Journal of King Saud University – Engineering Sciences (2015) **xxx**, xxx–xxx

King Saud University

Journal of King Saud University – Engineering Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River

A.A. Masrur Ahmed *, **Syed Mustakim Ali Shah** [1]

Department of Civil Engineering, Leading University, Sylhet, Bangladesh

Received 22 July 2014; accepted 17 February 2015

KEYWORDS

Water quality;
Adaptive neuro-fuzzy inference system;
Biochemical oxygen demand;
Surma River;
Sylhet;
Bangladesh

Abstract This paper describes the application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River of Bangladesh. The data sets consist of 10 water quality parameters which include pH, alkalinity (mg/L as $CaCO_3$), hardness, total solids (TS), total dissolved solids (TDS), potassium (K^+), PO_4^{-3} (mg/l), NO_3^- (mg/l), BOD (mg/l) and DO (mg/l). The performance of the ANFIS models was assessed through the correlation coefficient (R), mean squared error (MSE), mean absolute error (MAE) and Nash model efficiency (E). Study results show that the adaptive neuro-fuzzy inference system is able to predict the biochemical oxygen demand with reasonable accuracy, suggesting that the ANFIS model is a valuable tool for river water quality estimation. The result shows that, ANFIS-I has a high prediction capacity of BOD compared with ANFIS-II. The results also suggest that ANFIS method can be successfully applied to establish river water quality prediction model.

1. Introduction

Biochemical oxygen demand (BOD) is the amount of dissolved oxygen required for aerobic biological organisms in a water body to break the organic components available in a given water sample at certain temperature over a particular time period. The BOD is an approximate measure of the amount of biochemical degradable organic matter present in a water sample (Abyaneh, 2014). It is defined by the amount of oxygen required for the aerobic microorganisms present in the sample to oxidize the organic matter to a stable organic form (Chapman, 1992). Nutrients and light in the phytoplankton growth, the relationship between DO and phytoplankton concentrations and ammonia affect the BOD degradation (Lopes et al., 2005). Several water quality models such as traditional mechanistic approaches, artificial neural network have been successfully applied to accomplish the best practices for preserving the quality of water. Most of these models required several input parameters which are not easily accessible and

* Corresponding author at: Department of Civil Engineering, Leading University, Sylhet 3100, Bangladesh. Cell: +880 1912140065.
E-mail addresses: engr.masrur@outlook.com (A.A.M. Ahmed), mustakim.bd@gmail.com (S.M.A. Shah).
[1] Cell: +880 1717679169.

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksues.2015.02.001
1018-3639

Please cite this article in press as: Ahmed, A.A.M., Shah, S.M.A. Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. Journal of King Saud University – Engineering Sciences (2015), http://dx.doi.org/10.1016/j.jksues.2015.02.001

make it a very expensive and time consuming process (Suen et al., 2003). In recent years, artificial intelligence (AI) such as artificial neural network (ANN) and adaptive neuro fuzzy inference system (ANFIS) have become gradually popular for prediction and forecasting in a number of areas like water resources (Zhu and Fujita, 1994; Icaga, 2007; Dahiya et al., 2007; Lermontov et al., 2009) and environmental science. Areerachakul (2012) applied artificial neural networks and adaptive neuro fuzzy inference system model to estimate the biochemical oxygen demand (BOD) of Saen Saep canal in Bangkok. The five input parameters were namely dissolved oxygen (DO), chemical oxygen demand (COD), ammonia nitrogen, nitrate nitrogen, and total coliform bacteria (T-coliform). The experimental outcomes demonstrate that the artificial neural network model provided higher correlation coefficients ($R = 0.73$) and a lower mean square error (RMSE = 4.53) than the adaptive neuro-fuzzy inference system model ($R = 0.6768$; RMSE = 4.8182). Another study was carried out by Safavi (2012) where he applied the adaptive neuro-fuzzy inference system (ANFIS) for river quality predictions with an emphasis on DO and BOD for the Zayandehroud River. The revealed result shows that BOD predictions were obtained by the proposed system with a correlation coefficient of 0.953 in the calibration stage and 0.931 in the validation stage and DO predictions were obtained with a correlation coefficient of 0.921 in the calibration stage and 0.904 in the validation stage. Aslan (2008) applied the back-propagation ANN technique to predict DO in Eymir Lake located in Ankara using artificial neural networks and adaptive neuro fuzzy inference system. The results of this study indicated that it is possible to predict the dissolved oxygen concentrations in Lake Eymir using a limited number of input parameters. The results of ANFIS models were not as successful as expected. The main objective of this paper is to analyze and model BOD concentration by ANFIS modeling techniques of Surma River in Bangladesh.

2. Materials and methods

2.1. Study area and water quality data

The Surma River is located in the northeastern region of Bangladesh within the administrative districts of Sylhet and Sunamganj. The Surma River is the longest river that flows through Surma flood plain toward the downstream of Chattak. The Surma River has many tributaries, canals and hilly streams joining it from the Khasia–Jaintia hills of the Indian State of Meghalaya (Ahmed et al., 2010). The canals are normally originated from the hills of the city and tea gardens. The canals are responsible for surface runoff conveyance from its urban catchments to the receiving Surma and they crisscross Sylhet city and contribute the hydrographic position to the River (Alam et al., 2007). But these canals are being filled for housing, roads and urban infrastructure causing serious drainage congestion in these natural water flows. These canals are partly or fully clogged or extinct by filling activities of urbanization and siltation by the city's garbage and wastes.

Data sets of six monitoring stations (Fig. 1), comprising 10 water quality parameters monthly over 3 years (2010–2012), were obtained from the department of environment, Bangladesh. Total numbers of available observations are 120. The measured parameters are pH, alkalinity (mg/l as $CaCO_3$), hardness, total solids (TS), total dissolved solids (TDS), potassium (K^+), PO_4^{-3} (mg/l), NO_3^- (mg/l), BOD (mg/l) and DO (mg/l). In Table 1, the summarized basic statistics of the dataset is presented. The water quality parameters were tested using the standard methods developed by APHA–AWWA–WPCF (1998).

2.2. Data preparation and input selection

Normalization of the water quality parameters within a uniform range before applying them to the ANFIS model is important to prevent larger numbers from prevailing smaller ones. Normalization can be used to scale data in the same range of values for each input feature in order to minimize bias within the dataset for one feature to another. Initially all input data are normalized using the formula used by Wu et al. (2005) (Eq. (2)) but when the ANFIS model has been successfully executed, model outputs in the form of normalized values are converted to original values by inverse transformation using Eq. (3). At the beginning the standard deviation was calculated using Eq. (1)

$$\text{Std.} = \sqrt{\frac{\sum (X_i - \bar{X}_i)^2}{N-1}} \tag{1}$$

$$S_i = \frac{(X_i - \bar{X}_i)}{\text{Std.}} \tag{2}$$

$$Q_i = S_i * \text{Std.} + \bar{Q}_i \tag{3}$$

where, Std. = standard deviation of X_i; X_i = input data; $\bar{X}_i$ = arithmetic average of X; S_i = normalized X variable as input to ANN; N = total number of observations in the data set; Q_i = predicted value; S_i = normalized model output from ANN; and Q_i = arithmetic average of observed value in the set.While modeling BOD value for the Surma River, it is important to select the variables to have a positive relationship with one another. For this reason, a coefficient of correlation with BOD was prepared which is tabulated in Table 1. It is found that hardness, alkalinity, pH and DO have a significant relationship with BOD. Hence, they were selected as an input combination (ANFIS-II). Moreover, ANFIS-I contains all the water quality parameters tabulated in Table 2.

2.3. Adaptive neuro-fuzzy inference system (ANFIS)

Adaptive neuro fuzzy inference system (ANFIS) is a type of neural network that is focused on Takagi–Sugeno fuzzy inference system. ANFIS is a well-known artificial intelligence technique that has been used currently in hydrological processes (Bisht and Jangid, 2011). Several well-known neuro-fuzzy modeling algorithms are available in the literature, such as fuzzy inference networks, fuzzy aggregation networks, neural network-driven fuzzy reasoning, fuzzy modeling networks, fuzzy associative memory systems, etc. (Keller et al., 1992; Kosko, 1999). Adaptive neuro-fuzzy inference system (ANFIS), proposed by Jang (1993), is based on the first-order Sugeno fuzzy model. Generally, ANFIS uses either back-propagation or a combination of least square estimation and back-propagation for membership function parameter

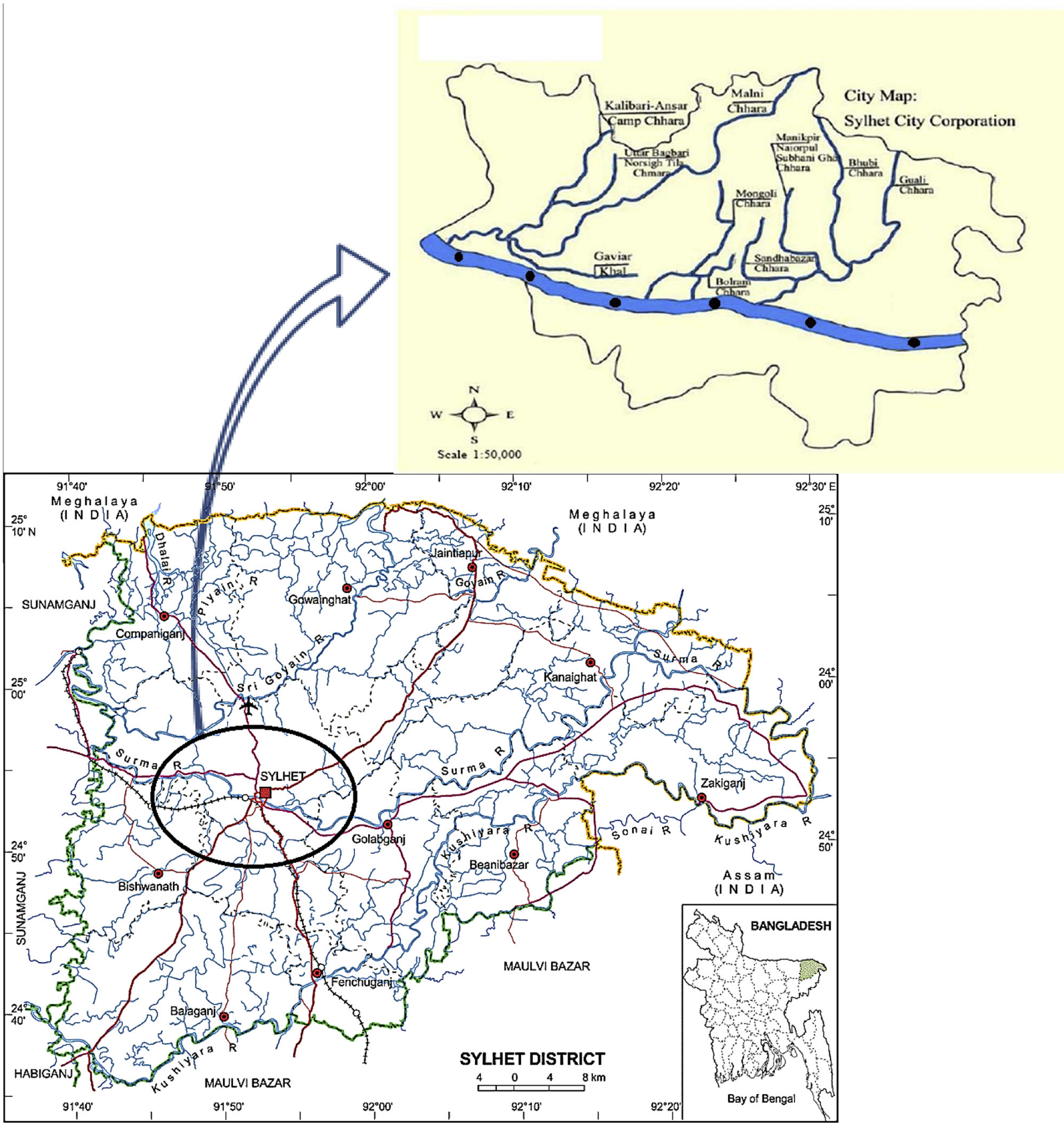

Figure 1 Sampling locations of water samples of Surma River.

estimation (Jang and Sun, 1997). The most significant goal of integrating fuzzy systems with neural networks is to practice learning ability of neural network while the learning capability is an improvement from the view point of a fuzzy system, from the perspective of a neural network there are additional benefits to a combined system. In ANFIS, Takagi–Sugeno type fuzzy inference system is used where the output of each rule can be a linear combination of input variables plus a constant term or can be only a constant term. The final output is the weighted average of every single rule's output. Basic ANFIS architecture that has two inputs x and y and one output z is shown in Fig. 2. The rule base contains two Takagi–Sugeno if-then rules as follows:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$
Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$

Yilmaz (2003) describes the typical ANFIS structure as follows:

The neural network structure contains 5 layers excluding the input layer (Layer 0):

(1) *Layer 0*, input layer, has n nodes where n is number of inputs to the system.
(2) *Layer 1* is the *fuzzification* layer in which each node represents a membership value to a linguistic term as a Gaussian function with the mean:

Table 1 Basic statistics of the measured water quality variables in the Surma River.

Variable	Unit	Minimum	Maximum	Mean	SD	CBOD
Nitrates	mg/l	0.17	4.0	1.54	1.06	0.109
Alkalinity	mg/l	24	194	60.22	30.00	−0.166
Hardness	mg/l	43	260	118.4	44.44	−0.116
TS	mg/l	51	937	295.4	163.2	0.045
TDS	mg/l	9.5	519	143	104.2	−0.043
pH	–	5.8	8.5	6.98	0.57	0.221
Turbidity	NTU	4.10	42.2	12.00	7.65	−0.103
K^+	mg/l	1.47	35.2	6.45	5.77	−0.098
DO	mg/l	1.80	17	5.44	2.38	0.727
BOD	mg/l	0.50	17	3.73	2.88	1

SD: standard deviation; CBOD: correlation with BOD.

Table 2 Data combination using correlation co-efficient between BOD and other variables.

Model name	Data combination
ANFIS-I	All variables
ANFIS-II	Hardness + alkalinity + pH + DO

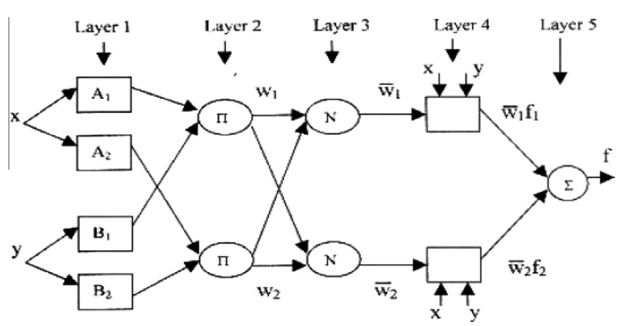

Figure 2 Basic structures of ANFIS.

$$\mu Ai(x) = \frac{1}{1 + \left[\frac{x - ci}{ai}\right]^{2bi}} \tag{4}$$

where ai, bi, ci are parameters for the function. These are adaptive parameters. Their values are adapted by means of the back-propagation algorithm during the learning stage. As the values of the parameters change, the membership function of the linguistic term Ai changes.

(3) In *Layer 2*, each node provides the strength of the rule by means of multiplication operator. It performs *min* (*AND*) operation. The membership values represented by μAi (x_0) and μBi (x_1) are multiplied in order to find the firing strength of a rule where the variable x_0 has a linguistic value of Ai, and xi has a linguistic value of Bi, in the antecedent part of Rule i. There are p^n nodes denoting the number of rules in Layer 2. Each node represents the antecedent part of rule (*if* part of an *if-then* rule). Here, n is the number of input variables and p is the number of membership functions.

$$wi = \mu Ai(x_0) * \mu Bi(x_1) \tag{5}$$

(4) *Layer 3* is the normalization layer which normalizes the strength of all rules according to the equation given below:

$$\bar{W}i = \frac{Wi}{\sum_{j=1}^{R} Wj} \tag{6}$$

where wi is the firing strength of the i^{th} rule which was computed in Layer 2. Node i computes the ratio of the ith rule's firing strength to the sum of all rule's firing strengths. There are p^n nodes in this layer.

(5) *Layer 4* is a layer of adaptive nodes. Every node in this layer computes a linear function where the function coefficients are adapted by using the error function of the multi-layer feed-forward neural network

$$\bar{W}ifi = \bar{W}i(p_0x_0 + p_1x_1 + p_2) \tag{7}$$

pi's are the parameters where $i = n + 1$ and n is the number of inputs to the system (i.e. number of nodes in Layer 0). This equation was given for two inputs. Finally $\bar{w}i$ is the output of Layer 3. These parameters are updated by a learning step. The least squares approximation is used in ANFIS. In a temporal model, back-propagation algorithm is used for training.

(6) *Layer 5* is the output layer whose function is the summation of net outputs of the nodes in Layer 4. The output is computed as

$$\sum_i \bar{w}ifi = \frac{\sum_i wifi}{\sum_i wi} \tag{8}$$

where $\bar{w}ifi$ is the output of node i in Layer 4. It denotes the consequent part of Rule i. The overall output of the neuro-fuzzy system is the summation of the rule consequents.

2.4. Evaluation of performance

After the training is complete, the performance of the ANFIS is determined. The performance of the forecasting of both the training and the testing sets is evaluated by the following measure of goodness-of-fit, mean squared error (MSE), mean absolute error (MAE), model efficiency (EFF) (Nash and Sutcliffe, 1970), and correlation coefficient (R) were performed. However, to activate the over fitting condition, the input arrays each consisting of 120 elements were separated into two parts randomly for training (70%) and testing (30%). Correlation coefficient is defined as the degree of correlation between the experimental and modeled values (Rankovic et al., 2012)

Please cite this article in press as: Ahmed, A.A.M., Shah, S.M.A. Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. Journal of King Saud University – Engineering Sciences (2015), http://dx.doi.org/10.1016/j.jksues.2015.02.001

$$R = \sum_{k=1}^{N}(y_k - \bar{y})(tk - \bar{z}) \Bigg/ \sqrt{\sum_{k=1}^{N}(y_k - \bar{y})^2 \sum_{k=1}^{N}(t_k - \bar{z})^2} \qquad (9)$$

Mean squared error (MSE) measures the average of the squares of the errors. The smaller values of MSE ensure the better performance. The MSE is calculated as:

$$\mathrm{MSE} = \frac{1}{N} * \sum_{k=1}^{N}(t_{k-}y_k)^2 \qquad (10)$$

The coefficient of efficiency (E) has been widely used to evaluate the performance of hydrologic models. Nash and Sutcliffe (1970) defined the coefficient of efficiency which ranges from minus infinity to 1.0, with higher values indicating a better agreement. It is calculated as:

$$E = \left(1 - \frac{\sum_{k=1}^{N}(t_k - y_k)^2}{\sum_{k=1}^{N}(tk - \bar{z})^2}\right) \qquad (11)$$

However, the performance of the model was also analyzed using mean absolute error (MAE).

$$\mathrm{MAE} = \frac{1}{N} * \sum_{k=1}^{N}|t_{k-}y_k| \qquad (12)$$

where, y_k and t_k denote the network output and measured value from the kth element; $\bar{y}$ and $\bar{z}$ denote their average respectively, and N represents the number of observations.

Table 3 The performance parameters of the ANFIS model for test, training and whole phase.

Model	Structure	Phase	MAE	MSE	EFF	R
ANFIS-I	09-15-1	Training	1.81	1.29	87.43	0.920
		Testing	1.01	1.67	83.12	0.885
		Whole	0.89	1.37	86.13	0.906
ANFIS-II	04-20-1	Training	2.16	1.97	85.85	0.915
		Testing	3.12	2.55	79.87	0.830
		Whole	1.23	2.64	73.23	0.842

3. Results and discussion

The ANFIS models that had a higher R value were screened out for training, testing, and whole array (training + testing) using various number of membership functions for different data combinations. The five performance parameters were calculated to evaluate the BOD prediction capability of the proposed models. Table 3 presents the correlation coefficient (R), Nash–Sutcliffe efficiency coefficient (E), mean absolute error (MAE), and mean squared error (MSE) for two different data combinations. Modeled and observed values of biochemical oxygen demand (BOD) are shown in Fig. 3. Note that although the horizontal axis represents time, it does not necessarily indicate equal time intervals, thus referred to as "order of data". Overall the ANN model performs quite well; and exceptionally well for some water quality parameters regardless of the watershed. The developed ANFIS model performs quite well at each data combinations for BOD prediction with *adequate* performance. The *ANFIS* model developed using all the variables work exceptionally well with a *high* coefficient of determination ranging from 0.842 to 0.920 which is tabulated in Table 3.

The scatter plot of observed versus modeled dissolved oxygen concentration for adaptive neuro fuzzy inference system have been illustrated in Fig. 4. Two types of ANFIS structures are established to evaluate the performance of the training and testing of the proposed model. The respective correlation coefficient for ANFIS-I model is 0.920, 0.885 and 0.906 for training, testing, and whole array. However, the MSE and MAE for the three data sets are 1.29 and 1.81 for training, 1.67 and 1.01 for testing, and 1.37 and 0.89 for whole set respectively. The Nash efficiency (E) was also evaluated and found as effective as other performance parameters (E = 0.874 for training; E = 0.831 for testing; E = 0.861 for

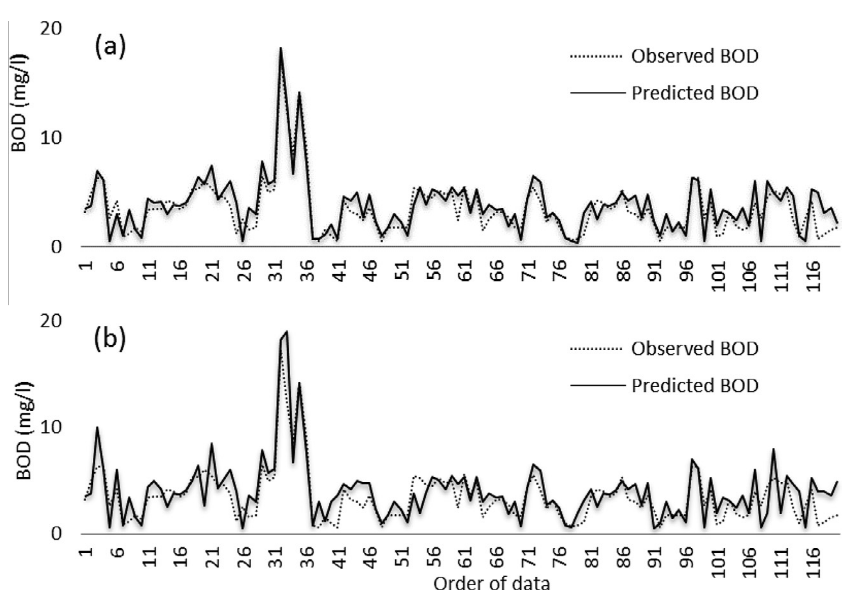

Figure 3 Comparison of the model computed and measured BOD levels in the Surma River water (a) ANFIS-I, (b) ANFIS-II.

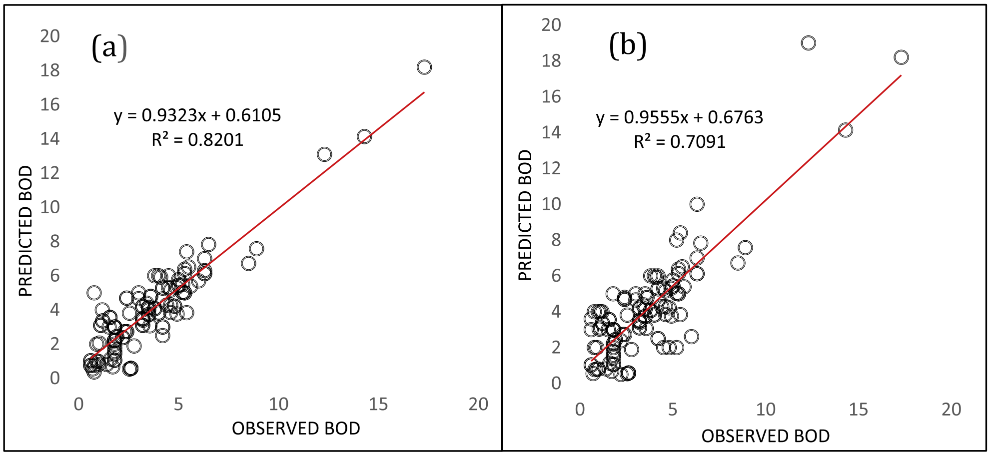

Figure 4 Scatter plot for the model predicted and observed BOD levels in the Surma River water (a) ANFIS-I, (b) ANFIS-II.

whole array). The ANFIS-II model is composed of Nitrates, TS, pH, DO and it gives a sufficient BOD prediction ability. The correlation coefficient (R) is found as 0.915, 0.830 and 0.842 for training, testing and whole array accordingly using ANFIS-II model. The respective values of MSE and MAE of ANFIS-II for the three data sets are 1.97 and 2.16 for training, and 2.55 and 3.12 for test, and 2.64 and 1.23 for the whole set. However, the model efficiency (E) for the test set was found below 0.80. Moreover, considering the correlation coefficient (R), the ANFIS-II model is not found as successful as ANFIS-I.

The achieved results of the study can be compared with some relevant studies discussed in the literature. The obtained result of the research identified that the proposed ANFIS models are capable to predict the water quality parameters utilizing limiting data sets.

5. Conclusion

From the research, it is found that using adaptive neuro-fuzzy inference system for water quality modeling could be successfully demonstrated. This may not necessarily mean that ANFIS always performs better than other conventional conceptual models; however, in this particular case the proposed ANFIS could be applied for biochemical oxygen demand forecasting with a reasonable performance. The ANFIS-I model with all the variables shows slightly better performance both in the training and testing periods than the ANFIS-II model. From the study, it is resolved that the developed ANFIS-I model is better than ANFIS-II model in prediction of biochemical oxygen demand (BOD).

References

Abyaneh, Z., 2014. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. J. Environ. Health Sci. Eng. 12 (40), 1–8.

Ahmed, M., Ahmed, A.A.M., Mazumder, R.K., 2010. Deterioration of water quality of Surma River influenced by natural canals passing through Sylhet City of Bangladesh. In: Proceedings of Intl. Conf. on Env. Aspects of Bangladesh Sept. 4. pp. 182–185.

Alam, M.J.B., Muyen, Z., Islam, M.R., Islam, S., Mamun, M., 2007. Water quality parameters along rivers. Int. J. Environ. Sci. Technol. 4 (1), 159–167.

APHA-AWWA-WPCF, 1989. Standard Methods for the Examination of Water and Wastewater. American Public Health Association, New York.

Areerachakul, S., 2012. Comparison of ANFIS and ANN for estimation of biochemical oxygen demand parameter in surface water. Int. J. Chem. Biol. Eng. 6, 286–290.

Aslan, M., 2008. Modeling the Water Quality of Lake Eymir using Artificial Neural Networks (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS) (M.Sc thesis). Submitted to the Department of Environmental Engineering, METU, Ankara.

Bisht, D.C.S., Jangid, A., 2011. Discharge modelling using adaptive neuro-fuzzy inference system. Int. J. Adv. Sci. Technol. 31, 99–114.

Chapman, D., 1992. Water Quality Assessments, first ed. Chapman and Hall Ltd., London, pp. 80–81.

Dahiya, S., Singh, B., Gaur, S., Garg, V.K., Kushwaha, H.S., 2007. Analysis of groundwater quality using fuzzy synthetic evaluation. J. Hazard. Mater. 147, 938–946.

Icaga, Y., 2007. Fuzzy evaluation of water quality classification. Ecol. Indic. 7, 710–718.

Jang, J.S.R., Sun, C.T., 1997. Neuro-Fuzzy and Soft Computing: A Computing Approach to Learning and Machine Intelligence. Prentice Hall, Englewood Cliffs, NJ.

Jang, S., 1993. Adaptive network-based Fuzzy Inference System. IEEE Journal 23 (3), 665–685.

Keller, J., Krishnapuram, R., Rhee, F.C.H., 1992. Evidence aggregation networks for fuzzy logic interface. IEEE Trans. Neural Networks 3 (5), 761–769.

Kosko, B., 1999. Fuzzy systems as universal approximators. In: Proc. IEEE Int. Conf. Fuzzy Sys. pp. 1153–1162.

Lermontov, A., Yokoyama, L., Lermontov, M., Machado, M.A.S., 2009. River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. Ecol. Indic. 9, 1188–1197.

Lopes, J.F., Dias, J.M., Cardoso, A.C., Silva, C.I.V., 2005. The water quality of the Ria de Aveiro lagoon, Portugal: from the observations to the implementation of a numerical model. Mar. Environ. Res. 60, 594–628.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, Part I, A discussion of principles. J. Hydrol. 10 (3), 282–290.

Rankovic, V., Radulovi, J., Radojevic, I., Ostojic, A., Comic, L., 2012. Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. Ecol. Modell. 221, 1239–1244.

Please cite this article in press as: Ahmed, A.A.M., Shah, S.M.A. Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. Journal of King Saud University – Engineering Sciences (2015), http://dx.doi.org/10.1016/j.jksues.2015.02.001

Safavi, H.R., 2012. Prediction of river water quality by adaptive neuro fuzzy inference system (ANFIS). J. Environ. Stud. 36 (1).

Suen, J.P., Eheart, J.W., Asce, M., 2003. Evaluation of neural networks for modelling nitrate concentration in rivers. J. Water Resour. Plann. Manage. 129, 505–510.

Wu, J., Han, J., Annambhotla, S., Bryant, S., 2005. Artificial neural networks for forecasting watershed runoff and stream flows. J. Hydrol. Eng. 10 (3), 216–222.

Yilmaz, N.A.S., 2003. A Temporal Neuro-fuzzy Approach for Time Series Analysis (PhD thesis paper). Department of Computer Engineering, The Middle East Technical University, pp. 34–35.

Zhu, M.L., Fujita, M., 1994. Comparison between fuzzy reasoning and neural networks methods to forecast runoff discharge. J. Hydrosci. Hydraul. Eng. 12 (2), 131–141.

Please cite this article in press as: Ahmed, A.A.M., Shah, S.M.A. Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. Journal of King Saud University – Engineering Sciences (2015), http://dx.doi.org/10.1016/j.jksues.2015.02.001