15:59	320	653	327	7	418	98	317	−3	7.69	6.953
10/20/2008 10:00	230	483	242	12	309	79	232	2	9.32	9.461
11/3/2008 8:59	250	515	258	8	330	80	249	−1	8.21	8.665
11/12/2008 0:29	370	743	372	2	476	106	366	−4	11.51	9.725
12/21/2008 1:00	420	844	422	2	540	120	416	−4	11.12	10.97
12/24/2008 12:15	460	923	462	2	591	131	457	−3	11.07	10.93
1/1/2009 16:30	480	969	485	5	620	140	481	1	10.43	9.987
1/16/2009 5:30	530	1,061	531	1	679	149	523	−7	11.76	9
2/11/2009 23:16	360	739	370	10	473	113	363	3	11.07	9.64
2/17/2009 12:01	320	654	327	7	419	99	321	1	11.03	9.633
3/13/2009 15:30	430	863	432	2	552	122	426	−4	11.01	9.725
3/24/2009 16:01	210	432	216	6	276	66	209	−1	10.12	9.541
4/2/2009 15:16	180	364	182	2	233	53	174	−6	9.63	9.296
4/10/2009 4:46	190	401	201	11	257	67	193	3	8.98	9.284
5/8/2009 10:00	200	424	212	12	271	71	204	4	8.67	7.759
5/18/2009 13:45	180	371	186	6	237	57	176	−4	6.25	6.953
6/10/2009 1:45	220	464	232	12	297	77	223	3	7.4	7.911
6/14/2009 1:30	160	343	172	12	220	60	164	4	8.4	7.995
7/12/2009 15:00	310	631	316	6	404	94	307	−3	5.25	6.953
7/19/2009 3:45	110	226	113	3	145	35	107	−3	7.37	6.953
8/12/2009 16:45	340	698	349	9	447	107	336	−4	6.33	6.953
8/14/2009 12:45	290	607	304	14	388	98	295	5	5.88	6.953
9/6/2009 18:14	230	473	237	7	303	73	228	−2	7.89	6.953
9/25/2009 5:00	260	541	271	11	346	86	261	1	7.47	6.953

[a]Pilot salinity data.
[b]Formula 1 (Salinity = 0.5 × EC) errors.
[c]Formula 2 (Salinity = 0.64 × EC) errors.
[d]Calculated data.
[e]Calculated data errors.
[f]Pilot oxygen data.
[g]Calculated oxygen data.

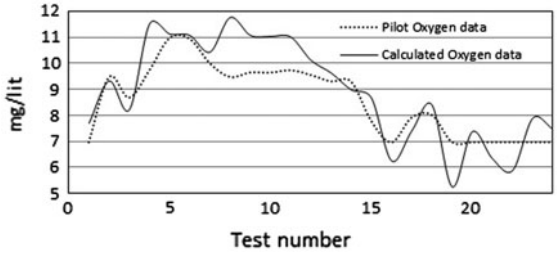

Fig. 13. Comparison between model, pilot test, and oxygen (mg/l) data.

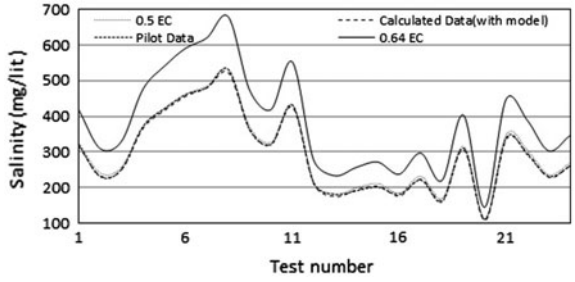

Fig. 14. Comparison between model, pilot test, and formula salinity (ppm) data.

EC/salinity relationship depends on the place and the water characteristics and have found various values for EC/salinity. McNeil and Cox [10] showed that EC/salinity could be somewhere between lower than 0.5 to higher than 1. But the most common formula in the literature is: [13]

$$\text{Salinity (ppm)} = 0.64 \times \text{EC}(\mu\text{s/cm}) \quad (13)$$

Tested data were then verified by the model, which showed that they were very different from the used data. Therefore, we optimized the value of EC/salinity relation to 0.5:

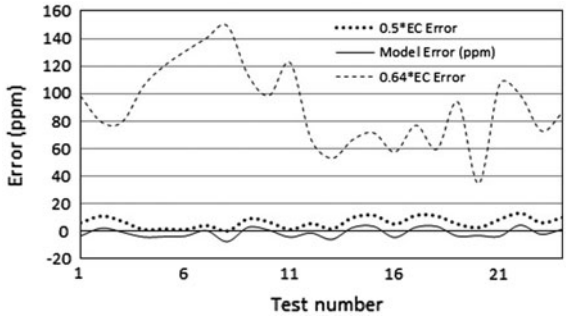

Fig. 15. Error analysis performance of the model.

$$\text{Salinity (ppm)} = 0.5 \times \text{EC}(\mu\text{s/cm}) \quad (14)$$

We reached the following optimum relations according to as Eq. (14), which are shown in Table 4. Table 4 shows the results of this formula as well. In Table 4, errors for each approach are summed for comparison.

Table 4 shows the precision of two models, especially for salinity, you can see that the result of the models is much closer to the pilot tests data than the result of traditional formulas. For a better understanding of this issue, you can see Figs. 13 and 14.

As Figs. 13 and 14 show, data were generated by models which were then verified by pilot data. For salinity problem, the quantity of MSE for Eq. (14) is considered to be 60.7. This value (MSE) is set to 9,061 for Eq. (13), while for the model, MSE stands at 12.5. The results from Fig. 14 show how the ANN model can work better than a common formula. The precision is higher because the model covers all the effective data-sets. Fig. 15 shows higher precision of the model.

It is evident that the error of the model is very small when compared to the equations. In fact, some of the existing errors result from low precision of the testing equipment and human errors. In other words, the model has a lower possibility of fault than the pilot data when more than 5,000 data points are used. The current model is not just a reliable model; it can be used to enhance the precision of the measurement equipment and their operators. Also, it can generate the missing data. It should be noted that the reliability of these two models for other water body depends on the properties of those particular water.

7. Conclusions

It is known that the ANN model can be used in many practical and scientific projects. In this work, we focused on the ANN model and its ability to simulate river water quality data. The results show that applications of the ANN model are useful in the following cases: (1) A reliable replacement for salinity test; (2) To control equipment and operators; (3) A proper tool for estimating the missing data; (4) Calibration of measurement tools; (5) The ability to predict quality data; (6) The ability to do sensitivity analyzes on the data generated by the model for scientific applications; (7) Suitable for conditions with experimental difficulties; (8) A highly precise replacement for traditional equations; (9) We also found a predicting equation (Salinity = 0.5*EC) which can estimate the salinity according to EC.

Several projects can be suggested for future works: (1) To present model(s) for estimating other parameters like DO; (2) To present model(s) for other stations or places and comparing the results with our results or combining two models to create a more powerful model; (3) Conducting sensitivity analyses based on the data generated by our model; (4) Conducting a comprehensive comparison between our results and some results obtained through other similar models; (5) Assessing model applicability and efficiency for a specific site in order to develop a useful global model for water quality analyses; (6) To formulate universal equations such as EC, TDS formulas. The current method of ANN modeling can be applied easily to other cases and data-sets.

Acknowledgment

The authors are grateful to Dr H. Moradi and Dr Sohrab Soori for their editorial and revision assistance. They are also thankful to San Joaquin River Monitoring Stations Control Board for providing data for the current analyses.

References

[1] Stockholm International Water Institute and Elsevier, The Water and Food Nexus: Trends and Development of the Research Landscape, 2012.
[2] L.M. Varnell, D.A. Evans, D.M. Bilkovic, J.E. Olney, Estuarine surface water allocation: A case study on the interactive role of science in support of management, Environ. Sci. Policy 11 (2008) 602–612.
[3] I.M. Schleiter, D. Borchardt, R. Wagner, T. Dapper, K. Schmidt, H. Schmidt, H. Werner, Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks, Ecol. Modell. 120 (1999) 271–286.
[4] I. Donohue, K. Irvine, Quantifying variability within water samples: The need for adequate subsampling, Water Res. 42 (2008) 476–482.
[5] H. Scholten, A. Kassahun, J.C. Refsgaard, T. Kargas, C. Gavardinas, A.J.M. Beulens, A methodology to support multidisciplinary model-based water management, Environ. Modell. Softw. 22 (2007) 743–759.

[6] U.S. McKnight, S.G. Funder, J.J. Rasmussen, M. Finkel, P.J. Binning, P.L. Bjerg, An integrated model for assessing the risk of TCE groundwater contamination to human receptors and surface water ecosystems, Ecol. Eng. 36 (2010) 1126–1137.
[7] T. Haughey, The Return on Investment (ROI) of Data Modeling, CA Erwin, March, 2010 1–18.
[8] M.J. Diamantopoulou, V.Z. Antonopoulos, D.M. Papamichail, The use of a neural network technique for the prediction of water quality parameters of Axios River in Northern Greece, Eur. Water 11(12) (2005) 55–62.
[9] C. Lihua, M. Shengquan, L. Li, A model to evaluate DO of river based on artificial neural network and style book, J. Hainan Normal Univ. (Nat. Sci.) 21(4) (2008) 372–376.
[10] V.H. McNeil, M.E. Cox, Relationship between conductivity and analysed composition in a large set of natural surface-water samples, Queensland, Australia, Environ. Geol. 39(12) (2000) 1325–1333.
[11] N. Granlund, A. Lundberg, D. Gustafsson, Laboratory study of salinity influence on the relationship between electrical conductivity and wetness of snow, in: 65th Eastern Snow Conference, Fairlee (Lake Morey), VT, USA, 2008, pp. 301–308.
[12] The Clean Water Team (CWT), Guidance Compendium for Watershed Monitoring and Assessment State Water Resources Control Board FS-3.1.3.0., (EC) V2e (2004) 1–5.
[13] R.S. Ayers, D.W. Westcot, Water Quality for Agriculture, FAO Irrigation and Drainage Paper 29, 1994.
[14] T. Koncsos, The application of neural networks for solving complex optimization problems in modeling, Conference of Junior Researchers in Civil Engineering (2010) 97–102.
[15] V.K. Patki, S. Shrihari, B. Manu, Water quality prediction in distribution system using cascade feed forward neural network, Int. J. Adv. Technol. Civ. Eng. 2(1) (2010) 84–91 (2231–5721).
[16] S.A. Rounds, Development of a neural network model for dissolved oxygen in the Tualatin river, Oregon, in: Second Federal Interagency hydrologic modeling conference, Las Vegas, NV 1 (2002) 1–13.
[17] I.G.A.C. Cordoba, Using of artificial neural network for evaluation and prediction of some drinking water quality parameters within a water distribution system, Water Management and Water Structures, Juniorstav, 2011, pp. 1–11.
[18] R.A. Aziz, K.F.V. Wong, A neural-network approach to the determination of aquifer parameters, Ground Water 30(2) (1992) 164–166.
[19] C. Paulin, Artificial neural network modeling of water table depth fluctuations, Water Resour. Res. 374 (2001) 885–896.
[20] M. Chitsazan, G. Rahmani, A. Neyamadpour, Groundwater level simulation using artificial neural network: A case study from Aghili plain, urban area of Gotvand, south-west Iran, J. Geope. 3 (2013) 35–46.
[21] A. Rak, Water turbidity modelling during water treatment processes using artificial neural networks, Int. J. Water Sci. 2 (2013) 1–10.
[22] R.K. Pandan, N. Pramanik, B. Bala, Simulation of river stage using artificial neural network and MIKE 11 hydrodynamic model, Comput. Geosci. 36 (2010) 735–745.
[23] H.B. Chu, W.X. Lu, L. Zhang, Application of artificial neural network in environmental water quality assessment, J. Agr. Sci. Technol. 15 (2013) 343–356.
[24] J. Ghazi Zade, R. Noori, Prediction of municipal solid waste generation by use of artificial neural network: A case study of Mashhad, Environ. Res. 2 (2008) 13–22.
[25] F. Nejadkoorki, S. Baroutian, Forecasting extreme PM10 concentrations using artificial neural networks, J. Environ. Res. 6 (2010) 277–284.
[26] G. Steyl, Application of Artificial Neural Networks in the Field Of Geohydrology, University of the Free State, Bloemfontein, 2009.
[27] P.D. Sreckanth, N. Geethanjali, P.D. Sreedevi, S. Ahmed, N. Ravi Kumar, P.D.K. Jayanthi, Forecasting groundwater level using artificial neural networks, J. Curr. Sci. 96(7) (2009) 933–939.
[28] A. Nadiri, Predicting Groundwater Level Surrounding Tabriz City, thesis, Tabriz University, Iran, 2011.
[29] C.M. Zealand, D.H. Burn, S.P. Simonovic, Short term streamflow forecasting using artificial neural networks, J. Hydrol. 214 (1999) 32–48.
[30] M.C. Demirel, A. Venancio, E. Kahya, Flow forecast by SWAT model and ANN in Pracana basin, Portugal, Adv. Eng. Softw. 40 (2009) 467–473.
[31] D. Svozil, V. Kvasnicka, J. Pospichal, Introduction to multi-layer feed-forward neural networks, Chemom. Intell. Lab. Syst. 39 (1997) 43–62.
[32] A.G. Carpenter, Neural network models for pattern recognition and associative memory, Neural Networks 2 (1989) 243–257.
[33] M.B. Menhaj, Fundamental of Neural Network, vol. 1. Industrial Amir Kabir University, Tehran, 2008.
[34] A. Abraham, Artificial Neural Networks. Oklahoma State University, Stillwater, OK, 2005, pp. 901–908.
[35] L. Fausett, Fundamentals of Neural Networks Architectures, Algorithms and Applications. Prentice Hall, Englewood Cliffs, NJ, 1994.
[36] S. Haykin, Neural networks: A comprehensive foundation, Macmillan, New York, NY, 1994.
[37] U.F. Dowla, L. Rogers, Solving Problems in Environmental Engineering and Geosciences with Artificial Neural Networks, MIT Press, Cambridge, MA, 1995.
[38] K. Gurney, An Introduction to Neural Network, UCL Press, London, 1999.
[39] D. Patterson, Artificial Neural Networks, Prentice Hall, Singapore, 1996.
[40] www.water.ca.gov\WaterDataLibraryContinuousTimeSeriesData.htm