Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2016.1167624

57 (2016) 27073–27084
December

Taylor & Francis
Taylor & Francis Group

Application of artificial neural networks and mathematical modeling for the prediction of water quality variables (case study: southwest of Iran)

E.S. Salami[a], M. Salari[a], M. Ehteshami[b,*], N.T. Bidokhti[c], H. Ghadimi[d]

[a]*Department of Environment Engineering, Shiraz University, Shiraz, Iran, emails: Esisalami@Yahoo.com (E.S. Salami), Salari.marjan@gmail.com (M. Salari)*
[b]*Department of Environment Engineering, KN Toosi University of Technology, P.O. Box 1587-544-16, Tehran, Iran, Tel. +98 21 88770006; Fax: +98 21 88779476; email: Maehtesh@gmail.com*
[c]*Department of Civil and Environment Engineering, Shiraz University, Shiraz, Iran, email: Nassertaleb@gmail.com*
[d]*Department of Civil and Environment Engineering, Sharif University of Technology, Tehran, Iran, email: h_ghadimi@mehr.sharif.ir*

Received 26 June 2015; Accepted 15 March 2016

ABSTRACT

River water quality monitoring using traditional water sampling and laboratory analyses is expensive and time-consuming. The application of artificial neural network (ANN) models to simulate water quality parameters is cost-effective, quick, and reliable. This study provides two methods of mathematical and ANN modeling to simulate and forecast five important river water quality indicators (DO, TDS, SAR, BOD_5, HCO_3) correlated with variables such as EC, temperature, and pH which can be measured easily and almost with no cost. The mathematical method is based on polynomial fitting with least square method and the neural network model was developed using a feed-forward algorithm. The 35 years' data were collected from 7 monitoring stations and 5 rivers located in southwest of Iran. DO concentration was simulated using an equation and a neural model. Two equations were calibrated for estimating TDS while SAR is simulated using a mathematical and neural model. Another two neural models were developed for BOD_5 and HCO_3 simulations. An acceptable precision was achieved, as shown in model verification results. Presented models and equations are reliable/useable tools for studying in similar locations (rivers), as a proper replacement for traditional water quality measuring practices.

Keywords: Water quality; Artificial neural networks; Modeling; Monitoring; Polynomial

1. Introduction

Rivers are an important source of water for mankind and other living creatures [1–3]. Assessment of river water quality properties and processes is a major issue in the management of aquatic environments [4]. Models are widely used to solve water quality management problems. Water management decisions are increasingly dependent on modeling studies [5], while modeling tools are becoming progressively more sophisticated [6]. Modeling for simulating river water quality properties has many advantages such as low/no simulation cost, short time required, decreased needs for laboratory and measuring equipment/operators, possibility of synthetic data generation in large quantities for data analyses, regenerating missed data, control and calibrating measurement equipments and

*Corresponding author.

many others, etc. [7]. Models can simulate the most vital quality parameters for users like human and the other living creatures, by an ingenious use of easily measured input parameters. That can help to assess and predict more complicated parameters which are not easily obtainable [8]. Temperature (T) can be measured quickly by a thermometer with no cost as well as the pH [9]. Electrical Conductivity (EC) itself is not a human or aquatic health concern, but it can be measured easily as a useful indicator to assess other water quality problems. Therefore, in this study, values of T, pH, and EC will be the inputs of our models/equations. This study aims to model essential water quality parameters of rivers whose measurement is difficult, costly, and time-consuming.

This study aims to develop equations/models to predict concentrations of DO, BOD, TDS, HCO_3, and SAR index by the use of data obtained from the specified stations. Many previous studies tried to analyze these types of data to assess (river) water quality or developing/calibrating models for simulating properties of (river) water, some of which are summarized in Table 1.

This study aims to develop mathematical equations and ANN models for simulating values of DO, TDS, BOD, SAR, and HCO_3 using some (or all) of the parameters of T, EC, and pH as simple and directly measurable input parameters. The main highlights of the present study are as follows: (1) analyzing several rivers with completely different conditions and water qualities; (2) analyzing a wide range of parameters; (3) presenting simple equations and ANN models to simulate important quality parameters; (4) including collected water quality data from more than 30 years of observation (from 1981 to 2011); (5) parameters that are chosen as inputs to equations/models are appropriate and easy to measure. Furthermore, the model outputs are essential parameters whose direct experimental measurement has inherent difficulties which make it prohibitively costly and time-consuming. In general, the presented equations/models are extremely useful and practical.

1.1. Dissolved oxygen

Dissolved oxygen (DO) is one of the most important parameters in water quality assessment [10]. It is difficult to simulate due to its influential parameters, however, it would be possible by calibrating to local conditions [11]. DO also can be measured using a modern dissolved oxygen meter [12], whereas lack of its availability makes a reliable mathematical or evolutionary model as a valuable asset to generate data at any number and conditions that we need for data analyses.

1.2. Biological oxygen demand

Biological oxygen demand (BOD_5) is a measure of the oxygen in the water that is required by the aerobic organisms. Measurement of BOD needs especial devices and is time-consuming [13,14]. Therefore, it's not an easy parameter for measurement in water quality analyses [15]. To develop a model to assess BOD would be very useful.

1.3. Total dissolved solid

Total Dissolved Solid (TDS) is a measurement of inorganic salts, organic matter, and other dissolved materials in water. TDS measurement can be performed by filtering the sample through a 2.0 μm pore size filter, evaporating the remaining filtrate and then drying the residuals to a constant weight at 180°C [16]. Many studies have shown the linear relationship between EC and TDS [17–19]. In the present study, this relationship will be investigated for the study area.

1.4. Sodium adsorption ratio

Sodium adsorption ratio (SAR) evaluates the danger of high-sodium content in water for soil structure and it is the major water quality index (WQI) for irrigation purposes [20–22]:

$$\mathrm{SAR} = \frac{\mathrm{Na}^{+}}{\sqrt{\frac{(\mathrm{Ca}^{2+} + \mathrm{Mg}^{2+})}{2}}} \tag{1}$$

where Na^+, Mg^{2+}, and Ca^{2+} are ionic concentrations expressed in meq/L. In order to calculate SAR, these concentrations should be measured first. This study presents an equation for estimating SAR and also an ANN model with higher precision.

1.5. HCO_3

HCO_3 is one of the most important components sources of carbon transport in rivers. Bicarbonate (HCO_3^-) can be measured directly by acid titration using methyl orange as an indicator [23], however, we find it useful to train an ANN model for simulating HCO_3 values in rivers.

Table 1
Summary of some modeling studies

Author(s)	Year	Study parameters	Result(s)
[4]	1999	pH, EC, TDS, Turbidity (Tu), COD, BOD, DO, Alkalinity, Ammonia-N, Cl, S, and Total hardness (TH)	Collected samples from a river and measured parameters then presented Pearson correlation coefficients between each pair of variables, and also showed low quality of water caused by industrial activities
[24](*)	2014	Salinity, EC, pH, T, DO, DO%, Total chlorophyll, and turbidity	A mathematical/experimental equation corresponding to EC presented for simulating salinity. An ANN model trained to simulate salinity with the use of seven other parameters as inputs, an ANN model also presented predicts DO corresponding to temperature changes
[10]	2015	pH, EC, T, TDS, NTU, BOD, DO, Nitrate, Cl, F, TH, and iron	Found correlation between parameters, at first. Then used (weighted) parameters to calculate the water quality index (WQI)
[20]	2011	SAR	Directly calculated SAR to show groundwater quality in the study area(s)
[21](*)	2013	SAR, Cl, Na^+, Cl^-, SO_4^-, HCO_3^- and TDS	Developed a model based on CoKriging optimization of monitoring network combined with optimized ANN by genetic algorithm to forecast Cl and SAR corresponding to salinity parameter
[25]	2005	COD, BOD, NO_2^-, NO_3^-, NH_4^+ and PO_3	Reviewed three river water qualities by checking each of them for the level of contaminations
[23]	2011	Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, Cl^- and SO_4^{2-}	Standard methods were used for the measurement of parameters of collected samples from surface water of a reservoir and deduce that quality of the water is acceptable for uses such as drinking water and irrigation
[15]	2013	BOD, UV light absorbance, and dissolved organic carbon (DOC)	Used multiple regression analysis to predict BOD with sensors that measured UV absorption and TOC. Both river water and waste water have been tested in their system and correlation factors were equal to 0.78 and 0.9, respectively
[26]	2013	SAR, Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, Cl^-, pH, EC, T, TDS, Tu, BOD, DO, Nitrate, F, TH and SiO_2^{2-}	At first, presented correlation coefficients between parameters then checked water quality for drinking and irrigation and showed that some of the quality parameters are adequate for drinking and some are not. The calculated SAR values denoted an acceptable quality for irrigation purposes
[27]	2013	Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, Cl^-, H_4SiO_4, SO_4^{2-}, NO_2^- and PO_4^{3-}	Investigated parameters responsible for variations in river water quality represent the correlation factor for parameters and hence that water is less contaminated anthropogenically than by natural weathering. By the way, the water had sufficient quality for drinking and/or for other purposes
[28]	2012	TS, DS, SS, pH, COD, BOD, DO, Cl and TH	Found correlation between parameters, and developed multiple linear regression analysis for mapping and area estimation of water quality parameters

(*Continued*)

Table 1 (*Continued*)

Author(s)	Year	Study parameters	Result(s)
[29]	2012	pH, EC, T, TDS, Tu, BOD, DO, NH_3, Mg, Cl, F, TH, Fe, Zn, As, Total coliform bacteria, *E. coli* bacteria, SS and NO_3-N	Used parameters to train a NN for simulating WQI and recommended ANN models for analysis of long-term environmental monitoring records
[30](*)	2010	River water level (stage)	Compared the precision of an ANN model and MIKE11 model for predicting the river water level and showed that the ANN model has a better precision
[31]	2009	Stream flow, precipitation, flux and soil moisture index	Applied NNs to simulate daily nitrate—nitrogen and suspended sediment fluxes
[32]	2013	EC, pH, TH, TDS, SAR	Developed a Kriging—ANN (compound model) to simulate SAR using four easily measured parameters

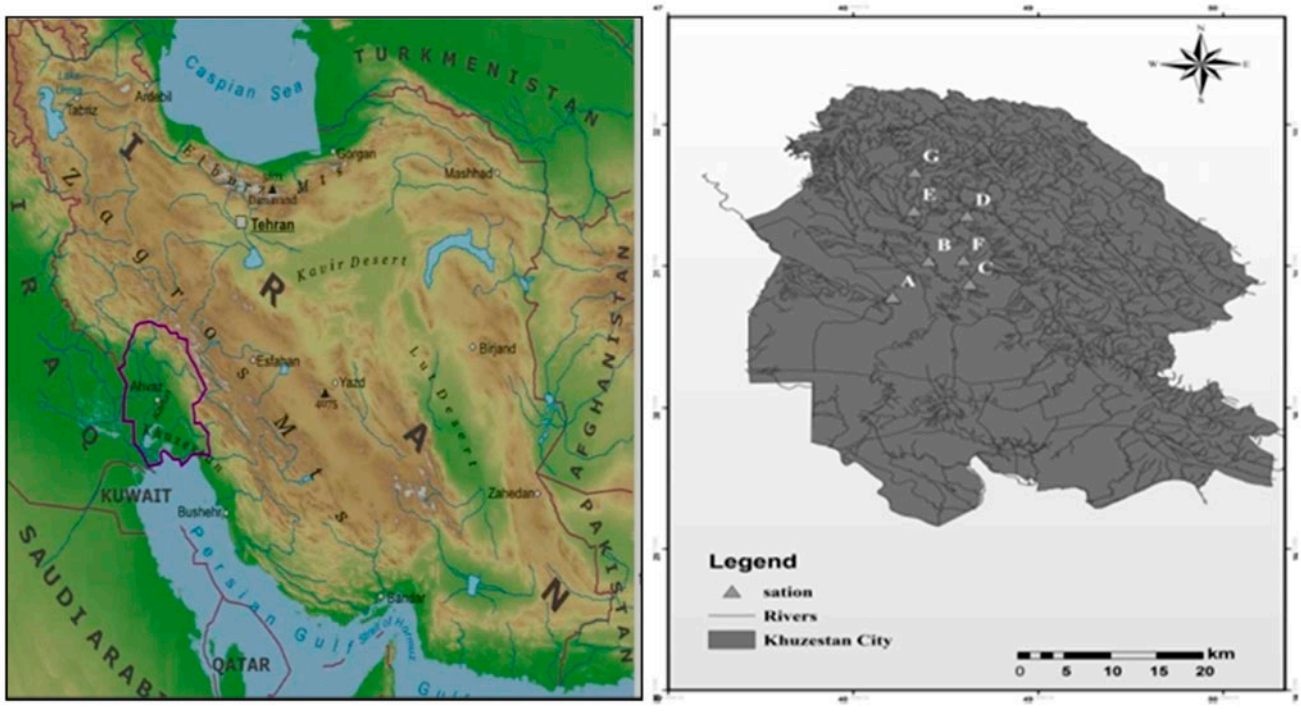

Fig. 1. Location of monitoring stations (www.ngdir.org).

1.6. Study area

Khuzestan Province is in the southwest of Iran with an area of about 64,236 square kilometers, between 47˚41′50′′ and 39˚58′33′′ degrees E. Its importance in terms of water resources is having the highest share (33%) of fresh water in the country with five major rivers of (Karun, Karkheh, Dez, Hendijan, and Jarahi). Data from five different rivers in the southwest of Iran are used in this study. Fig. 1 and Table 2 show the details of each monitoring station's data used in this study.

2. Method and materials

Table 3 shows the variation, range, mean, minimum, and maximum values for each parameter. In order to provide a sufficient data-set with an adequate number of predictors for ANN model development, a higher

Table 2
Properties of monitoring stations

Station number	River name	Station name	Longitude	Latitude	Elevation (m)	Years
S#1	Karoon	Molla-Saani	48-52-00	31-35-00	23	1976–2011
S#2	Gorgor	Band-Ghir	48-53-00	31-39-00	20	1986–1995
S#3	Dez	Baamdej	48-41-00	31-41-00	25	2004–2010
S#4	Gorgor	Vali-Aabaad	48-53-00	31-39-00	22	2004–2008
S#5	Dez	Harmaleh	48-33-00	31-57-00	38	1976–2011
S#6	Shatiyat	Arab-Asad	48-53-00	31-51-00	27	1981–2011
S#7	Shoor	Dasht-Bozorg	48-55-00	32-06-31	63	1986–2011

Table 3
Properties of variables that are used in the study

Parameter	Min.	Max.	Mean	Stations used for modeling	Number of existed data
T (°C)	3	33	20.6	–	–
EC (us/cm)	341	50,240	4,800	–	–
pH	4.45	10.2	7.9	–	–
DO (mg/L)	2	11.8	7.4	S#1, S#3, S#4	238
TDS (mg/L)	280	31,250	3,065	S#1, S#2, S#3, S#4, S#5, S#6, S#7	1,328
SAR	0.55	98.56	10.78	S#1, S#2, S#5, S#6, S#7	1,202
BOD (mg/L)	0.28	6.52	2.65	S#1, S#3, S#4	238
HCO_3 (mg/L)	0.5	5.86	2.69	S#1, S#2, S#3, S#5	1,257

precision is achieved by applying an interpolation method for increasing the number of data for some input parameters (that are marked in Table 6 by "*").

3. Mathematical method

In this study, the least squares method is used to solve the overconstrained linear system to obtain the coefficients of the fitting polynomial [33]. The method is based on linear algebra. Let y be the nth degree polynomial of x:

$$y_i = f(x_i, n) = \sum_{j=0}^{n} a_j \cdot x_i^n \tag{2}$$

where a_j = coefficient of polynomial, y_i = state (dependent variable) variable, and x_i = independent variable.

Eq. (2) can be also expressed in the form of a matrix:

$$\overbrace{\begin{bmatrix} a_0 & a_1 & \dots & a_n \end{bmatrix}}^{A_{n+1}} \cdot \overbrace{\begin{bmatrix} x_1^0 = 1 & 1 & \dots & 1 \\ x_1^1 & x_2^2 & \dots & x_i^2 \\ . & . & & . \\ . & . & \dots & . \\ . & . & & . \\ x_1^n & x_2^n & \dots & x_i^n \end{bmatrix}}^{X_{n+1,i}} = \overbrace{\begin{bmatrix} y_1 & y_2 & \dots & y_i \end{bmatrix}}^{Y_i} \tag{3}$$

Or:

$$A_{n+1} \cdot X_{n+1,i} = Y_i \text{ or } A \cdot X = Y \tag{4}$$

For every matrix $X \in R^{n\times m}$, a unique matrix $X^+ \in R^{m\times n}$, which is called generalized (right) inverse, exists satisfying [34]:

$$X \cdot X^+ = I_n \tag{5}$$

where I_n is Identity matrix of size n. By multiplying X^+ on (right hand of) both sides of Eq. (4) we have:

$$A \cdot X \cdot X^+ = A \cdot I = A = Y \cdot X^+ \tag{6}$$

And by this method "A" which is polynomial, coefficients matrix can be calculated. The analyses of the present study are performed using MATLAB software.

4. Neural network method

The artificial neural network (ANN), as its name implies, is a technique for simulation of the human brain functions during the problem-solving process, which has been developed and originated about 60 years ago. The neural network approach can be

applied to powerful computation of complex nonlinear relationships, just as humans apply knowledge gained from past experience to new problems or situations [34,35]. Thus, for modeling parameters that don't have a simple (linear) relationship with input data, ANN method can be employed effectively. The multilayer feed-forward networks trained with backpropagation algorithm are the most popular type of networks [36,37]. For example, models that are marked by (*) in Table 1 have used feed-forward networks for their development. It must mention that other possible ANN network types (that existed in NN tool of MATLAB) tested too, and with reviewing the related literatures "feed-forward back-propagation" was selected as optimum network.

5. Structure of the networks

The basic architecture consists of three types of neuron layers: input, hidden, and output layers. Fig. 2 shows a two-layer network. In feed-forward ANN networks, the signal flows from input to output units, strictly in a feed-forward direction [38]. Hidden layers consist of different number of neurons. Fig. 2 shows a parameter such as "a" is the output of neuron and "p" is the input. Parameters w and p are weight and bias, respectively. All parameters are denoted as matrices, and can be expressed as:

$$a = f(\text{net}) = f(n) = f(w^T \cdot p + b) = f\left(\sum_{i=1}^{R} w_R^T \cdot p_R + b\right) \tag{7}$$

$$p = [p_1, p_2, ..., p_R],\ w = [w_1, w_2, ..., w_R] \tag{8}$$

where a, b, w, and p are output, bias, weight and input matrices (for each layer), respectively. The most common "f" functions are presented in Fig. 3. These transfer functions transfer output of each layer to a simpler more useful expression for calibrating the w_i and b_i (s) in next layer/step.

Tan-sigmoid transfers inputs between $[-\infty, +\infty]$ into real numbers between $[-1, 1]$. In this study, all ANN models have tansig transfer function.

$$\tan\text{sig}(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}} \tag{9}$$

Three transfer functions that are shown in Fig. 3 are available/useable for "feed-forward back propagation" NN type with MATLAB. The training process determines the ANN weights and is similar to the calibration of a mathematical model. In order to perform training correctly, we must iteratively continue and repeat the process of calibrating and optimizing the w_i, b_i (s), with the final target of minimizing the mean square error (MSE) value as possible as it is, the process will continue until the required precision is reached. In the following procedure, weights and biases will change every time the process is repeated. The calibration process for w_i, b_i (s) is as [39,40]:

$$w_{i,j}^{(l+1)} = w_{i,j}^{(l)} - \alpha \frac{\partial e(w,b)}{\partial w_{i,j}^{(l)}} \tag{10}$$

$$b_{i,j}^{(l+1)} = b_{i,j}^{(l)} - \alpha \frac{\partial e(w,b)}{\partial b_{i,j}^{(l)}} \tag{11}$$

$$\frac{\partial e(w,b)}{\partial w_{i,j}^{(l)}} = \left[\frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial w_{i,j}^{(l)}} e(w,b;x^{(i)},y^{(i)})\right] + \alpha w_{i,j}^{(l)} \tag{12}$$

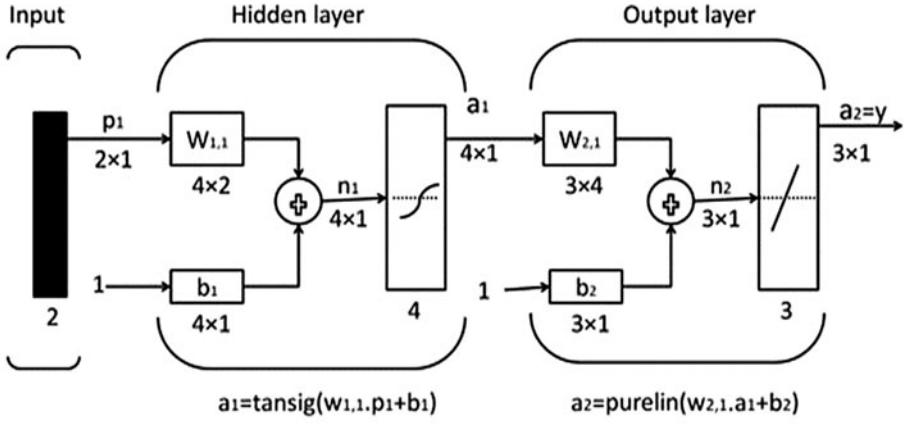

Fig. 2. A two-layer feed-forward network.

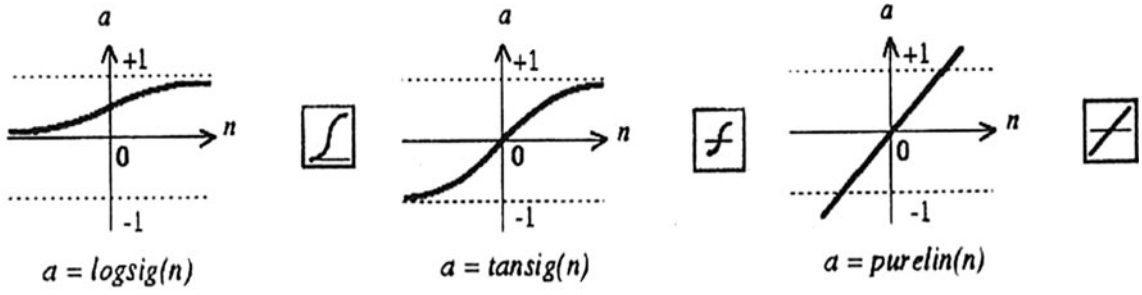

Fig. 3. Transfer functions.

$$\frac{\partial e(w,b)}{\partial b_{i,j}^{(l)}} = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b_{i,j}^{(l)}} e(w,b;x^{(i)},y^{(i)})\right] + \alpha b_{i,j}^{(l)} \tag{13}$$

where $e(w,b)$ = error and α = the learning rate.

Fig. 2 shows that output of the previous layer will be the input of the neurons in the next layer and the result(s) of the output layer will be compared with the target values. In this study, the MSE is the criterion for comparing the outputs.

$$\text{mse} = \frac{1}{m}\sum_{i=1}^{m} e^2 = \frac{1}{m}\sum_{i=1}(t_i - a_i)^2 \tag{14}$$

where t_i = the target (real) value, a_i = the network output and e = error.

Number of layers and neurons in each layer are designed such as: first layer of network contains inputs (input layer) and thus has a number of neurons same as the number of input parameter. This layer does not count in layer numbering (see Fig. 2) and the number of neurons in left layer (output layer) equals to the number of target parameter(s). The number of (hidden) layers in modeling projects like present study is four or less, but there is no specific rule for calculating the proper number of (hidden) layer and neurons in each layer [41].

The validation of all equations/models after development, have been tested with precision parameters such as R or R^* and MAE. R^* is used in cases where we had negative values of R. In other words, when $y_j > \bar{y}_j$ R will be negative.

$$R = \frac{\sum_{j=1}^{i} 1 - \frac{|\bar{y}_j - y_j|}{\bar{y}_j}}{i} \tag{15}$$

$$\begin{cases} \text{for } y_j > \bar{y}_j \rightarrow R_1 = \sum \dfrac{\bar{y}_j}{y_j} \\ \text{for } y_j < \bar{y}_j \rightarrow R_2 = \sum \dfrac{y_j}{\bar{y}_j} \\ R^* = \dfrac{R_1 + R_2}{i} \end{cases} \tag{16}$$

$$\text{MAI} = \frac{\sum_{j=1}^{i} |\bar{y}_j - y_j|}{i} \tag{17}$$

where i = the number of existed data for each parameter, y_j = estimated data number "j", and $\bar{y}_j$ = real data number "j".

6. Results

Five feed-forward networks with back-propagation learning rule (Eqs. (10)–(13)) are used to develop the models in MATLAB environment. The design parameters of the networks have been represented in Table 4.

Best (possible) number of (hidden) layers and their neurons found by try-and-error method, from 1 neuron to 250 neurons in each layer, and best conditions (due to minimizing MSE value) were selected. For smoothing and avoiding the uncertainty, which is caused by data with great values (much greater than average), tan sig(n) was selected as transfer function for "feed-forward back-propagation" network. For all (target) parameters both mathematical and ANN models are tested, in a wide range of possible design parameters. For mathematical modeling all types of curve fittings (Exponential, Fourier, Gaussian, Interpolant, Polynomial, Power, sum of sin functions) were tested with every possible (and rational) range of fitting parameters, for example for testing Polynomial fitting method orders from 1 to 6 are tested and best order (due to minimizing MSE) is selected. Both of

Table 4
Training parameters

$\alpha 0$	0.001	Network type	Feed-forward back-propagation
α decrease	0.1	Training function	Trainlm (Levenberg–Marquardt)
α increase	10	Adaptive learning function	Train GDM
Maximum α	1E+10	Performance function	MSE
Min. grad	1.00E-10	Transfer function	Tansig(x)

ANN and mathematical models were created for all target parameters but for some cases (the best obtained) mathematical or ANN models didn't have good precision (R and/or $R^* < 0.7$); thus it is not presented here.

DO seems to have a linear relation with T˚C [15], therefore, an equation is presented that simulates DO with an acceptable precision:

$$\mathrm{DO\,(mg/L)} = 0.0005 \cdot T^3 - 0.0318 \cdot T^2 + 0.529 \cdot T + 5.2022 \quad (18)$$

Also an ANN model is trained for estimating DO (Model 1). Three parameters of T˚C, EC (uS/cm), and pH are input parameters. The obtained precision value for Model 1 (0.913) is better than that of Eq. (18) with (0.871) as it is illustrated in Table 6 and Fig. 5.

6.1. TDS

The linear Correlation between EC and TDS is stated by several researches [24,42]. TDS/EC is always linear but the coefficient(s) varies in time and space [42,24,17]:

$$\mathrm{TDS\,(mg/L)} = 0.64\,\mathrm{EC\,(\mu S/cm)} \quad (19)$$

$$\mathrm{TDS\,(mg/L)} = 0.50\,\mathrm{EC\,(\mu S/cm)} \quad (20)$$

$$\mathrm{TDS\,(mg/L)} = 0.781\,\mathrm{EC\,(\mu S/cm)} + 70 \quad (21)$$

Results of this study exactly matched with Eq. (19) and an extra equation (polynomial) was also developed to calculate TDS/EC corresponding to T:

$$\frac{\mathrm{TDS\,(mg/L)}}{\mathrm{EC\,(\mu S/cm)}} = 0.0001 \cdot T^2 - 0.005 \cdot T + 0.71 \quad (22)$$

As it is expressed in Table 6, both equations (Eqs. (19) and (22)) have the same precision value. However, Eq. (22) has a vision of temperature effect on TDS/EC relation that is shown in Fig. 4. Fig. 6 shows the comparison between models/equations and real data for TDS.

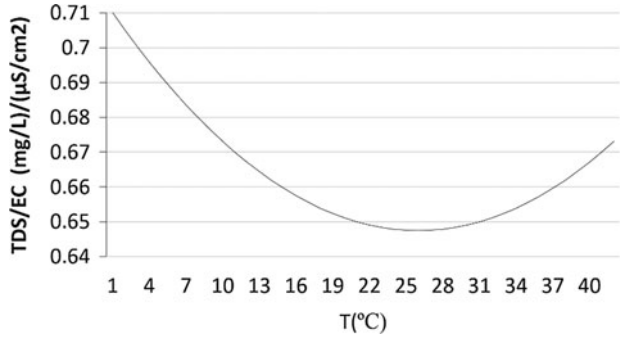

Fig. 4. Effect of temperature on TDS/EC.

6.2. SAR

A sensitivity analysis demonstrated that pH has no significant correlation with SAR and therefore, it is not a "helpful" parameter in mathematical modeling. However, EC, according to its (linear) relation with TDS, can lead us to simulate SAR value [32]. The parameter α is defined as:

$$\alpha = EC\,(0.0001 \cdot T^2 - 0.005 \cdot T + 0.71) \quad (23)$$

Using mathematical method, SAR can be expressed as:

$$\mathrm{SAR} = 0.00281 \cdot \alpha + 0.9978 \quad (24)$$

According to Table 6, Eq. (24) can simulate SAR with a sufficiently high precision. Also an ANN model (Model 2) is trained using EC, T. and pH data as input variables. As it is shown, compared with Eq. (24), model 2 has the advantage of higher precision (Table 6 and Fig. 7).

6.3. BOD

Model 3 is presented to simulate BOD_5. It uses EC, T, DO, and pH as input parameters to model the

Table 5
Design parameters of ANN models

Model #	Number of hidden layer(s)	Number of neurons in each layer
Model 1	2	10–40
Model 2	4	15–20–30–10
Model 3	3	30–25–20
Model 4	2	20–30

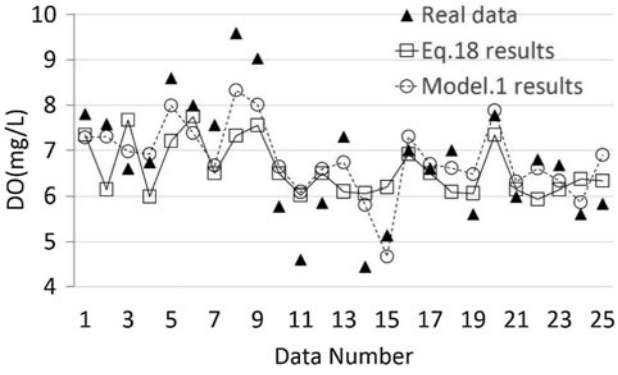

Fig. 5. Comparison between models/equations and real data for DO.

BOD_5. Fig. 8 shows the simulation results compared with real data.

6.4. HCO_3

Similar to BOD, an ANN model (Model 4) is trained for modeling HCO_3. The model accurately simulates HCO_3 concentrations corresponding to values of EC, T, and pH (as inputs) with R^2 of 0.86. Table 5 shows the number of (hidden) layers and the number of neurons for each model. Fig. 5 and Table 6

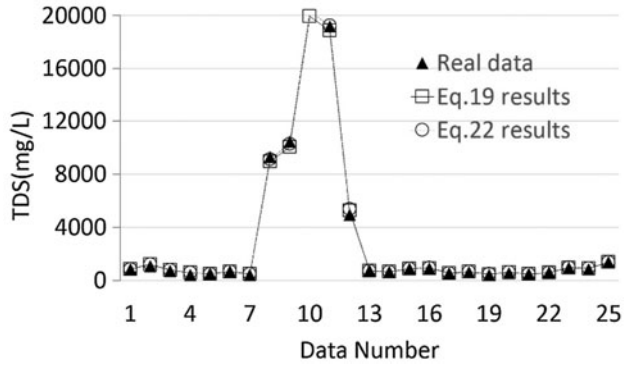

Fig. 6. Comparison between models/equations and real data for TDS.

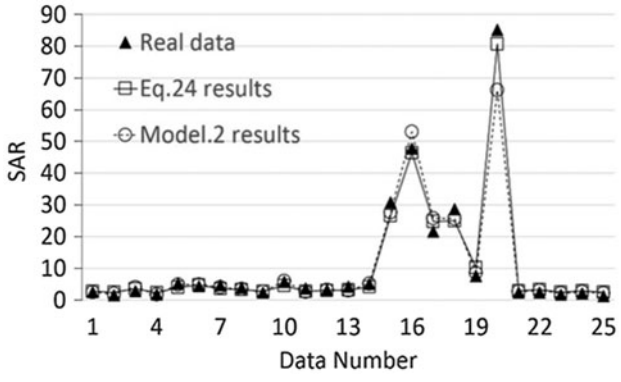

Fig. 7. Comparison between models/equations and real data for SAR.

Table 6
A summary of equations and models properties and comparing their precisions values

	Target of simulation	Number of used data	Input parameters	R	R^*	MAE	Average parameter
Eq. (18)	DO (mg/L)	238	T	0.871	–	0.94 mg/L	7.21 mg/L
Model 1	DO (mg/L)	2,380 (×10)(*)	T, EC, pH	0.913	–	0.59 mg/L	7.21 mg/L
Eq. (19)	TDS/EC (mg-μS/L-cm^2)	1,328	0.64	0.957	–	0.02 mg/L	0.642 (mg-μS/L-cm^2)
Eq. (22)	TDS/EC (mg-μS/L-cm^2)	1,328	T	0.95	–	0.03 mg/L	0.642 (mg-μS/L-cm^2)
Eq. (19)	TDS (mg/L)	1,328	EC	0.954	–	168.9 mg/L	3,093 mg/L
Eq. (22)	TDS (mg/L)	1,328	T, EC	0.948	–	187.9 mg/L	3,094 mg/L
Eq. (24)	SAR	1,202	EC,T	0.688	0.79	1.63	10.78
Model 2	SAR	4,808 (×4)(*)	EC, T, pH	0.744	0.819	1.35	10.78
Model 3	BOD_5 (mg/L)	2,380 (×10)(*)	EC, T, pH, DO	0.852	–	0.35 mg/L	2.65 mg/L
Model 4	HCO_3 (mg/L)	1,257	EC, T, pH	0.821	0.862	0.40 mg/L	2.69 mg/L

*The number of data increased n times, by interpolation method. n values are shown in Parenthesis (×n).

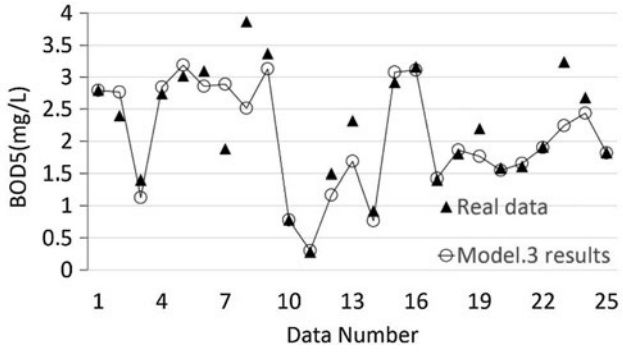

Fig. 8. Comparison between models/equations and real data for BOD_5.

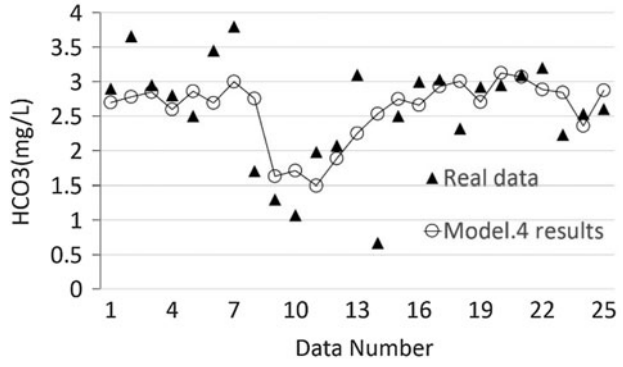

Fig. 9. Comparison between models/equations and real data for HCO_3.

demonstrate all models properties and equations presented in this study. Fig. 9 shows the real data that is used for verifying the models and equations.

All existed data (Table 3) are used for testing the models/equations. Table 6 shows the results of these tests. Moreover, typical results of the models/equations are shown in Fig. 5–9 using 25 randomly selected data.

7. Conclusions

This study presented simple and useful models/equations that can be used for many purposes such as regenerating missing data, assessing measurement devices and operators, calibrating measuring equipment's, and as a reliable replacement for the measurement of equipment/processes that can significantly reduce costs and save time. The models/equations are developed based on data from 7 different monitoring stations in 5 rivers for 30 years. Therefore, we consider/cover almost every possible conditions/circumstances that can be useful in other similar locations (rivers). Table 6 shows high precision of the models/equations that makes them extremely reliable. This study also shows the adequacy of ANN and mathematical methods for modeling that can simply be used for other similar purposes. For parameters such as BOD and HCO_3, mathematical models couldn't reach an acceptable precision (R and/or $R^* < 0.7$). In the case of TDS, all possible ANN models were tested and the optimum ANN model, have much less precision (comparing to mathematical Equations). Therefore, simple linear modeling could be more effective than other more comprehensive procedure.

Main advantages of mathematical modeling are: (1) simplicity of expression/proving and using them and (2) the form of solution and its sensitivity toward two/multi parameters could be observed and analyzed. The advantages of ANN modeling are: (1) the ability of modeling complex relationships which is not possible to model with ordinary mathematical modeling methods [43–45]; (2) in most cases, ANN models have higher precision compared with mathematical methods; (3) ANN models could be adapted with new data and upgraded with them; (4) best combination of design parameters in ANN models could be achieved by try and error method. This process also revealed that the number of hidden layers and their neurons should be chosen wisely, both lack and excess of it, could decrease the model precision. One of the most important finding in this study is when the number of data is sufficient for modeling, the precision of the model directly depends on (and limited by) data that are being used. In other words, the most uncertainty would be caused by uncertainty of the input data. This means that even if we use the other methods of modeling (such as genetic algorithm, dynamic NN, etc.), better result won't be achieved because at this point errors depend on error of data that are used for making the models. Table 6 shows that each (target) parameter depends on (input) parameters; in other words, it shows highest sensitivities and dependency between "target" and "input" parameters.

Acknowledgment

The authors are grateful to Dr Mohammad Rabbani for their editorial and revision assistance. They are also thankful to the Department of Water Resources, Khuzestan Province authority, for their logistical assistance in the process of data collection and analyses.

Nomenclature

DO	—	dissolved oxygen (mg/L)
SAR	—	sodium adsorption rate (unit less)
TDS	—	total dissolved solids (mg/L)
BOD, BOD_5	—	biological oxygen demand (in 5 d) (mg/L)
T	—	temperature
EC	—	electrical conductivity
Na^+	—	sodium concentration (mg/L)
Ca^{2+}	—	calcium concentration (mg/L)
Mg^{2+}	—	Manganese concentration (mg/L)
MAE	—	Mean of absolute error (as same as target unit)
MSE	—	Mean of squared error (second power of target unit)
R, R^*	—	rate of precision (unitless number ≤ 1)

References

[1] P.N. Okeke, E.N. Adinna, Water quality study of Otamiri river in Owerri, Nigeria, Universal J. Environ. Res. Technol. 3(6) (2013) 641–649.

[2] P. Vaishali, P. Punita, Assessment of seasonal variation in water quality of River Mini, at Sindhrot, Vadodara, Int. J. Environ. Sci. 3(5) (2013) 1424–1436.

[3] F. Al-Badaii, M. Shuhaimi-Othman, M.B. Gasim, Water Quality Assessment of the Semenyih River, Selangor, Malaysia, J. Chem. 2013 (2013) 10.

[4] 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.

[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. Model. Soft. 22 (2007) 743–759.

[6] U.S. Mcknight, S. 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(9) (2010) 1126–1137.

[7] 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.

[8] I. Sevostianov, M. Shrestha, Cross-property connections between overall electric conductivity and fluid permeability of a random porous media with conducting skeleton, Int. J. Eng. Sci. 48 (2010) 1702–1708.

[9] J.A. Dean, What is pH, and How Is It Measured? Hach Company, McGraw-Hill Book Company, New York, NY, 2003, pp. 1–25.

[10] K. Kumar N. Yadav Gupta, V. Kumar, S. Sharma, S. Arya, Water Quality Assessment of Pahuj River using Water Quality Index at Unnao Balaji, M.P., India, Int. J. Sci.: Basic and Appl. Res. (IJSBAR) 19(1) (2015) 241–250.

[11] C. Lihua, M. Shengquan, L. Li, A model to evaluate DO of river based on artificial neural network and stylebook, J. Hainan Normal Univ. (Nat. Sci.) 21(4) (2008) 372–376.

[12] B. Jones, Oxygen—The most important water quality parameter? Water Column 23 (2011) 1–6.

[13] K. Venkatesharaju, P. Ravikumar, R.K. Somashekar, K.L. Prakash, Physic-chemical and bacteriological investigation on the river Cauvery of Kollegal stretch in Karnataka, Kathmandu University J. Sci. Eng. Technol. 6(1) (2010) 50–59.

[14] T. Siwiec, L. Kiedrynska, K. Abramowicz, A. Rewicka, P. Nowak, BOD measuring and modelling methods—Review, Land Reclamation 43(2) (2011) 143–153.

[15] J. Kwak, B. Khang, E. Kim, H. Kim, Estimation of biochemical oxygen demand based on dissolved organic carbon, UV absorption, and fluorescence measurements, J. Chem. 1 (2013) 9.

[16] P.K. Weber-Scannell, L.K. Duffy, Effects of total dissolved solids on aquatic organisms: A review of literature and recommendation for salmonid species, Am. J. Environ. Sci. 3(1) (2007) 1–6.

[17] I.E. Uwidia, H.S. Ukulu, Studies on electrical conductivity and total dissolved solids concentration in raw domestic wastewater obtained from an estate in Warri, Nigeria, Greener J. Phys. Sci. 3(3) (2013) 110–114.

[18] N.R.G. Walton, Electrical conductivity and total dissolved solids—What is their precise relationship? Desalination 72 (1989) 275–292.

[19] S.A. Naqvi, M. Kyung, K. Moonil, A case study on the relationship between conductivity and dissolved solids to evaluate the potential for reuse of reclaimed industrial wastewater, Korean Soc. Civil Eng. 16 (2012) 708–713.

[20] M.R.G. Sayyed, G.S. Wagh, An assessment of groundwater quality for agricultural use: A case study from solid waste disposal site SE of Pune, India, Proc. Int. Acad. Ecol. Environ. Sci. 1 (2011) 195–201.

[21] A. Honarbakhsh, E. Kakaei Lafdani, Performance comparison of ANN and geo statistics methods for estimating spatial distribution of sodium adsorption ratio (SAR) in groundwater, Int. J. Agric. Crop Sci. 5 (2013) 2837–2847.

[22] K. Srinivasa Reddy, Assessment of groundwater quality for irrigation of Bhaskar Rao Kunta watershed, Nalgonda District, India, Int. J. Water Resour. Environ. Eng. 5 (2013) 418–425.

[23] K. Anim, G.O. Duodu, E.K. Ahialey, Y. Serfor-Armah, Assessment of surface water quality: The perspective of the Weija dam in Ghana, Int. J. Chem. 3(2) (2011) 32–39.

[24] E. Salami Shahid, M. Ehteshami, Application of artificial neural networks to estimating DO and salinity in San Joaquin River basin, Desalin. Water Treat. 57(11) (2016) 4888–4897, doi: 10.1080/19443994.2014.995713.

[25] M. Kotti, A. Vlessidis, N. Thanasoulias, N. Evmiridis, Assessment of river water quality in northwestern Greece, Water Resour. Manage. 19 (2005) 77–94.

[26] M.J. Wirmvem, T. Ohba, W.Y. Fantong, S.N. Ayonghe, J.Y. Suila, A. Nkengmatia, E. Asaah, G. Tanyileke, J.V. Hell, Hydrochemistry of shallow groundwater and surface water in the Ndop plain, North West Cameroon, Afr. J. Environ. Sci. Technol. 72 (2013) 3585–5398.

[27] P. Kumarasamy, R.A. James, H.U. Dahms, C.W. Byeon, R. Ramesh, Multivariate water quality assessment from the Tamiraparani river basin, Southern India, Environ. Earth Sci. 71 (2014) 2441–2451.

[28] S. Somvanshi, P. Kunwar, N.B. Singh, S.P. Shukla, V. Pathak, Integrated remote sensing and GIS approach for water quality analysis of Gomti river, Uttar Pradesh, Int. J. Environ. Sci. 3 (2012) 62–75.
[29] N.M. Gazzaz, K. Yusoff, A. Zaharinaris, H. Juahir, M.F. Ramli, Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors, Marine Pollution Bulletin 64(11) (2012) 2409–2420.
[30] R.K. Panda, N. Pramanik, B. Bala, Simulation of river stage using artificial neural network and MIKE 11 hydrodynamic model, Comput. Geosci. 36 (2010) 735–745.
[31] F. Anctil, M. Filion, J. Tournebize, A neural network experiment on the simulation of daily nitrate-nitrogen and suspended sediment fluxes from a small agricultural catchment, Ecol. Model. 220 (2009) 879–887.
[32] S.A. Moasheri, A.S. Goshki, A. Parsaie, SAR Qualities parameter persistence by a compound method of geostatic and artificial neural network (Case study of Jiroft plain), Int. J. Agric. Crop Sci. IJACS 6(3) (2013) 157–166.
[33] J. Verschelde, MATLAB Lecture 3. Polynomials and curve fitting, UIC, Department of Math, Stat & CS, Introduction to symbolic computation, Matlab Lecture 3 (2007) 1–3.
[34] N.P. Karampetakis, Computation of the generalized inverse of a polynomial matrix and applications, Elsevier Science Inc. 252 (1997) 35–61
[35] Y.M. Kuo, C.W. Liu, K.H. Lin, Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan, Water Res. 38 (2004) 148–158.
[36] N. Samani, M. Gohari-Moghadam, A.A. Safavi, A simple neural network model for the determination of aquifer parameters, J. Hydrol. 340 (2007) 1–11.
[37] D. Svozil, V. Kvasnicka, J. Pospichal, Introduction to multi-layer feed-forward neural networks, Chemom. Intell. Lab. Syst. 39(1) (1997) 43–62.
[38] K. Tamás, The application of neural networks for solving complex optimization problems in modeling, Conf. Junior Res. Civil Eng. 1 (2010) 97–102.
[39] Abraham, Handbook of Measuring System Design, Oklahoma State University, Stillwater, OK, 2005, pp. 901–908.
[40] M.B. Menhaj, Computational intelligence: Fundamental of neural network, second ed., Amir Kabir University of Technology publication, Tehran, 2005.
[41] M.T. Hagan, H.B. Demuth, M.H. Beale, O. De Jesús, Neural Network Design, second ed., 2014, 1012. Available from: <hagan.okstate.edu/nnd.html>.
[42] R.S. Ayers, D.W. Westcot, Water quality for agriculture, FAO irrigation and drainage. United Nations Rome, 29(1) (1994) 97 p.
[43] E.S. Salami, M. Ehteshami, Simulation, evaluation and prediction modeling of river water quality properties (case study: Ireland Rivers), Int. J. Environ. Sci. Technol. 12(10) (2015) 3235–3242, doi: 10.1007/s13762-015-0800-7.
[44] M. Ehteshami, N. Dolatabadi Farahani, S. Tavassoli, Simulation of nitrate contamination in groundwater using artificial neural networks, Model. Earth Syst. Environ. (Springer) 2(1) (2016) 1–10
[45] E.S. Salami, M. Ehteshami, Application of neural networks modeling to environmentally global climate change at San Joaquin old river station, Model. Earth Syst. Environ. (Springer). 2(1) (2016) 1–9.