

Comparison of ANFIS and ANN for Estimation of Biochemical Oxygen Demand Parameter in Surface Water

S. Areerachakul

Abstract—Nowadays, several techniques such as; Fuzzy Inference System (FIS) and Neural Network (NN) are employed for developing of the predictive models to estimate parameters of water quality. The main objective of this study is to compare between the predictive ability of the Adaptive Neuro-Fuzzy Inference System (ANFIS) model and Artificial Neural Network (ANN) model to estimate the Biochemical Oxygen Demand (BOD) on data from 11 sampling sites of Saen Saep canal in Bangkok, Thailand. The data is obtained from the Department of Drainage and Sewerage, Bangkok Metropolitan Administration, during 2004-2011. The five parameters of water quality namely Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Ammonia Nitrogen (NH_3N), Nitrate Nitrogen (NO_3N), and Total Coliform bacteria (T-coliform) are used as the input of the models. These water quality indices affect the biochemical oxygen demand. The experimental results indicate that the ANN model provides a higher correlation coefficient ($R=0.73$) and a lower root mean square error ($\text{RMSE}=4.53$) than the corresponding ANFIS model.

Keywords—adaptive neuro-fuzzy inference system, artificial neural network, biochemical oxygen demand, surface water.

I. INTRODUCTION

WATER quality is one of the main characteristics of a river, even when its purpose is other than human water supply [12]. Water, in the sources such as canals and rivers, is not only utilized as human water supply, but it also used for various activities such as transportation, tourism, and consumption. Water quality in the superficial waters has started to degenerate as a result of wastewater being let go to the receiving ground and surface water without any control.

Biochemical Oxygen Demand (BOD) is an important parameter for usage conditions of surface waters. It is an approximate measure of the amount of biochemical degradable organic matter presented in a water sample. 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 [1]. Excessive BOD loads damage the quality of canal water, causing low Dissolved Oxygen (DO) concentration and unsuitable living conditions for flora and fauna within. At the same time, BOD-DO relationships include exchanges within the canal bed as well as nitrification and denitrification [2]. Nutrients and light in the phytoplankton growth, the relationship between DO and phytoplankton concentrations, and ammonia affect the BOD degradation [3].

DO levels, water temperature and nutrient levels (ammonia, nitrite, nitrate) are among the most critical factors for BOD in the canals. This process requires 5 days, with data collection and evaluation occurring on the last day. The oxygen consumption from degradation of organic material is normally measured as BOD and COD, so there is an important relation between them. Performing the test for BOD requires significant time and commitment for preparation and analysis. A test is used to measure the amount of oxygen consumed by these organisms during a specified period of time (usually 5 days at 20°C). This is referred to as a BOD_5 measurement [4]. Several water quality models such as Multiple Linear Regression (MLR) and 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 [5]. The ANN is a potential approach for water quality modeling. ANN is a mathematical model that attempt to simulate the functionality and decision-making processes of the human brain [6] and have become a center of interest across many scientific studies, including water quality. The main purpose of this study is to analyze and compare the performance of ANFIS and ANN models in predicting of biochemical oxygen demand of Saen Saep canal in Bangkok.

II. RELATED WORK

Modeling of water resource variables is a very active field of study and there have been still a lot of researches to be done in this area. In the initial stages, modeling of water resource variables was done using the traditional statistical models. In recent years, modern techniques have been proposed as efficient modeling tools. Hence, there is always an attempt to investigate the most efficient technique for a particular application. Gamal El-Din [7] applied Artificial Neural Networks (ANNs) to model wastewater treatment processes. This was a comparative study between conventional deterministic models and ANNs. They observed, in addition to the information contained in the conventional models, that neural networks contained a great deal of extra information with regard to the system being modeled. [16] analyzed and compared the performances of neural network (NN) with Multiple Linear Regression (MLR) models in predicting of biochemical oxygen demand of canals in Bangkok. Jain [8] employed neural networks to model the short-term water demand at the Indian Institute of Technology (IIT), Kanpur,

S. Areerachakul is with the Faculty of science and technology, Suan Sunandha Rajabhat University, Dusit, Bangkok 10300 Thailand (phone: 662-160-1111; e-mail: sirilak.ar@ssru.ac.th, sirilak.ar@gmail.com).

India. Six network models, five regression models and two time series models were developed and compared. All of the network models, generally, displayed better performance when measured against other models. Maier [9] conducted a study reviewing 43 research papers in which neural networks were utilized for prediction and forecasting of water resources variables. They observed that network models always work well. Their usages in the study of water are growing due to their ability to handle large amounts of non-linear, non-parametric data. Civelekoglu and Kitis [10] conducted a comparative study between MLR and ANNs in the prediction of bromate formation. There is evidence from the survey literature necessary to compare the performance of various approximation techniques. This is motivated by the fact that merits are given to various methods and as a result, it is difficult to tell which method will be more suitable for a particular application [11]. This study carries an element of novelty since it is the first one to carry out parameter of water quality prediction using computational intelligence techniques for Saen Saep canal in Bangkok, Thailand.

III. DATA AND SITE DESCRIPTIONS

Saen Saep canal consists of 11 sampling sites. This network of canal is important for the daily life of the people in Bangkok. This canal is used for consumption, transportation and recreation. Therefore, the rapid growths of industry, condominiums, high-rise and low-rise buildings, and other infrastructures, have had a significant effect on the canal water quality. Biochemical oxygen demand is an important parameter for the condition of surface waters. The estimation of biochemical oxygen demand results can be utilized in water management and treatment systems.

In this study, water quality data are provided by the Department of Drainage and Sewerage, Bangkok Metropolitan Administration, during 2004-2011. There are 828 records of data. Each record consists of 6 parameters, namely: Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Ammonia Nitrogen (NH_3N), Nitrate Nitrogen (NO_3N), Total Coliform bacteria (T-coliform) and Biochemical Oxygen Demand (BOD).

IV. METHODOLOGY

A. Artificial Neural Network

Artificial neural network models are generally comprised of three independent layers: input, hidden, and output. Each layer consists of several processing neurons. Each neuron in a layer operates in logical similarity. Information is transmitted from one layer to others in serial operations. The neurons in the input layer include the input values. Each neuron in the hidden layer processes the inputs into the neuron outputs. The pattern of hidden layers to be applied in the modeling can be either multiple layers or a single layer. The most widely used training algorithm for neural networks is the back-propagation algorithm [10].

Fig.1 shows the architecture of a Multilayer Perceptron (MLP) network. The MLP is one of artificial neural networks that are extensively used to solve a number of different problems, including pattern recognition and interpolation [13], [14]. Each layer is composed of neurons, which are interconnected with each other in a previous layer by weights. In each neuron, a specific mathematical function called the activation function accepts a weighted sum of the outputs from a previous layer as the function's input, and generates the function's output. In the experiment, the hyperbolic tangent sigmoid transfer function [15] is used as the activation function. It is defined by

$$f(s) = \frac{1 - e^{-2s}}{1 + e^{-2s}} \quad (1)$$

where $s = \sum_{i=1}^n w_i x_i + b$, in which w_i are weights, x_i are inputs of neuron, b is bias and n is the number of variables.

The MLP is trained by using the Levenberg–Marquardt technique. This technique is more powerful than the conventional gradient descent techniques [13].

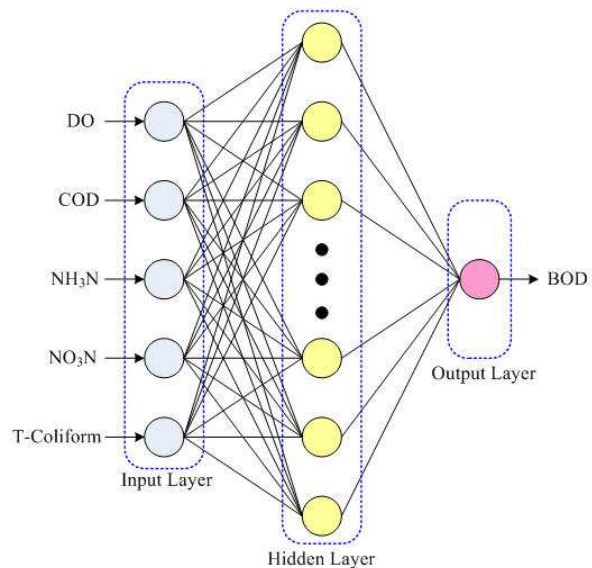


Fig. 1 The Architecture of Multilayer Perceptron Neural Network

B. Adaptive neuro-fuzzy inference system (ANFIS)

ANFIS is a method based on the input–output data of the system under consideration. Success in obtaining a reliable and robust ANFIS network depends heavily on the choice of process variables involved as well as the available data set and the domain used for training purposes [19]. Basically, a fuzzy inference system is composed of five function blocks [17]:

- (i) a rule base containing a number of fuzzy if-then rules.
- (ii) a database which defines the membership function of the fuzzy sets used in the fuzzy rules.

- (iii) a decision-making unit which perform the inference operation on the rules.
- (iv) a fuzzification inference which transforms the crisp inputs into degrees of match with linguistic values.
- (v) a defuzzification inference which transforms the fuzzy results of the inference into a crisp output.

For simplicity, a fuzzy inference system with two inputs x and y , and one output is assumed. For a first-order Sugeno fuzzy model, a common rule set with two fuzzy if-then rules is defines as follows:

Rules 1: If x is A_1 and y is B_1 , then $f_1 = p_1x + q_1y + r_1$ (2)

Rules 2: If x is A_2 and y is B_2 , then $f_2 = p_2x + q_2y + r_2$ (3)

Here type-3 fuzzy inference system proposed by Takagi and Sugeno [17] is used. In this inference system the output of each rule is a linear combination of input variables added by a constant term. The final output is the weighted average of each rule's output. The corresponding equivalent ANFIS structure is shown in Fig. 2.

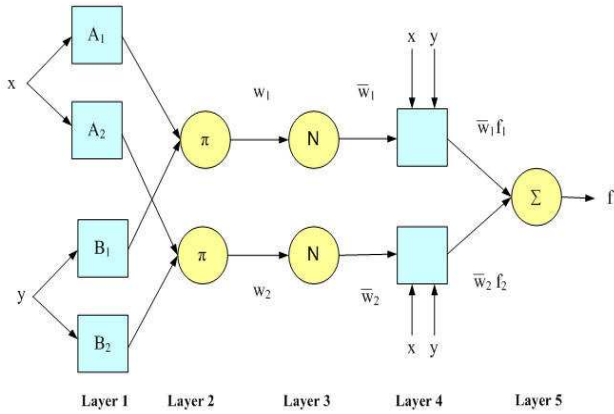


Fig. 2 A typical ANFIS architecture

Figure 2 shows a typical ANFIS architecture [17]. Every node in layer 1 is an adaptive node with a node function that may be a Gaussian membership function or any membership functions. Every node in layer 2 is a fixed node labeled Π , representing the firing strength of each rule. Every node in layer 3 is a fixed node labeled N , representing the normalized firing strength of each rule. Every node in Layer 4 is an adaptive node with a node function. The single node in layer 5 is a fixed node labeled Σ , indicating the overall output (Z) as the summation of all incoming signals [18]. In this study, Gaussian membership function is used for the input variable. The hybrid learning algorithm is employed to determine the parameters of Sugeno-type fuzzy inference systems. For a given training dataset, the combination of the least-squares method and the back-propagation gradient descent method is utilized to update FIS membership function parameters.

The performance of both ANFIS and NN models is assessed by evaluating the differences between the observed and predicted values by the correlation coefficient (R) and root

mean square error (RMSE). The correlation coefficient is defined as in equation (4) and the RMSE is calculated using equation (5):

$$R = \frac{\sum_{i=1}^N (t_p - \bar{t})(o_p - \bar{o})}{\sqrt{\sum_{i=1}^N (t_p - \bar{t})^2 \sum_{i=1}^N (o_p - \bar{o})^2}} \quad (4)$$

and

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_p - o_p)^2} \quad (5)$$

where N is amount of data, t_p is the observed value, o_p is the predicted value, \bar{t} and \bar{o} are the averages of the observed and predicted values respectively.

V. EXPERIMENT AND RESULTS

A. Data Preprocessing

At the initial stage of the experiment, data is scaled or normalized to within the range 0.1-0.9 using following equation (6):

$$x_{new} = 0.8 \frac{x - x_{min}}{x_{max} - x_{min}} + 0.1 \quad (6)$$

where x_{new} is the normalized value of an original parameter, x is the original data point, x_{min} and x_{max} are the minimum and maximum values in the data set, respectively. This normalized form [12] is chosen because it tends to provide a better outcome on the water quality application.

B. Artificial neural network model

The ratio of training to test data records employed in the experiment is 70:30. This means that with 828 data records, there are 580 records for the training set and 248 records for the test set.

For learning process, the input vectors and corresponding target vectors are used to train the neural networks by applying Levenberg-Marquardt algorithm. The number of hidden units directly affects the performance of the network. Therefore, many experimental investigations are conducted. The number of hidden nodes determined to provide the optimal result is 8.

Finally, the architecture of the network is 5-8-1. The number of input nodes is 5, representing the parameters of water quality that affect biochemical oxygen demand. The number of the hidden nodes is set to 8. The number of output nodes is 1, representing biochemical oxygen demand (BOD).

The stopping criteria for the training are: MSE below 0.002 or a number of epochs greater than 10000. Fig. 3 shows the plot comparing the observed (actual) values of BOD with the predicted BOD from the artificial neural network. The correlation coefficient, which measures the strength and

direction of the linear relation between two variables (actual and predicted values) is $R=0.7279$. The RMSE of the model is equal to 4.5348.

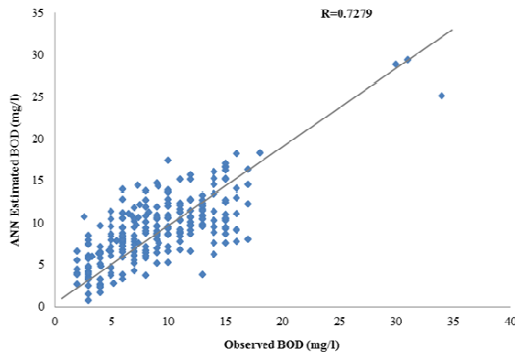


Fig. 3 Scatter plot between observed and estimated BOD values using the ANN model

C. Adaptive neuro-fuzzy inference system model

The training and test data sets for the adaptive neuro-fuzzy inference system model are the same as employed in the artificial neural network model. The ANFIS model performance based on the correlation coefficient and the root mean square error. ANFIS is trained with the help of MATLAB version 7.8 (2009). The parameter types and their values used in ANFIS model represent in Table I.

TABLE I
THE PARAMETER TYPES AND THEIR VALUES USED IN ANFIS MODEL

ANFIS parameters type	Value
Number of fuzzy rule	3
Number of linear parameters	18
Number of nonlinear parameters	30
Total number of parameter	48
Number of training data pair	580
Number of testing data pair	248
Number of node	44

In this study, the initial FIS and the number of the membership functions of each input and output variables are determined by the subtractive clustering algorithm [19]. The radius of regions is set to 0.25. By comparing the observed (actual) values of BOD with the predicted BOD from ANFIS model, the experimental result shows that the correlation coefficient and RMSE values are 0.6768 and 4.8182, respectively.

Fig. 4 shows the scatter plot between observed (actual) values and predicted BOD values from the ANFIS model. The correlation coefficient of the MLR equals to 0.6768. The RMSE of the model is 4.8182.

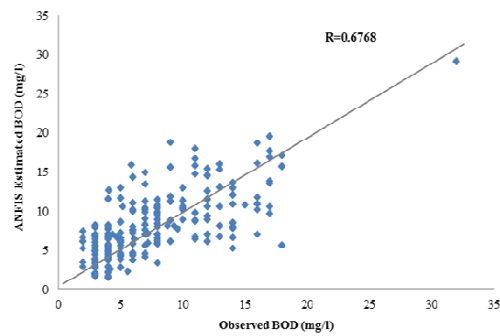


Fig. 4 Scatter plot between BOD observed values and estimated values using ANFIS model

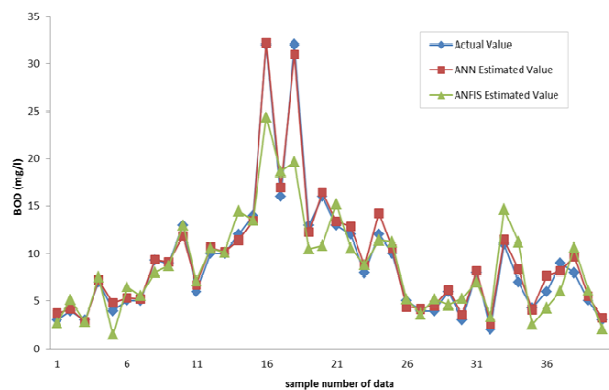


Fig. 5 Comparison of actual BOD value, estimated BOD from the ANN and estimated BOD from the ANFIS on select sample points

Fig. 5 demonstrates the sample points of actual BOD values compared with the estimated value from the neural network and adaptive neuro-fuzzy inference system models. The graph illustrates that predicted values from the artificial neural network is closer to the BOD actual value than that from the adaptive neuro-fuzzy inference system.

VI. DISCUSSION AND CONCLUSION

In this paper, we compare two models, artificial neural network and adaptive neuro-fuzzy inference system, for the estimation of biochemical oxygen demand of Saen Saep canal in Bangkok, Thailand. The experimental results show that the artificial neural network model provided higher correlation coefficients ($R=0.73$) and lower mean square error ($RMSE=4.53$) than the adaptive neuro-fuzzy inference system model. Biochemical oxygen demand is an important parameter for usage conditions of surface waters. Typically, this process requires 5 days, with data collection and evaluation occurring on the last day. The test is used to measure the amount of oxygen consumed by the organisms during a specified period of time (usually 5 days at 20 °C). This result may be applied to automate BOD estimations.

As a result, cost and time to analyst BOD could be minimized. However, it is recommended that other techniques and/or other statistical models should also be approached in the experiments.

ACKNOWLEDGMENT

The authors would like to thank Suan Sunandha Rajabhat University for scholarship support. Thanks to the Department of Drainage and Sewerage, Bangkok Metropolitan Administration, for the provided data.

REFERENCES

- [1] D. Chapman, "Water Quality Assessment," 1st ed. London: Chapman and Hall, 1992, pp. 80–81.
- [2] M. Radwan, P.Willems, and J.Berlamomnt, "Modeling of dissolved oxygen and biochemical oxygen demand in river water using a detailed and simplified model," 2nd ed. *River Basin Manage*, 2003, pp. 97–103.
- [3] J.F. Lopes, J.M Dias and A.C Cadoso, "The water quality of the Ria de Aveiro lagoon, Portugal: from the observations to the implementation of a numerical model," *Mar Environ Res*, pp.594-628.
- [4] G.C Delzer, S.W Mckenzie, "Five -day Biochemical Oxygen Demand," U.S .Geological Survey TWRI Book vol.9, 1999 pp.25-99.
- [5] J.P Suen, J.W Eheart and M.Asce, "Evaluation of neural networks for modeling nitrate concentration in rivers," *J Water Res Plan Manage*, pp.505-510.
- [6] P.A Aguilera, A.G Frenich and J.A Torres, "Application of the Konohen neural network in coastal water management: methodological development for the assessment and prediction of water quality," *Water Res*, 2001, pp.4053-4062.
- [7] A. Gamal El-Din, D.W Smith and M. Gamal El-Din, "Application of artificial neural networks in wastewater treatment," *J. Environ. Eng Sci.*, pp.81-95, Jan 2004.
- [8] A. Jain, A.K. Varshney and U.C. Joshi , "Short-term Water Demand Forecast Modeling ai IIT Kanpur Using Artificial Neural Networks," *IEEE Transactions on Water Resources Management*, vol. 15, no.1, pp.299-321, Aug 2001.
- [9] H.R. Maier and G.C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications," *Environmental Modeling and Software*, pp.101-124, Jan 2000.
- [10] G.Civelekoglu, N.O. Yigit, E.Diamadopoulos and M.Kitis, "Prediction of Bromate Formation Using Multi-Linear Regression and Artificial Neural Networks," *Journal of Ozone Science and Engineer*, Taylor&Francis, vol.5, no.5, pp.353-362, Sep 2007.
- [11] S.Msiza, V.Nelwamondo and T.Marwala, "Water Demand Prediction using Artificial Neural Networks and Support Vector Regression," *Journal of Computers*, vol. 3, no.11, Nov 2008.
- [12] Dogan, E., Sengorur, B., Koklu, Robia. :Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. :*Journal of Environmental Management* (2008)
- [13] Simon Haykin, "Neural Networks:A Comprehensive foundation second edition", Pearson Prentice Hall, Delhi India, 2005.
- [14] S.H.Musavi and M.Golabi "Application of Artificial Neural Networks in the River Water Quality Modeling: Karoon River,Iran", *Journal Of Applied Sciences*, Asian Network for Scientific Information, 2008, pp. 2324-2328.
- [15] L.Fausett, "Fundamentals of Neural Networks Architecture.Algorithms and Applications", Pearson Prentice Hall, USA, 1994.
- [16] S.Areerachakul and S.Sanguansintukul "A Comparison between the Multiple Linear Regression Model and Neural Networks for Biochemical Oxygen Demand Estimations", *The Eight International Symposium on Natural Language Processing*, Thailand, 2009.
- [17] Jang, J.S.R., 1993. ANFIS: "Adaptive-network-based fuzzy inference systems". *IEEE Trans. Syst. Man Cybern*, 23: 665-685.
- [18] Dastorani, M.T., A. Moghadamnia, J. Piri and M. Rico-Ramirez, "Application of ANN and ANFIS models for reconstructing missing flow data". *Environ. Monitoring Assess*, 2009.
- [19] M. Buragohain and C. Mahanta "A novel approach for ANFIS modelling based on full factorial design" *Applied Soft Computing* (2008), pp. 609–625.