

Research Article

Predicting the Water Level Fluctuation in an Alpine Lake Using Physically Based, Artificial Neural Network, and Time Series Forecasting Models

Chih-Chieh Young,¹ Wen-Cheng Liu,^{2,3} and Wan-Lin Hsieh²

¹Hydrotech Research Institute, National Taiwan University, Taipei 10617, Taiwan

²Department of Civil and Disaster Prevention Engineering, National United University, Miao-Li 36063, Taiwan

³Taiwan Typhoon and Flood Research Institute, National Applied Research Laboratories, Taipei 10093, Taiwan

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

Received 3 October 2014; Revised 9 March 2015; Accepted 20 March 2015

Academic Editor: Ufuk Yolcu

Copyright © 2015 Chih-Chieh Young et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate prediction of water level fluctuation is important in lake management due to its significant impacts in various aspects. This study utilizes four model approaches to predict water levels in the Yuan-Yang Lake (YYL) in Taiwan: a three-dimensional hydrodynamic model, an artificial neural network (ANN) model (back propagation neural network, BPNN), a time series forecasting (autoregressive moving average with exogenous inputs, ARMAX) model, and a combined hydrodynamic and ANN model. Particularly, the black-box ANN model and physically based hydrodynamic model are coupled to more accurately predict water level fluctuation. Hourly water level data (a total of 7296 observations) was collected for model calibration (training) and validation. Three statistical indicators (mean absolute error, root mean square error, and coefficient of correlation) were adopted to evaluate model performances. Overall, the results demonstrate that the hydrodynamic model can satisfactorily predict hourly water level changes during the calibration stage but not for the validation stage. The ANN and ARMAX models better predict the water level than the hydrodynamic model does. Meanwhile, the results from an ANN model are superior to those by the ARMAX model in both training and validation phases. The novel proposed concept using a three-dimensional hydrodynamic model in conjunction with an ANN model has clearly shown the improved prediction accuracy for the water level fluctuation.

1. Introduction

Accurate predictions of water level fluctuation that results from hydrometeorological variations and anthropogenic disturbances [1] are needed for sustainable development and management of lake water usage [2–8]. As a complicated mathematical problem in water resources planning and engineering, water level variations in a lake represent the balance of water components such as direct precipitation or evaporation at a lake's surface, groundwater exchange, and incoming or outgoing river flows. Besides, water level can be influenced by the near surface wind field [9]. Particularly, during severe rain storms (or floods), the water level of a thermally stratified lake can rise sharply with vertical mixing, leaving significant socioeconomic and ecological impacts [10].

To date, lake level variations at various time intervals have drawn more and more research attentions. In general, lake water level changes seasonally (e.g., high in the wet summer and low in the dry winter) with sharp rising/falling limbs during typhoon events, but not in a simple periodic mode (except for the seiche oscillation that occurs mainly in large lakes). Effective prediction tools play an important role in the studies of lakes. They can be used to simulate the lake water level variations based upon the available measured data and predict the possible responses under different scenarios, supporting management decisions of valuable water resources. In the past decades, several types of methodology have been developed. For example, physically based numerical models are constructed using a set of governing equations that address conservation laws of mass/momentun and transport

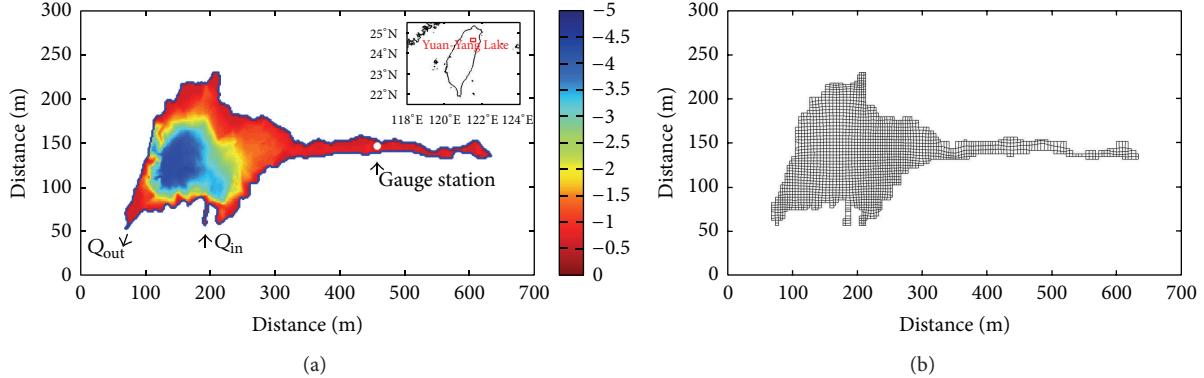


FIGURE 1: (a) Location of Yuan-Yang Lake in Taiwan, bathymetry (contours in m), and gauge location and (b) horizontal grid of Yuan-Yang Lake for three-dimensional hydrodynamic model.

processes. In contrast to a simple zero-dimensional mass-balance computation for elevation and storage, these models can provide more detailed spatial-temporal patterns of water level, circulation, vertical stratification/mixing, and other aspects of lake physics [10–14].

Time series forecasting methods that define the trend or stochastic processes of variables have also been applied to predict water level fluctuation in lakes. One of the most commonly used approaches is the autoregressive integrated moving average (ARIMA) model, where the autoregressive and moving average filters account for systematic effects and internal shock effects in the endogenous variable, respectively. ARIMA models require a stationary time series without data missing [15]. Based on the observations time series, an identified underlying process can be constructed under the assumptions of linearity, normality, and homoscedasticity. For better prediction accuracy and variability in endogenous variables, an autoregressive moving average with exogenous (ARMAX) model [16] further includes the explanatory variables in the cause-effect technique.

Alternatively, artificial neural networks (ANNs) provide a novel computational approach for one-step-ahead forecasting. ANNs are widely accepted by cross disciplines and suitable for the application of hydrological processes due to their informative processing characteristics, for example, nonlinearity, parallelism, and noise tolerance, as well as learning and generalization capabilities [17–23]. Altunkaynak [19] used an artificial neural network model to accurately predict dynamic changes of surface water level in the lake. Van. Yarar et al. [20] estimated level changes of Lake Beysehir in Turkey using the adaptive neurofuzzy inference system (ANFIS), ANN, and seasonal autoregressive integrated moving average approaches. Kisi et al. [22] forecasted daily lake levels by artificial intelligence approaches. However, detailed comparison of the time series forecasting and ANN models with the physically based models has not yet been studied in lakes.

The purpose of this study is to establish a physically based three-dimensional hydrodynamic model to simulate the water level fluctuation in the Yuan-Yang Lake (YYL) in Taiwan. An artificial neural network model is adopted to predict the lake water level and to improve the calculations

by the hydrodynamic model. Compared with increasingly applied data assimilation methodologies [24, 25], the novelty in our proposed concept shows another way for accuracy improvement in lake water level prediction. A time series forecasting (ARMAX) model has also been developed. Four quantitative statistical indices including mean absolute error, root mean square error, correlation coefficient, and Nash Sutcliffe efficiency/skill score are used to evaluate the prediction performance. Model accuracy is investigated and carefully discussed.

2. Description of Study Site and Data Collection

Yuan-Yang Lake (YYL, $24^{\circ}34'60.00''N$, $121^{\circ}24'0.00''E$, area = $3.7 \times 10^4 m^2$, and maximum depth = 4.5 m), located in north-central Taiwan, is a subtropical, subalpine, humic lake in the Cilan National Forest Preserve (Figure 1(a)). The lake is located 1670 m above mean sea level. The geography of the drainage basin allows large quantities of terrestrial runoff from the surrounding mountains to enter YYL. The mean annual air temperature is approximately $13^{\circ}C$ (monthly averages range from -5 to $15^{\circ}C$), and annual precipitation can exceed 4000 mm. The depth of the euphotic layer is approximately 1.5 m [26]. The lake experiences three to seven typhoons annually during the summer and fall; more than 40% of all annual precipitation can occur during the typhoons. Wind speeds over the lake are relatively weak. The dominant wind directions are from the east and southwest because the V-shaped valley faces east to west. Kimura et al. [10] developed a nowcasting system to predict the spatial distributions and temporal variations in the dynamics of YYL when subjected to severe rainstorms. They found that the lake water temperature dynamics were subject to meteorological conditions during normal conditions, while the strong winds and high inflows caused high water levels and affected vertical mixing during the typhoon events.

A gauge station was established in April 2004 to collect the water level data hourly; these data are accessible from the Global Ecological Lakes Observatory Network (GLEON) website. All data from the gauge station (Figure 1(a)) and

inflow and outflow rates were downloaded from the GLEON publicly accessible database (<https://lter.limnology.wisc.edu/project/global-lakes-ecological-observatory-network-gleon>). The hourly water level data collected from August 1 to December 31, 2009, and from January 1 to May 31, 2010, were utilized to calibrate/train and validate the hydrodynamic, ANN, and time series forecasting models, respectively. The measured time-series data of inflow discharge, outflow discharge, precipitation, and water level in the YYL for the model calibration and validation phases are shown in Figure 3. It can be observed that there are two water level peaks (Figure 3(d)) during the periods of August 5 to 10 and October 3 to 6 in 2009, corresponding to Typhoon Morakot and Typhoon Parma, respectively. Note that the calibration samples including both smooth variations and typhoon-induced peaks ensure the prediction/generalization capabilities of these models in the validation phase.

3. Methodology

3.1. Hydrodynamic Model Description and Model Setup. In this study, we adopted the well-known three-dimensional hydrodynamic model EFDC (Environmental Fluid Dynamics Code) [27, 28] which has been widely used for describing flow and transport processes in surface water systems, for example, rivers, lakes, estuaries, reservoirs, wetlands, and coastal regions [29–33]. This model solves the 3D equations of motion and continuity equation with the Mellor-Yamada level 2.5 turbulence closure scheme [34–36] under the stretched (or sigma) vertical coordinates and Cartesian (or curvilinear) orthogonal horizontal coordinates. For numerical aspect, it employs a second-order accurate, three-time-level finite difference scheme with an internal-external mode splitting procedure to separate internal baroclinic mode from the external free surface gravity wave [37]. Further details of the EFDC model are given in Hamrick [27, 28].

In this study, EFDC for lake water level prediction was driven by freshwater flow and atmospheric forcing (wind-shear stress, precipitation/evaporation, and heat flux). An orthogonal boundary-fitted computational grid (containing 700 horizontal cells and five vertical layers) was developed to represent the irregular shorelines and bottoms in YYL study area (Figure 1(b)) based on the topography data measured in August 2007. The nonuniform grid sizes are approximately $6\text{ m} \times 7\text{ m}$ along the horizontal directions. The simulations were carried out using $\Delta t = 1$ second without the sign of numerical instability.

3.2. Artificial Neural Networks (ANNs). An ANN model uses a multilayered technique to approximate complex mathematical functions or data. In this study, the back propagation neural network (BPNN), a learning algorithm based on back propagation of error gradient, was utilized to amend the water level results from hydrodynamic model for better prediction accuracy.

The BPNN proposed by Rumelhart et al. [38] is a multiple-layer network (see an input layer, a hidden layer, and an output layer in Figure 2). Each layer contains a number of neurons with nonlinear differentiable transfer functions.

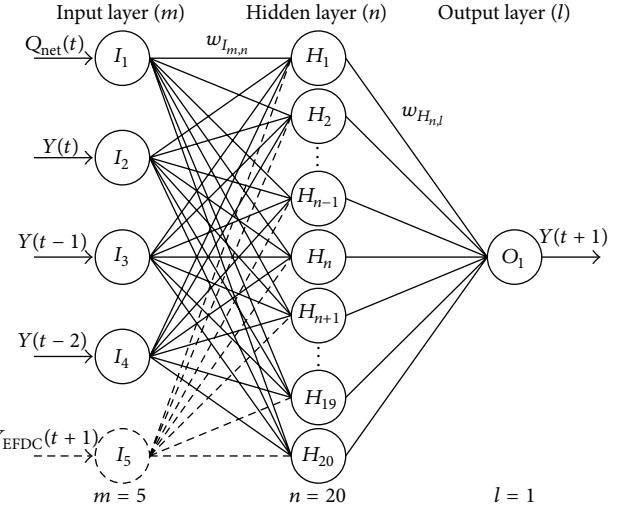


FIGURE 2: Architecture of back propagation neural network (BPNN).

Each neuron receives inputs from neurons in the previous layers or external inputs, and input is converted to either an output signal or another input signal to be used by neurons in the next layers. Weighted connections between neurons in successive layers represent their relative importance in the network. The neurons execute a weighted summation of all inputs and further assess the weighted sum by activation function f :

$$H_n = f(w_{I_m,n} I_m), \quad (1a)$$

$$O_l = f(w_{H_n,l} H_n), \quad (1b)$$

where H_n is the hidden layer with neuron n , O_l is the output of neuron l , I_m is the input to the neuron m , and $w_{I_m,n}$ and $w_{H_n,l}$ are the synaptic weights.

A hyperbolic tangent sigmoid transfer function in (2) is used in the hidden layer:

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

A linear transfer function in (3) is applied in the output layer:

$$f(x) = x. \quad (3)$$

To scale the inputs and the targets, normalized equation (4) is often used, thus forcing the data to fall within a specified range:

$$Y_N = (y_{\max} - y_{\min}) \times \left(\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \right) + y_{\min}, \quad (4)$$

where Y_N is the value after normalization, x_{\min} and x_{\max} are the minimum and maximum data values, respectively, and y_{\min} and y_{\max} are -1 and 1 , respectively.

The training process of a neural network involves tuning the network weight and bias values through minimization of the cost function C:

$$C = \frac{1}{P} \sum_{p=1}^P \sum_{l=1}^L e_l^2(p), \quad (5)$$

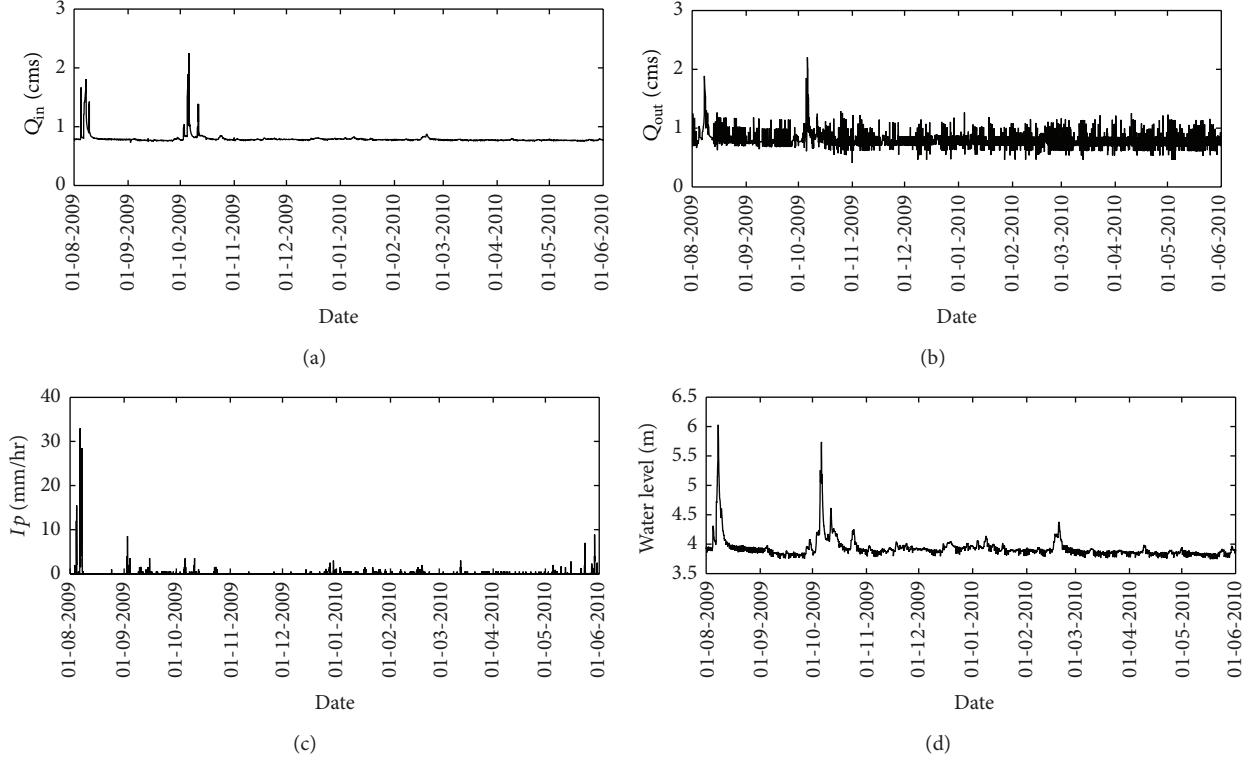


FIGURE 3: Hydrological conditions for three-dimensional hydrodynamic, BPNN, and time series forecasting models: (a) inflow discharge, (b) outflow discharge, (c) precipitation, and (d) water level in the lake from August 1, 2009, to May 31, 2010.

where P is the number of input-output training patterns, $e_l(p)$ is the difference between the output and target of neuron l for the p th input-output pattern, and L is the number of neurons in the output layer.

The back propagation method was accomplished using the Levenberg-Marquardt (LM) algorithm [39] that optimizes the network parameters with a nonlinear least squares error minimization technique. The LM algorithm combines the capability of the gradient descent method and Gauss-Newton method in terms of converge and speed. The ANN model together with the LM technique was implemented in the Neural Network Toolbox in Matlab (e.g., 2009 or latest version).

3.3. Time Series Forecasting Model. ARMAX models have been used extensively to represent the relationship of system output and input in the presence of noise in many linear dynamic systems [16]. In the case of lake water balance, a simple ARMAX (p, q, r) model can be expressed as

$$\begin{aligned} Y(t+1) + a_1 Y(t) + \dots + a_p Y(t-p) \\ = b_1 Q_{\text{net}}(t) + \dots + b_q Q_{\text{net}}(t-q) + e(t) + \dots + c_r e(t-r), \end{aligned} \quad (6)$$

where $Q_{\text{net}}(t) = I_p(t) \cdot A + Q_{\text{in}}(t) - Q_{\text{out}}(t)$, $Q_{\text{net}}(t)$ is the net volume flux in hour t (m^3/hr), $Q_{\text{in}}(t)$ and $Q_{\text{out}}(t)$ are the discharge rates of inflow and outflow, respectively, $I_p(t)$ is the rainfall intensity (mm/hr), A is the surface area of the YYL,

$Y(t+1)$ is the water level in hour $t+1$, $Y(t)$ is the water level in hour t , a_p , b_q , and c_r are the weights of the autoregressive and moving average processes and the weight noise, and e is normally distributed with zero mean and unit variance.

3.4. Indices of Simulation Performance. To evaluate the performances of the three-dimensional hydrodynamic model, the BPNN model, and the time series forecasting model, three different criteria were used to compare the predicted results with the observed data: mean absolute error (MAE), root mean square error (RMSE), coefficient of correlation (R), and the skill score (SS), as based on the following equations:

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{i=1}^N |(Y_m)_i - (Y_o)_i|, \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N [(Y_m)_i - (Y_o)_i]^2}, \\ R &= \frac{(1/N) \sum_{i=1}^N [(Y_m)_i - \bar{Y}_m] [(Y_o)_i - \bar{Y}_o]}{\sqrt{(1/N) \sum_{i=1}^N [(Y_m)_i - \bar{Y}_m]^2} \times \sqrt{(1/N) \sum_{i=1}^N [(Y_o)_i - \bar{Y}_o]^2}}, \\ \text{SS} &= 1 - \frac{\sum_{i=1}^N [(Y_o)_i - (Y_m)_i]^2}{\sum_{i=1}^N [(Y_o)_i - \bar{Y}_o]^2}, \end{aligned} \quad (7)$$

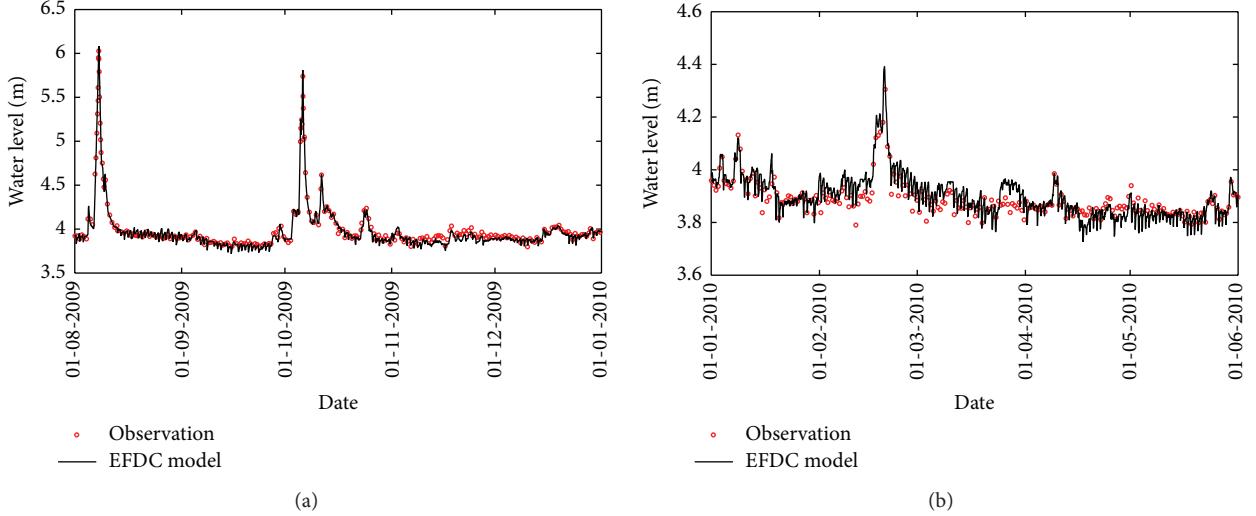


FIGURE 4: Comparison of predicted hourly water levels with three-dimensional hydrodynamic model and measured data for (a) model calibration and (b) model validation.

where N is the total number of data, Y_m is the predicted water level, Y_o is the observed water level, $\bar{Y}_m = (\sum_{i=1}^N (Y_m)_i)/N$, and $\bar{Y}_o = (\sum_{i=1}^N (Y_o)_i)/N$.

4. Results

4.1. Calibration and Validation of the Three-Dimensional Hydrodynamic Model. Figure 4 shows the calibration and validation results for the three-dimensional hydrodynamic model. In the calibration phase, the model captures the observed water level fluctuation with satisfactory predictions of peak water levels associated with the two typhoon events. Interestingly, both the observation and simulation results show high-frequency oscillation, which might be attributed to the diurnal cycles from the atmospheric forcing. Further studies are required to clarify the mechanisms for the temporal variability of water level fluctuations at different scales and will be reported separately in the future. For the validation phase, however, it somewhat fails to mimic the observed water levels (see Figure 4(b) with a zoom-in vertical axis). Table 1 presents the performance assessment of the three-dimensional hydrodynamic model. The MAE, RMSE, R , and SS values are 2.56 cm, 3.70 cm, 0.990, and 0.974 (or 2.75 cm, 3.53 cm, 0.923, and 0.769) for model calibration (or validation) phase, respectively. The results indicate that the model calibration performance is better than the model validation performance.

4.2. ANN Training and Validation. To forecast the water level, three cases were established using two to four input nodes, that is, the net volume flux $Q_{\text{net}}(t)$ with the measured water levels at 1 to 3 hr lag time $Y(t)$, $Y(t - 1)$, $Y(t - 2)$ for cases 1, 2, and 3, respectively. The output of these three cases is the predicted water level $Y(t + 1)$. For the hidden layer, the selection of an appropriate number of nodes is quite important because an excessively large number of nodes may

result in overfitting, while an insufficient number of nodes may not capture the information adequately [17–23]. The best network architecture (number of hidden nodes, number of iterations, learning rate, and momentum coefficient) was obtained by trial and error based on RMSE in the training and validation phases. Figure 5 shows the RMSE trend with the increasing number of hidden nodes in the training and validation phases. In this study, twenty hidden nodes in the BPNN were selected to yield the optimal performance. Other ANN training parameters include a learning rate of 0.01, a momentum coefficient of 0.3, and maximum iterations of 400.

A comparison of water level between the observation data and simulated results from ANN case 3 is shown in Figure 6. In contrast to the hydrodynamic model, ANN is able of forecast the measured water levels in both training and validation phases. Based on Table 1, case 3 gives the best performance among these three model settings with MAE, RMSE, R , and SS equal to 0.77 cm, 1.17 cm, 0.999, and 0.997 (or 0.77 cm, 1.12 cm, 0.988, and 0.977) in the training (or validation) phase, respectively.

4.3. Training and Validation of the Time Series Forecasting Model. For time series forecasting, ARMAX (1, 1, 1), ARMAX (2, 1, 1), and ARMAX (3, 1, 1) were used in cases 4, 5, and, 6 to predict the water level $Y(t + 1)$, respectively.

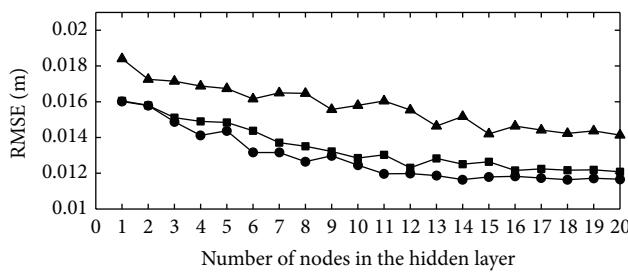
A comparison of water level between the measured data and predicted results from case 5 is shown in Figure 7. According to the performance assessment in Table 1, it can be found that ARMAX (2, 1, 1) yields the best water level prediction with MAE, RMSE, R , and SS equal to 0.96 cm, 1.64 cm, 0.997, and 0.995 (or 0.83 cm, 1.18 cm, 0.987, and 0.974) in the training (or validation) phase, respectively.

4.4. Water Level Predictions Using the Combination of Hydrodynamic and ANN Models. To improve the water level

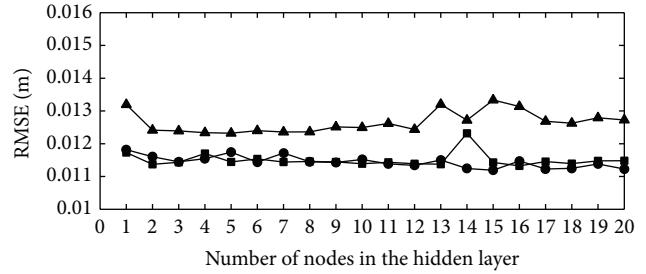
TABLE 1: Assessment of water level prediction performance for three-dimensional hydrodynamic model (EFDC), artificial neural network model (ANN), time series forecasting model (ARMAX), and the combination model (EFDC + ANN).

| Statistical parameter | MAE (cm) | RMSE (cm) | R | SS |
|--|-------------------|-----------|------|-------|
| Calibration | EFDC model | 2.56 | 3.70 | 0.990 |
| | ANN case 1 | 0.87 | 1.41 | 0.998 |
| | ANN case 2 | 0.81 | 1.21 | 0.999 |
| | ANN case 3 | 0.77 | 1.17 | 0.999 |
| | ARMAX case 4 | 0.98 | 1.71 | 0.997 |
| | ARMAX case 5 | 0.96 | 1.64 | 0.997 |
| | ARMAX case 6 | 0.96 | 1.64 | 0.997 |
| | EFDC + ANN case 7 | 0.81 | 1.26 | 0.998 |
| | EFDC + ANN case 8 | 0.74 | 1.08 | 0.999 |
| Validation | EFDC + ANN case 9 | 0.72 | 1.05 | 0.999 |
| | EFDC model | 2.75 | 3.53 | 0.923 |
| | ANN case 1 | 0.84 | 1.27 | 0.985 |
| | ANN case 2 | 0.78 | 1.15 | 0.988 |
| | ANN case 3 | 0.77 | 1.12 | 0.988 |
| | ARMAX case 4 | 0.83 | 1.18 | 0.987 |
| | ARMAX case 5 | 0.83 | 1.18 | 0.987 |
| | ARMAX case 6 | 0.83 | 1.18 | 0.987 |
| | EFDC + ANN case 7 | 0.85 | 1.29 | 0.986 |
| | EFDC + ANN case 8 | 0.82 | 1.19 | 0.988 |
| | EFDC + ANN case 9 | 0.80 | 1.17 | 0.989 |
| Case 1: the input nodes include $Y(t)$ and $Q_{net}(t)$; case 2: the input nodes include $Y(t - 1)$, $Y(t)$, and $Q_{net}(t)$; and case 3: the input nodes include $Y(t - 2)$, $Y(t - 1)$, $Y(t)$, and $Q_{net}(t)$. Case 4: ARMAX (1, 1, 1); case 5: ARMAX (2, 1, 1); case 6: ARMAX (3, 1, 1). Case 7: the input nodes include $Y(t)$, $Q_{net}(t)$, and $Y_{EFDC}(t + 1)$; case 8: the input nodes include $Y(t - 1)$, $Y(t)$, $Q_{net}(t)$, and $Y_{EFDC}(t + 1)$; case 9: the input nodes include $Y(t - 2)$, $Y(t - 1)$, $Y(t)$, $Q_{net}(t)$, and $Y_{EFDC}(t + 1)$. | | | | |

Case 1: the input nodes include $Y(t)$ and $Q_{net}(t)$; case 2: the input nodes include $Y(t - 1)$, $Y(t)$, and $Q_{net}(t)$; and case 3: the input nodes include $Y(t - 2)$, $Y(t - 1)$, $Y(t)$, and $Q_{net}(t)$. Case 4: ARMAX (1, 1, 1); case 5: ARMAX (2, 1, 1); case 6: ARMAX (3, 1, 1). Case 7: the input nodes include $Y(t)$, $Q_{net}(t)$, and $Y_{EFDC}(t + 1)$; case 8: the input nodes include $Y(t - 1)$, $Y(t)$, $Q_{net}(t)$, and $Y_{EFDC}(t + 1)$; case 9: the input nodes include $Y(t - 2)$, $Y(t - 1)$, $Y(t)$, $Q_{net}(t)$, and $Y_{EFDC}(t + 1)$.



(a)



(b)

FIGURE 5: The effect of number of nodes in the hidden layer on the root mean square error (RMSE) for the (a) ANN training phase and (b) ANN validation phase. Note that the triangle, square, and circle marks represent case 1, case 2, and case 3, respectively, for the ANN model.

prediction quality, an ANN model was utilized to amend the prediction results from the three-dimensional hydrodynamic model. The proposed approach offers another way for accuracy improvement, in comparison to widely used data assimilation [24, 25]. Besides the original inputs, we similarly constructed cases 7, 8, and 9 with an additional input $Y_{EFDC}(t + 1)$ that represents the water level predicted by the EFDC hydrodynamic model in hour $t + 1$. Exactly the same learning parameters (i.e., a learning rate of 0.01, a momentum coefficient of 0.3, and maximum iterations of 400) and twenty hidden nodes were chosen for the optimal performance (see the RMSE trend in Figure 8).

Figure 9 presents a comparison of water level between the measured data and case 9 results obtained by the combination model for training and validation phases. Table 1 shows that case 9 gives the best performance assessment with MAE, RMSE, R, and SS equal to 0.72 cm, 1.05 cm, 0.999, and 0.998 (or 0.80 cm, 1.17 cm, 0.989, and 0.975) in the training (or validation) phase, respectively.

4.5. Comparison of the Predicted Water Levels from All Models. To evaluate the model performance of predicting hourly water levels, the statistical indices for the three-dimensional hydrodynamic model, ANN model, ARMAX model, and

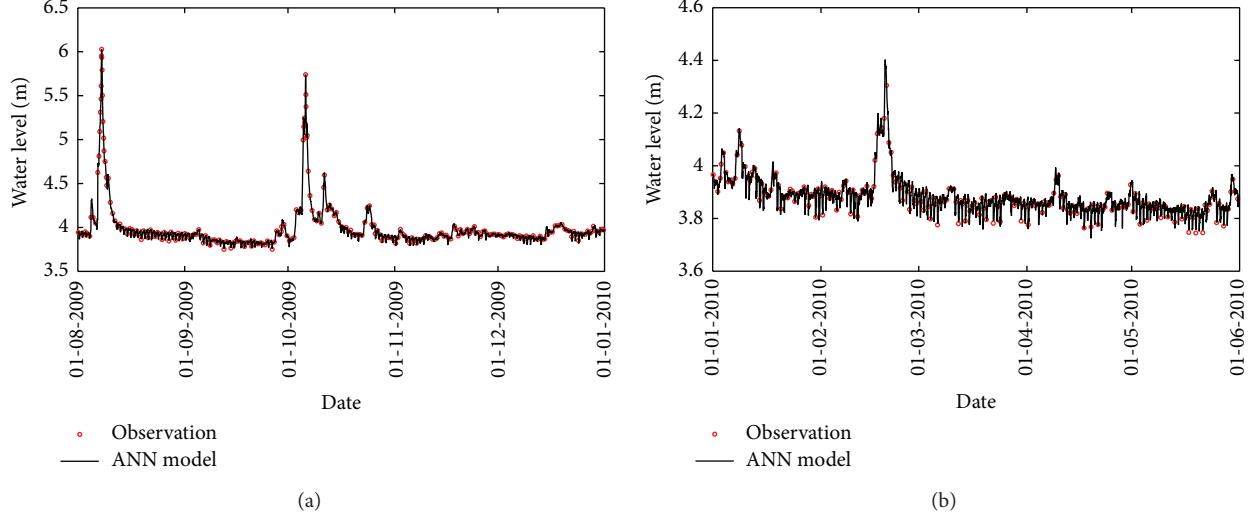


FIGURE 6: Comparison of predicted hourly water levels with the ANN model and measured data for (a) training phase and (b) validation phase.

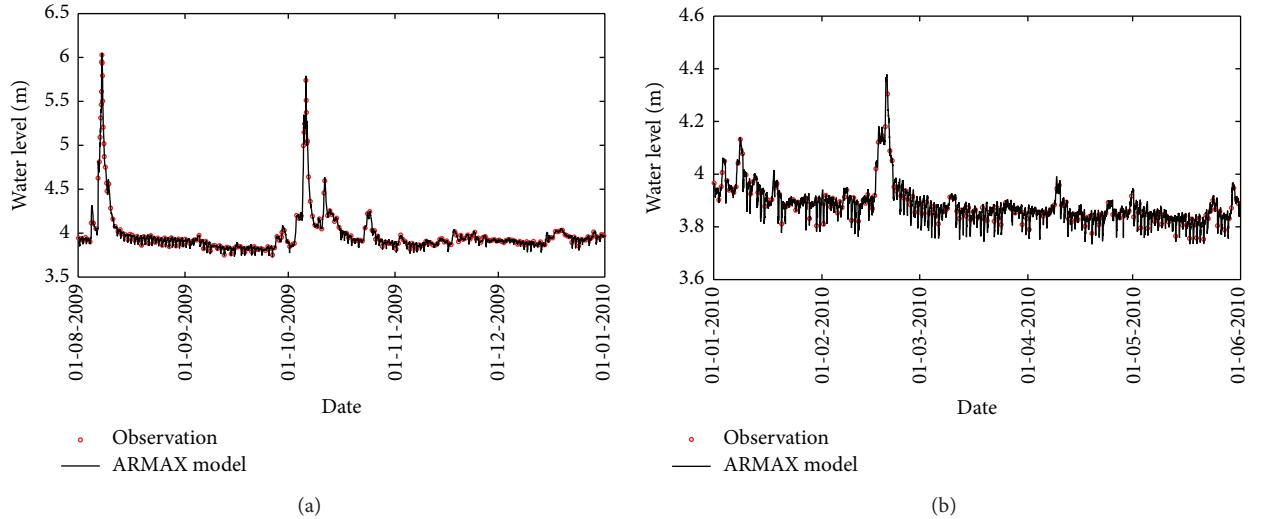


FIGURE 7: Comparison of predicted hourly water levels with the ARMAX model and measured data for (a) training phase and (b) validation phase.

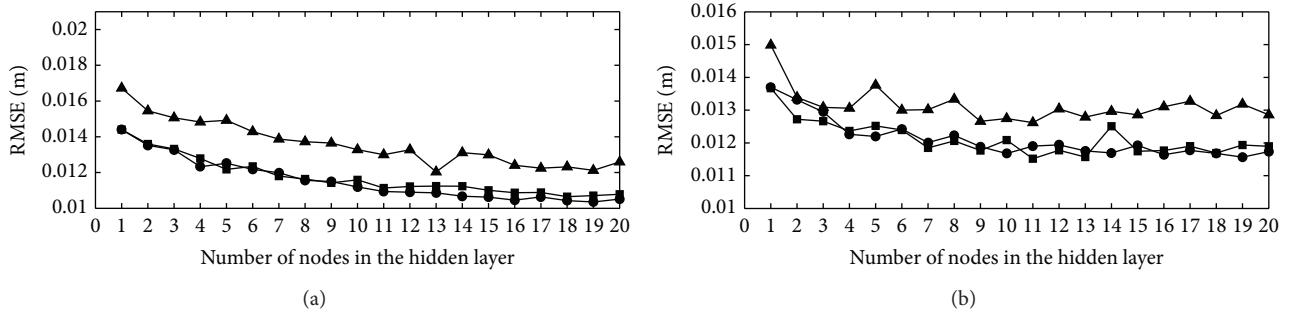


FIGURE 8: The effect of number of nodes in the hidden layer on the root mean square error (RMSE) for the (a) ANN training phase and (b) ANN validation phase. Note that the triangle, square, and circle marks represent case 7, case 8, and case 9, respectively, for the combination of the three-dimensional hydrodynamic model and ANN model.

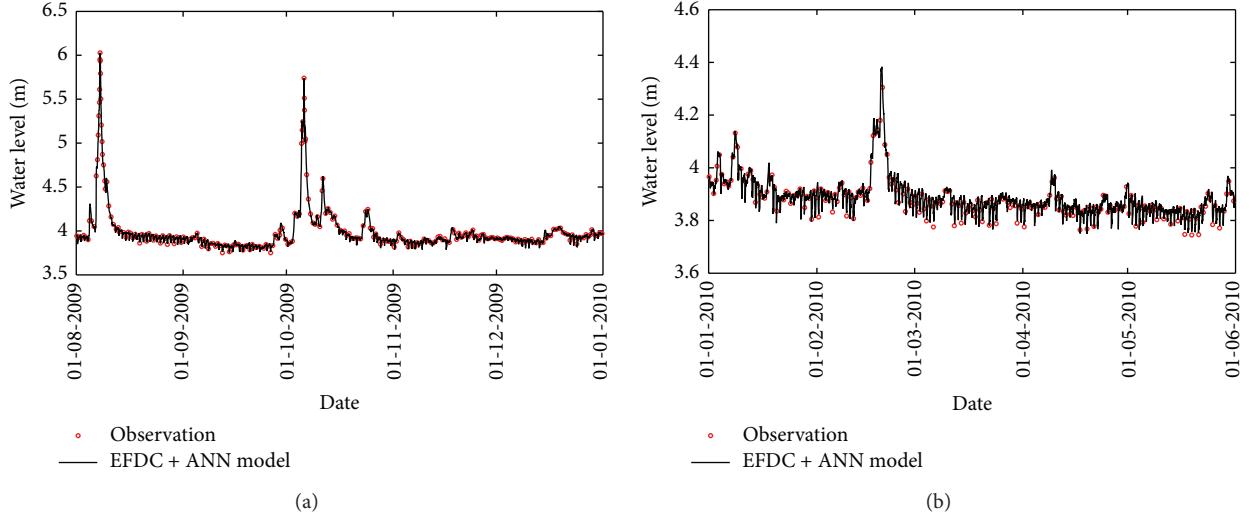


FIGURE 9: Comparison of predicted hourly water levels with the combination of the three-dimensional hydrodynamic and ANN models and measured data for the (a) training phase and (b) validation phase.

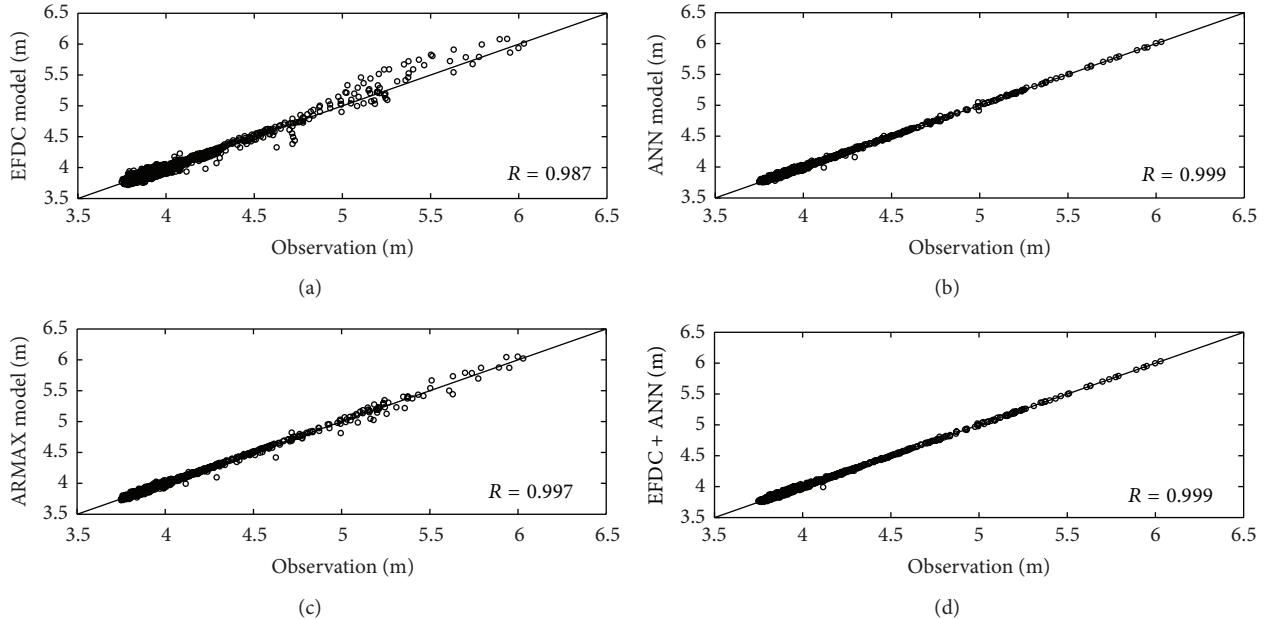


FIGURE 10: Scatter plots of predicted and measured water levels with (a) three-dimensional hydrodynamic model calibration, (b) ANN model training, (c) ARMAX model training, and (d) combination of three-dimensional hydrodynamic and ANN models training. The solid lines represent the parity lines.

the combination model were compared. Figures 10 and 11 further present scatter plots for the measurement-prediction pairs of the hourly water level in the calibration (training) and validation phases. The three-dimensional hydrodynamic model did not show good predictions for lake water level, compared to other prediction approaches (Figures 10(a) and 11(a)). The ANN model better predicts water level variations than the ARMAX model. Overall, the combination model provides further accuracy for hourly water level predictions. The five input nodes consisting of $Y(t - 2)$, $Y(t - 1)$, $Y(t)$, $Q_{\text{net}}(t)$, and $Y_{\text{EFDC}}(t + 1)$ are the best model setting to predict hourly water levels $Y(t + 1)$ in this study.

5. Discussions

In terms of the computational expense for one-month simulation, a three-dimensional hydrodynamic model requires about 3.5 hours of CPU time on an Intel Core i5 PC, while the ANN and ARMAX models only need 1.5 and 1.2 minutes, respectively. The hydrodynamic model takes much longer computational time, compared with the ANN and ARMAX models. For the simulation aspect, the ARMAX model uses a linear stochastic approach to represent the relationship between the system input and output. Consequently, the non-linear ANN model is better than the ARMAX model. Besides,

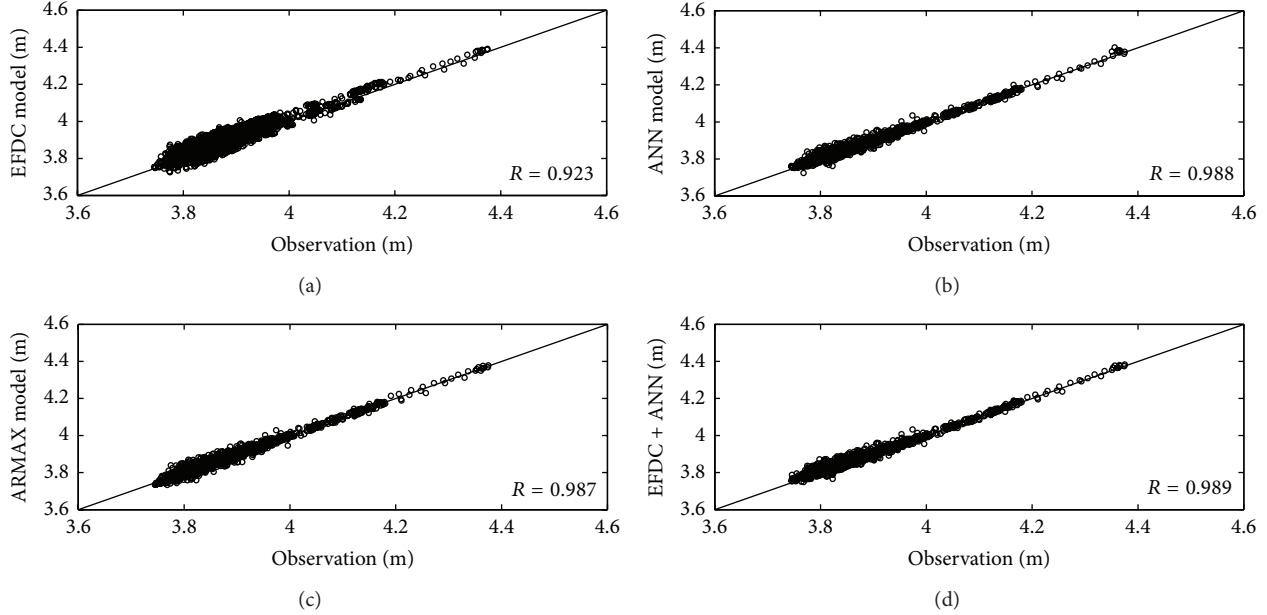


FIGURE 11: Scatter plots of predicted and measured water levels with (a) three-dimensional hydrodynamic model validation, (b) ANN model validation, (c) ARMAX model validation, and (d) combination of three-dimensional hydrodynamic and ANN models validation. The solid lines represent the parity lines.

the predictability of the data-driven ANN model could be increased by providing a large number of appropriate input-output data sets during the training and validation phases [40]. Nevertheless, the black-box features would hinder the understanding of physical processes in a lake, for example, conservation of mass/moment. Simulation of physical processes is of critical importance for lake management. The hydrodynamic model can be used to physically describe the water level fluctuation in response to the high inflow discharge during the typhoon events with certain accuracy. For lake water level prediction, this study proposed a novel and accurate approach which combines the three-dimensional hydrodynamic model (i.e., a physically based model) with the ANN model (i.e., a black-box data-driven model), providing an alternative way for accuracy improvement in addition to the data assimilation methodology [24, 25]. The comparison between the present hybrid approach and the data assimilation method deserves further discussion and leaves an interesting topic for future research.

6. Conclusions

Four modelling approaches (the three-dimensional hydrodynamic model, ANN model, ARMAX model, and the combination model) have been implemented to predict water level fluctuation of YYL in north-central Taiwan. The measured data from August 1 to December 31, 2009 (a total of 3672 data sets), was used for calibration (training), while the data from January 1 to May 31, 2010 (a total of 3624 data sets), was adopted for validation. The performance of these models was

comprehensively evaluated by various statistical indices (i.e., MAE, RMSE, R , and SS).

The three-dimensional hydrodynamic model satisfactorily presents the temporal variability of water level fluctuation in the calibration phase but somewhat fails to mimic the water level in the validation phase. The ANN model preserving nonlinear characteristics between input and output variables can predict water level fluctuation in both training and validation phases. The ARMAX model predicts the water levels using linear stochastic characteristics. Overall, the accuracy of the three-dimensional hydrodynamic model and ARMAX model is less than that of the ANN model. The development of an ANN model simply requires the time-series of field measurement data, and the implementation of an ANN model does not require iterative computation. Therefore, the ANN model can be accomplished with much less effort in comparison to the three-dimensional hydrodynamic model for a lake. However, the ANN is a black-box model, while the hydrodynamic model is a physically based model. One novel idea is proposed to combine the three-dimensional hydrodynamic and ANN models to predict water level fluctuation in the lake. The water level results predicted by the three-dimensional hydrodynamic model were amended using the ANN model for further accuracy (rather than the widely used data assimilation methods [24, 25]). The overall results show that the combination model yields better prediction results for water level fluctuation. Currently, a systematic study for more detailed internal physics (circulation and vertical mixing/stratification), water quality, and ecological processes in YYL is ongoing based upon this modelling methodology and will be reported soon in the near future.

Conflict of Interests

The authors declare that there is not conflict of interests regarding the publication of this paper.

Acknowledgments

This research was conducted with the support of the National Science Council and Academia Sinica, Taiwan, Grants nos. 101-2625-M-239-001 and AS-103-TP-B15, respectively. This financial support is greatly appreciated. The authors sincerely thank two anonymous reviewers for their valuable comments to substantially improve this paper.

References

- [1] H. Hofmann, A. Lorke, and F. Peeters, "Temporal scales of water-level fluctuations in lakes and their ecological implications," *Hydrobiologia*, vol. 613, no. 1, pp. 85–96, 2008.
- [2] M. T. Coe and J. A. Foley, "Human and natural impacts on the water resources of the Lake Chad basin," *Journal of Geophysical Research D: Atmospheres*, vol. 106, no. 4, pp. 3349–3356, 2001.
- [3] R. M. Usmanova, "Aral Sea and sustainable development," *Water Science and Technology*, vol. 47, no. 7-8, pp. 41–47, 2003.
- [4] J. K. Bourne, K. Joel, and G. Ludwig, "Eccentric Salton sea," *Natural Geographic*, vol. 207, no. 6, pp. 88–107, 2005.
- [5] M. Brauns, X. F. Garcia, and M. T. Pusch, "Potential effects of water-level fluctuations on littoral invertebrates in lowland lakes," *Hydrobiologia*, vol. 613, no. 1, pp. 5–12, 2008.
- [6] K. M. Wantzen, K.-O. Rothhaupt, M. Mörtl, M. Cantonati, L. G.-Tóth, and P. Fischer, "Ecological effects of water-level fluctuations in lakes: an urgent issue," *Hydrobiologia*, vol. 613, no. 1, pp. 1–4, 2008.
- [7] C. R. Singh, J. R. Thompson, D. G. Kingston, and J. R. French, "Modelling water-level options for ecosystem services and assessment of climate change: Loktak Lake, northeast India," *Hydrological Sciences Journal*, vol. 56, no. 8, pp. 1518–1542, 2011.
- [8] J. Xiang, H. Chen, and H. Ishikawa, "A mathematical model for the transmission of *Schistosoma japonicum* in consideration of seasonal water level fluctuations of Poyang Lake in Jiangxi, China," *Parasitology International*, vol. 62, no. 2, pp. 118–126, 2013.
- [9] M. S. Brennwald, F. Peeters, D. M. Imboden et al., "Atmospheric noble gases in lake sediment pore water as proxies for environmental change," *Geophysical Research Letters*, vol. 31, no. 3, Article ID L04202, 2004.
- [10] N. Kimura, W.-C. Liu, C.-Y. Chiu, T. K. Kratz, and W.-B. Chen, "Real-time observation and prediction of physical processes in a typhoon-affected lake," *Paddy and Water Environment*, vol. 10, no. 1, pp. 17–30, 2012.
- [11] R. O. Anyah and F. Semazzi, "Idealized simulation of hydrodynamic characteristics of Lake Victoria that potentially modulate regional climate," *International Journal of Climatology*, vol. 29, no. 7, pp. 971–981, 2009.
- [12] Y. R. Rao and J. Zhao, "Numerical simulation of the influence of a Red River flood on circulation and contaminant dispersion in Lake Winnipeg," *Natural Hazards*, vol. 55, no. 1, pp. 51–62, 2010.
- [13] D. Trolle, D. P. Hamilton, C. A. Pilditch, I. C. Duggan, and E. Jeppesen, "Predicting the effects of climate change on trophic status of three morphologically varying lakes: implications for lake restoration and management," *Environmental Modelling & Software*, vol. 26, no. 4, pp. 354–370, 2011.
- [14] L. Liu, Z. X. Xu, N. S. Reynard, C. W. Hu, and R. G. Jones, "Hydrological analysis for water level projections in Taihu Lake, China," *Journal of Flood Risk Management*, vol. 6, no. 1, pp. 14–22, 2013.
- [15] V. S. Ediger and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," *Energy Policy*, vol. 35, no. 3, pp. 1701–1708, 2007.
- [16] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland, Calif, USA, 1976.
- [17] I. A. Basheer and M. Hajmeer, "Artificial neural networks—fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, 2000.
- [18] T. A. Awchi and D. K. Srivastava, "Analysis of drought and storage for Mula project using ANN and stochastic generation models," *Hydrology Research*, vol. 40, no. 1, pp. 79–91, 2009.
- [19] A. Altunkaynak, "Forecasting surface water level fluctuations of lake van by artificial neural networks," *Water Resources Management*, vol. 21, no. 2, pp. 399–408, 2007.
- [20] A. Yasar, M. Onucyildiz, and N. K. Copty, "Modelling level change in lakes using neuro-fuzzy and artificial neural networks," *Journal of Hydrology*, vol. 365, no. 3-4, pp. 329–334, 2009.
- [21] M. A. Ghorbani, R. Khatibi, A. Aytek, O. Makarynskyy, and J. Shiri, "Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks," *Computers & Geosciences*, vol. 36, no. 5, pp. 620–627, 2010.
- [22] O. Kisi, J. Shiri, and B. Nikoofar, "Forecasting daily lake levels using artificial intelligence approaches," *Computers & Geosciences*, vol. 41, pp. 169–180, 2012.
- [23] O. Kisi and J. Shiri, "Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations," *Hydrology Research*, vol. 43, no. 3, pp. 286–300, 2012.
- [24] J. V. T. Sørensen, H. Madsen, and H. Madsen, "Efficient Kalman filter techniques for the assimilation of tide gauge data in three-dimensional modeling of the North sea and Baltic sea system," *Journal of Geophysical Research-Oceans*, vol. 109, p. C03017, 2004.
- [25] J. V. T. Sørensen and H. Madsen, "Parameter sensitivity of three Kalman filter schemes for assimilation of water levels in shelf sea models," *Ocean Modelling*, vol. 11, no. 3-4, pp. 441–463, 2006.
- [26] J. W. Tsai, T. K. Kratz, P. C. Hanson et al., "Metabolic changes and the resistance and resilience of a subtropical heterotrophic lake to typhoon disturbance," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 68, no. 5, pp. 768–780, 2011.
- [27] J. M. Hamrick, "A three-dimensional environmental fluid dynamics computer code: theoretical and computational aspects," Special Report 317, The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, Va, USA, 1992.
- [28] J. M. Hamrick, "User's manual for the environmental fluid dynamics computer code," Special Report in Marine Science and Ocean Engineering no. 331, The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, Va, USA, 1996.
- [29] Z.-G. Ji, G. Hu, J. Shen, and Y. Wan, "Three-dimensional modeling of hydrodynamic processes in the St. Lucie Estuary," *Estuarine, Coastal and Shelf Science*, vol. 73, no. 1-2, pp. 188–200, 2007.

- [30] W.-C. Liu, J.-T. Kuo, C.-C. Young, and M.-C. Wu, "Evaluation of marine outfall with three-dimensional hydrodynamic and water quality modeling," *Environmental Modeling and Assessment*, vol. 12, no. 3, pp. 201–211, 2007.
- [31] S. Elçi, A. Bor, and A. Calskan, "Using numerical models and acoustic methods to predict reservoir sedimentation," *Lake and Reservoir Management*, vol. 25, no. 3, pp. 297–306, 2009.
- [32] Y. Li, K. Acharya, and Z. Yu, "Modeling impacts of Yangtze River water transfer on water ages in Lake Taihu, China," *Ecological Engineering*, vol. 37, no. 2, pp. 325–334, 2011.
- [33] S. Sinha, X. Liu, and M. H. Garcia, "Three-dimensional hydrodynamic modeling of the Chicago River, Illinois," *Environmental Fluid Mechanics*, vol. 12, no. 5, pp. 471–494, 2012.
- [34] G. L. Mellor and T. Yamada, "Development of a turbulence closure model for geophysical fluid problems," *Reviews of Geophysics & Space Physics*, vol. 20, no. 4, pp. 851–875, 1982.
- [35] A. F. Blumberg and G. L. Mellor, "A description of a three-dimensional coastal ocean circulation model," in *Three-Dimensional Coastal Ocean Models*, N. S. Heaps, Ed., Coastal and Estuarine Sciences, Book 4, pp. 1–19, American Geophysical Union, 1987.
- [36] B. Galperin, L. H. Kantha, S. Hassid, and A. Rosati, "A quasi-equilibrium turbulent energy model for geophysical flows," *Journal of the Atmospheric Sciences*, vol. 45, no. 1, pp. 55–62, 1988.
- [37] A. Arakawa and V. R. Lamb, "Computational design of the basic dynamical processes of the UCLA general circulation model," in *Methods in Computational Physics*, vol. 17, pp. 174–265, Academic Press, New York, NY, USA, 1973.
- [38] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [39] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
- [40] R. K. Panda, N. Pramanik, and B. Bala, "Simulation of river stage using artificial neural network and MIKE 11 hydrodynamic model," *Computers & Geosciences*, vol. 36, no. 6, pp. 735–745, 2010.

