

## Application of an Artificial Neural Network Model to Rivers Water Quality Indexes Prediction – A Case Study

Hossein Banejad<sup>1</sup>, Ehsan Olyaie<sup>1</sup>

<sup>1</sup>: Department of Water Engineering, Faculty of Agriculture, Bu-Ali Sina University of Hamedan, Iran  
[Hossein\\_banejad@yahoo.com](mailto:Hossein_banejad@yahoo.com)

**Abstract:** Taxonomic Recent trends in the management of water supply have increased the need for modeling techniques that can provide reliable, efficient, and accurate representation of the nonlinear dynamics of water quality within water distribution systems. Since artificial neural networks have been widely applied to the nonlinear transfer function approximation, in this study we present an empirical multi layer perceptron neural network to estimate water quality indexes (BOD, Do) in Morad Big River in the western part of Iran. In this paper, the information and data including 10 monthly parameters of water quality in the Hamedan Morad Big River in duration of one year and six stations were used for modeling biological oxygen demanded (BOD) and dissolved oxygen (DO) as indices affecting water quality. To validate the performance of the trained ANN, it was applied to an unseen data set from a station in the region. Performance of the model was evaluated by statistical criteria includes correlation coefficient (r), root mean square error (RMSE) and mean absolute error (MAE). In the optimum structure of neural network correlation coefficient for BOD and DO are 0.986 and 0.969, also root mean square error are 8.42 and 0.84 respectively. The results show the identified ANN's great potential to simulate water quality variables.

[Hossein Banejad, Ehsan Olyaie. Application of an Artificial Neural Network Model to Rivers Water Quality Indexes Prediction – A Case Study. Journal of American Science 2011;7(1):60-65]. (ISSN: 1545-1003). <http://www.americanscience.org>.

**Keywords:** Artificial Neural Networks; Predicting; Water Quality Index; BOD; DO

### 1. Introduction

As population and industrial development increases, so does the need for water. The surface water quality in a region largely depends on the nature and extent of the industrial, agricultural and other anthropogenic activities in the catchments. Human beings take water from the hydrologic cycle for their vital and economic needs and give it to the same cycle after using it (Durdu, 2009). The substances that mixed with water during this cycle bring out the concept of water pollution as they change the physical, chemical, and biological properties of water after natural refining (Maier et al., 2004; May and Sivakumar, 2009). In consequence of this, the suitable part of water from which humans cannot cease having for their lives becomes less and less. To prevent this unwanted trend, control of water pollution seriously has become very essential to maintain the sustainability of water resources. River pollution parameters exhibit different properties (Wu et al., 2009; Sahoo et al., 2006). Since the change of pollution in time at any point of a river can affect the downstream of that point at a certain rate, there is a relation also between the change of pollution in time of any point and the change of positional pollution along the river. The river systems are most adversely affected due to their dynamic nature and an easy accessibility for the waste disposal directly or indirectly through drains/tributaries. Since, the rivers and streams are among most important sources of

water for irrigation, industrial and other uses, these serve as the lifelines of the population staying in the basins (May et al., 2008). In general, the organic pollution in an aquatic system is measured and expressed in terms of the biochemical oxygen demand (BOD) and declined dissolved oxygen (DO) level. The BOD measures an approximate amount of bio-degradable organic matter present in water and serves as an indicator parameter for the extent of water pollution. The BOD of any aquatic system is the foremost parameter needed for assessment of the water quality as well as development of management strategies for the protection of water resources. This warrants for a foolproof method for its determination. It causes low DO (dissolved oxygen) concentration and unsuitable life conditions for flora and fauna in the river (Dogan et al., 2009). At the same time, BOD–DO relationships include exchange with the river bed and nitrification and denitrification (Sengorur et al., 2006; Singh et al., 2009). Nutrients and light in the phytoplankton growth, the relationship between DO and phytoplankton concentrations and ammonia affect the BOD degradation. Currently available method for BOD determination is very tedious and prone to measurement errors. Since, BOD is inversely related to the dissolved oxygen in water, the high values of the earlier indicate for a low level of the dissolved oxygen (DO) or even anoxic conditions in water. Therefore, both these parameters (DO-BOD) are

generally needed to be determined simultaneously and there is a need to devise some suitable secondary (indirect) method for predicting these variables in a large number of samples for water quality assessment (Singh et al, 2009). In recent years, several water quality models such as traditional mechanistic approaches have been developed in order to manage the best practices for conserving the quality of water. Most of these models need several different input data which are not easily accessible and make it a very expensive and time consuming process ANN is the suitable approach for water quality modeling (Chen et al., 2003; Jan-Tai et al., 2006).

The main aim of the present work is to construct an artificial neural network (ANN) model of the Morad Big River water quality (DO,BOD) and demonstrate its application to complex water quality data as how it can improve the interpretation of the results.

## 2. Material and Methods

### 2.1. Study Area and Water Quality Data

Morad Big River is in the western part of Iran in Hamedan. The water quality data of 6 stations were used in this study. Those stations located between latitudes  $34^{\circ} 44' N$  and  $34^{\circ} 51' N$  and longitudes  $48^{\circ} 30' E$ . The length of river is 13949 m. The river during its course receives low to very high pollution load from various diffuse and point sources in its different stretches while flowing through urban townships, thus exhibiting very large variations in water quality variables. The locations of these stations are illustrated in Figure 1.

For the stations, the data for September 23 2008 to May 23 2009 were chosen for training and data for May 24 2009 to September 22 2009 were chosen for validation, arbitrarily.

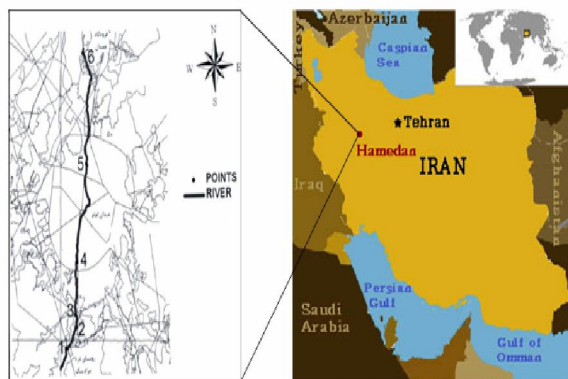


Figure 1. The locations of the stations of Morad Big River (Iran)

At each station, we have measured for 12 different quality parameters are carried out with monthly periods. The measured parameters at these stations in the basin are shown in Table1.

Table 1. Parameters measured at quality observation stations in Morad Big River.

Parameters	Unit
pH	-
Electrical Conductivity	Micromhos/cm
Total Dissolved Solid	mg/lit
Total Suspended Solid	mg/lit
Turbidity	NTU
Sodium	mg/lit
Bicarbonate	mg/lit
Nitrate	mg/lit
Ammoniac	mg/lit
Phosphate	mg/lit
Dissolved oxygen	mg/lit
Biological Oxygen Demand	mg/lit

### 2.2. Artificial Neural Networks Modeling

In this study, an empirical neural network algorithm was applied to estimate surface water quality parameters (BOD, DO). ANN models are highly flexible function-approximators that have shown their utility in a broad range of water resources applications. Most of these studies showed that ANNs performed better than classical modeling methods (Zhang et al, 2002). ANN is a logical programming technique developed by imitating the working mechanism of human brain. An ANN algorithm can make the brain operations, decide, reach the result in the condition of insufficient data with the help of present information and accept continuous data input, train, and remember. ANN aims to develop the basic operations-which a human brain does biologically-with a definite algorithm. The greatest advantage of a neural network is its ability to model complex nonlinear relationship without a priori assumptions of the nature of the relationship. The ANN model performs a nonlinear functional mapping from the past observations ( $X_{t-1}, X_{t-2} \dots X_{t-p}$ ) to the future value  $X_t$ , i.e (Durdu, 2009).

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + et \quad (1)$$

Where  $w$  is a vector of all parameters and  $f$  is a function determined by the network structure and connection weights. Thus, the neural network is equivalent to a nonlinear autoregressive model. Training a network is an essential factor for the success of the neural networks. Training the problem by adjusting the weights in itself as using the present information and remembering later are the most important properties of ANN. ANN is a technique that is used in recognizing video and audio, speaking,

analyzing, deciding, defining complex models, and controlling. It differs from the classical programming from very different points of view and is developed by considering working principles of a human brain. In this technique, neuro-physiological working principles (working with electro-chemical principles) of human brains are applied by artificial neurons in computer. The neurons, the basic structural element of a human brain, are electro-chemical operation elements and make operations in degrees of milliseconds. Among the several learning algorithms available, back-propagation has been the most popular and most widely implemented learning algorithm of all neural networks paradigms (Sundarambal et al., 2008; Wu et al., 2008). Feed forward, back propagation networks have previously been identified as the most common type of ANN models used in water resources applications. Therefore, such networks were used in the current study. Networks constructed in the current study comprised of three layers: an input layer, a hidden layer and an output layer. An example of a network topology is shown in Figure 2.

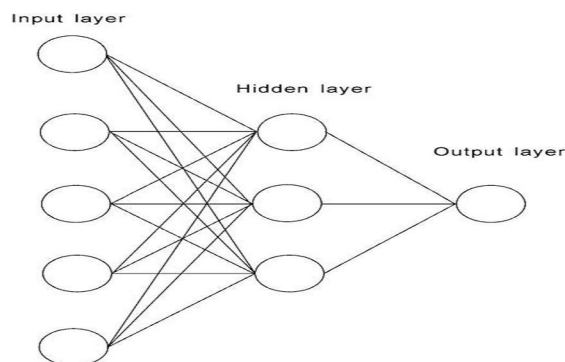


Figure 2. An example of an artificial neural network topology with one input layer, one hidden layer and one output layer

### 2.2.1. Back Propagation Neural Networks Learning Algorithm

The back propagation (BP) is a commonly used learning algorithm in ANN application (Ying et al, 2007). The back-propagation algorithm based upon the generalized delta rule proposed by Rumelhart et al. (1986) was used to train the ANN in this study. In the back-propagation algorithm, a set of inputs and outputs is selected from the training set and the network calculates the output based on the inputs. This output is subtracted from the actual output to find the output-layer error. The error is back-propagated through the network, and the weights are suitably adjusted. This process continues for the number of prescribed sweeps or until a pre

specified error tolerance is reached. The mean square error over the training samples is the typical objective function to be minimized. It uses the back propagation (BP) of the error gradient. This training algorithm is a technique that helps distribute the error in order to arrive at a best fit or minimum error. After the information has gone through the network in a forward direction and the network has predicted an output, the back propagation algorithm redistributes the error associated with this output back through the model, and weights are adjusted accordingly. Minimization of the error is achieved through several iterations. After training is complete, the ANN performance is validated. Depending on the outcome, either the ANN has to be retrained or it can be implemented for its intended use.

In this study, before the training of the network both input and output variables were normalized within the range 0.1–0.9 as follows:

$$x = 0.8 \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} + 0.1 \quad (2)$$

is the normalized value of a certain parameter,  $x$  is the measured value for this parameter,  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum values in the database for this parameter, respectively (Dogan et al, 2009).

### 2.2.2. Input Variable and Data processing

The monthly data of twelve water quality parameters measured over a period of one year at all the six sampling station were selected for this analysis. The DO and BOD are two major parameters in water quality assessment. Based on existing measured values of different variables and their correlative analysis, total 10 factors (variables) including pH, EC, HCO<sub>3</sub>, TDS, TSS, Turbidity, NO<sub>3</sub>, PO<sub>4</sub>, Na and NH<sub>3</sub> were identified which affect the water quality (DO and BOD) to certain degree and finally selected for the model development (table 1). Subsequently, two different ANN models were constructed for the computation of DO and BOD in the river water. The network was trained using the training data set, and then it was validated with the validation data set. The development of ANN models required that the available modeling data be sampled into two smaller subsets for training and validating the network. In this study, the perspective proportions of samples allocated to each of these subsets were as follow:

The data for 23 September 2008 to 23 May 2009 were chosen for training, and data for 24 May 2009 to 22 September 2009 were chosen for validation.

All the computations were performed using the Qnet2000 and EXCEL.

### 2.3. Performance Criteria

A model trained on the training set can be evaluated by comparing its predictions to the measured values in the over fitting test set. These values are calibrated by systematically adjusting various model parameters. A multi-criteria approach was adopted for assessing the models developed, in which model performance was evaluated using several statistical error and goodness-of-fit measures, including the root mean squared error (RMSE), the mean absolute error (MAE), the correlation coefficient (r). Scatter plots and time series plots are used for visual comparison of the observed and predicted values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (4)$$

$$r = \sqrt{1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n O_i^2 - \frac{\sum_{i=1}^n P_i^2}{n}}} \quad (5)$$

Here  $P_i$  and  $O_i$  are the predicted and observed values, respectively. And  $n$  is the total number of observations.

### 3. Results

Different ANN models were constructed and tested in order to determine the optimum number of nodes in the hidden layer and transfer functions. Selection of an appropriate number of nodes in the hidden layer is very important aspect as a larger number of these may result in over-fitting, while a smaller number of nodes may not capture the information adequately. Subsequently, two different ANN models were constructed for the computation of DO and BOD in the river water. The network was trained using the training data set, and then it was validated with the validation data set. The optimal network size was selected from the one which resulted in minimum mean absolute error (MAE) in training and validation data sets. The model features for both the ANNs are given in Table 2.

The selected ANN for the DO model is composed of one input layer with ten input variables, one hidden layer with sixteen nodes and one output layer with one output variable, whereas, the BOD model differed in number of nodes in the hidden layer, as it optimized with ten nodes in this layer. The constructed ANN models (DO and BOD) were trained using the Back Propagation algorithm (BP).

The correlation coefficient (r), RMSE, and the MAE as computed for the training and validation data sets used for the two models (DO and BOD) are presented in Table 2. Fig.3 shows the plots between measured and model computed values of DO in training and validation sets. The selected ANN (10 nodes in input layer, 16 nodes in hidden layer, and single node in output layer) provided a best fit model for all the three data sets. The correlation coefficient (r) values for the training and validation sets were 0.956 and 0.969 respectively. The respective values of RMSE for the two data sets are 1.02 for training and 0.84 for validation (Table 2). A closely followed pattern of variation by the measured and model computed DO concentrations in river water (Fig. 3), r, RMSE and MAE values suggest for a good-fit of the DO model to the data set.

### 4. Discussions

In case of the BOD, the selected ANN (10 nodes each in input and 20 nodes in hidden layers and single node in output layer) provided a best fit model for all the two (training and validation) sets. Fig.4 shows the plots between the measured and model computed values of BOD in training and validation sets. The correlation coefficient (r) values for the training and validation sets were 0.969 and 0.986, respectively. The respective values of RMSE for the two data sets are 9.01 for training and 8.42 for validation (Table 2). A closely followed pattern of variation by the measured and model computed BOD values (Fig. 4), r, RMSE and MAE values suggest for a good-fit of the selected BOD model to the data set.

In Fig. 5, the predicted DO and BOD of BPNN with architectures 10-16-1 (ten neurons in the input layer, sixteen neurons in the hidden layer and one neuron in the output layer) and 10-26-1 (ten neurons in the input layer, twenty six neurons in the hidden layer and one neuron in the output layer) are compared with corresponding measured DO and BOD respectively. The figure reveals that an acceptable agreement between the simulations and observations can be achieved. The correlation coefficient values between the ANN models predicted values and observed data for dissolved oxygen and biological oxygen demand are 0.969 and 0.986, respectively, which are satisfactory in common model applications. These results indicate that the neural network model is able to recognize the pattern of the water quality parameters to provide good predictions of the monthly variations of water quality data (BOD and DO) of the Morad Big River.

These results clearly indicate the performance of the neural network. This is expected because of the nonlinear nature of the transfer



function between the water quality characteristics such as BOD, Do and other water quality parameters.

The present study shows that the optimal networks are capable to capture long-term trends observed for the tedious water quality variables (DO and BOD), both in time and space. We propose the neural networks as effective tool for the computation of river water quality and it could also be used in other areas to improve the understanding of river pollution trends. Thus, the ANN can be seen to be a

powerful predictive alternative to traditional modeling techniques.

#### Corresponding Author:

Dr. Hossein Banejad

Department of Water Engineering, Faculty of Agriculture, Bu-Ali Sina University of Hamedan, Iran

E-mail: [Hossein\\_banejad@yahoo.com](mailto:Hossein_banejad@yahoo.com)

Table 2. Performance parameters of the artificial neural network models for computation of the DO and BOD in Morad Big River water (Iran)

Model	ANN-Structure	data sets	RMSE (mg/lit)	MAE (mg/lit)	r
DO	10-16-1	Training	1.02	0.78	0.956
		Validation	0.84	0.69	0.969
BOD	10-20-1	Training	9.01	6.71	0.969
		Validation	8.42	6.27	0.986

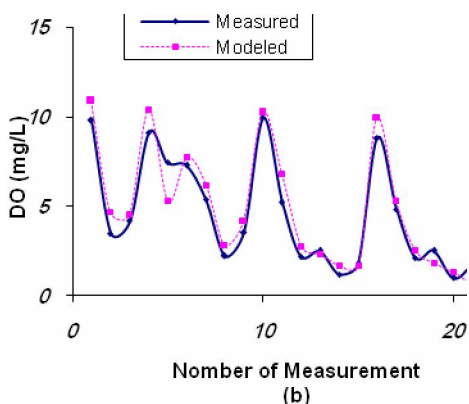
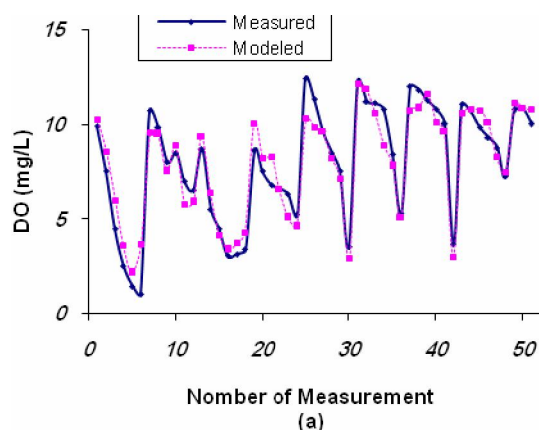


Fig 3. Comparison of the model computed and measured DO levels in the river water (a) training and (b) validation using DO-ANN model.

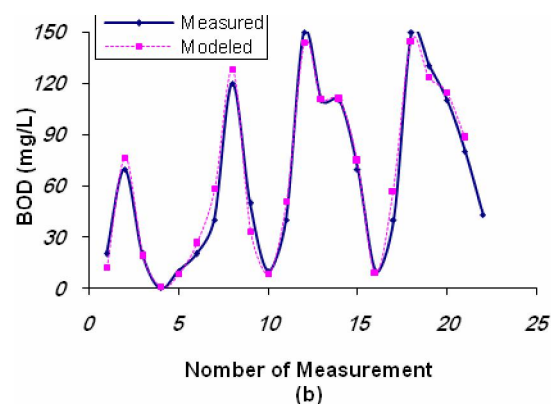
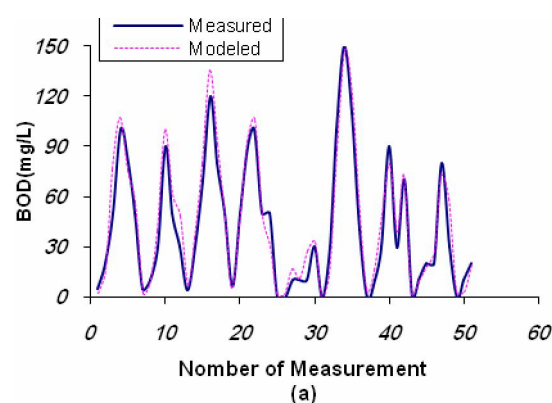


Fig 4. Comparison of the model computed and measured BOD values in the river water (a) training and (b) validation sets using BOD-ANN model

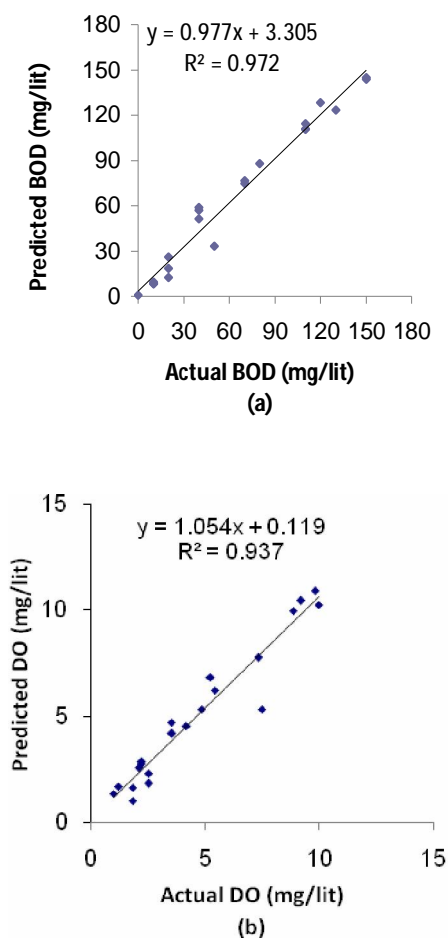


Fig 5. Scatter plots of actual versus predicted BOD values (a) and DO value (b) for the validation data obtained using ANN model.

#### References

1. Durdu OF. A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence* 2009; 32: 1025-1034.
2. Maier HR, Morgan N, Chow CW. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling & Software* 2004; 19: 485-494.
3. May DB, Sivakumar M. Prediction of urban stormwater quality using artificial neural networks. *Environmental Modelling & Software* 2009; 24: 296-302.
4. Wu J, Lu J, Wang J. Application of chaos and fractal models to water quality time series prediction. *Environmental Modelling & Software* 2009; 24:632-636.
5. Sahoo GB, Ray C, Carlo EH. Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. *Journal of Hydrology* 2006; 327: 525-538.
6. May RJ, Dandy GC, Maier HR, Nixon JB. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environmental Modelling & Software* 2008; 23: 1289-1299.
7. Dogan E, Sengorur B, Koklu R. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *Journal of Environmental Management* 2009; 90: 1229-1235.
8. Sengorur B, Dogan E, Koklu R, Samandar A. Dissolved oxygen estimation using artificial neural network for water quality control. *Fresen. Environ. Bull* 2006; 15: 1064-1067.
9. Singh KP, Basant A., Malik A., Jain G. Artificial neural network modeling of the river water quality—A case study. *Ecological Modeling* 2009; 220: 888-895.
10. Chen JC, Chang NB, Shieh WK. Assessing wastewater reclamation potential by neural network model. *Eng. Appl. Artif. Intell* 2003; 16: 149-157.
11. Jan-Tai Ku, Ying-Yi W, Wu-Seng L. A hybrid neural-genetic algorithm for reservoir water quality management. *Water Research* 2006; 40:1367 - 1376.
12. Zhang Y, Pulliainen J, Koponen S, Hallikainen M. Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sensing of Environment* 2002; 81: 327- 336.
13. Sundarambal P, Shie-Yui L, Pavel T. An ANN application for water quality forecasting. *Marine Pollution Bulletin* 2008; 56:1586-1597.
14. Wu GD, Lo SL. Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system. *Engineering Applications of Artificial Intelligence* 2008; 21: 1189- 1195.
15. Ying Z, Jun N, Fuyi C, Liang G. Water quality forecast through application of BP neural network at Yuquio reservoir. *J. Zhejiang Univ. Sci* 2007; 8: 1482-1487.

11/22/2010