Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Environmental Sciences 18 (2013) 310 – 316
Procedia Environmental Sciences

2013 International Symposium on Environmental Science and Technology (2013 ISEST)

Using artificial neural network models for eutrophication prediction

Shouliang Huo, Zhuoshi He, Jing Su, Beidou Xi*, Chaowei Zhu

[a]*State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science, Beijing 100012, China*

Abstract

Artificial neural network (ANN), a data driven modeling approach, is proposed to predict the water quality indicators of Lake Fuxian, the deepest lake of southwest China. To determine the non-linear relationships between the water quality factors and the eutrophication indicators, several ANN models was chosen for the investigation. A commonly used back-propagation neural network model was used to relate the key factors that influence a number of water quality indicators such as dissolved oxygen (DO), total phosphorus (TP), chlorophyll-a (Chl-a), and secchi disk depth (SD) in Lake Fuxian. The measured data were fed to the input layer, representing forcing functions to control the in-lake bio-chemical processes. Eutrophication indicators such as DO, TN, Chl-a and SD were represented in the output layers. The results indicated that the back-propagation neural network model performs good in ten months prediction and the neural network is able to predict these indicators with reasonable accuracy. This study also suggested that the neural network is a valuable tool for lake management.

Keywords: artificial neural network; eutrophication; water quality; lake management

1. Introduction

Lake eutrophication is one of the most important water environmental problems in China, which would lead to abundant development of aquatic plants, growth of algae, and disturb the balance of organisms in the water [1]. Water quality indicators (WQIs) are widely used to predict the eutrophication levels of lake waters. However, there is still difficulty in prediction because of two reasons. One the one hand, the spatial and temporal distributions has been affected by various climatic, geographical and ecological factors. On the other hand, the indicators are inter-dependent and interrelated, which further increase the complexity in prediction [2]. Unlike many statistically based water quality models, which

* Corresponding author. *E-mail address:* xibeidou@263.net

doi:10.1016/j.proenv.2013.04.040

assume the relationship between response variable and prediction variables are linear and distributed normally, ANNs are able to map the non-linear relationships among variables that are characteristic of ecosystems [3-6]. Previous studies indicated that ANNs are capable of simulating trends of algal growth dynamics and predicting algal blooms based on water quality monitoring data [6-8]. In recent years, a number of efforts have been taken on the use of ANN models for predicting the water quality of river [9,10], coastal [7-11], reservoirs [12] and shallow-lake [13,14]. However, their limitations have not received sufficient attention [15]. Moreover, little was applied in deep plateau lake because of the complex characteristics such as the interactions among different indicators. The purpose of this study is to apply artificial neural networks (ANNs) as an alternative modeling approach to simulate the eutrophication processes. Furthermore, the results from the back-propagation neural network algorithms were performed and analyzed.

2. Study area

Lake Fuxian, located in Yunnan–Guizhou Plateau in southwest China, is the deepest freshwater plateau lake in this area with a surface area of 216.6 km^2 and a mean depth of 95.2 m. The storage capacity of the lake is 206.2×108 m^3, which occupies 9.16% of the total storage capacity of freshwater lake in China. The amount of phytoplankton in Lake Fuxian increased by 2.6 times, Chl-a increased by 3 times and SD almost decreased by 50% in the past two decades [16]. The water quality data used in this study were from 2003 to 2008 collected from the ambient monitoring network maintained by the Department of Environmental Protection of the Yunnan Provinces, China. Total phosphorus (TP), total nitrogen (TN), secchi depth (SD), dissolve oxygen (DO) and chlorophyll-a (Chl-a) were used as indicators of lake eutrophication.

3. Modeling methods

3.1. Artificial neural networks (ANN)

ANNs constitute an information-processing paradigm that is inspired by biological nervous systems [17]. The key element is the novel structure of the paradigm. It is made up by a large number of highly interconnected processing neurons working in unison to solve specific problems. In this study, the back-propagation neural network ANNs was employed.

Standardization. To eliminate the impact of variables' dimension, standardization needs to be done for the variables. There are many ways of standardization. One of the ways is linear transformation, which transfers the original data to a specified interval by a linear formula. The specified interval is usually preset to (0.15, 0.85) [6].

Back-propagation neural network. Back-propagation is a commonly used learning algorithm in ANN applications. It uses the gradient descent algorithm to determine the weights in the network. An ANN consists of three or more layers: an input layer, a hidden layer(s), and an output layer. The input layer contains input nodes (neurons), i.e. the input variables for the network. The output layer contains the desired output of the system, and the hidden layer usually contains a series of nodes associated with a transfer function. Each layer of the network is linked by weights that need to be determined through a learning algorithm. Sigmoid function is a commonly used transfer function, which is adopted as the property that it squashes the independent variable, which may have a range from $-\infty$ to ∞, to the range 0-1 [8]. At the output layer where the network output has to be compared with the target output, the target values need to be standardized to the range of (0, 1).

The ANN models are applied to establish the relationship between the inputs and outputs through the data collected over several years in Lake Fuxian. Once trained, the weights in the ANN were fixed and the model was validated by assessing its predictive performance on a set of testing data excluding the training data.

3.2. Choice of network variables

Input variables. Total nitrogen is a common indicator for the lake eutrophication. The concentration of TN correlates with NH_3-N and SD well. So NH_3-N and SD were chosen as the input variable of the TN models. TP is another important indicator, but its models were not built because of the lack of data. SD is usually influenced by water color, turbidity and scum. SD also changes along with the change of season and Chl-a, so they were used as the input variables in the SD prediction model. DO concentration is another common indicator of the quality of aquatic ecosystems. The sources of DO in a water body include re-aeration from the atmosphere, photosynthetic oxygen production and DO loading. The sinks include oxidation of carbonaceous and nitrogenous material, sediment oxygen demand and respiration by aquatic plants. DO models were established by using inputs of month, pH, Chl-a, NH_3-N, BOD_5 and temperature. The growth rate of algae is influenced by sunlight, water temperature, and nutrients. The Chl-a model is related to month, water temperature, pH, DO, SD, TN and TP. So they were used as the input variables to build the Chl-a prediction model. To sum, four models were built to predict four indicators (DO, Chl-a, TN and SD). Fig. 1 show the BP structure, which inter-connects the input and output layer via a hidden layer consisting of a number of neurons.

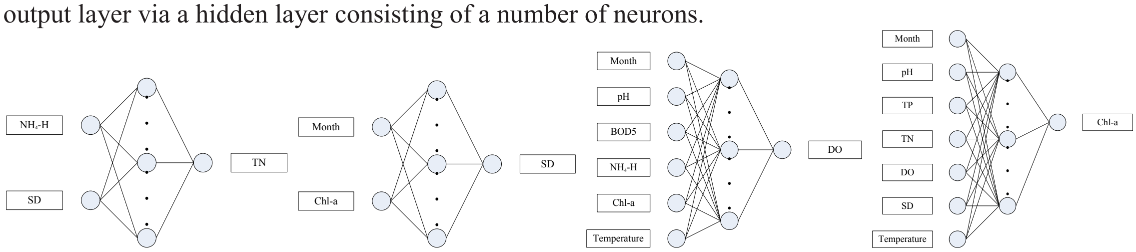

Fig. 1. ANN structures of TP, SD, DO, and Chl-a models for the Lake Fuxian.

Parameters of training and testing. There are a number of key parameters in the BP neural network. The initial values of the weights were assigned randomly based on an input random number seed. The initial values of the weights were determined by a random starting, and randomly set between -0.1 and 0.1. This range was chosen where all the models created can be run. All the BP neural network models used in this study were composed of three layers with nodes in adjacent layers fully connected, i.e., only one hidden layer with 20-60 hidden nodes was employed.

To speed up the training and convergence of the weight values, the learning rate should be adjustably chosen. It is necessary to have a learning rate that is small enough to converge but large enough that the computing time is reasonable. So the learning rate was chosen as 0.05. And the momentum was chosen as 0.9.

Performance assessment of the ANN models. The performance of the ANN models was assessed using the root mean square error (RMSE) as a measure of goodness-of-fit. RMSE was chosen because it best describes an average measure of the error in predicting the changes of the eutrophication indicators. The RMSE and correlation coefficient of the models are shown in table 1, respectively. Moreover, further analysis is needed to assess the effect of the input variables and their contribution to the network output[8].

Table 1. The root mean square error and the correlation coefficient of the models.

	RMSE				R			
	TN	SD	DO	Chl-a	TN	SD	DO	Chl-a
Train	0.15	0.09	0.04	0.07	0.72	0.82	0.96	0.09
Test	0.14	0.12	0.14	0.14	0.69	0.74	0.65	0.76

Sensitivity analysis. To evaluate the effect of each input variable to the ANN models, the most commonly used method is the sensitivity analysis, which was carried on a relatively complicated network [18]. The sensitivity demonstrates how the trained network reacts to the change of each input. Each input of the four models was altered by 5, 10 and 20%. The change in the output caused by the change in the input is calculated. Sensitivity of each input variables of four models is shown in table 2. The input indicators with sensitivity higher than 100% were supposed to be very sensitivity as model inputs.

Table 2. The sensitivity of each input variables of the models.

	Month	Chl-a	SD	NH3-N	Temperature	pH	BOD5	DO	TN	TP
SD Model	115	140	-	-	-	-	-	-	-	-
TN Model	-	-	103	164	-	-	-	-	-	-
DO Model	74	99	-	79	68	121	81	-	-	-
Chl-a Model	36	-	79	-	63	36	-	48	45	30

4. Model results and discussions

4.1. TN model

Fig. 2 (a) and (b) shows the results of the back-propagation neural network for TN models training and testing, respectively. The results show that factors correlated with the TN in the Lake Fuxian are NH_3-N and SD， which is correlated with suspend solid (SS). The RMSE of the training and testing are 0.15 and 0.14, respectively. The correlation coefficients of training and testing are 0.72 and 0.69, respectively. Although this model is constructed for TN, it also is used to predict NH_3-N. If we only know the result of TN and SD, we may use the try-and-error method to find the concentration of NH_3-N.

4.2. SD model

SD can cheaply provide a great deal of information on lake water quality and, together with chlorophyll a, has become routinely used as a measure of lake trophic status. Fig. 2 (c) and (d) shows the results of the back-propagation neural network for SD models training and testing, respectively. The RMSE of the training and testing are 0.09 and 0.12, respectively. The correlation coefficient of training and testing are 0.82 and 0.74, respectively. According to the SD model developed herein, there are two factors that influence the SD in Lake Fuxian. One is Chl-a, and the other is month. As with other variables, SD may vary considerably in a given lake between and within seasons; therefore, it is desirable to have an “indicator season.” The month factor shows that the SD changes seasonally. This factor is used to indicate that some variables change monthly, such as temperature, average rain, and pollutant loading, etc.

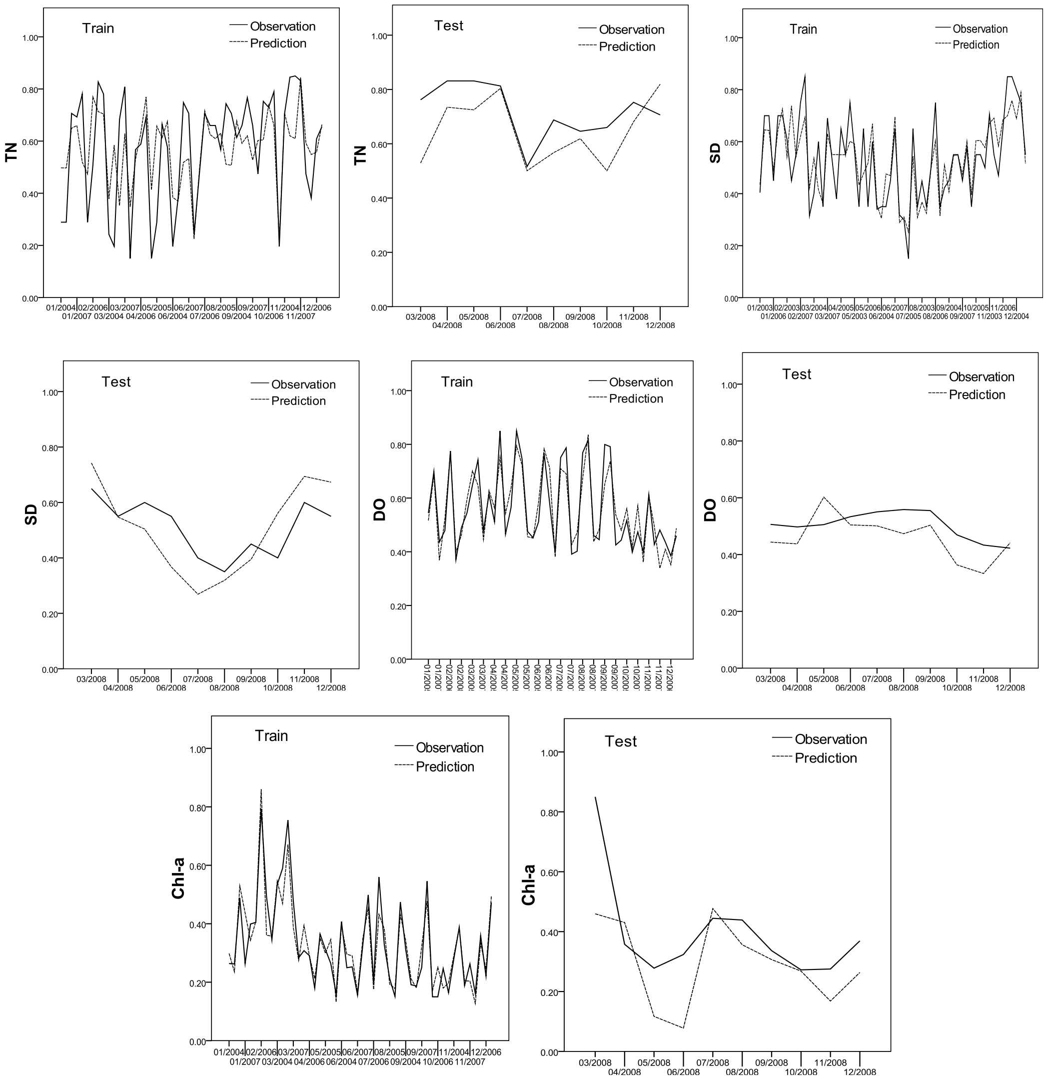

Fig. 2. (a) Predicted and measured data of TN for training; (b) Predicted and measured data of TN for testing; (c) Predicted and measured data of SD for training; (d) Predicted and measured data of SD for testing; (e) Predicted and measured data of DO for training; (f) Predicted and measured data of DO for testing; (g) Predicted and measured data of Chl-a for training; (h) Predicted and measured data of Chl-a for testing.

4.3. DO model

Oxygen concentrations and rates of depletion have been used to characterize lakes and can in some instances be related back to nutrient status. According to DO mechanism, we can establish the ANN models by using input of month, pH, Chl-a, NH_3-H, BOD_5, and temperature. Fig. 2 (e) and (f) shows the results of the back-propagation neural network for DO models training and prediction, respectively. The RMSE of the training and testing are 0.04 and 0.14, respectively. The correlation coefficient of training and testing are 0.96 and 0.65, respectively.

4.4. Chl-a model

Chl-a model is related to month, water temperature, pH, DO, SD, TN, and TP. According to these factors, we can see that the limiting nutrients of Chl-a are TN, and TP. Month and SD would influence the sunlight intensity in water. Water temperature and DO may indicate how much oxygen is produced by Chl-a. All factors fit the principle theory of Chl-a growth dynamics. Fig. 2 (g) and (h) shows the results of the back-propagation neural network for Chl-a models training and prediction, respectively. The RMSE of the training and testing are 0.07 and 0.14, respectively. The correlation coefficient of training and testing are 0.90 and 0.76, respectively.

5. Conclusions

Four ANN models were developed to predict the major constituent concentrations of water quality of Lake Fuxian. The ANN models can preserve nonlinear characteristics between input and output variables, which are superior to traditional statistical models. The RMSE between predicted values and measured data was less than 0.2 for four models. The correlation coefficients between predicted values and measured data are well above 0.7 for these models. In the present case, the performance of the ANN model is most encouraging and the complex mechanism in the Lake Fuxian can be quantified and expressed by these four back-propagation neural network models. The modeling approach described in this study for analysis of the eutrophication problem in the Lake Fuxian has yielded useful information for effective water quality management.

Acknowledgements

This study is supported by the Mega-projects of Science Research for Water Environment Improvement (Program No.2009ZX07106-001; 2012ZX07101-002), the China Basic Research Program (Program No. 2008CB418206), and the National Natural Science Foundation of China (No. 40901248).

References

[1] Gao JQ, Xiong ZT, Zhang JD, Zhang WH. Phosphorus removal from water of eutrophic Lake Donghu by five submerged macrophytes. *Desalination* 2009;242 (1- 3):193-204.
[2] Kurunça A, Yüreklia K, Çevikb O. Performance of two stochastic approaches for forecasting waterquality and streamflow data from Yeşilırmak River, Turkey. *Environmental Modelling & Software* 2005;20 (9):1195-1200.
[3] Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S. Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling* 1996;90 (1):39-52.
[4] Maier HR, Morgan N, Chow CWK. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling & Software* 2004;19:485-494.
[5] Wieland R, Wilfried M. Adaptive fuzzy modeling versus artificial neural networks. *Environmental Modelling & Software* 2008;23:215-224.
[6] Kuo JT, Hsieh MH, Lung WS, She N. Using artificial neural network for reservoir eutrophication prediction. *Ecological Modelling* 2007;200 (1-2):171-177.
[7] French M, Recknagel F. Modeling of algal blooms in freshwaters using artificial neural networks. *Computational Mechanics* 1994;2:87-94.

[8] Lee JHW, Yan H, Dickman M, Jayawardena AW. Neural network modelling of coastal algal blooms. *Ecological Modelling* 2003;159 (2-3):179-201.

[9] Zhang Q, Stanley SJ. Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling. *Water Research* 1997;31 (9):2340-2350.

[10] Singh KP, Basant A, Malik A, Jain G. Artificial neural network modeling of the river waterquality—A case study. *Ecological Modelling* 2009;220 (6):888-895.

[11] Barciela RM, García E, Fernández E. Modelling primary production in a coastal embayment affected by upwelling using dynamic ecosystem models and artificial neural networks. *Ecological Modelling* 1999;120 (2-3):199-211.

[12] Adeloye A. The relative utility of regression and artificial neural networks models for rapidly predicting the capacity of water supply reservoirs. *Environmental Modelling & Software* 2009;24:1233-1240.

[13] Cüneyt K, Selçuk S, Ahmet FÇ, Nihat A, Emin G. Case studies on the use of neural networks in eutrophication modeling. *Ecological Modelling* 2000;134:145–152.

[14] Zeng WH, Song QL, Liu HC, Wang T. Research on ANN-based Pre-warning Water Bloom Model of LiuHai Lake in Beijing. *Procedia Environmental Sciences* 2010;2:625-635.

[15] Young WA, Millie DF, Weckman GR, Anderson JS, Klarer DM, Fahnenstiel GL. Modeling net ecosystem metabolism with an artificial neural network and Bayesian belief network. *Environmental Modelling & Software* 2011;26:1199-1210.

[16] Cui YD, Liu XQ, Wang HZ. Macrozoobenthic community of Fuxian Lake, the deepest lake of southwest China. *Limnologica* 2008;38:116-125.

[17] Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. New Jersey: Prentice-Hall, Englewood Cliffs; 1998.

[18] Maier HR, Dandy GC, Burch MD. Use of artificial neural networks for modeling cyanobacteria Anabaena spp. in the River Murray, South Australia. *Ecological Modelling* 1998;105:257-272.