Research Article

Water Quality Modeling in Reservoirs Using Multivariate Linear Regression and Two Neural Network Models

Wei-Bo Chen[1] and Wen-Cheng Liu[2,3]

[1]*National Science and Technology Center for Disaster Reduction, New Taipei City 23143, Taiwan*
[2]*Department of Civil and Disaster Prevention Engineering, National United University, Miaoli 36063, Taiwan*
[3]*Taiwan Typhoon and Flood Research Institute, National Applied Research Laboratories, Taipei 10093, Taiwan*

Correspondence should be addressed to Wen-Cheng Liu; wcliu@nuu.edu.tw

In this study, two artificial neural network models (i.e., a radial basis function neural network, RBFN, and an adaptive neurofuzzy inference system approach, ANFIS) and a multilinear regression (MLR) model were developed to simulate the DO, TP, Chl *a*, and SD in the Mingder Reservoir of central Taiwan. The input variables of the neural network and the MLR models were determined using linear regression. The performances were evaluated using the RBFN, ANFIS, and MLR models based on statistical errors, including the mean absolute error, the root mean square error, and the correlation coefficient, computed from the measured and the model-simulated DO, TP, Chl *a*, and SD values. The results indicate that the performance of the ANFIS model is superior to those of the MLR and RBFN models. The study results show that the neural network using the ANFIS model is suitable for simulating the water quality variables with reasonable accuracy, suggesting that the ANFIS model can be used as a valuable tool for reservoir management in Taiwan.

1. Introduction

Water is an important resource for the survival and health of humans and ecosystems. The quality of water is also crucial. The wording, water quality, can be used to address the condition in water column which included the characteristics of physical, chemical, and biological characteristics. Assessing the water quality variables is needed to develop best planning and management for water resources [1, 2].

The definition of eutrophication is the nutrient enrichment which means the excessive phosphorus and nitrogen loads in lakes and reservoirs, resulting in a serious problem. The natural or artificial enrichment of inland water bodies can cause eutrophication with algal blooms to deteriorate the water quality for human use and the decrease of dissolved oxygen levels, resulting in adverse effects on fisheries. The application of some modeling techniques to predict the behavior of enriched water bodies allows for combating adverse effects [3–5].

Due to the advanced computing capabilities, two-dimensional/three-dimensional reservoir hydrodynamic and water quality models have been developed and widely applied to resolve water quality problems [6–9]. Although deterministic models have been adopted for modeling water quality, these models require input data, model parameters, and extensive information to obtain results [10]. Because a large number of factors affecting the water quality have complicated nonlinear relationships with the variables, traditional deterministic models are not easy to handle.

Because of the existing difficulties and challenges in the simulating of water quality conditions using the hydrodynamic and water quality model, relatively novel computational approaches, artificial neural networks (ANNs), which have found wide acceptance in many disciplines, provide an alternative method for understanding and managing the water quality in reservoirs. ANNs are well-suited for this application because of their informative processing characteristics, such as nonlinearity, parallelism, noise tolerance, and learning and generalization capabilities [11, 12].

One of the successful applications of artificial intelligence techniques including knowledge-based systems, genetic
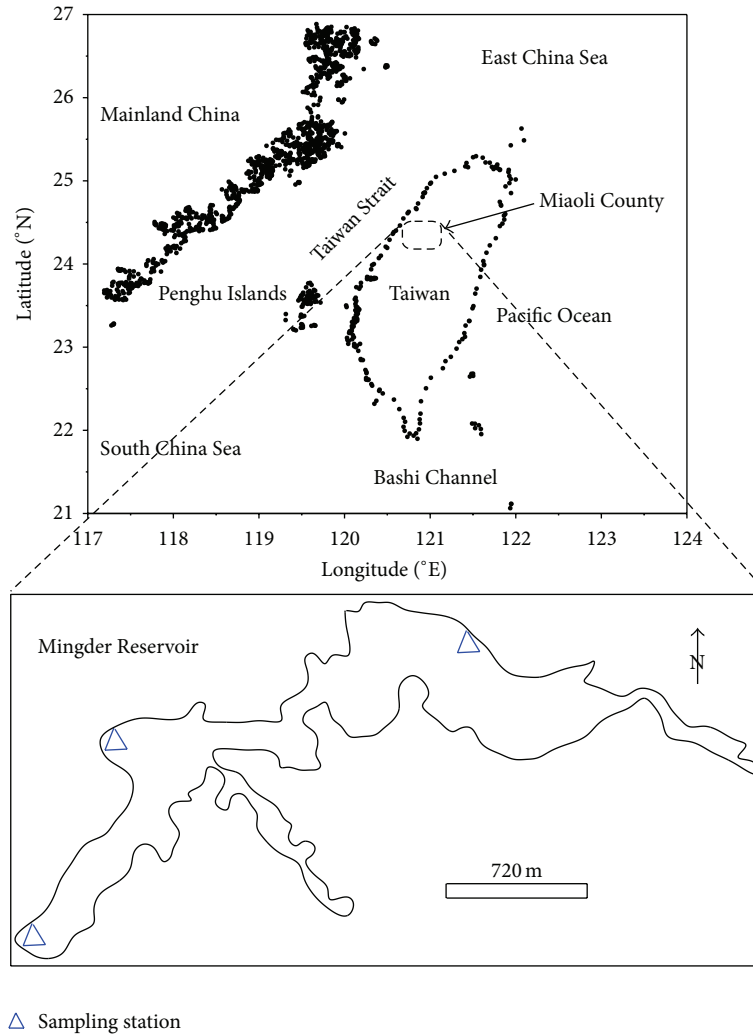
FIGURE 1: The Mingder Reservoir and water quality sampling stations.

algorithms, artificial neural networks, and fuzzy inference systems has been simulating complex nonlinear systems [13]. In particular, artificial neural networks have been successfully used as tools to simulate and predict water quality in water bodies [14–18]. The ANN model which is a black-box model does not require knowledge of any of the parameters [19]. ANNs have several advantages including learning ability, dealing with complex nonlinear data, and parallel processing ability. Prediction with ANN can be performed by the network learning experimentally generated data or using validated models [20].

Recently, Ranković et al. [21] developed a feed-forward neural network model to simulate the dissolved oxygen in the Gruza Reservoir in Serbia. Soltani et al. [22] combined a water quality simulation model and a hybrid genetic algorithm to determine the optimal operating polices for different reservoir outlets. The water quality simulation model is based on an adaptive neural fuzzy inference system that was trained using the results of a numerical water quality simulation model. Again, Ranković et al. [15] used an adaptive network-based fuzzy inference system model to simulate the dissolved oxygen in the Gruza Reservoir in Serbia. However,

evaluating the efficiency of multivariate linear regression and artificial neural network models to predict the water quality parameters in reservoirs has not been reported.

The main objective of this study is to establish a multivariate linear regression (MLR) model and two artificial neural network models, including a radial basis function neural network (RBFN) and an adaptive neurofuzzy inference system (ANFIS), to simulate the dissolved oxygen (DO), the total phosphorus (TP), the chlorophyll a (Chl a) content, and the Secchi disk depth (SD), which are commonly used as indicators of the eutrophication of reservoirs. A further aim is to demonstrate the application of these models to identify complex nonlinear relationships between the inputs and outputs in the Mingder Reservoir, Taiwan. The simulated and measured DO, TP, Chl a, and SD values were compared using the MLR, RBFN, and ANFIS models.

2. Materials and Methods

2.1. Description of Study Area and Data Collection. The Mingder Reservoir (Figure 1) is located in central Taiwan. The reservoir, completed in 1970, has a total watershed area
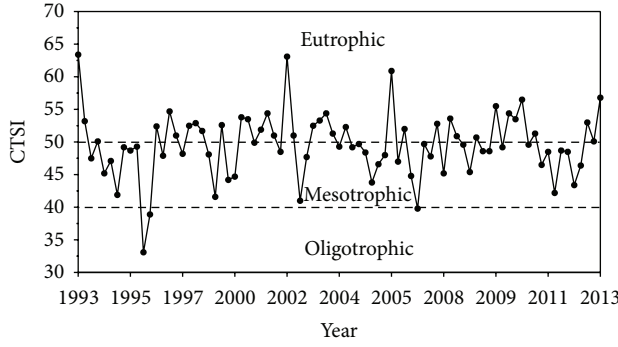
FIGURE 2: Carlson trophic state index (CTSI) for the Mingder Reservoir.

of $61 \, \text{km}^2$ and an effective storage volume of $1.65 \times 10^8 \, \text{m}^3$ and was built as a water supply source to Miaoli County for industrial supply, irrigation, and flood control. The dam site of the Mingder Reservoir is located at the tributary of the Houlong River and is approximately 7 km from Miaoli City.

The Mingder Reservoir is an in-channel reservoir that was formed by the construction of an earth dam and has a surface area of $1.62 \, \text{km}^2$. Its water quality has been routinely monitored over the past two decades, once per season since 1993. There are three water quality sampling stations in the reservoir.

Carlson [23] proposed a trophic state index (CTSI) related to the diverse aspects of the trophic state found in multiparameter indices. The CTSI has the simplicity of a single parameter index for the water quality assessment of impounded water bodies. The CTSI has been adopted by the Taiwan Environmental Protection Administration (TEPA) to determine eutrophication status. It is calculated with three variables: the Secchi disk depth (SD), the total phosphorus (TP), and the chlorophyll a concentration (Chl-a). For this reason, these three water quality variables were predicted using three approaches presented in this study.

Figure 2 shows the CTSI values of the Mingder Reservoir from 1993 to 2013. A CTSI value above 50 indicates that the water body is eutrophic, as shown in Figure 2 for the Mingder Reservoir. This figure shows that the reservoir is in a status between mesotrophic and eutrophic. Agricultural activities are the major sources of nutrients and provide excessive nutrient loads to the reservoir.

The historical water quality data, including the water temperature, pH, electrical conductivity, turbidity, suspended solids, total hardness, total alkalinity, Secchi disk depth, dissolved oxygen, chlorophyll a concentration, total phosphorus, nitrate nitrogen, and biochemical oxygen demand from 1993 to 2013, were collected from the TEPA and analyzed. A statistical summary of the water quality variables is shown in Table 1.

2.2. Adaptive Neurofuzzy Inference System (ANFIS). The ANFIS is a fuzzy Sugeno model with multilayer feed-forward network which potentially captures the benefits of artificial neural networks and fuzzy logic approaches in a single framework. ANFIS provides easy learning and adaptation

since it is a framework of adaptive systems. The Sugeno system is the most commonly adapted fuzzy model for ANFIS framework to deal with the complicated nonlinear problems because of less computational time [24].

To easily understand the architecture of ANFIS, we assume that the fuzzy inference system has two inputs, x_1 and x_2, and only one output, y. Therefore two fuzzy if-then rules based on a first-order Sugeno fuzzy model is provide as below [25]:

Rule 1: if x_1 is A_1 and x_2 is B_1, then $y_1 = p_1 x_1 + q_1 x_2 + r_1$,

Rule 2: if x_1 is A_2 and x_2 is B_2, then $y_2 = p_2 x_1 + q_2 x_2 + r_2$,

where A_i and B_i are the fuzzy sets and p_i, q_i, and r_i are the parameters that will be determined during the training and testing processes. The architecture of ANFIS with two rules is shown in Figure 3, in which circles represent fixed nodes and squares indicate adaptive nodes. A brief introduction of the ANFIS model follows.

Layer 1. Each node i in this fuzzy layer is an adaptive node with a node function

$$O_{1,i} = \mu_{A_i}(x_1), \quad i = 1, 2,$$
$$O_{1,i} = \mu_{B_{i-2}}(x_2), \quad i = 3, 4, \tag{1}$$

where x_1 and x_2 are the inputs to node i; A_i and B_i are the linguistic labels; and μ_{A_i} and $\mu_{B_{i-2}}$ are the membership functions for the A_i and B_i linguistic labels, respectively. Although many membership functions such as the trapezoidal membership function, the Gaussian membership function, the Gaussian combination membership function, the spline-based membership function, and the sigmoidal membership function can be used. Chau [26] investigated the effect of different membership functions on the performance of ANFIS model and found that the differences are insignificant. In this study, we adopted the generalized bell-shaped membership function which is a commonly used function [26]:

$$\mu_{A_i} = \frac{1}{1 + \left| (x_1 - c_i) / a_i \right|^{2b_i}},$$
$$\mu_{B_{i-2}} = \frac{1}{1 + \left| (x_2 - c_i) / a_i \right|^{2b_i}}, \tag{2}$$

where a_i, b_i, and c_i are the parameter sets. The parameters in this layer are the premise parameters.

Layer 2. Each node in this layer is a fixed node which is marked with circle and labeled with M in Figure 3. The outputs of this layer, which are called the firing strengths ($O_{2,i}$), are the products of the corresponding degrees obtained from Layer 1 (the input layer):

$$O_{2,i} = w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_1), \quad i = 1, 2. \tag{3}$$

Layer 3. Every node in this layer is a circle node labeled N (Figure 3). The third layer contains fixed nodes that calculate

Table 1: The statistic summary of the water quality variables in the Mingder Reservoir.

Variables	Minimum	Maximum	Mean	St. Dev.	C. V
Water temperature (°C)	13.70	33.60	23.46	5.18	0.22
pH	5.28	9.50	8.11	0.67	0.08
Electrical conductivity (μmho/cm 25°C)	78.00	447.00	234.74	52.33	0.22
Turbidity (NTU)	0.87	750.00	13.64	58.63	4.30
Suspended solids (mg/L)	1.48	438.00	10.49	33.31	3.18
Total hardness (mg/L)	1.20	190.00	96.67	22.70	0.23
Total alkalinity (mg/L)	29.40	174.00	90.80	22.40	0.25
Secchi disk depth (m)	0.15	3.40	1.50	0.52	0.35
Dissolved oxygen (mg/L)	0.00	14.80	7.68	3.15	0.41
Chlorophyll a (μg/L)	0.10	52.80	7.96	6.84	0.86
Total phosphorus (μg/L)	3.00	191.00	22.01	16.97	0.77
Nitrate nitrogen (mg/L)	0.01	1.90	0.12	0.21	1.69
Biochemical oxygen demand (mg/L)	0.69	50.98	8.67	5.78	0.67

St. Dev.: standard deviation.
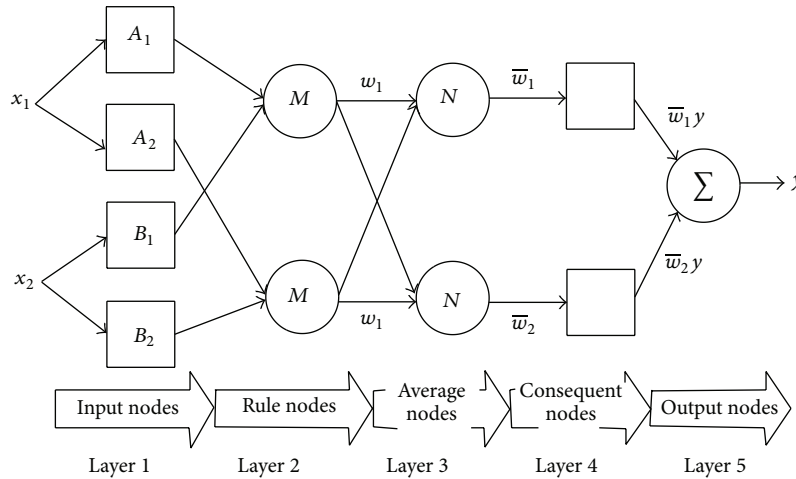C. V.: coefficient of variation (St. Dev./Mean).

Figure 3: Architecture of the adaptive network-based fuzzy interface system (ANFIS).

the ratio of the firing strengths of the ith rule to the sum of all rules' firing strength:

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \tag{4}$$

Layer 4. The nodes in this layer are adaptive and adjustable. The output of each node is the product of normalized firing strength, $\overline{w}_i$, from the third layer:

$$O_{4,i} = \overline{w}_i y_i = \overline{w}_i \left(p_i x + q_i y + r_i \right), \quad i = 1, 2, \tag{5}$$

where p_i, q_i, and r_i in this layer are consequent parameters.

Layer 5. The single node computes the overall output by summing all of the incoming signals:

$$O_{5,1} = \sum_{i=1}^{2} \overline{w}_i y_i = \frac{\sum_{i=1}^{2} w_i y}{w_1 + w_2}. \tag{6}$$

The details and mathematical background for these algorithms can be found in Jang [24].

2.3. Radial Basis Function Neural Network (RBFN). A radial basis function neural network (RBFN) contains a feedforward structure including one input layer, a single hidden layer, and one output layer, as shown in Figure 4. The concept of the RBFN is to establish a radial basis function (ϕ) and to determine the relationship between the input and the output using curve-fitting approaches [27, 28]. A single output RBFN with m hidden layer neurons can be described as

$$y = \sum_{i=1}^{m} \omega_i \phi_i (x), \tag{7}$$

where x is the input vector; ω_i is the connecting weights between the hidden layer and the output layer; and $\phi_i(x)$

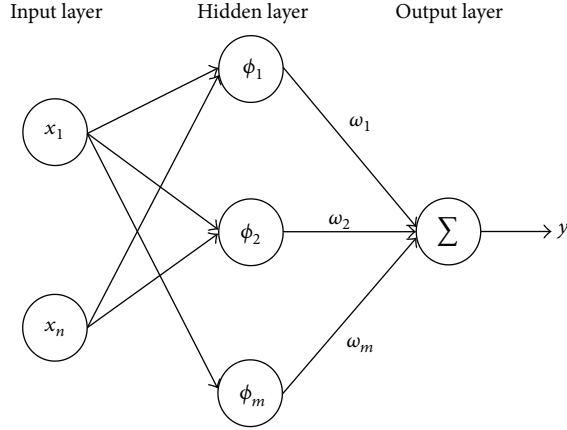Input layer   Hidden layer   Output layer

Figure 4: Architecture of the radial basis function neural network (RBFN).

is the output value of the neuron in the hidden layer after transfer by the radial basis function. Many radial basis functions can be used in RBFN model. The most common radial function is the Gaussian function [29] which is adopted in the present study:

$$\phi_i(x) = \exp\left(-\frac{\|x - c_i\|}{2\sigma^2}\right), \qquad (8)$$

where c_i is center vector of the ith neuron in the hidden layer; σ is defined as $\sigma = d_{max}/\sqrt{nc}$, d_{max} is the maximum distance among c_i and nc is the number of the center vectors; and $\|x - c_i\|$ denotes the Euclidean distance between x and c_i. The approach used to determine nc is the orthogonal least squares (OLS) approach developed originally by Chen et al. [30] and improved by Ham and Kostanic [31] and Kecman [32]. The simplest way to select RBFN centers is random method [33]. However, the OLS algorithm provides an optimum number of centers (nc) in RBFN model from the training patterns [34].

2.4. Multilinear Regression. Multilinear regression (MLR) is used to model the linear relationship between a dependent variable and one or more independent variables. MLR is based on least squares: the model is fit such that the sum of the squares of the differences between the observed and simulated values is minimized [35]. The model expresses the value of a predicted variable as a linear function of one or more predictor variables:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_i x_i, \qquad (9)$$

where x_i is the value of the ith predictor, b_0 is the regression constant, and b_i is the coefficient of the ith predictor.

2.5. Indices of Simulation Performance. Many statistical indexes can be used to evaluate the performance of models. The most common indexes for evaluating the performance of ANN models are the root mean square error (RMSE) and the correlation coefficient (R). To strictly evaluate the performance of MLR, RBFN model, and ANFIS model, three

criteria, mean absolute error (MAE), RMSE, and R, are employed in this study. These criteria can be defined by the following equations:

$$\text{MAE} = \frac{1}{N}\sum_N \left|(C_p)_N - (C_m)_N\right|,$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_N \left[(C_p)_N - (C_m)_N\right]^2},$$

$$R = \frac{\sum_N \left[(C_p)_N - \overline{C_p}\right]\left[(C_m)_N - \overline{C_m}\right]}{\sqrt{\sum_N \left[(C_p)_N - \overline{C_p}\right]^2 \sum_N \left[(C_m)_N - \overline{C_m}\right]^2}},$$

$$(10)$$

where N is the total number of data points, C_p is a simulated water quality variable, C_m is a measured water quality variable, and $\overline{C_p}$ and $\overline{C_m}$ denote the average simulated and measured water quality variables, respectively.

3. Results and Discussion

3.1. Selecting the Input Variables. The selection of an appropriate set of input variables for the MLR, RBFN, and ANFIS models is important for predicting the water quality variables in reservoirs [36]. It is difficult to determine how many input variables and which input variables should be adopted in the MLR, RBFN, and ANFIS models. The straight and useful way is to find the correlation coefficient between the individual dependent variables and water quality (DO, TP, Chl a, and SD). A total of 389 data sets collected from 1993 to 2013 were used for the analyses. Table 2 shows the correlation coefficient (R) between the individual dependent variables and DO, TP, Chl a, and SD, respectively. A correlation coefficient greater than 0.1 was set as the threshold for selecting the input variables. It indicates that the most effective inputs to affect DO are the pH value, chlorophyll a concentration, water temperature, total phosphorus, nitrate nitrogen, and biochemical oxygen demand. The most important water quality variables that affect the TP are the turbidity, suspended solids, pH value, dissolved oxygen, electrical conductivity, and nitrate nitrogen. It also illustrates that the most important input variables that affect Chl a and SD are, respectively, the six variables of the pH value, dissolved oxygen, water temperature, biochemical oxygen demand, nitrate nitrogen, and total hardness, and the six variables of the suspended solids, electrical conductivity, turbidity, total hardness, total alkalinity, and total phosphorus.

3.2. Water Quality Prediction Using the Multilinear Regression Model. The multilinear regression models were used to predict DO, TP, Chl a, and SD. The multilinear regression models in (11) to (14) were obtained from the input water quality variables:

$$\begin{aligned} \text{DO} = &\ 2.8083 \times \text{pH} + 0.0887 \times \text{Chl } a - 0.074 \\ &\times \text{WT} - 0.0045 \times \text{TP} - 0.4928 \times \text{NO}_3 \\ &- 0.0055 \times \text{BOD} - 13.851, \end{aligned} \qquad (11)$$

TABLE 2: The correlation coefficient (R) between individual dependent variable and DO, TP, Chl a, and SD.

Variables	R value between dependent variable and DO	R value between dependent variable and TP	R value between dependent variable and Chl a	R value between dependent variable and SD
Water temperature (WT)	0.2025	0.0861	0.3642	0.0571
pH	0.6339	0.2401	0.4090	0.0765
Electrical conductivity (EC)	0.0890	0.1769	0.0938	0.2164
Turbidity (TB)	0.0715	0.3505	0.0003	0.2140
Suspended solids (SS)	0.0812	0.3505	0.0094	0.2394
Total hardness (TH)	0.0289	0.0145	0.1051	0.1989
Total alkalinity (TA)	0.0250	0.0239	0.0488	0.1707
Secchi disk depth (SD)	0.0062	0.0982	0.0916	—
Dissolved oxygen (DO)	—	0.1917	0.3963	0.0062
Chlorophyll a (Chl a)	0.3963	0.0453	—	0.0916
Total phosphorus (TP)	0.1917	—	0.0453	0.1080
Nitrate nitrogen (NO_3)	0.1763	0.1740	0.1410	0.0978
Biochemical oxygen demand (BOD)	0.1199	0.0507	0.1777	0.0528

$$TP = 0.039 \times TB + 0.07 \times SS - 4.4717 \times pH$$
$$- 0.2463 \times DO - 0.0553 \times EC + 70.2402, \quad (12)$$

$$Chl\ a = 1.1423 \times pH + 0.5528 \times DO + 0.3228 \times WT$$
$$+ 0.1079 \times BOD - 1.7734 \times NO_3 \quad (13)$$
$$- 0.0103 \times TH - 12.8466,$$

$$SD = -0.0119 \times SS - 0.0025 \times EC + 0.0042 \times TB$$
$$- 0.0016 \times TH - 0.0010 \times TA - 0.0017 \quad (14)$$
$$\times TP + 2.4449,$$

where BOD is the biochemical oxygen demand (mg/L); EC is the electrical conductivity (μmho/cm 25°C); NO_3 is the nitrate nitrogen (mg/L); pH is the pH value; SS is the suspended solids (mg/L); TA is the total alkalinity (mg/L); TB is the turbidity (NTU); TH is the total hardness (mg/L); and WT is the water temperature (°C).

Equation (11) was used to predict the DO concentration for the training and testing phases. 272 and 117 data sets were used for training and testing the MLR model, respectively. Table 3 shows the performance evaluation using the MLR model for DO, TP, Chl a, and SD. The MAE, RMSE, and R values for the DO training phase were 1.81 mg/L, 2.33 mg/L, and 0.67, and these values for the DO testing phase were 1.81 mg/L, 2.41 mg/L, and 0.64, respectively. The MAE, RMSE, and R values for the TP training phase were 9.82 μg/L, 15.29 μg/L, and 0.55, and these values for the TP testing phase were 9.99 μg/L, 14.70 μg/L, and 0.31, respectively. The MAE, RMSE, and R values for the Chl a training phase were 3.96 μg/L, 5.85 μg/L, and 0.50, and these values for the Chl a testing phase were 4.18 μg/L, 5.92 μg/L, and 0.55, respectively. The MAE, RMSE, and R values for the SD training phase were 0.34 m, 0.45 m, and 0.43, and these values for the SD testing phase were 0.37 m, 0.53 m, and 0.35, respectively.

3.3. Water Quality Prediction Using the RBFN Model.

Due to the poor performance using the multilinear regression model, the RBFN model was applied to predict the DO, TP, Chl a, and SD in the reservoir. The same data sets adopted in the multilinear regression model were used for training and testing the RBFN model.

Table 4 shows the performance evaluation for the DO, TP, Chl a, and SD during the training and testing phases. The MAE, RMSE, and R values for the DO during the training phase were 1.49 mg/L, 1.97 mg/L, and 0.77, and these values during the testing phase were 1.45 mg/L, 1.96 mg/L, and 0.79, respectively. The MAE, RMSE, and R values for the TP during the training phase were 8.89 μg/L, 11.77 μg/L, and 0.75, and these values during the testing phase were 7.95 μg/L, 9.91 μg/L, and 0.73, respectively. The MAE, RMSE, and R values for Chl a during the training phase were 3.09 μg/L, 4.39 μg/L, and 0.79, and these values during the testing phase were 2.86 μg/L, 4.08 μg/L, and 0.75, respectively. The MAE, RMSE, and R values for SD during the training phase were 0.29 m, 0.39 m, and 0.68, and these values during the testing phase were 0.23 m, 0.31 m, and 0.76, respectively. The performances for the DO, TP, Chl a, and SD using the RBFN model were better than using the MLR model during the training and testing phases.

Some specific parameters may cause significant changes in ANN model [37]. In order to avoid overfitting or underfitting, we repeated ANN model until the minimum of errors occurred. Through the reiterative runs, the sum of error reduction ratio in RBFN model is set to 0.85 for providing the best performances. However, the statistical errors for the R value were still lower than 0.8 using the RBFN model, which means that the RBFN model is not good enough for predicting the water quality variables.

3.4. Water Quality Prediction Using the ANFIS Model.

An alternative approach, the ANFIS model, was used to predict the water quality variables. The network was trained using

TABLE 3: Performance evaluation during the training and testing phases for multilinear regression model.

Water quality variable	Training phase			Testing phase		
	MAE	RMSE	R	MAE	RMSE	R
DO	1.81 mg/L	2.33 mg/L	0.67	1.81 mg/L	2.41 mg/L	0.64
TP	9.82 μg/L	15.29 μg/L	0.55	9.99 μg/L	14.70 μg/L	0.31
Chlorophyll a	3.96 μg/L	5.85 μg/L	0.50	4.18 μg/L	5.92 μg/L	0.55
SD	0.34 m	0.45 m	0.43	0.37 m	0.53 m	0.36

TABLE 4: Performance evaluation during the training and testing phases for RBFN model.

Water quality variable	Training phase			Testing phase		
	MAE	RMSE	R	MAE	RMSE	R
DO	1.49 mg/L	1.97 mg/L	0.77	1.45 mg/L	1.96 mg/L	0.79
TP	8.89 μg/L	11.77 μg/L	0.75	7.95 μg/L	9.91 μg/L	0.73
Chl a	3.09 μg/L	4.39 μg/L	0.79	2.86 μg/L	4.08 μg/L	0.75
SD	0.29 m	0.39 m	0.68	0.23 m	0.31 m	0.76

TABLE 5: Performance evaluation during the training and testing phases for ANFIS model.

Water quality variable	Training phase			Testing phase		
	MAE	RMSE	R	MAE	RMSE	R
DO	1.30 mg/L	1.74 mg/L	0.85	0.92 mg/L	1.32 mg/L	0.88
TP	6.45 μg/L	9.92 μg/L	0.86	5.44 μg/L	7.42 μg/L	0.86
Chl a	3.19 μg/L	4.67 μg/L	0.77	2.25 μg/L	3.05 μg/L	0.83
SD	0.21 m	0.31 m	0.80	0.15 m	0.24 m	0.89

the training data set (i.e., 272 data sets), and then it was tested with the testing data set (i.e., 117 data sets). Figures 5 and 6 show the comparison of the model-predicted DO and model-measured DO for the training and testing phases and scatter plots, respectively. The error values, representing the predicted DO minus the measured DO, are also shown in Figure 5. The MAE, RMSE, and R values for the training phase were 1.30 mg/L, 1.74 mg/L, and 0.85, while these values for the testing phase were 0.92 mg/L, 1.32 mg/L, and 0.88, respectively, as shown in Table 5. Figures 7 and 8 compare the model-predicted and model-measured TP concentrations for the training and testing phases and scatter plots, respectively. The MAE, RMSE, and R values for the training phase were 6.45 μg/L, 9.22 μg/L, and 0.86, while these values for the testing phase were 5.44 μg/L, 7.42 μg/L, and 0.86, respectively. Number of membership functions of ANFIS model is specified as 2 to obtain the best performances.

Figures 9 and 10 compare the model-predicted and model-measured chlorophyll a concentrations for the training and testing phases and scatter plots, respectively. The MAE, RMSE, and R values for the training phase were 3.19 μg/L, 4.67 μg/L, and 0.77, and these values for the testing phase were 3.25 μg/L, 3.05 μg/L, and 0.83, respectively. The performance evaluation for predicting the Chl a concentration using the ANFIS model during the training phase is slightly inferior to that using RBFN model, but the performance evaluation for predicting the Chl a concentration using the ANFIS model during testing phase is superior to that using RBFN model (Table 4). This result may be the reason that the predicted Chl a concentration using the ANFIS

model underestimated the high Chl a concentration in the observational data during the training phase.

Figures 11 and 12 illustrate the comparison of the model-predicted and model-measured Secchi disk depth for the training and testing phases and scatter plots, respectively. The MAE, RMSE, and R values for the training phase were 0.21 m, 0.31 m, and 0.80, while these values for the testing phase were 0.15 m, 0.24 m, and 0.89, respectively. Overall, the performances of the DO, TP, Chl a, and SD with the ANFIS model are better than those with the multilinear regression model and the RBFN model. The performance evaluations with ANFIS during the testing phase are superior to those during the training phase. A similar result is also found in Ranković et al. [15].

Akkoyunlu et al. [14] applied two ANN models and an MLR model to estimate the DO concentration in Lake Iznik, Turkey. They found that the MLR model performed less accurately to predict the DO. In the current study, the MLR model not only is less accurate for predicting the DO concentration but also does not simulate the TP, Chl a, or SD well.

3.5. Discussion. In general, the neural network models provide better performance than the MLR model in this study. The MLR model is still a useful tool for simple and fast predicting water quality, although the predicted results by MLR model are of insufficient accuracy. For instance, Chenini and Khemiri [38] adopted MLR approach to evaluate the ground water quality in Maknassy Basin, central Tunisia. Thoe and Lee [39] used MLR to forecast the daily water quality of Hong Kong Beach.
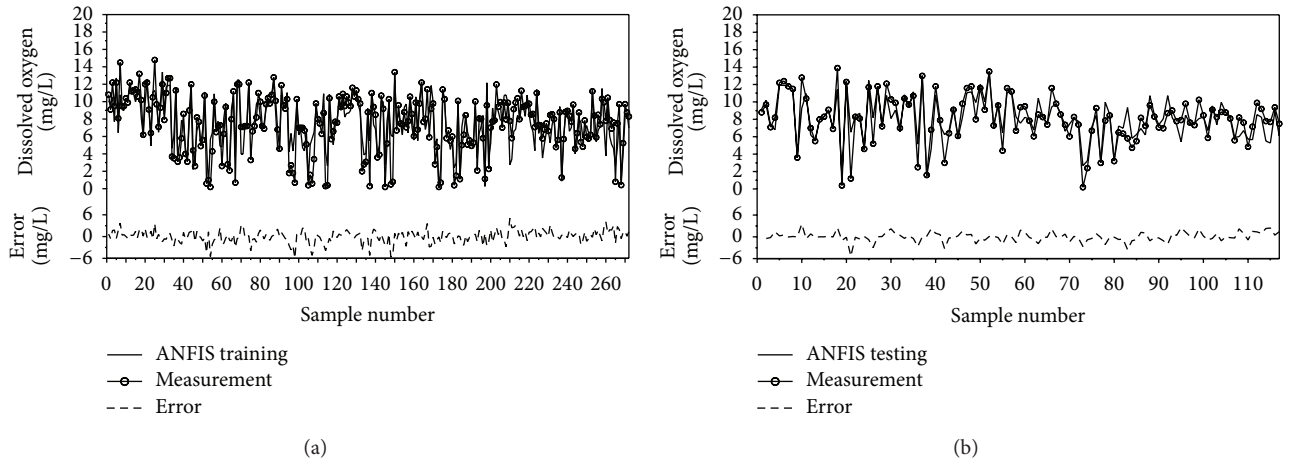
(a)   (b)

FIGURE 5: Comparison of predicted and measured dissolved oxygen using ANFIS model for the (a) training phase and (b) testing phase.

$R = 0.85$
MAE = 1.30 mg/L
RMSE = 1.74 mg/L

$R = 0.88$
MAE = 0.92 mg/L
RMSE = 1.32 mg/L

(a)   (b)
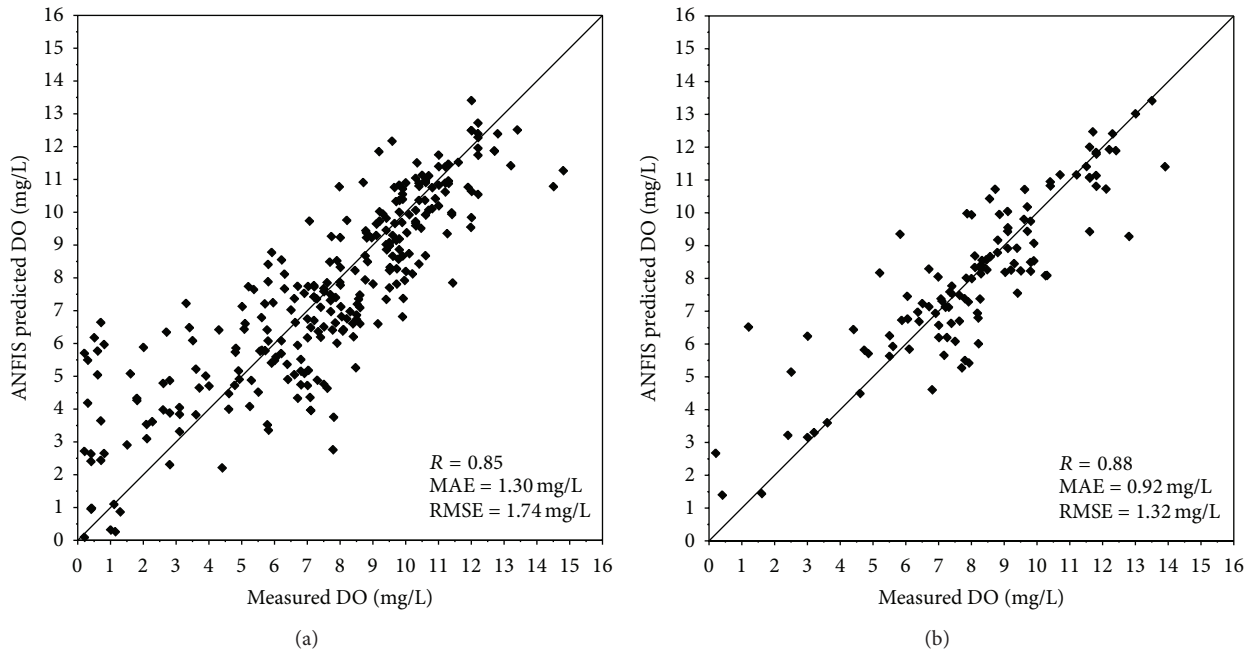
FIGURE 6: Scatter plots of predicted and measured dissolved oxygen using ANFIS model for the (a) training phase and (b) testing phase.
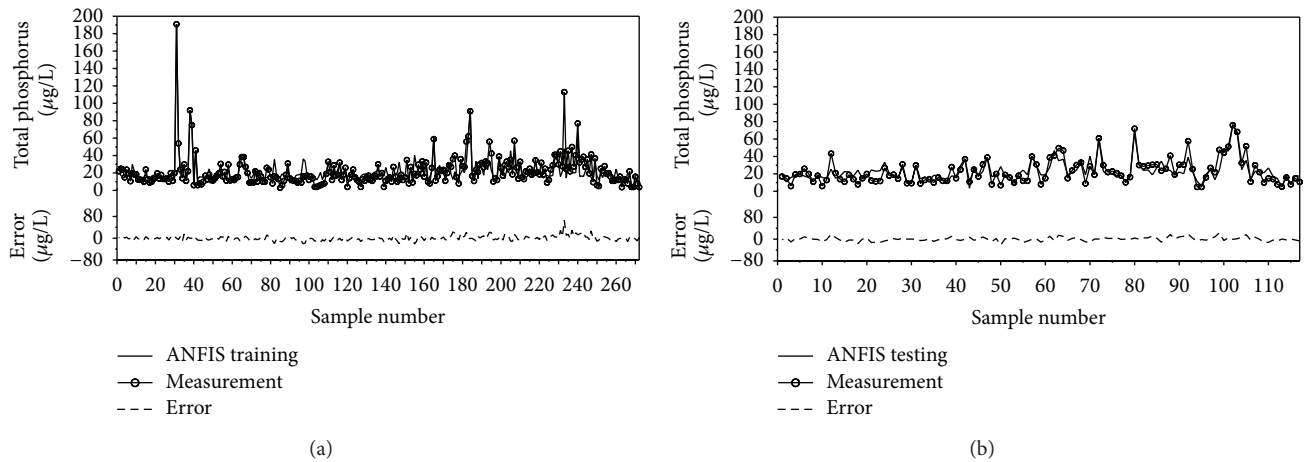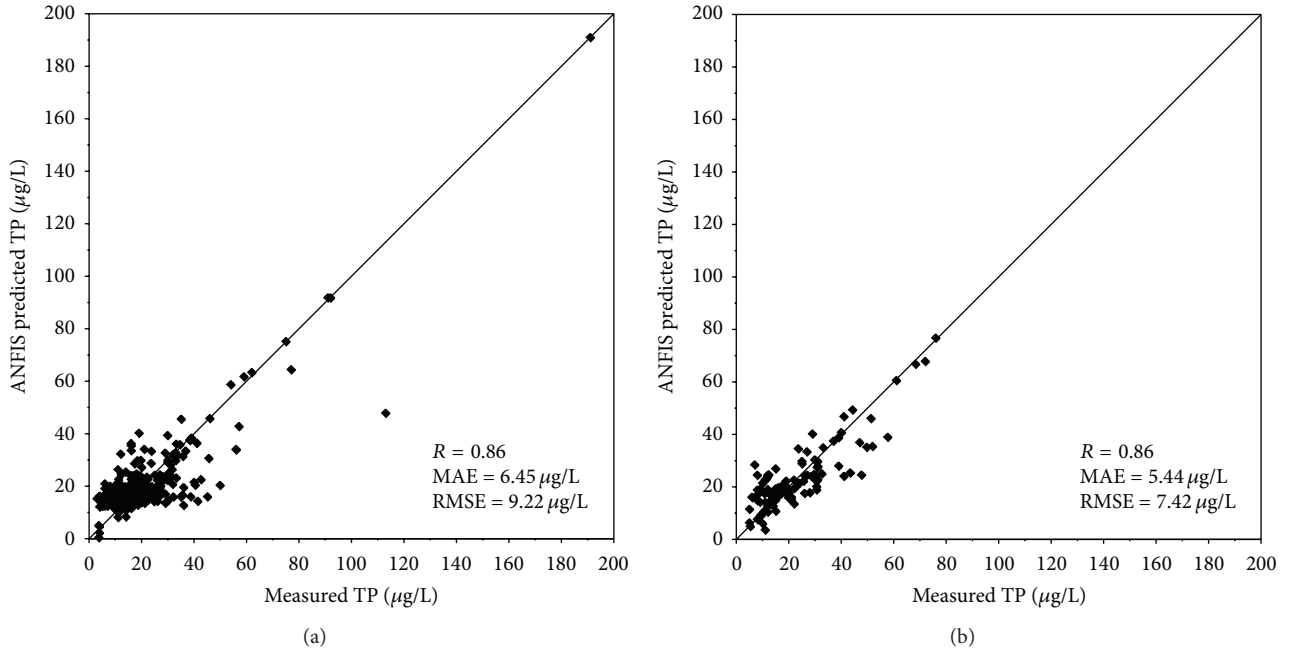

(a)   (b)

FIGURE 7: Comparison of predicted and measured total phosphorus using ANFIS model for the (a) training phase and (b) testing phase.

(a)

(b)

FIGURE 8: Scatter plots of predicted and measured total phosphorus using ANFIS model for the (a) training phase and (b) testing phase.
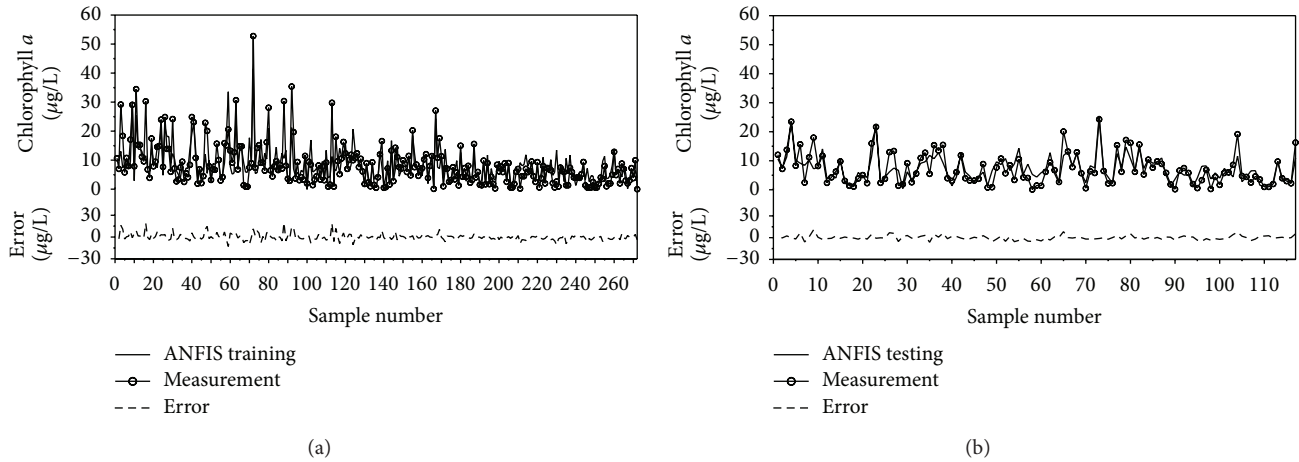

(a)

(b)

FIGURE 9: Comparison of predicted and measured chlorophyll *a* using ANFIS model for the (a) training phase and (b) testing phase.

Chau [26] performed a benchmarking comparison for predicting the river flow discharge using ANN and ANFIS models. His study indicated that the ANFIS model exhibited better performance than ANN model, but the performance of these two models were very close. Our study showed that the predictions of water quality using RBFN and ANFIS models had significant differences. It may be the reason that the characteristics of input data are different. Chau [26] used antecedent flow data (i.e., Q_t, Q_{t-1}, and Q_{t-2}) as input vector, while input data with different independent variables is adopted in the present study.

4. Conclusions

ANN models (i.e., RBFN and ANFIS models) and an MLR model were developed to predict the DO, TP, Chl *a*, and SD

in the reservoir. The performances of the RBFN, ANFIS, and MLR models were evaluated using the mean absolute error, the root mean square error, and the correlation coefficient. The linear regression between the water quality variables (DO, TP, Chl *a*, and SD) and the individual dependent variables was used to select the major input variables for the RBFN, ANFIS, and MLR models.

These models were then constructed to predict the DO, TP, Chl *a*, and SD in the Mingder Reservoir of central Taiwan. The water quality variables predicted using the MLR model did not yield good results. The correlation coefficient between the predicted water quality variables and the measured data was less than 0.8 using the RBFN model. In general, the ANFIS model better predicts the water quality variables than the RBFN model does. The ANN model, including the RBFN and ANFIS model, can preserve the nonlinear characteristics
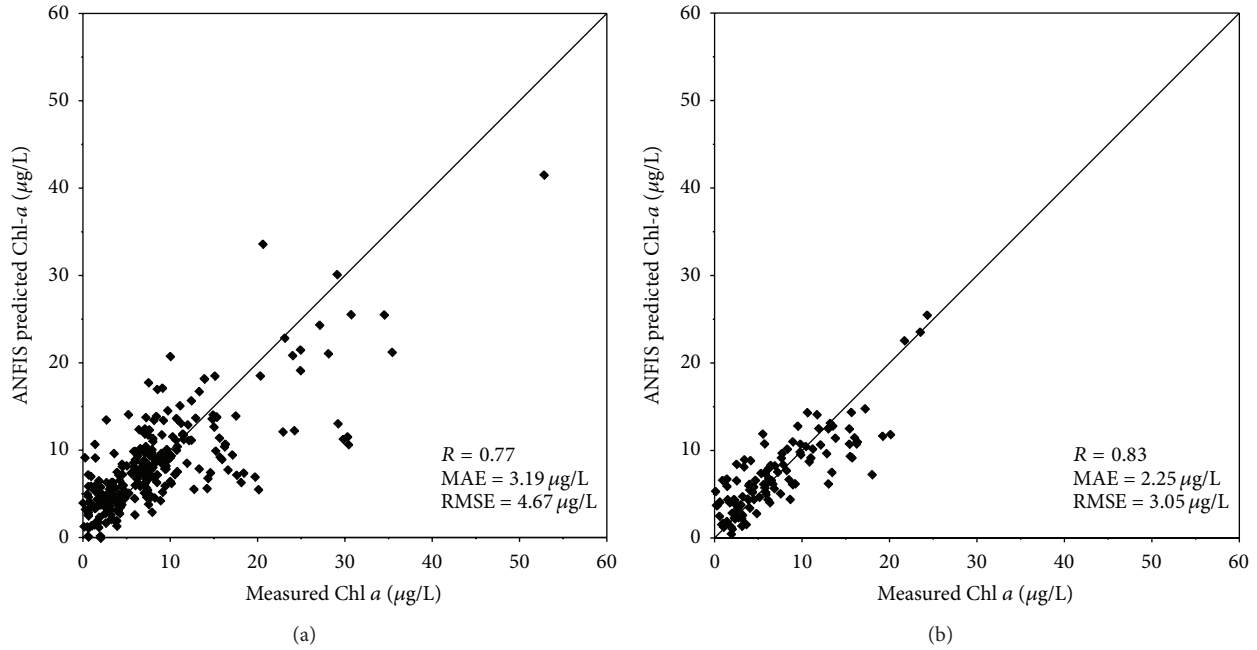
FIGURE 10: Scatter plots of predicted and measured chlorophyll *a* using ANFIS model for the (a) training phase and (b) testing phase.
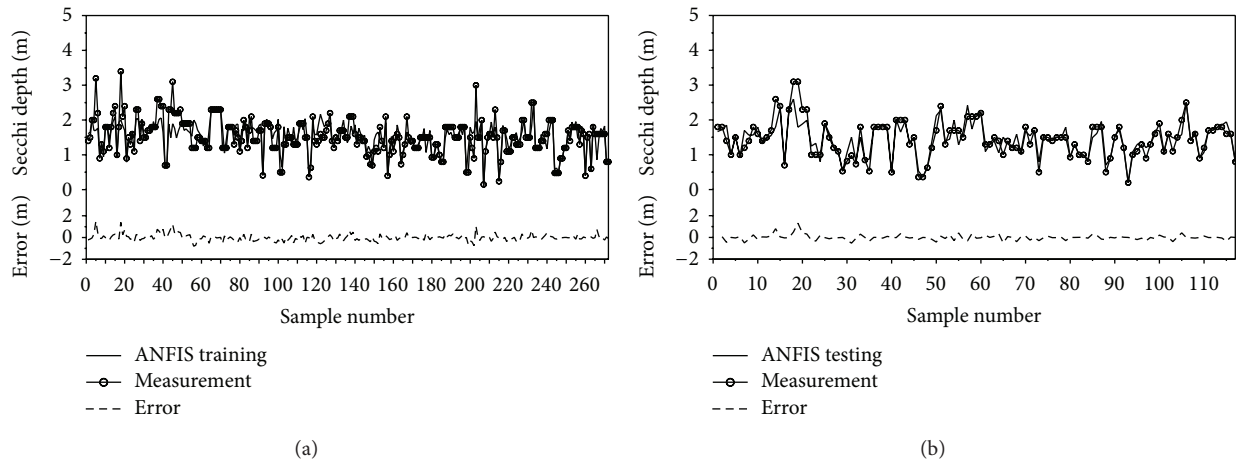

FIGURE 11: Comparison of predicted and measured Secchi disk depth using ANFIS model for the (a) training phase and (b) testing phase.

between the input and output variables, which are superior to conventional statistical approaches (i.e., the MLR model). In the real world, temporal and spatial distributions in observational data do not exhibit simple regularities; therefore, they are difficult to accurately predict. It is necessary to use nonlinear models, such as the ANN model, which are suitable for complex nonlinear systems. The proposed approach using the ANFIS model has yielded valuable information that can be used by decision-makers for aiding reservoir water quality management.

In the present study, we focus on the prediction of water quality instead of forecast. In a future study, different lead-time forecasts in the water quality can be developed to assist the local authorities for water quality management. The soft computing techniques, such as the combining fuzzy optimal

model with genetic programming [40, 41], support vector machine [42, 43], and particle swarm optimization training algorithm for a neural network [44], can also be develop to improve the prediction of water quality in the reservoirs.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
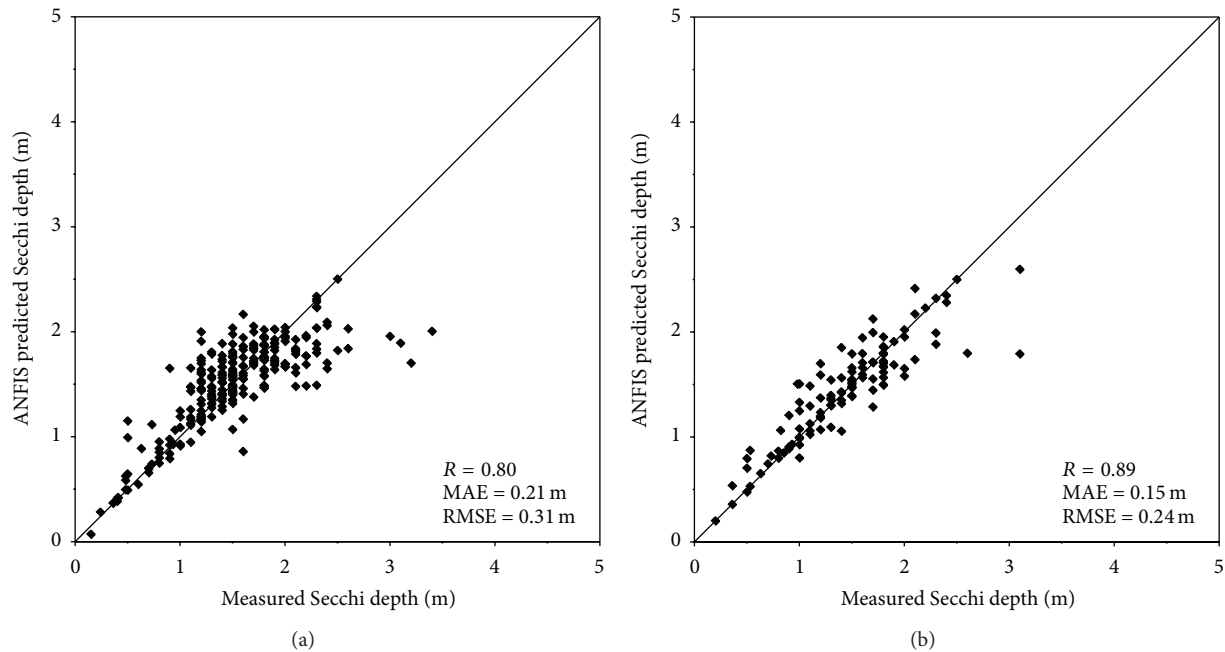
Acknowledgments

FIGURE 12: Scatter plots of predicted and measured Secchi disk depth using ANFIS model for the (a) training phase and (b) testing phase.

References

[1] W.-C. Liu, W.-B. Chen, and N. Kimura, "Impact of phosphorus load reduction on water quality in a stratified reservoir-eutrophication modeling study," *Environmental Monitoring and Assessment*, vol. 159, no. 1–4, pp. 393–406, 2009.

[2] A. Najah, A. El-Shafie, O. A. Karim, and A. H. El-Shafie, "Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring," *Environmental Science and Pollution Research*, vol. 21, no. 3, pp. 1658–1670, 2014.

[3] F. Recknagel, A. Talib, and D. van der Molen, "Phytoplankton community dynamics of two adjacent Dutch lakes in response to seasons and eutrophication control unravelled by non-supervised artificial neural networks," *Ecological Informatics*, vol. 1, no. 3, pp. 277–285, 2006.

[4] R. O. Strobl, F. Forte, and L. Pennetta, "Application of artificial neural networks for classifying lake eutrophication status," *Lakes & Reservoirs: Research and Management*, vol. 12, no. 1, pp. 15–25, 2007.

[5] A. Afshar and M. Saadatpour, "Reservoir eutrophication modeling, sensitivity analysis, and assessment: application to karkheh reservoir, Iran," *Environmental Engineering Science*, vol. 26, no. 7, pp. 1227–1238, 2009.

[6] P. Kudu, A. Debsarkar, and S. Mukherjee, "Artificial neural network modeling for biological removal of organic carbon and nitrogen from slaughterhouse wastewater in a sequencing batch reactor," *Advances in Artificial Neural Systems*, vol. 2013, Article ID 268064, 15 pages, 2013.

[7] G. He, H. Fang, S. Bai, X. Liu, M. Chen, and J. Bai, "Application of a three-dimensional eutrophication model for the Beijing Guanting Reservoir, China," *Ecological Modelling*, vol. 222, no. 8, pp. 1491–1501, 2011.

[8] C. Lindim, J. L. Pinho, and J. M. P. Vieira, "Analysis of spatial and temporal patterns in a large reservoir using water quality and hydrodynamic modeling," *Ecological Modelling*, vol. 222, no. 14, pp. 2485–2494, 2011.

[9] C. Zhang, X. Gao, L. Wang, and Y. Chen, "Analysis of agricultural pollution by flood flow impact on water quality in a reservoir using a three-dimensional water quality model," *Journal of Hydroinformatics*, vol. 15, no. 4, pp. 1061–1072, 2013.

[10] J. M. P. Vieira, J. L. S. Pinho, N. Dias, D. Schwanenberg, and H. F. P. van den Boogaard, "Parameter estimation for eutrophication models in reservoirs," *Water Science and Technology*, vol. 68, no. 2, pp. 319–327, 2013.

[11] C. T. Cheng, K. W. Chau, Y. G. Sun, and J. Y. Lin, "Long-term prediction of discharges in Manwan Reservoir using artificial neural network models," in *Advances in Neural Networks—ISNN 2005*, vol. 3498 of *Lecture Notes in Computer Science*, pp. 1040–1045, Springer, Berlin, Germany, 2005.

[12] W. C. Liu and C. E. Chung, "Enhancing the predicting accuracy of the water stage using a physical-based model and an artificial neural network-genetic algorithm in a river system," *Water*, vol. 6, no. 6, pp. 1642–1661, 2014.

[13] K.-W. Chau, "A review on integration of artificial intelligence into water quality modelling," *Marine Pollution Bulletin*, vol. 52, no. 7, pp. 726–733, 2006.

[14] A. Akkoyunlu, H. Altun, and H. K. Cigizoglu, "Depth-integrated estimation of dissolved oxygen in a lake," *Journal of Environmental Engineering*, vol. 137, no. 10, pp. 961–967, 2011.

[15] V. Ranković, J. Radulović, I. Radojević, A. Ostojić, and L. Čomić, "Prediction of dissolved oxygen in reservoirs using adaptive network-based fuzzy inference system," *Journal of Hydroinformatics*, vol. 14, no. 1, pp. 167–179, 2012.

[16] H. B. Chu, W. X. Lu, and L. Zhang, "Application of artificial neural network in environmental water quality assessment,"

Journal of Agricultural Science and Technology, vol. 15, no. 2, pp. 343–356, 2013.

[17] O. Kisi, N. Akbari, M. Sanatipour, A. Hashemi, K. Teimourzadeh, and J. Shiri, "Modeling of dissolved oxygen in river water using artificial intelligence techniques," *Journal of Environmental Informatics*, vol. 22, no. 2, pp. 92–101, 2013.

[18] M. Ay and O. Kisi, "Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques," *Journal of Hydrology*, vol. 511, pp. 279–289, 2014.

[19] K. P. Singh, A. Basant, A. Malik, and G. Jain, "Artificial neural network modeling of the river water quality-A case study," *Ecological Modelling*, vol. 220, no. 6, pp. 888–895, 2009.

[20] E. Dogan, B. Sengorur, and R. Koklu, "Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique," *Journal of Environmental Management*, vol. 90, no. 2, pp. 1229–1235, 2009.

[21] V. Ranković, J. Radulović, I. Radojević, A. Ostojić, and L. Čomić, "Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia," *Ecological Modelling*, vol. 221, no. 8, pp. 1239–1244, 2010.

[22] F. Soltani, R. Kerachian, and E. Shirangi, "Developing operating rules for reservoirs considering the water quality issues: application of ANFIS-based surrogate models," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6639–6645, 2010.

[23] R. E. Carlson, "A trophic state index for lakes," *Limnology and Oceanography*, vol. 22, no. 2, pp. 361–369, 1977.

[24] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[25] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[26] K. W. Chau, *Modelling for Coastal Hydraulics and Engineering*, Spon Press, 2010.

[27] J. C. Mason, R. K. Price, and A. Tem'me, "A neural network model of rainfall-runoff using radial basis functions," *Journal of Hydraulic Research*, vol. 34, no. 4, pp. 537–548, 1996.

[28] Y. B. Dibike, D. P. Solomatine, and M. B. Abbott, "On the encapsulation of numerical-hydraulic models in artificial neural network," *Journal of Hydraulic Research*, vol. 37, no. 2, pp. 147–161, 1999.

[29] M. Verleysen and K. Hlavackova, "Learning in RBF networks," in *Proceedings of the International Conference on Neural Network (ICNN '96)*, pp. 199–204, Washington, DC, USA, June 1996.

[30] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.

[31] F. M. Ham and I. Kostanic, *Principles of Neurocomputing for Science and Engineering*, McGraw-Hill, New York, NY, USA, 2001.

[32] V. Kecman, *Learning and Soft Computing: Support Vector Machine, Neural Networks, and Fuzzy Logic Models*, MIT Press, Cambridge, Mass, USA, 2001.

[33] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[34] M. M. Zirkohi, M. M. Fateh, and A. Akbarzade, "Design of radial basis function network using adaptive particle swarm optimization and orthogonal least squares," *Journal of Software Engineering and Applications*, vol. 3, no. 7, pp. 704–708, 2010.

[35] S. Weisberg, *Applied Linear Regression*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1985.

[36] S. Soyupak, F. Karaer, H. Gürbüz, E. Kivrak, E. Sentürk, and A. Yazici, "A neural network-based approach for calculating dissolved oxygen profiles in reservoirs," *Neural Computing & Applications*, vol. 12, no. 3-4, pp. 166–172, 2003.

[37] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural network Studies. 1. Comparison of overfitting and overtraining," *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 5, pp. 826–833, 1995.

[38] I. Chenini and S. Khemiri, "Evaluation of ground water quality using multiple linear regression and structural equation modeling," *International Journal of Environmental Science and Technology*, vol. 6, no. 3, pp. 509–519, 2009.

[39] W. Thoe and J. H. W. Lee, "Daily forecasting of Hong Kong beach water quality by multiple linear regression models," *Journal of Environmental Engineering*, vol. 140, no. 2, Article ID 04013007, 2014.

[40] O. Kisi and T. M. Cengiz, "Fuzzy genetic approach for estimating reference evapotranspiration of Turkey: Mediterranean region," *Water Resources Management*, vol. 27, no. 10, pp. 3541–3553, 2013.

[41] C. T. Cheng, C. P. Ou, and K. W. Chau, "Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration," *Journal of Hydrology*, vol. 268, no. 1–4, pp. 72–86, 2002.

[42] J.-Y. Lin, C.-T. Cheng, and K.-W. Chau, "Using support vector machines for long-term discharge prediction," *Hydrological Sciences Journal*, vol. 51, no. 4, pp. 599–612, 2006.

[43] H. Tabari, O. Kisi, A. Ezani, and P. Hosseinzadeh Talaee, "SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment," *Journal of Hydrology*, vol. 444-445, pp. 78–89, 2012.

[44] K. W. Chau, "Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River," *Journal of Hydrology*, vol. 329, no. 3-4, pp. 363–367, 2006.