

A Comparison of Three Approaches to Predict Phytoplankton Biomass in Gonghu Bay of Lake Taihu

J. C. Huang¹, J. F. Gao^{1,*}, W. M. Mooij^{2,3}, G. Hörmann⁴, and N. Fohrer⁴

¹Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

²Netherlands Institute of Ecology (NIOO-KNAW), Department of Aquatic Ecology, Wageningen 6700 AB, Netherlands

³Wageningen University, Department of Aquatic Ecology and Water Quality Management, Wageningen 6700 AA, Netherlands

⁴Department of Hydrology and Water Resources Management, Institute of Natural Resources Conservation, Kiel University, Kiel 24118, Germany

Received 21 August 2012; revised 20 December 2012; accepted 20 March 2013; published online 12 June 2014

ABSTRACT. Algal blooms have caused severe problems in Lake Taihu, China. Early warning of phytoplankton accumulation can support decision-making against harmful algal bloom events. To investigate the performance of different models in forecasting high phytoplankton biomass, we developed a mechanistic, a regression and three artificial neural network (ANN) models to predict short-term (3 days) changes of phytoplankton biomass (expressed as chlorophyll-*a* concentration) in Gonghu Bay of Lake Taihu. We determined the input variables of the ANN models with a sensitivity analysis, and optimized their parameters with a trial-and-error approach. The sensitivity analysis revealed the effects of the input variables on phytoplankton biomass. To calibrate and validate the models, we collected two data sets of Lake Taihu in 2009: hourly-averaged data collected by an automatic monitoring system and field data with a sampling interval of twice a week. Although the sensitivity analysis results vary among the five models, there is a general consensus that phytoplankton changes are significantly affected by water temperature in Gonghu Bay. The ANN models obtained good model fit indicating their practical values in predicting non-linear phytoplankton dynamics for water management purpose. The mechanistic model predicted the phytoplankton distribution dynamically and described the variable interactions explicitly. The regression model is characterized by its easy development. This comparison study assists the modelers in selecting an approximate model for their specific purposes.

Keywords: artificial neural network, Gonghu Bay, mechanistic model, phytoplankton biomass, regression model

1. Introduction

Severe algal blooms have occurred and caused deterioration of water quality in Lake Taihu, a large and shallow lake in China (Guo, 2007; Qin et al., 2007; Yang et al., 2008). A previous study based on two-decade satellite images revealed that algal blooms in the Lake Taihu were overspreading from the northwest to the southeast, from the border area to the center area, with a peak distribution area (979.1 km², 41.9% area of the lake) of algal blooms in late June, 2007 (Duan et al., 2009). In late May 2007, algal bloom accumulation in drinking water intake areas of Gonghu Bay (a bay in north-eastern Lake Taihu) led to a major drinking water crisis in surrounding cities (Guo, 2007; Qin et al., 2010). Early warnings of phytoplankton accumulation can help lake managers to take timely emergency measures to avoid or minimize the negative impact of harmful algal bloom events in sensitive

areas (e.g., drinking water intakes) (Huang et al., 2012a). Many models have been developed for predicting phytoplankton biomass (e.g., Jørgensen, 1976; Hamilton and Schladow, 1997; Recknagel et al., 1997; Jeong et al., 2006; Mooij et al., 2010), which is a biological indicator of algal blooms. Regression, mechanistic and artificial neural network (ANN) models are most widely used.

Mechanistic models are most commonly used to predict phytoplankton biomass (Hamilton and Schladow, 1997; Jørgensen, 2001; Arhonditsis and Brett, 2005; Jørgensen, 2010). They provide a better understanding of biological and physical processes. In order to develop a good mechanistic model, all underlying physical and biological processes should be understood adequately (Maier et al., 1998). However, in reality, many of the complex and highly nonlinear processes affecting phytoplankton biomass are not adequately understood. Therefore, it is still challenging to predict phytoplankton biomass with mechanistic models. Moreover, mechanistic models may contain a large number of parameters, which increase model uncertainty (Maier et al., 1998).

Regression models have been widely used to predict phytoplankton biomass due to their simplicity and straight-

* Corresponding author. Tel.: +86 25 86882124; fax: +86 25 86882124.
E-mail address: gaojunf@niglas.ac.cn (J. F. Gao).

forward interpretation (Lek et al., 1996; Whitehead et al., 1997; Jeong et al., 2006). However, they perform poorly when relationships between variables are nonlinear (Lek et al., 1996), and their applications are constrained by the complex ecological data (Graham, 2003).

Artificial neural network (ANN) models have been increasingly used to forecast phytoplankton biomass of rivers (e.g., Whitehead et al., 1997; Maier et al., 1998; Jeong et al., 2001), lakes (e.g., Recknagel et al., 1997; Recknagel et al., 1998; Karul et al., 2000) and coastal systems (e.g., Barciela et al., 1999; Lee et al., 2003). ANN models commonly achieve a better model fit than conventional regression or mechanistic models (Lek et al., 1996; Barciela et al., 1999), mainly due to their potential for modelling highly nonlinear and complex natural systems (Recknagel, 2001; Lee et al., 2003). However, due to the lack of time series data with an appropriate resolution, ANN has been rarely used to model phytoplankton biomass of Lake Taihu. Automatic monitoring systems in the lake can overcome this lack of data by monitoring water quality in real time (Lee and Lee, 1995; Hull et al., 2008).

Given the different performances of the above-mentioned approaches, a comparison based on the same case study is needed. The main objectives of this paper are to forecast short-term (a few days) changes of phytoplankton biomass in Gonghu Bay with a mechanistic, a regression and three ANN models, and to compare their performance in predicting phytoplankton biomass. We developed three ANN models with different network structures and input variables. We compared the model fits in predicting dynamics of phytoplankton biomass in Gonghu Bay, a bay with several drinking water intakes located in it. The effects of the input variables on phytoplankton were discussed based on a sensitivity analysis.

2. Materials and Methods

2.1. Study Area

Lake Taihu is a large (surface area, 2,338 km²), shallow (mean depth, 1.9 m) and eutrophic lake in China (Figure 1). The water volume of the lake is 4.4 × 10⁹ m³ and the mean hydraulic retention time is 284 days (Hu et al., 2006). The lake is located in a plain with a complicated river and channel

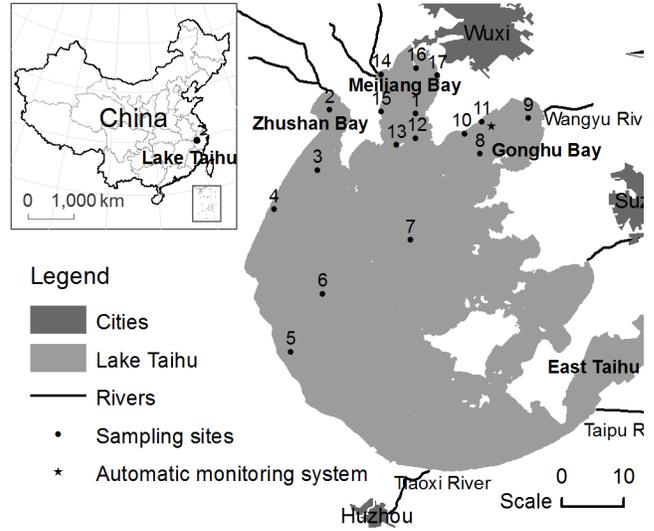


Figure 1. Locations of Gonghu Bay, automatic monitoring system and sampling sites.

network and is important for water supply, flood control, tourism and recreation, fisheries, shipping and aquaculture and irrigation (Hu et al., 2006; Qin et al., 2007). Harmful algal blooms in Lake Taihu have posed a direct threat to public health (Guo, 2007). Gonghu Bay is located in northeast of Lake Taihu (Figure 1). Since Gonghu Bay is a drinking water intake, increasing attention has been paid to its water quality.

2.2. Data Collection and Preprocessing

Two data sets collected in Lake Taihu between April 22 and September 30, 2009 were used for model training and testing (Table 1). One data set (Data set 1) was obtained every five minutes by an automatic monitoring system (Figure 1) located in Gonghu Bay. The data set includes eight variables, namely chlorophyll-*a* concentration (µg L⁻¹), wind direction (°), solar radiation (W m⁻²), water temperature (°C), precipitation (mm), dissolved oxygen (mg L⁻¹), turbidity (NTU) and pH. Wind speed was not included in this data set. However, previous studies show that it is positively correlated with tur-

Table 1. List of Datasets Used in This Study

Data	Item	Time period	Temporal resolution	Sampling sites
Data set 1	(1) chlorophyll- <i>a</i> concentration (2) wind direction (3) solar radiation (4) water temperature (5) precipitation (6) dissolved oxygen (7) turbidity (8) pH	Apr. 22, 2009-Sep. 30, 2009	Once every five minutes	1
Data set 2	(1) chlorophyll- <i>a</i> concentration (2) total dissolved phosphorus (3) total dissolved nitrogen	Apr. 22, 2009-Sep. 30, 2009	Twice a week Once a week Once a week	17

bidity (e.g., Li et al., 2007). Daily-averaged data in 158 days were then derived for the ANN, regression and mechanistic models.

The second data set (Data set 2) includes vertically-averaged chlorophyll-*a* concentration ($\mu\text{g L}^{-1}$), total dissolved phosphorus and nitrogen (mg L^{-1}) for 17 sites in Lake Taihu (sites 1 to 17 in Figure 1), with seven sites in Meiliang Bay, four sites in Gonghu Bay, one site in Zhushan Bay, three sites in the western area and two sites in the center area of Lake Taihu. Chlorophyll-*a* was measured at approximately 10:00 h, with a sampling frequency of 3 and 4 days. Total dissolved phosphorus and nitrogen were measured once a week. The spatial distribution of these three variables in Lake Taihu was interpolated using the inverse distance weighted (IDW) method. The chlorophyll-*a* data from 17 sampling sites and the automatic monitoring system (Figure 1) were used as the initial data of the mechanistic model. Daily nutrient data (i.e., dissolved phosphorus and nitrogen) in Gonghu Bay were obtained based on linear interpolation.

2.3. Artificial Neural Network Models

The ANN models simulated a three-day forward chlorophyll-*a* concentration (chlorophyll-*a* concentration at time $t+3$) with a time step of one day. In order to ensure that all variables receive equal attention during the training, testing and validation process, both input and output variables were scaled to the range [0.1, 0.9] with the following linear transformation:

$$x'_i = x'_{min} + \frac{(x_i - x_{min})(x'_{max} - x'_{min})}{x_{max} - x_{min}} \quad (1)$$

where x_i and x'_i are the measured and normalized values, respectively; x_{max} and x_{min} are the maximum and minimum values of the measured data, respectively; x'_{max} and x'_{min} are the maximum and minimum values of the normalized range ($x'_{max} = 0.9$ and $x'_{min} = 0.1$), respectively.

Data set 1 (158 days) was randomized and divided into three subsets. The training, testing and validation data sets comprised of the data in 121 (76.6%), 13 (8.2%) and 24 (15.2%) days, respectively. The data partition proportion has been commonly used in previous studies (Maier et al., 2000). Cross-validation was used in this study with the selection of the training, testing and validation data sets showed in Figure 2. The data in the first 24 days were firstly used for validation. Then, the data in sequence days (25 to 48, 49 to 72, 73 to 96, 97 to 120, 121 to 144, and 145 to 158) were used in turn for validation. This data selection method ensured that the data in each day have been used for validation. This cross-validation method resulted in seven ANN models (f_i in Figure 2, $i=1, 2, \dots, 7$) for each network structure. The model output (Chl_{t+3}) was the mean output of these seven ANN models, i.e.,

$$Chl_{t+3} = \sum_{i=1}^7 f_{ANN_i} / 7 \quad (2)$$

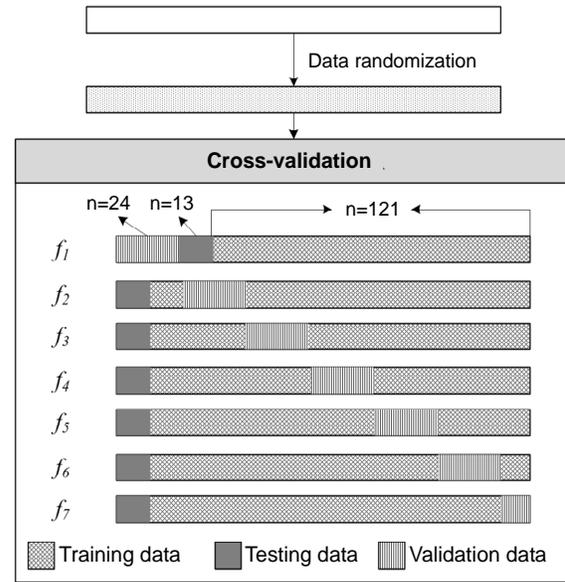


Figure 2. Procedure of the cross-validation used in ANN models.

Further details about the cross-validation method could be found in previous studies (e.g., Burden et al., 1997; Stone, 1974).

An autocorrelation analysis (Figure 3) suggests that the coefficient of determination (r^2) between Chl_{t+3} (chlorophyll-*a* concentration at time $t+3$) and Chl_{t-i} (chlorophyll-*a* concentration at time $t-i$ ($i = 0, 1, 2, \dots, 13$)) decreases when i is increased. Three time-lagged chlorophyll-*a* concentrations (Chl_t , Chl_{t-1} and Chl_{t-2}) with r^2 -values higher than 0.4 were selected to develop ANN models. In order to compare model performance with different structures, we developed three ANN models with input and output variables shown in Figure 4. Twelve variables were used as input variables of the most complicated ANN model (ANN 1). In order to compare the performance between the ANN and regression models, five input variables (Chl_t , Chl_{t-1} , Chl_{t-2} , total dissolved phosphorus and water temperature) used in the regression model were selected to develop ANN 2. The variable (Chl_t) with the highest sensitivity in the regression model was selected to

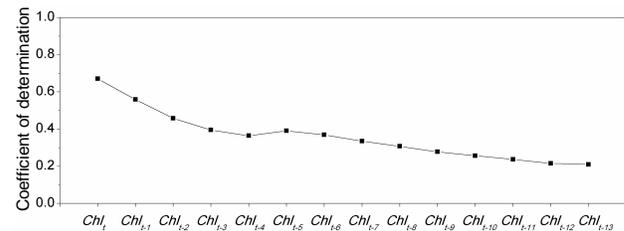


Figure 3. The coefficient of determination between chlorophyll-*a* concentration at time $t+3$ (Chl_{t+3}) and chlorophyll-*a* concentration at time $t-i$ ($i = 0, 1, 2, \dots, 13$) (Chl_{t-i}).

develop ANN 3.

All the three ANN models consist of an input layer, a hidden layer and an output layer (Figure 4). This structure is the most flexible neural network with good performance (Hornik et al., 1989; Scardi and HardingJr, 1999; Maier and Dandy, 2000). The input layer consists of the input variables with a linear transfer function, and the output layer includes a single node of chlorophyll-*a* concentration at time $t + 3$ with a sigmoid transfer function. The hidden layer links the input and output variables by its nodes (neurons). The node number of the hidden layer was determined by a trial-and-error approach, and varied from 2 to 20. Eight, six and two hidden layers for the three ANN models were found to be the optimal selection.

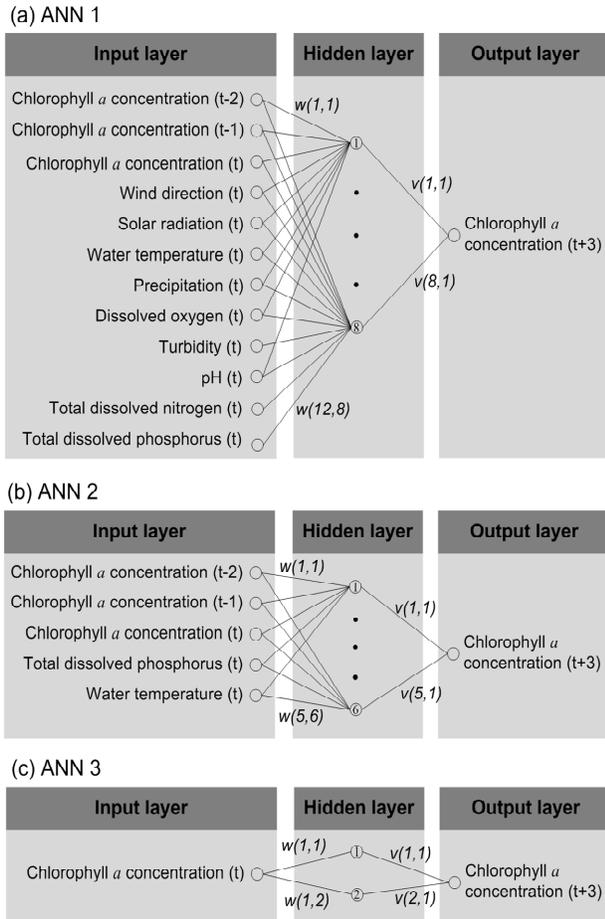


Figure 4. Neural network structures for ANNs 1-3. (a) ANN 1, twelve input variables; (b) ANN 2, six input variables; (c) ANN 3, one input variable.

The ANN models were trained with Levenberg-Marquardt backpropagation algorithm, in which the weights (w_{n+1}) were adjusted during training using the following equation (Parisi et al., 1996):

$$w_{n+1} = w_n - \eta_n d_n \quad (3)$$

where w_n is the weight vector of the network, η_n is the step size, d_n is a vector defining the direction of descent, and n is the iteration number. The training process was stopped when the errors for the testing set began to rise, or the training error (mean squared error) was lower than the goal value (0.0005), or a maximum of 1000 iterations was reached. This early stopping rule prevents the ANN models from overfitting. Seven ANN models of ANN 1 and two ANN models of ANN 2 reached the goal. None of the ANN models reached the maximum iterations.

2.4. Mechanistic Model

A coupled hydrodynamic-phytoplankton model was used to predict the three-day-ahead chlorophyll-*a* concentration at the location of the automatic monitoring system (Figure 1). It includes a phytoplankton module, a two-dimensional hydrodynamic module and a mass-transport module with their interactions given in Figure 5. The model has been calibrated by Huang et al. (2012a) with its parameters listed in Table 2. Data sets 1 and 2, including spatial information, were used for the mechanistic model due to its requirement of spatial input data.

The phytoplankton module includes phytoplankton growth, mortality, respiration, excretion, grazing and sinking with a time step of one day (Figure 5 (a)). This module requires spatial distribution of chlorophyll-*a* concentration as initial data. Forcing data of solar radiation, water temperature, total dissolved nitrogen and phosphorus were required to drive this module. Phosphorus and nitrogen conditions were not simulated in this module, but taken from measured data.

The two-dimensional hydrodynamic module simulated the vertically-averaged water velocity (referred to as “water flows” in Figure 5) with the forcing data of wind conditions. Water temperature was not simulated in the hydrodynamic module because measured data of water temperature were available. The mass-transport module describes phytoplankton transport between grid cells based on the water flows generated by the hydrodynamic module. A time step of 200 seconds was used in the hydrodynamic module and the phytoplankton transport module. Further details of this hydrodynamic-phytoplankton model can be found in Huang et al. (2012a).

2.5. Regression Models

The environmental variables of ANN 1 (Figure 4a) were used as independent variables to predict the dependent variable (chlorophyll-*a* concentration at time $t + 3$). Input and output variables were linearly scaled to the range [0.1, 0.9] by Equation 1. We used the forward stepwise selection method to

select the variables for the regression models. To compare the performance with ANN models, the regression model used the same validation data set (Figure 2) as ANN models. Other data (training and testing data set in Figure 2) was used for calibration. The cross-validation method described in Figure 2 were used in developing regression models, and resulted in the following seven linear equations:

$$Chl_{t+3} = f_{RM_1} = 0.941Chl_t + 0.112TDP_t - 0.425Chl_{t-1} - 0.129T_t + 0.107 \quad (4)$$

$$Chl_{t+3} = f_{RM_2} = 0.735Chl_t + 0.106TDP_t - 0.248Chl_{t-2} - 0.12T_t + 0.104 \quad (5)$$

$$Chl_{t+3} = f_{RM_3} = 0.726Chl_t + 0.114TDP_t - 0.232Chl_{t-2} - 0.154T_t + 0.124 \quad (6)$$

$$Chl_{t+3} = f_{RM_4} = 0.812Chl_t + 0.095TDP_t - 0.305Chl_{t-2} - 0.147T_t + 0.125 \quad (7)$$

$$Chl_{t+3} = f_{RM_5} = 0.893Chl_t + 0.118TDP_t - 0.37Chl_{t-1} - 0.119T_t + 0.095 \quad (8)$$

$$Chl_{t+3} = f_{RM_6} = 0.738Chl_t + 0.034 \quad (9)$$

$$Chl_{t+3} = f_{RM_7} = 0.969Chl_t + 0.093TDP_t - 0.433Chl_{t-1} - 0.104T_t + 0.094 \quad (10)$$

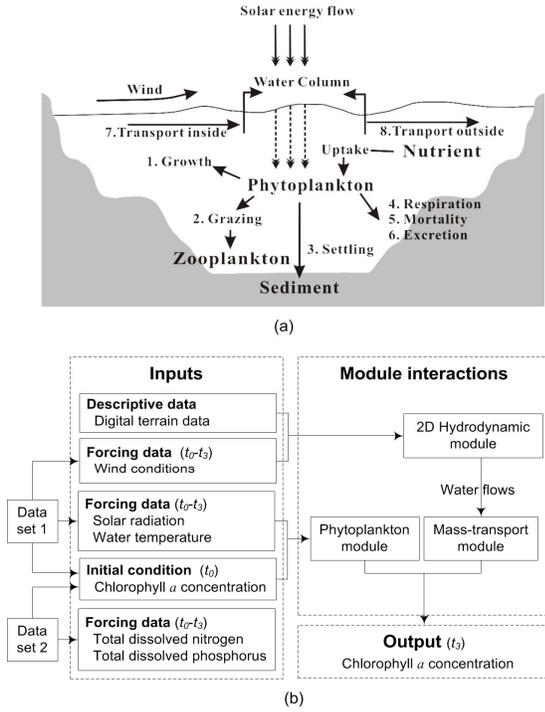


Figure 5. Conceptual diagram (a) (redrawn from Huang et al. (2012a)) and module interactions (b) of the hydrodynamic-phytoplankton model.

where Chl_{t+3} , Chl_t , Chl_{t-1} and Chl_{t-2} are the chlorophyll-*a* concentration at time $t + 3$, t , $t - 1$ and $t - 2$, respectively, TDP_t and T_t are the total dissolved phosphorus and water temperature at time t , respectively. The coefficients in these equations were determined by minimizing the sum of squared differences between simulated and measured chlorophyll-*a* in the calibration period. The validation data set used in ANNs was used to validate the regression model. We compared the pre-

Table 2. Parameters Used in the Hydrodynamic-phytoplankton Model

Symbols	Parameters	Values
U_{max}	Maximum growth rate of the phytoplankton	1.145 d ⁻¹
T_{opt}	Optimum water temperature	27.5 °C
α	Average spectral extinction coefficient for the absorption of short-wave radiation by water and its non-phytoplankton component	0.45 m ⁻¹
β	Average spectral extinction coefficient for the absorption of short wave radiation by the phytoplankton	0.016 m ² (mg Chl) ⁻¹
I_{opt}	Saturation light intensity	12 MJ m ⁻² d ⁻¹
KP	Michaelis constant for phosphorus uptake	0.01 mg L ⁻¹
KN	Michaelis constant for nitrogen uptake	0.022 mg L ⁻¹
K	Sinking velocity of the phytoplankton	0.0864 m d ⁻¹
GR_{max}	Maximum grazing rate of the zooplankton	0.09 d ⁻¹
F_{min}	Minimum concentration of phytoplankton available for grazing	100 µg L ⁻¹
F_s	Michaelis-Menten constant of phytoplankton available for grazing	500 µg L ⁻¹
k_m	Rate coefficient for the chlorophyll α loss caused by mortality	0.027 d ⁻¹
k_r	Rate coefficient for the chlorophyll α loss caused by respiration	0.17 d ⁻¹
k_e	Rate coefficient for the chlorophyll α loss caused by excretion	0.01 d ⁻¹
ϑ	Temperature multipliers	1.08

dictive power of the regression model with other models based on the mean simulation results of Equations 4 to 10, i.e.,

$$\begin{aligned} Chl_{t+3} &= \sum_{i=1}^7 f_{RM_i} / 7 \\ &= 0.831Chl_t - 0.175Chl_{t-1} - 0.131Chl_{t-2} \\ &\quad + 0.091TDP_t - 0.110T_t + 0.098 \end{aligned} \quad (11)$$

In order to investigate the correlation between dependent variable and independent variables, we used a partial correlation analysis to calculate the correlation coefficient of an individual variable with dependent variable, the other variables being constant.

2.6. Model Simulation and Sensitivity Analysis

The ANN, regression and mechanistic models were implemented in Matlab, SPSS and a phytoplankton prediction system for Lake Taihu (Taihu PPS), respectively (Huang et al., 2012b). All these models simulated three-day-ahead chlorophyll-*a* concentration between April 22 and September 30, 2009 at the location of the automatic monitoring system in Gonghu Bay (Figure 1). This three-day-ahead prediction aims to provide early warning of algal blooms for water resource managers, and allows them to take precautionary measures in time to reduce loss. To compare their model fits, root mean square error (*RMSE*), relative absolute error (*RE*) and the coefficient of determination (r^2) were calculated using:

$$RMSE = \sqrt{\sum_{i=1}^n (Chl_i^S - Chl_i^O)^2 / n} \quad (12)$$

$$RE = \sum_{i=1}^n \frac{|Chl_i^S - Chl_i^O|}{Chl_i^O} \times 100 / n \quad (13)$$

$$r^2 = \left(\frac{\sum_{i=1}^n (Chl_i^O - Chl_{Mean}^O)(Chl_i^S - Chl_{Mean}^S)}{\sqrt{\sum_{i=1}^n (Chl_i^O - Chl_{Mean}^O)^2} \sqrt{\sum_{i=1}^n (Chl_i^S - Chl_{Mean}^S)^2}} \right)^2 \quad (14)$$

where Chl_i^S and Chl_i^O are the simulated and observed chlorophyll-*a* concentrations on day *i*, respectively, Chl_{Mean}^S and Chl_{Mean}^O are the mean values of simulated and observed chlorophyll-*a* concentrations, respectively, and *n* is the number of observations. We aim to obtain a good-fit model with low *RMSE* and *RE*, and r^2 approaching 1.

Furthermore, we carried out a sensitivity analysis to quantify the effects of input variables on the short-term phytoplankton dynamics. Each input variable was increased or decreased by 10%. All other variables were kept fixed. A total of 2*m* (*m* is the number of the testing variables) scenarios

were simulated and compared with the base scenario. All the input variables in ANNs 1 and 2 and regression model were tested. In the mechanistic model, five input variables (i.e., water temperature, wind speed, wind direction, dissolved phosphorus and nitrogen) were tested. The sensitivity of each testing variable was given by:

$$\text{Sensitivity} = \frac{\text{change in output (\%)}}{\text{change in input (\%)}} \times 100\% \quad (15)$$

3. Results

3.1. Chlorophyll-*a* Prediction

The agreement between the measured and simulation chlorophyll-*a* varied considerably among the ANN, mechanistic and regression models. Figure 6 shows the chlorophyll-*a* changes at the automatic monitoring system (Figure 1). In order to compare the model performance at the low chlorophyll-*a* concentration period (between June 1 and September 30), figures with a Y-axis scale from 0 to 9 were given in Figure 6. The absolute errors of these five models were given in Figure 7.

ANN 1 has the best model fit with the mean root mean square error (*RMSE*) of 3.14 $\mu\text{g L}^{-1}$ in validation periods. The ANN model captured the chlorophyll-*a* peaks on April 28 and May 23, and simulated the chlorophyll-*a* changes in low chlorophyll-*a* period (between June 1 and September 30) better than other models. ANN 2 simulated the chlorophyll-*a* peak on April 28 adequately. However, the ANN model failed to capture the chlorophyll-*a* peak on May 23 and June 30. Both ANN 3 and the regression model failed to reproduce chlorophyll-*a* peaks well in the modeling periods. The performance of the mechanistic model is not as good as ANN 1 and ANN 2, considering its large *RMSE* (5.44 $\mu\text{g L}^{-1}$) and low r^2 value ($r^2 = 0.65$).

The model fits of these models were further evaluated with three measures of goodness-of-fit, *RMSE*, *RE* and r^2 . Table 3 shows that ANN 1 has a higher degree of conformity than other models.

ANN 1 reproduced the chlorophyll-*a* changes well in the modeling periods, with lower *RE* and *RMSE* values, and higher r^2 values than other models. The *RE*, r^2 and *RMSE* values show that ANN 2 did not perform as well as ANN 1, but better than the regression model. ANN 3 was not well trained (*RMSE* = 6.08 $\mu\text{g L}^{-1}$). This resulted in a large deviation in the validation period (*RMSE* = 5.84 $\mu\text{g L}^{-1}$). The *RMSE* values in the training and validation periods (5.79 and 6.24 $\mu\text{g L}^{-1}$, respectively) show that the regression model did not simulate chlorophyll-*a* adequately. The large deviation of the regression model was mainly due to the failure to capture the chlorophyll-*a* peaks in the modeling period (Figure 6).

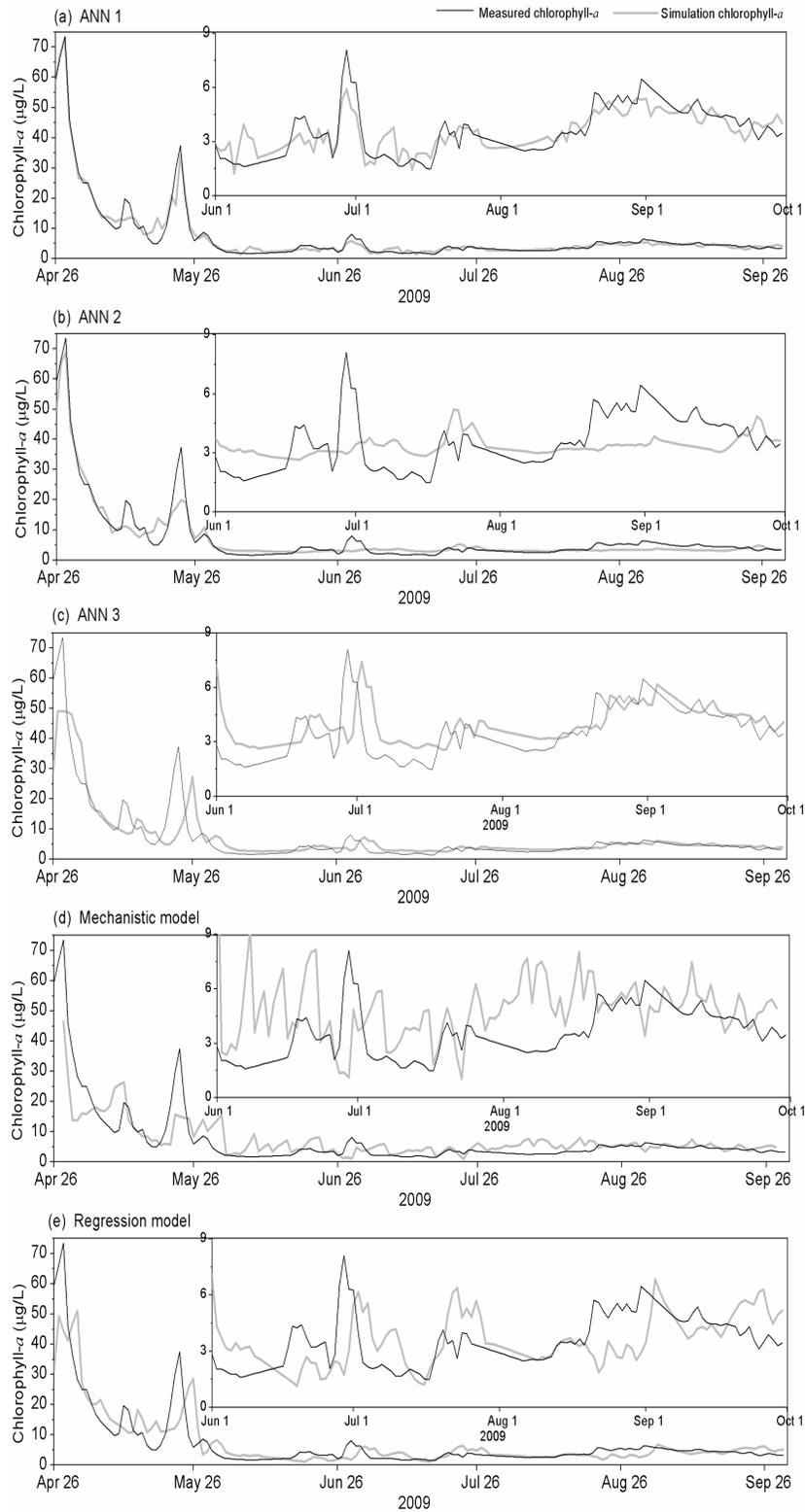


Figure 6. Measurement and simulation of chlorophyll- a concentration between April 26 and September 30, 2009 in (a) the artificial neural network model with 12 input variables (ANN 1), (b) the artificial neural network model with five input variables (ANN 2), (c) the artificial neural network model with one input variables (ANN 3), (d) the mechanistic model and (e) the regression model.

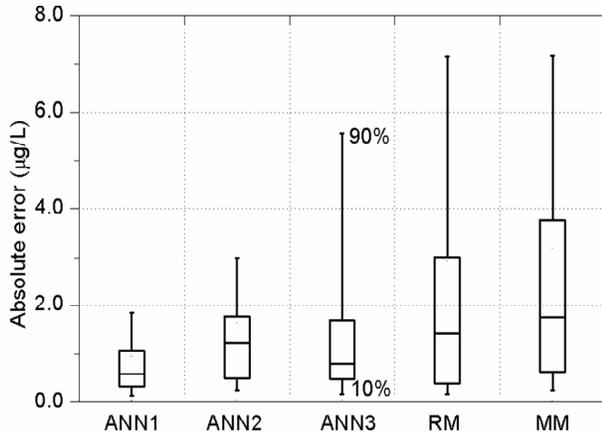


Figure 7. The absolute errors of ANNs 1-3, the regression and mechanistic models during the modeling period. RM and MM represent the regression and mechanistic models, respectively.

Table 3. Simulation Errors of the Five Models Measured by Root Mean Square Error (RMSE), Relative Absolute Error (RE) and coefficient of Determination (r^2)

		Training	Testing	Validation
ANN 1	RMSE ($\mu\text{g L}^{-1}$)	1.89	1.94	3.14
	RE (%)	25.4	40.8	38.6
	r^2	0.97	0.66	0.85
ANN 2	RMSE ($\mu\text{g L}^{-1}$)	3.23	2.59	4.23
	RE (%)	41.1	42.8	46.3
	r^2	0.92	0.57	0.78
ANN 3	RMSE ($\mu\text{g L}^{-1}$)	6.08	1.89	5.91
	RE (%)	36.1	42.8	41.2
	r^2	0.71	0.56	0.76
Regression model	RMSE ($\mu\text{g L}^{-1}$)	5.79	---	6.24
	RE (%)	45.1	---	47.3
	r^2	0.72	---	0.74
Mechanistic model	RMSE ($\mu\text{g L}^{-1}$)	---	---	5.44
	RE (%)	---	---	65.1
	r^2	---	---	0.65

3.2. Sensitivity Analysis

Sensitivities of the input variables in ANNs 1 and 2, the regression and mechanistic models are given in Figure 8. The coefficients and significance levels of the five variables in the regression models (Equations 4 to 10) are given in Table 4. Partial correlation coefficient of an individual variable is calculated to represent the contribution of the variable to chlorophyll-*a* concentration at time $t + 3$ (Table 4).

Five input variables (water temperature, dissolved oxygen, turbidity, total dissolved nitrogen and phosphorus) in ANN 1 had a sensitivity value higher than 60%. In ANN 2, water temperature is the most sensitive variables with a sensitivity value of 69.4% (Figure 8). In the mechanistic model, we tested sensitivities of the six variables: water temperature, solar radiation, wind speed, wind direction, total dissolved nitrogen and phosphorus. Chlorophyll-*a* concentration at time t (Chl_t), $t-1$ (Chl_{t-1}) and $t-2$ (Chl_{t-2}), total dissolved phos-

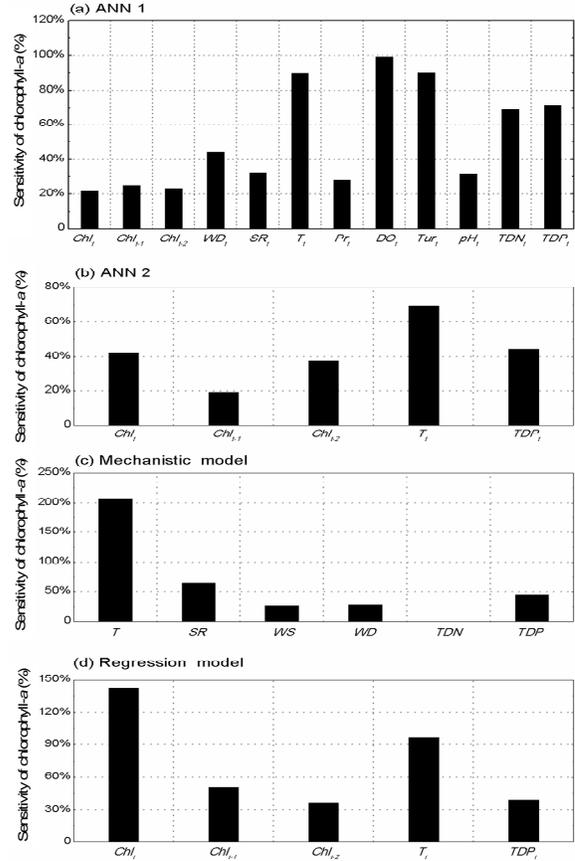


Figure 8. Sensitivity analysis of (a) ANN 1, (b) ANN 2, (c) the mechanistic and (d) regression models. Chl_t , Chl_{t-1} , and Chl_{t-2} represent chlorophyll-*a* concentration at time t , $t-1$, $t-2$; WD , wind direction; WS , wind speed; SR , solar radiation; T , water temperature; Pr , precipitation; DO , dissolved oxygen; Tur , turbidity; TDN , total dissolved nitrogen; TDP , total dissolved phosphorus.

phorus (TDP) and water temperature (T) were included in the regression model (Equation 11). The sensitivity value of Chl_t (142.5%) in the regression model and its partial correlation coefficient (0.452) show high sensitivity of Chl_t on phytoplankton.

Phytoplankton responds differently to different meteorological conditions. Phytoplankton growth is highly sensitive to water temperature in these four models shown in Figure 8. Wind direction and solar radiation have relative low sensitivities in ANN 1 and the mechanistic model. Sensitivity of precipitation is also low in ANN 1.

Sensitivities of the nutrient variables, i.e., total dissolved phosphorus and nitrogen, are significantly different between the ANN, regression and mechanistic models. In ANN 1, chlorophyll-*a* change is highly sensitive to total dissolved phosphorus and nitrogen. In the mechanistic and regression model, phytoplankton growth is slightly limited by total dissolved phosphorus. Phytoplankton growth is almost not limited by total dissolved nitrogen (sensitivity approximately equal to 0) in the mechanistic model.

4. Discussion

4.1. Performance of the ANN Models

The performance of the ANN models are significantly different mainly due to the difference in selection of the input parameters. ANN 1 reached a better fit than the mechanistic and regression models. This finding is in accordance with other studies using ANN approaches (e.g., Lek et al., 1996; Barciela et al., 1999). The r^2 values of the “best” ANN model (i.e., ANN 1) for training and validation periods are higher than 0.85, higher than the ones of the mechanistic models used in previous studies to predict phytoplankton biomass (e.g., Hu et al., 2006; Trolle et al., 2011). The good model fit of ANN model is probably due to its capability to describe complex nonlinear relationships. Other advantages of ANN models are the easy set up and that little priori knowledge is required to develop ANN models.

In this study, an exhaustive search approach was not used to obtain an optimal ANN model. This is because there are large subsets of different variable combinations ($2^n - 1$, n is the number of input variable). Instead, we selected the input variables used in the regression model to develop ANN 2 for performance comparison between ANN and regression models. The better performance of ANN 2 than the regression model demonstrates the usefulness of artificial neural network techniques in nonlinear prediction. ANN 2 was well trained with an r^2 value of 0.92 in the training period. However, ANN 2 did not perform as well as ANN 1. The different performances indicate that the seven excluded variables in ANN 2 (i.e., total dissolved nitrogen, wind direction, solar radiation, precipitation, dissolved oxygen, turbidity and pH) contribute to achieve a good model fit of chlorophyll-*a* prediction.

The model fit of ANN 3 is not as good as other two ANN models. This is probably due to the strong impact of environmental factors (e.g., meteorological conditions) on chlorophyll-*a* dynamics in Lake Taihu. Therefore, chlorophyll-*a* changes cannot be explained only by its time-lagged state.

This conclusion is different from Lee et al. (2003), that predicted chlorophyll-*a* changes in the coastal area with time-lagged chlorophyll-*a* as inputs, and achieved acceptable results.

The model fits did not vary considerably among the seven ANN models (Figure 2) for each of ANNs 1 to 3, suggesting the reliability of ANNs 1 to 3. Although some methodological issues (e.g., choice of performance criteria, data pre-processing, selection of input variables and determination of network architecture) have not yet been fully addressed (Bowden et al., 2002; Bowden et al., 2005; Srinivasulu and Jain, 2006; Maier et al., 2010), they are gradually overcome by the ANN modelers with different backgrounds (Kumar et al., 2011; Marini, 2009; Mohanraj et al., 2012). ANN approach is now increasingly applied in various areas, and is particularly promising for non-linear problems.

4.2. Performance of the Mechanistic Model

The mechanistic model has been calibrated by Huang et al. (2012a) with an additional data set of Lake Taihu collected in 2008. The model fit of the mechanistic model is not as high as that of ANNs 1 and 2 (Table 3). The different deviations may partly result from the different data sets used in the ANN and mechanistic models. ANN 1 included more environmental variables than the mechanistic model. These variables (e.g., precipitation), excluded in the mechanistic model, may be useful in predict phytoplankton dynamics.

The deviation of the mechanistic model is mainly due to the uncertainties of model input, structure and parameters (Huang et al., 2012a). For example, spatial and temporal resolution of input variables was inadequate for an appropriate representation of spatial heterogeneity. Moreover, modelers select the processes and algorithms based on their experience, which is not necessarily objective and optimal. Finally, parameter values were taken from the scientific literature and

Table 4. Correlation Analysis between Chlorophyll-*a* Concentration at Time $t+3$ and Other Environmental Variables

	Regression analysis		Partial correlation analysis	
	Coefficient	Significance level	Partial correlation coefficient	Significance level
Chl_t	[0.726, 0.969]	(0, 0.001)	0.452	< 0.001
Chl_{t-1}	[-0.37, -0.425]	(0, 0.006)	-0.070	0.400
Chl_{t-2}	[-0.305, -0.232]	(0, 0.012)	-0.083	0.320
PAR_t			-0.102	0.219
WD_t			0.158	0.057
T_t	[-0.154, -0.119]	(0, 0.015)	-0.162	0.050
pH_t			-0.027	0.743
DO_t			0.151	0.069
ZD_t			0.094	0.256
Pr_t			-0.009	0.912
TDN_t			-0.089	0.287
TDP_t	[0.093, 0.118]	(0.001, 0.003)	0.223	0.007

* Chl_t , Chl_{t-1} , and Chl_{t-2} represent chlorophyll-*a* concentration at time t , $t-1$, $t-2$; WD , wind direction; WS , wind speed; SR , solar radiation; T , water temperature; Pr , precipitation; DO , dissolved oxygen; Tur , turbidity; TDN , total dissolved nitrogen; TDP , total dissolved phosphorus.

calibrated with limited field data. Given the large variability of the water quality in Lake Taihu, the use of different model structures and parameters in different modeling periods may improve the model, a strategy similar to the one used in structural dynamic models (e.g., Jørgensen, 1986) and machine learning (e.g., Volf et al., 2011).

Although the model fit of the mechanistic model is not as good as that of the ANN models in this study, mechanistic models are more widely used for phytoplankton prediction than ANN models. They describe phytoplankton processes (e.g., physical and biological processes) explicitly, and are thus commonly used to investigate the phytoplankton dynamics in lakes. One of the shortcomings of mechanistic models is the need to estimate large amount of model parameters. Moreover, the difficulty to collect the model input data limits the application of the mechanistic model. The two-dimensional mechanistic model describe phytoplankton transport based on hydrodynamic conditions. Therefore, although we targeted here only the water quality of Gonghu Bay in Lake Taihu, spatial data of chlorophyll-*a* distribution in the whole lake are required as initial conditions of the model. These data are difficult to collect, especially for a large lake like Lake Taihu with a high heterogeneity of water quality.

4.3. Performance of the Regression Model

The predictive ability of the regression model is mainly because the three-day-ahead chlorophyll-*a* concentration is linearly related to chlorophyll-*a* concentration at time *t* (the coefficient of determination = 0.67). This simple model is widely used for identifying the relationship between predictive variables and input variables. However, it fails to capture the chlorophyll-*a* peaks and shows a lower model fit than the ANN 1 and 2, because some input variables are not linearly related to the dependent variable. For instance, it is well known that the growth rate of phytoplankton decreases either above or below an optimal water temperature (Jørgensen, 2001). However, the regression coefficient of water temperature in Equations 4 to 10 (ranging from -0.104 to -0.154) indicates a negative correlation between water temperature and three-day-ahead chlorophyll-*a* concentration in the regression model. The negative coefficient is probably because that the water temperature is above the optimal temperature for phytoplankton growth.

The *RMSE* value of the regression model in the validation period ($6.24 \mu\text{g L}^{-1}$) shows that the regression model did not perform as good as other models. The relatively low model fit indicates that the regression model is not suitable for predicting chlorophyll-*a* changes in Gonghu Bay. Moreover, several shortcomings should be mentioned. First, it is difficult to select suitable input variables, the same challenge as the ANN model. We used the forward stepwise method to select the optimal subset of variables. However, the method may eliminate variables that are significantly correlated with the dependent variable. The selected input variables may be different if other methods (e.g., Partial least squares regression) were used. It is thus difficult to identify the optimal variable

set for regression model. Besides the challenge to select input variables, other problems, such as multicollinearity (Graham, 2003) and the problem of limited information that can be extracted from the model results (Gevrey et al., 2003), are still waiting to be solved. Finally, other non-linear modelling techniques, such as generalized additive models and decision trees, are worth to be used in describing complex phytoplankton dynamics.

4.4. Effects of the Input Variables on Phytoplankton Biomass

The sensitivity analysis demonstrates that the influences of the input variables on phytoplankton dynamics are significantly different between the ANN, mechanistic and regression models (Figure 8).

The relatively high sensitivity of Chl_t in the regression model indicates that all the past information may be embodied in the chlorophyll-*a* concentration itself, a conclusion that is supported by previous studies (e.g., Lee et al., 2003). Sensitivity of Chl_t , Chl_{t-1} and Chl_{t-2} are relatively low in ANN 1 and ANN 2. This may be because that environmental variables (e.g., T_t and TDP_t) in these ANN models could explain the changes of Chl_{t+3} . The low partial correlation coefficient of Chl_{t-1} is because Chl_{t-1} is highly related to Chl_t (see autocorrelation analysis in Figure 3), and is thus a redundant variable for the prediction of Chl_{t+3} . This explains the contradiction that although Chl_{t-1} is a significant variable in the regression model, its partial correlation coefficient is as low as -0.070.

The sensitivity analysis results show that phytoplankton in Lake Taihu depends strongly on water temperature, which is in accordance with previous studies in Lake Taihu (e.g., Chen et al., 2003). Chlorophyll-*a* concentration can be significantly affected by heavy rainfall, which dilutes chlorophyll-*a* considerably. However, sensitivity of precipitation is not high in ANN 1. This may be because the ratio of the rainy days to the modeling period is as low as 24.7%.

River inflows may change the water turbidity in Gonghu Bay, especially when large amount of water was transferred from Yangtze River to improved the water quality of Lake Taihu (Yangtze River-TaiHu Lake Water Transfer Project). This Water Transfer Project resulted in a large flow of Wangyu River (Figure 1), and might change chlorophyll-*a* concentration significantly. Thus, the chlorophyll-*a* change is sensitive to water turbidity in ANN 1.

Phytoplankton responded differently to nutrient conditions in the ANN, regression and mechanistic models. The mean total dissolved phosphorus and nitrogen concentration in Gonghu Bay during the modeling period were 0.030 and 1.73 mg/L, respectively. These nutrient levels are much higher than the Michaelis constants for the phosphorus and nitrogen uptake (0.01 and 0.022 mg/L as shown in Table 2). Thus, phytoplankton growth rate would not change significantly in case of increasing or decreasing nutrient concentration by 10%. This resulted in low sensitivity values of nutrient con-

ditions in the mechanistic model. However, the ANN models show higher sensitivities of total dissolved phosphorus and nitrogen. The contradiction between the ANN and mechanistic models indicates that ANN models may fail to identify the sensitive variables, especially when redundant and highly correlated variables are included in the ANN models.

4.5. Possible Applications to Lake Management

These short-term models presented here attempts to predict phytoplankton accumulation in the ensuing few days for early warning of algal blooms. This prediction can help lake managers to take precautionary measures to minimize the harmful effects of severe algal blooms (Reynolds, 1999; Lee et al., 2003). For early warning of algal blooms, lake managers are more interested in the predictive power of the model than an explicit representation of the biological processes. Multi-dimensional models have been widely used for modeling water quality in a large lake (e.g., Hu et al., 2006; Huang et al., 2012a). In this study, the zero-dimensional ANN model achieved a good model fit, and thus appeared to be a good alternative in short-term prediction of phytoplankton biomass for management purpose.

There is a trend of phytoplankton accumulation in the north bays due to the dominate wind conditions. The mechanistic model presented in this study has the capacity to describe the horizontal transport of phytoplankton in the ensuing few days by coupling with hydrodynamic model, and thus to identify locations with severe algal blooms in Lake Taihu. Moreover, this mechanistic model represents physical and biological processes explicitly, and thus could be used for investigating the responses of phytoplankton to environmental changes (e.g., Mooij et al., 2007). Therefore, it can be used to evaluate different lake restoration scenarios and find a cost-effective strategy. However, the challenge to obtain adequate data for model calibration and validation, and the relatively large deviations limit the use of the mechanistic model in early warning of severe algal blooms.

5. Conclusions

We compared the performance of ANN, mechanistic and regression models to predict phytoplankton biomass (expressed as chlorophyll-*a* concentration) in Gonghu Bay. Two data sets collected in 2009 were used to drive a total of five models, namely three ANN models, one mechanistic model and one regression model. Although some issues (e.g., data pre-processing and selection of input variables) have not yet been adequately addressed, one of these ANN models obtained the “best” results, implying its potential in short-term prediction of phytoplankton biomass. Therefore, the black-box model (ANN) could be considered as an option for water resource management purpose without an adequate understanding of the mechanisms involved. Similar model fits were obtained when using regression and mechanistic models. Mechanistic model could be used for spatial prediction and investigating variable interactions. It is so far the most widely used model

in predicting phytoplankton dynamics. The regression model is characterized by its easy development.

Acknowledgements. The project was financially supported by National Basic Research Program of China (No 2012CB417006) and Major Science and Technology Program for Water Pollution Control and Treatment (2012ZX07529-001 and 2012ZX07529-002). The authors would like to thank Taihu Laboratory for Lake Ecosystem Research for providing the data for model calibration and validation.

References

- Arhonditsis, G.B., and Brett, M.T. (2005). Eutrophication model for Lake Washington (USA): Part I. Model description and sensitivity analysis. *Ecol. Model.*, 187(2-3), 140-178. <http://dx.doi.org/10.1016/j.ecolmodel.2005.01.040>.
- Barciela, R.M., García, E., and Fernández, E. (1999). Modelling primary production in a coastal embayment affected by upwelling using dynamic ecosystem models and artificial neural networks. *Ecol. Model.*, 120(2-3), 199-211. [http://dx.doi.org/10.1016/S0304-3800\(99\)00102-7](http://dx.doi.org/10.1016/S0304-3800(99)00102-7).
- Bowden, G.J., Dandy, G.C., and Maier, H.R. (2005). Input determination for neural network models in water resources applications. Part 1--background and methodology. *J. Hydrol.*, 301(1-4), 75-92. <http://dx.doi.org/10.1016/j.jhydrol.2004.06.021>.
- Bowden, G.J., Maier, H.R., and Dandy, G.C. (2002). Optimal division of data for neural network models in water resources applications. *Water Resour. Res.*, 38(2), 1010. <http://dx.doi.org/10.1029/2001wr000266>.
- Burden, F.R., Brereton, R.G., and Walsh, P.T. (1997). Cross-validation selection of test and validation sets in multivariate calibration and neural networks as applied to spectroscopy. *Analyst*, 122(10), 1015-1022. <http://dx.doi.org/10.1039/A7035651>.
- Chen, Y., Qin, B., Teubner, K., and Dokulil, M.T. (2003). Long-term dynamics of phytoplankton assemblages: Microcystis-domination in Lake Taihu, a large shallow lake in China. *J. Plankton Res.*, 25(4), 445-453. <http://dx.doi.org/10.1093/plankt/25.4.445>.
- Duan, H., Ma, R., Xu, X., Kong, F., Zhang, S., Kong, W., Hao, J., and Shang, L. (2009). Two-decade reconstruction of algal blooms in China's Lake Taihu. *Environ. Sci. Technol.*, 43(10), 3522-3528. <http://dx.doi.org/10.1021/es8031852>.
- Gevrey, M., Dimopoulos, I., and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.*, 160(3), 249-264. [http://dx.doi.org/10.1016/S0304-3800\(02\)00257-0](http://dx.doi.org/10.1016/S0304-3800(02)00257-0).
- Graham, M.H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, 84(11), 2809-2815. <http://dx.doi.org/10.1890/02-3114>.
- Guo, L. (2007). Doing battle with the green monster of Lake Taihu. *Science*, 317, 1166. doi:10.1126/science.317.5842.1166.
- Hamilton, D.P., and Schladow, S.G. (1997). Prediction of water quality in lakes and reservoirs. Part I -- Model description. *Ecol. Model.*, 96(1-3), 91-110. [http://dx.doi.org/10.1016/S0304-3800\(96\)00062-2](http://dx.doi.org/10.1016/S0304-3800(96)00062-2).
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feed-forward networks are universal approximators. *Neural Networks*, 2(5), 359-366. [http://dx.doi.org/10.1016/0893-6080\(89\)90020-8](http://dx.doi.org/10.1016/0893-6080(89)90020-8).
- Hu, W., Jørgensen, S.E., and Zhang, F. (2006). A vertical-compressed three-dimensional ecological model in Lake Taihu, China. *Ecol. Model.*, 190(3-4), 367-398. <http://dx.doi.org/10.1016/j.ecolmodel.2005.02.024>.
- Huang, J., Gao, J., and Hörmann, G. (2012a). Hydrodynamic-phyto-

- plankton model for short-term forecasts of phytoplankton in Lake Taihu, China. *Limnologica*, 42(1), 7-18. <http://dx.doi.org/10.1016/j.limno.2011.06.003>.
- Huang, J., Gao, J., Hörmann, G., and Mooij, W.M. (2012b). Integrating three lake models into a Phytoplankton Prediction System for Lake Taihu (Taihu PPS) with Python. *J. Hydroinf.*, 14(2), 523-534. <http://dx.doi.org/10.2166/hydro.2011.020>.
- Hull, V., Parrella, L., and Falcucci, M. (2008). Modelling dissolved oxygen dynamics in coastal lagoons. *Ecol. Model.*, 211(3-4), 468-480. <http://dx.doi.org/10.1016/j.ecolmodel.2007.09.023>.
- Jørgensen, S.E. (1976). A eutrophication model for a lake. *Ecol. Model.*, 2(2), 147-165. [http://dx.doi.org/10.1016/0304-3800\(76\)90030-2](http://dx.doi.org/10.1016/0304-3800(76)90030-2).
- Jørgensen, S.E. (1986). Structural dynamic model. *Ecol. Model.*, 31(1-4), 1-9. [http://dx.doi.org/10.1016/0304-3800\(86\)90051-7](http://dx.doi.org/10.1016/0304-3800(86)90051-7).
- Jørgensen, S.E., and Bendoricchio, G. (2001). Fundamentals of ecological modelling. Elsevier.
- Jørgensen, S.E. (2010). A review of recent developments in lake modelling. *Ecol. Model.*, 221(4), 689-692. <http://dx.doi.org/10.1016/j.ecolmodel.2009.10.022>.
- James, F.C., and McCulloch, C.E. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annu. Rev. Ecol. Syst.*, 21, 129-166. <http://dx.doi.org/10.1146/annurev.ecolsys.21.1.129>.
- Jeong, K.S., Joo, G.J., Kim, H.W., Ha, K., and Recknagel, F. (2001). Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. *Ecol. Model.*, 146(1-3), 115-129. [http://dx.doi.org/10.1016/S0304-3800\(01\)00300-3](http://dx.doi.org/10.1016/S0304-3800(01)00300-3).
- Jeong, K.S., Kim, D.K., and Joo, G.J. (2006). River phytoplankton prediction model by Artificial Neural Network: Model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system. *Ecol. Inf.*, 1(3), 235-245. <http://dx.doi.org/10.1016/j.ecoinf.2006.04.001>.
- Karul, C., Soyupak, S., Cilesiz, A.F., Akbay, N., and Germen, E. (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecol. Model.*, 134(2-3), 145-152. doi:10.1016/S0304-3800(00)00360-4.
- Kumar, M., Raghuvanshi, N., and Singh, R. (2011). Artificial neural networks approach in evapotranspiration modeling: a review. *Irrig. Sci.*, 29(1), 11-25. <http://dx.doi.org/10.1007/s00271-010-0230-8>.
- Lee, H.S., and Lee, J.H.W. (1995). Continuous monitoring of short term dissolved oxygen and algal dynamics. *Water Res.*, 29(12), 2789-2796. [http://dx.doi.org/10.1016/0043-1354\(95\)00126-6](http://dx.doi.org/10.1016/0043-1354(95)00126-6).
- Lee, J.H.W., Huang, Y., Dickman, M., and Jayawardena, A.W. (2003). Neural network modelling of coastal algal blooms. *Ecol. Model.*, 159(2-3), 179-201. [http://dx.doi.org/10.1016/s0304-3800\(02\)00281-8](http://dx.doi.org/10.1016/s0304-3800(02)00281-8).
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., and Aulagnier, S. (1996). Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model.*, 90(1), 39-52. [http://dx.doi.org/10.1016/0304-3800\(95\)00142-5](http://dx.doi.org/10.1016/0304-3800(95)00142-5).
- Li, T., Wang, D., Zhang, B., Liu, H., and Tang, H., (2007). Morphological characterization of suspended particles under wind-induced disturbance in Taihu Lake, China. *Environ. Monit. Assess.*, 127(1-3), 79-86. <http://dx.doi.org/10.1007/s10661-006-9261-2>.
- Maier, H.R., and Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Model. Software*, 15(1), 101-124. [http://dx.doi.org/10.1016/s1364-8152\(99\)00007-9](http://dx.doi.org/10.1016/s1364-8152(99)00007-9).
- Maier, H.R., Dandy, G.C., and Burch, M.D. (1998). Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecol. Model.*, 105(2-3), 257-272. [http://dx.doi.org/10.1016/s0304-3800\(97\)00161-0](http://dx.doi.org/10.1016/s0304-3800(97)00161-0).
- Maier, H.R., Jain, A., Dandy, G.C., and Sudheer, K.P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Software*, 25(8), 891-909. <http://dx.doi.org/10.1016/j.envsoft.2010.02.003>.
- Marini, F. (2009). Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Anal. Chim. Acta*, 635(2), 121-131. <http://dx.doi.org/10.1016/j.aca.2009.01.009>.
- Mohanraj, M., Jayaraj, S., and Muraliedharan, C. (2012). Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systems-A review. *Renew. Sustain. Energy Rev.*, 16(2), 1340-1358. <http://dx.doi.org/10.1016/j.rser.2011.10.015>.
- Mooij, W., Janse, J., De Senerpont Domis, L., Hülsmann, S., and Ibelings, B. (2007). Predicting the effect of climate change on temperate shallow lakes with the ecosystem model PCLake. *Hydrobiologia*, 584(1), 443-454. <http://dx.doi.org/10.1007/s10750-007-0600-2>.
- Mooij, W.M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P.V., Chitamwebwa, D.B.R., Degermendzhy, A.G., DeAngelis, D.L., De Senerpont Domis, L.N., Downing, A.S., Elliott, J.A., Fragoso Jr., C.R., Gaedke, U., Genova, S.N., Gulati, R.D., Håkanson, L., Hamilton, D.P., Hipsey, M.R., 't Hoen, J., Hülsmann, S., Los, F.H., Makler-Pick, V., Petzoldt, T., Prokopkin, I.G., Rinke, K., Schep, S.A., Tominaga, K., Van Dam, A.A., Van Nes, E.H., Wells, S.A., and Janse, J.H. (2010). Challenges and opportunities for integrating lake ecosystem modelling approaches. *Aquat. Ecol.*, 44(3), 633-667. <http://dx.doi.org/10.1007/s10452-010-9339-3>.
- Parisi, R., Di Claudio, E.D., Orlandi, G., and Rao, B.D. (1996). A generalized learning paradigm exploiting the structure of feed-forward neural networks. *IEEE Trans. Neural Networks*, 7(6), 1450-1460. <http://dx.doi.org/10.1109/72.548172>.
- Qin, B., Xu, P., Wu, Q., Luo, L., and Zhang, Y. (2007). Environmental issues of Lake Taihu, China. *Hydrobiologia*, 581(1), 3-14. <http://dx.doi.org/10.1007/s10750-006-0521-5>.
- Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecol. Model.*, 146(1-3), 303-310. [http://dx.doi.org/10.1016/S0304-3800\(01\)00316-7](http://dx.doi.org/10.1016/S0304-3800(01)00316-7).
- Recknagel, F., French, M., Harkonen, P., and Yabunaka, K.-I. (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Model.*, 96(1-3), 11-28. [http://dx.doi.org/10.1016/S0304-3800\(96\)00049-X](http://dx.doi.org/10.1016/S0304-3800(96)00049-X).
- Recknagel, F., Fukushima, T., Hanazato, T., Takamura, N., and Wilson, H. (1998). Modelling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by artificial neural networks. *Lakes Reserv.: Res. Manage.*, 3(2), 123-133. <http://dx.doi.org/10.1111/j.1440-1770.1998.tb00039.x>.
- Reynolds, C.S. (1999). Modelling phytoplankton dynamics and its application to lake management. *Hydrobiologia*, 396, 123-131. <http://dx.doi.org/10.1023/A:1017039900307>.
- Robson, B.J., Hamilton, D.P., Webster, I.T., and Chan, T. (2008). Ten steps applied to development and evaluation of process-based biogeochemical models of estuaries. *Environ. Model. Software*, 23(4), 369-384. <http://dx.doi.org/10.1016/j.envsoft.2007.05.019>.
- Scardi, M., and Harding Jr, L.W. (1999). Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecol. Model.*, 120(2-3), 213-223. [http://dx.doi.org/10.1016/s0304-3800\(99\)00103-9](http://dx.doi.org/10.1016/s0304-3800(99)00103-9).
- Srinivasulu, S., and Jain, A. (2006). A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Appl. Soft Comput.*, 6(3), 295-306. <http://dx.doi.org/10.1016/j.asoc.2005.02.002>.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)*, 36(2), 111-147.

- Trolle, D., Hamilton, D.P., Pilditch, C.A., Duggan, I.C., and Jeppesen, E. (2011). Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management. *Environ. Model. Software*, 26(4), 354-370. <http://dx.doi.org/10.1016/j.envsoft.2010.08.009>.
- Volf, G., Atanasova, N., Kompare, B., Precali, R., and Ozanic, N. (2011). Descriptive and prediction models of phytoplankton in the northern Adriatic. *Ecol. Model.*, 222(14), 2502-2511. <http://dx.doi.org/10.1016/j.ecolmodel.2011.02.013>.
- Whitehead, P.G., Howard, A., and Arulmani, C. (1997). Modelling algal growth and transport in rivers: a comparison of time series analysis, dynamic mass balance and neural network techniques. *Hydrobiologia*, 349(1), 39-46. <http://dx.doi.org/10.1023/a:1003089310834>.
- Yang, M., Yu, J., Li, Z., Guo, Z., Burch, M., and Lin, T. (2008). Taihu Lake not to blame for Wuxi's woes. *Science*, 319(5860), 158. <http://dx.doi.org/10.1126/science.319.5860.158a>.
- Young Li, W.A., Millie, D.F., Weckman, G.R., Anderson, J.S., Klarer, D.M., and Fahnenstiel, G.L. (2011). Modeling net ecosystem metabolism with an artificial neural network and Bayesian belief network. *Environ. Model. Software*, 26(10), 1199-1210. <http://dx.doi.org/10.1016/j.envsoft.2011.04.004>.