# Using Artificial Neural Network Models for Eutrophication Prediction

**Article** · December 2013

DOI: 10.1016/j.proenv.2013.04.040

# Prediction of lake eutrophication using artificial neural networks

Shouliang Huo, Zhuoshi He, Jing Su,
Beidou Xi*, Lieyu Zhang and Fengyu Zan

State Key Laboratory of Environmental Criteria and Risk Assessment,
Chinese Research Academy of Environmental Science,
Beijing 100012, China
Email: huoshouliang@126.com
Email: huosl@craes.org.cn
Email: zhuoshihe@163.com
Email: sujing169@163.com
Email: xibeidou@263.net
Email: zhanglieyu@163.com
Email: zanfengyu@126.com
*Corresponding author

**Abstract:** An artificial neural network (ANN), which is a data-driven modelling approach, is proposed to indicate the water quality of Lake Fuxian, the deepest lake of southwest China. To determine the nonlinear relationships between the water quality factors and eutrophication indicators, several ANN models were chosen. The back-propagation and radial basis function neural network models were applied to relate the key factors that influence a number of water quality indicators, such as total nitrogen (TN), secchi disk depth (SD), dissolved oxygen (DO) and chlorophyll-a (Chl-a) in Lake Fuxian. The measured data were fed to the input layer, representing forcing functions to control the in-lake biochemical processes. Eutrophication indicators (TN, SD, DO and Chl-a) were represented in the output layers. The results indicated that the back-propagation neural network model performed better than radial basis function neural network model in ten months prediction and was able to predict these indicators with reasonable accuracy. Such neural networks can be a valuable tool for lake water management.

**Keywords:** artificial neural network; ANN; eutrophication; water quality; lake management.

**Reference** to this paper should be made as follows: Huo, S., He, Z., Su, J., Xi, B., Zhang, L. and Zan, F. (2014) 'Prediction of lake eutrophication using artificial neural networks', *Int. J. Environment and Pollution*, Vol. 56, Nos. 1/2/3/4, pp.63–78.

**Biographical notes:** Shouliang Huo is an Associate Professor at the State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science. His work focuses on nutrient criteria, lake eutrophication control standards, sediment criteria, and composition and chemical characteristics of dissolved organic nitrogen in wastewater treatment process or surface water. These results and achievements have been published in more than 30 papers in international journals.

Zhuoshi He is an Assistant Researcher at the State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science. Her work focuses on environmental monitoring, water quality models and lake eutrophication models.

Jing Su is an Associate Professor at the State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science. His work focuses on lake eutrophication control standards, environmental models and environmental management. These results and achievements have been published in more than ten papers in international journals.

Beidou Xi is a Professor at the State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science. His work focuses on lake eutrophication control and management, solid waste treatment technology, and groundwater remediation technology. The results and achievement have been published in over 80 papers in international journals.

Lieyu Zhang is an Associate Professor at the State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science. His work focuses on lake eutrophication control technology, especially on constructed wetlands. These results have been published in more than ten papers in international journals.

Fengyu Zan is an Associate Professor at the State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science. His work focuses on lake management, sediment criteria, and composition and chemical characteristics of dissolved organic nitrogen in surface water. These results and achievements have been published in more than ten papers in international journals.

This paper is a revised and expanded version of a paper entitled 'Using artificial neural network models for eutrophication prediction' presented at 2013 International Symposium on Environmental Science and Technology (2013 ISEST), Dalian, China, 2013 June.

## 1 Introduction

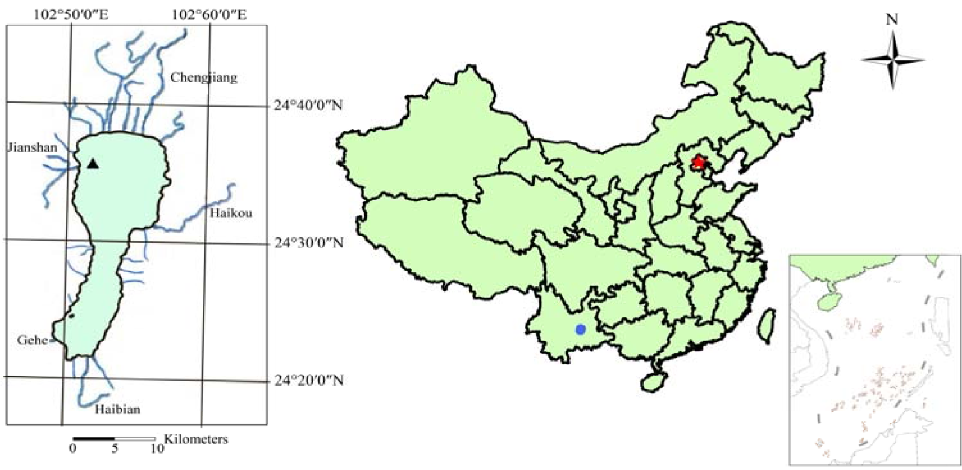
Lake eutrophication is one of the most critical environmental water problems in China, which would lead to excessive growth of aquatic plants and algae, hence disturbing the balance of aquatic life (Gao et al., 2009). Water quality indicators are widely used to predict eutrophication levels of lake waters (Zheng et al., 2011; Huang and Liu, 2008; Shuhaibar and Riffat, 2008; Bagheri and Yu, 2008). However, there is still difficulty in prediction for two reasons. Firstly, the spatial and temporal distributions are affected by changing climatic, geographical and ecological factors. Secondly, the indicators are interdependent and interrelated, which further increases the complexity of prediction (Kurunça et al., 2005). Hence exploring techniques that can quantitatively forecast eutrophication indicators is an important undertaking. Effective techniques would aid the design of efficient strategies to prevent eutrophication.

An artificial neural network (ANN) is a computational paradigm designed to mimic the human brain and nervous system. Unlike many statistically-based water quality models, which assume that the relationships between response variables and prediction variables are linear and normally distributed, ANNs can represent the nonlinear relationships among the variables that are characteristic of ecosystems (Lek et al., 1996; Maier et al., 2004; Wieland and Wilfried, 2008; Kuo et al., 2007). In addition, ANNs can use a known input data without prior assumptions (Gardner and Dorling, 1998). The ANN develops a mapping of input and output variables, which can subsequently be used to predict desired outputs as a function of suitable inputs (Schalkoff, 1992). A multi-layer neural network can approximate any smooth, measurable function between input and output vectors by selecting a suitable set of connecting weights and transfer functions (Gardner and Dorling, 1998).

Previous studies indicated that ANNs are capable of simulating trends of algal growth dynamics and predicting algal blooms based on water quality monitoring data (French and Recknagel, 1994; Lee et al., 2003; Kuo et al., 2007). In recent years, a number of efforts have focused on the use of ANN models for predicting river water quality (Zhang and Stanley, 1997; Singh et al., 2009), coastal areas (Barciela et al., 1999; Lee et al., 2003, Young et al., 2011), reservoirs (Adeloye, 2009) and shallow lakes (Zeng et al., 2010). However, few were applied to deep plateau lakes due to various complex characteristics such as the interactions among different indicators.

The purpose of this study is to apply ANNs as an alternative modelling approach to simulate the eutrophication processes for Lake Fuxian in the Yunnan-Guizhou Plateau Lake region. Two types of neural network models were used to measure the predictive performance efficiencies of each network. The results from two different neural network algorithms were compared and analysed.

**Figure 1** Map of the study area (see online version for colours)




## 2 Study area

Lake Fuxian, located in the Yunnan-Guizhou Plateau, is the deepest freshwater plateau lake in southwest China 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 × 10$^8$ m$^3$, which represents 9.16% of the total storage capacity of freshwater lakes in China. In the past two decades, the amount of phytoplankton in Lake Fuxian has increased by 2.6 times, chlorophyll-a (Chl-a) increased by three times and secchi depth (SD) decreased by almost 50% (Cui et al., 2008). The water quality data used in this study were from 2003 to 2008 and were collected by the ambient monitoring network maintained by the Department of Environmental Protection of the Yunnan Provinces. Total nitrogen (TN), SD, dissolved oxygen (DO) and Chl-a were used as indicators of lake eutrophication.

## 3 Modelling methods

### *3.1 Artificial neural networks*

ANNs constitute an information processing paradigm that is inspired by biological nervous systems (Haykin, 1998). The key element is the structure of the paradigm. It is made up by a large number of highly interconnected processing neurons working in unison to solve specific problems. The present study uses two types of neural network models to measure the predictive performance efficiencies of each network. One is a back-propagation neural network, which is a feed-forward neural network with a back-propagation learning algorithm. The other is a neural network based on a radial basis function.

#### *3.1.1 Standardisation*

To eliminate the impact of variable dimensionality, standardisation is required for the variables. There are many methods of standardisation. We use the following linear transformation (Lee et al., 2003):

$$x_i' = lower + \frac{(x_i - x_{\min})(upper - lower)}{x_{\max} - x_{\min}} \tag{1}$$

where $x_i'$ is standardised data, $x_i$ is the original data, $x_{\min}$ and $x_{\max}$ are the maximum and the minimum values of the original data and lower and upper are the smallest and largest output values allowed in the network. The lower and upper values are usually preset to 0.15 and 0.85, respectively (Kuo et al., 2007).

#### *3.1.2 Back-propagation neural network*

Back-propagation is a commonly used learning algorithm in ANN applications (i.e., BP-ANN). It uses a gradient descent algorithm to determine the weights in the network. An ANN consists of three or more layers: an input layer, 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. The sigmoid function is a commonly used transfer function, which is described as follows (Lee et al., 2003):

$$f(x) = \frac{2}{1 + \exp(-2x)} - 1. \tag{2}$$

This function was adopted due to its ability to limit the independent variable $x$, which may range from $-\infty$ to $\infty$, to the range 0–1. At the output layer, where the network output is compared with the target output, the target values need to be standardised to the range of 0–1.

#### *3.1.3 Radial basis function neural network*

The radial basis function artificial neural network (RBF-ANN) is a two-layer network and the input and output layers are connected by an RBF. The first layer utilises nonlinear mapping based on an RBF. The RBF is based on radial symmetry, in which the most popular RBF is the Gaussian function (Jin and Wei, 2008):

$$R_i(x) = \exp\left[-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right], \quad i = 1, 2, \ldots, p, \tag{3}$$

where $x$ is the input vector, $c_i$ is the centre of the $i^{\text{th}}$ basis function, $\sigma_i$ is the $i^{\text{th}}$ perception variable, $p$ is the number of perception variables and $\|x - c_i\|^2$ is the normal form of $x - c_i$. The function calculates its weighted inputs by normalisation and its net inputs by combining its weighted inputs and biases. The second layer is a linear mapping from the radial basis to the outputs.

Both types of ANN models were used to establish the relationships between the inputs and outputs through data collected over several years in Lake Fuxian. Once trained, the weights in the ANNs were fixed and the models were validated by assessing its predictive performance on a set of testing data excluding the training data.

### *3.2 Choice of network variables*

#### *3.2.1 Input variables*

One of the main tasks of an ANN is to determine the model input variables that significantly affect the output variable(s). Generally, the choice of input variables is based on a priori knowledge of causal variables, inspections of time series plots and statistical analysis of potential inputs and outputs (Palani et al., 2008).

TN is a common indicator of lake eutrophication. The concentration of TN is significantly related to ammonia nitrogen ($NH_3$-N) and SD, which correlates with suspended solids. Therefore, $NH_3$-N and SD were chosen as the input variables for the TN models. Total phosphorus (TP) is another important indicator, but TP models were not built because of the lack of relevant data.

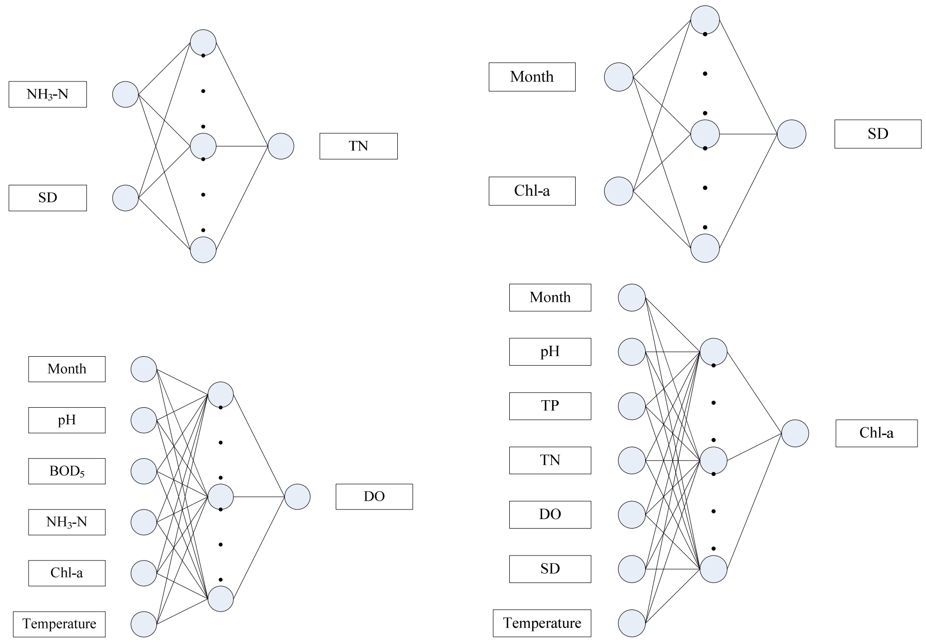
SD is usually influenced by water colon, turbidity and scum. SD also changes seasonally and with Chl-a levels. The monthly factor shows that the SD changes seasonally (Kuo et al., 2007); therefore, month and Chl-a were used as the input variables in the SD prediction model.

DO concentration is another common indicator of the quality of aquatic ecosystems. It is important for both respiration of organisms and some chemical reactions. The sources of DO in a water body include re-aeration from the atmosphere, photosynthetic oxygen production and DO loading. Based on existing measured values for different variables and their correlative analysis, a total of six factors (variables) that affect DO to a certain degree, including month, pH, Chl-a, $NH_3$-N, biochemical oxygen demand (BOD, specifically $BOD_5$) and temperature, were identified and selected as inputs for model development. Temperature, which is influenced by month, can dramatically affect the rates of chemical and biological reactions. BOD is inversely related to the DO in water; high BOD values indicate a low level of DO or even anoxic conditions in water (Kunwar et al., 2009).

The growth rate of algae is influenced by sunlight, water temperature and nutrients (Kunwar et al., 2009). Chl-a is significantly correlated with month, SD, water temperature, DO, pH, TN and TP. Month and SD influence the intensity of sunlight in water and water temperature and DO may indicate how much oxygen is produced by Chl-a. The pH factor which is a major factor controlling chemical reaction speed, may be influenced by algae production or upstream basin inflow. TN and TP are essential nutrients for algae growth (Lee et al., 1978; Portielje and van der Molen, 1999). Therefore, these seven variables were used as the input variables to build the Chl-a prediction model.

In summary, eight models were built to predict four indicators (DO, Chl-a, TN and SD). Figure 2 shows the BP structure, which interconnects the input and output layers via a hidden layer consisting of a number of neurons.

**Figure 2** ANN structures for TP, SD, DO and Chl-a models for Lake Fuxian (see online version for colours)




### 3.2.2 *Testing and training parameters*

#### 3.2.2.1 *Parameters for BP*

There are a number of key parameters in the BP neural network. The initial weight values were randomly assigned between −0.1 and 0.1 based on an input random number seed. This range was selected because it can be used with all of the proposed models. All BP neural network models used in this study were composed of three layers with fully connected nodes in adjacent layers, i.e., only one hidden layer was employed.

To speed up the training and convergence of the weight values, an appropriate learning rate must be selected. It is necessary to have a learning rate that is sufficiently small to converge but large enough that the computing time remains reasonable. Therefore, a learning rate of 0.05 was selected.

**Table 1** BP neural network parameters

| | *TN* | *SD* | *DO* | *Chl-a* |
|---|---|---|---|---|
| Learning rate | 0.05 | 0.05 | 0.05 | 0.05 |
| Momentum | 0.9 | 0.9 | 0.9 | 0.9 |
| Iteration | 800 | 3,000 | 5,000 | 5,000 |
| Input nodes | 2 | 2 | 6 | 7 |
| Output nodes | 1 | 1 | 1 | 1 |
| Hidden nodes | 20 | 20 | 60 | 30 |

#### *3.2.2.2 Parameters for RBF*

The key parameter of the RBF neural network model is the number of neurons. In this study, a loop programme was written to determine the number of neurons for the minimum mean square error. In addition, the distribution density of the RBF must be determined. A spread that is too large requires many neurons to fit a fast-changing function. A spread that is too small requires many neurons to fit a smooth function and the network may not generalise well. Therefore, the distribution density was determined as one.

### *3.2.3 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 when predicting changes of eutrophication indicators. The RMSEs and correlation coefficients of the models are shown in Tables 2 and 3, respectively. It should be noted that further analysis is required to assess the effects of the input variables and their contribution to the network output (Lee et al., 2003).

**Table 2** RMSE of the models

| *RMSE* | *BP* | | *RBF* | |
|---|---|---|---|---|
| | *Train* | *Test* | *Train* | *Test* |
| TN | 0.15 | 0.14 | 0.17 | 0.16 |
| SD | 0.09 | 0.12 | 0.09 | 0.12 |
| DO | 0.04 | 0.14 | 0.09 | 0.08 |
| Chl-a | 0.07 | 0.14 | 0.09 | 0.14 |

**Table 3** Correlation coefficient of the models

| *R* | *BP* | | *RBF* | |
|---|---|---|---|---|
| | *Train* | *Test* | *Train* | *Test* |
| TN | 0.72 | 0.69 | 0.62 | 0.69 |
| SD | 0.82 | 0.74 | 0.82 | 0.73 |
| DO | 0.96 | 0.65 | 0.81 | 0.29 |
| Chl-a | 0.90 | 0.76 | 0.81 | 0.49 |

### *3.2.4 Sensitivity analysis*

To evaluate the effect of each input variable on the ANN models, the most commonly used method is sensitivity analysis, when carried on a relatively complicated network (Maier et al., 1998). The sensitivity demonstrates how the trained network reacts to changes of each input. Each input of the eight models was altered by 5%, 10% and 20%. The change in the output caused by the change of input was then calculated. The sensitivity of each input can be given by (Lee et al., 2003):

$$Sensitivity\ (\%) = \frac{1}{N_p}\sum_{i=1}^{N_p}\left(\frac{Change\ In\ Output\ (\%)}{Change\ In\ Input\ (\%)}\right)_i \times 100, \tag{4}$$

where $N_p$ denotes the number of examples, constructed with inputs and corresponding outputs, in the training dataset. The sensitivity of each input variable for the eight models is shown in Table 4. The input indicators with sensitivity higher than 100% are assumed to be very sensitive model inputs.

**Table 4** Sensitivity of each input variables for eight models

| *Variable* | *BP* | | | | *RBF* | | | |
|---|---|---|---|---|---|---|---|---|
| | *TN model* | *SD model* | *DO model* | *Chl-a model* | *TN model* | *SD model* | *DO model* | *Chl-a model* |
| Month | - | **115** | 74 | 36 | - | **109** | 40 | 78 |
| Chl-a | - | **140** | 99 | - | - | **109** | 66 | - |
| SD | **103** | - | - | 79 | **411** | - | - | 102 |
| $NH_3$-N | **164** | - | 79 | - | **408** | - | 36 | - |
| Temperature | - | - | 68 | 63 | - | - | 38 | 94 |
| pH | - | - | **121** | 36 | - | - | 24 | 84 |
| $BOD_5$ | - | - | 81 | - | - | - | 33 | - |
| DO | - | - | - | 48 | - | - | - | 80 |
| TN | - | - | - | 45 | - | - | - | 91 |
| TP | - | - | - | 30 | - | - | - | 80 |

Table 4 shows that the BP-ANN is very sensitive to changes of $NH_3$-N in the TN model and in the RBF-ANN model, TN is very sensitive to changes in SD. The RBF-ANN model did not show significant sensitivity to changes of pH in the DO model, while the BP-ANN demonstrated an opposite reaction in which pH was the most sensitive variable. This indicates that the sensitivities of the different networks demonstrated significant differences.

## 4 Model results and discussions

### *4.1 TN model*

The ANN models based on two algorithms were developed to simulate monthly TN concentration in Lake Fuxian. Figures 3 and 4 show the comparison between the BP-ANN and RBF-ANN forecasts for the TN models with respect to the observed TN values. The results indicate that the phases and approximate magnitudes of the TN concentrations have been quite well predicted. The ANN models can simulate the TN concentration with an accuracy of a degree or less.

**Figure 3** Predicted and observed TN for training set of BP and RBF models

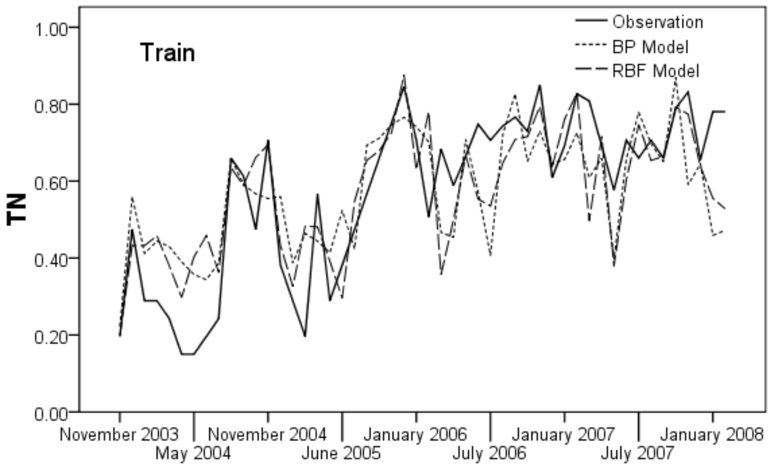
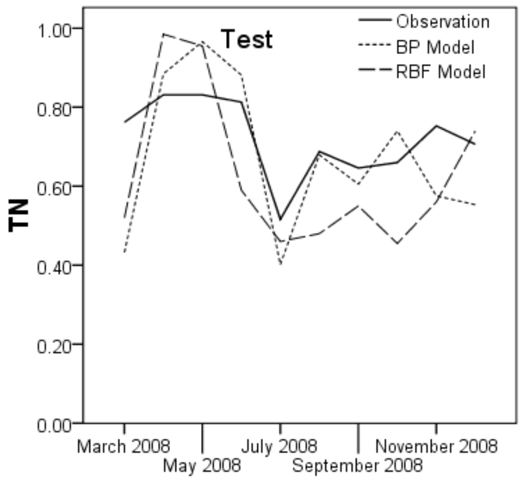



**Figure 4** Predicted and observed TN for testing set of BP and RBF models



Tables 2 and 3 show that the RMSEs of the training and testing data are 0.15 and 0.14, respectively, for the BP-ANN model. The correlation coefficients for the training and testing data are 0.72 and 0.69, respectively. For the RBF-ANN model, the RMSEs of the training and testing data are 0.17 and 0.16, respectively. The correlation coefficients for the training and testing data are 0.62 and 0.69, respectively. This indicates that the BP-ANN model predictions are better correlated with the actual TN concentration for both the training and testing data. Although this model is constructed for TN, it can also be used to predict $NH_3$-N. If we know the TN and SD values, we may use a trial-and-error method to determine the concentration of $NH_3$-N, i.e., TN and SD values can be used as input variables to the ANN models for output of a $NH_3$-N variable, which can then be used to predict $NH_3$-N concentrations.

## *4.2 SD model*

SD can provide a great deal of information about lake water quality and, together with Chl-a, has become a routine measure of lake trophic status. Hence, the proposed ANN models were also developed to simulate monthly SD values in Lake Fuxian. Figures 5 and 6 show the results of the BP-ANN and RBF-ANN models for SD training and testing data, respectively. It should be noted that the prediction trends obtained by the ANN models were mostly consistent. The results show that adequate SD prediction can be obtained with month and Chl-a data as the input layers.

**Figure 5** Predicted and observed SD for training set of BP and RBF models

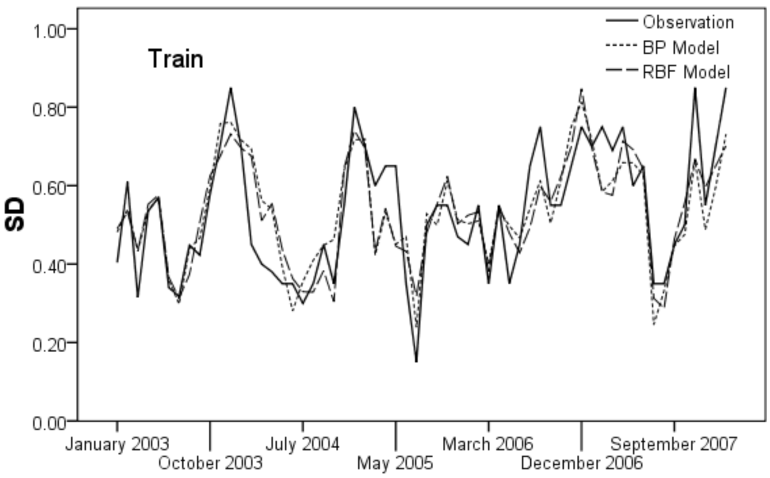
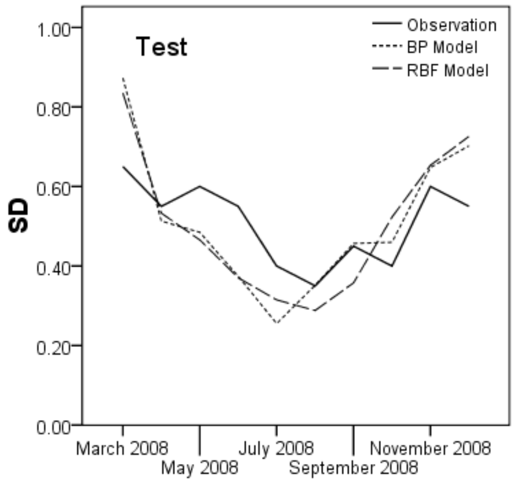


**Figure 6** Predicted and observed SD for testing set of BP and RBF models




Tables 2 and 3 show that the RMSEs of the training and testing data for the BP-ANN model are consistent with those of the RBF-ANN model at 0.09 and 0.12, respectively. Compared with RBF-ANN model, the correlation coefficients of the testing data for the BP-ANN were higher at 0.73 and 0.74, respectively and the correlation coefficients of the training set were the same at 0.82. This indicates that there were no significant differences of predictive ability between the two ANN models.

According to the developed SD model, there are two factors that influence SD in Lake Fuxian: Chl-a and month. As with other variables, SD may considerably vary in any given lake between and within seasons; therefore, it is desirable to have a 'season indicator'. The month factor shows that SD changes seasonally and is used to indicate that some variables, such as temperature, average rain and pollutant loading, change monthly.

### *4.3 DO model*

Oxygen concentrations and rates of depletion have been used to characterise lakes and, in some instances, can be related back to nutrient status. According to the DO mechanism, we can establish the ANN models using month, pH, Chl-a, $NH_3$-N, $BOD_5$ and temperature as input. Figures 7 and 8 show the results of the BP-ANN and RBF-ANN models for DO training and testing data, respectively. From Fig. 7, it can be seen that the BP-ANN model predictions are better correlated with the observed DO concentrations for the training data. However, it was difficult to maintain good consistency with the observed values with the RBF-ANN model. This indicates that the performance of the BP-ANN model is much better than the RBF-ANN model for DO prediction.

In addition, the RMSEs and correlation coefficients of the two data sets for the BP-ANN model were 0.04 and 0.96 for the training set and 0.14 and 0.65 for the testing set, respectively. For the RBF-ANN model, the RMSEs and correlation coefficients of the two datasets were 0.09 and 0.81 for the training set and 0.08 and 0.29 for the testing set, respectively. These results indicate that the performance of the BP-ANN model was much better than that of the RBF-ANN model for predicting DO concentrations.

**Figure 7** Predicted and observed do for training set of BP and RBF models

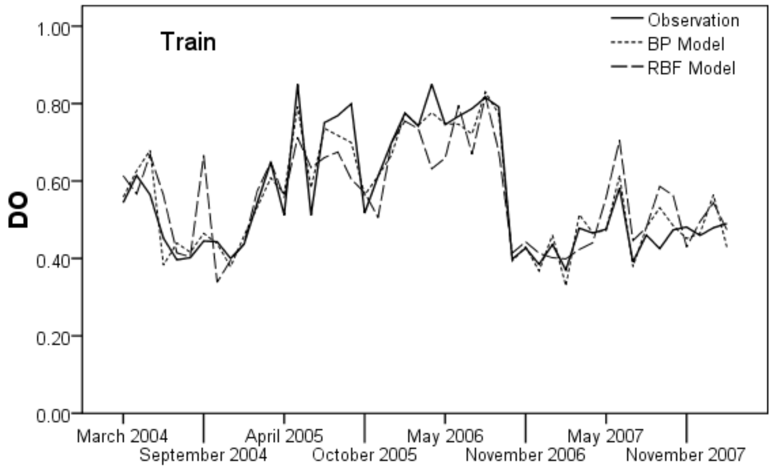
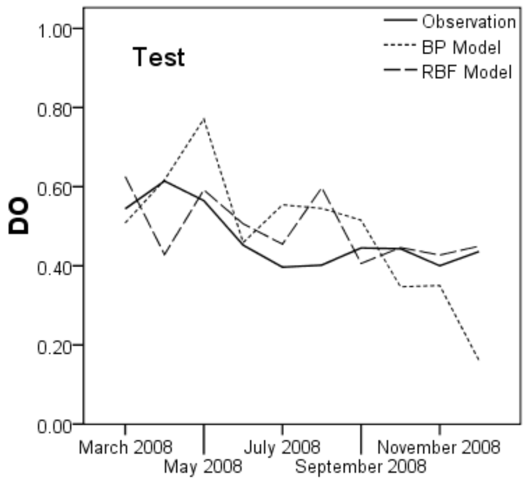



**Figure 8** Predicted and observed do for testing set of BP and RBF models




## *4.4 Chl-a model*

Chl-a is a very important indicator of the existence and degree of eutrophication in water bodies. Chl-a levels have a close and significant correlation with both nutrients (TN and TP) and water quality variables such as pH, temperature, DO and SD. Figures 9 and 10 show the results of the two ANN models for Chl-a training and testing data, respectively. Both model predictions seem to closely follow the actual Chl-a concentrations except for some peak values.

**Figure 9** Predicted and observed Chl-a for training set of BP and RBF models

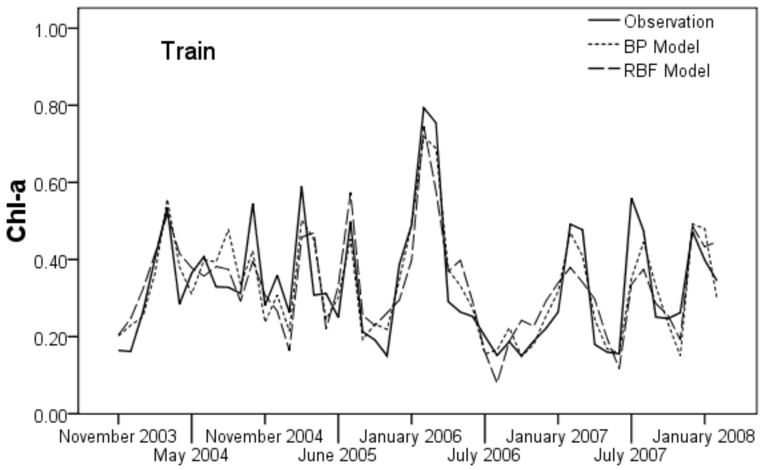
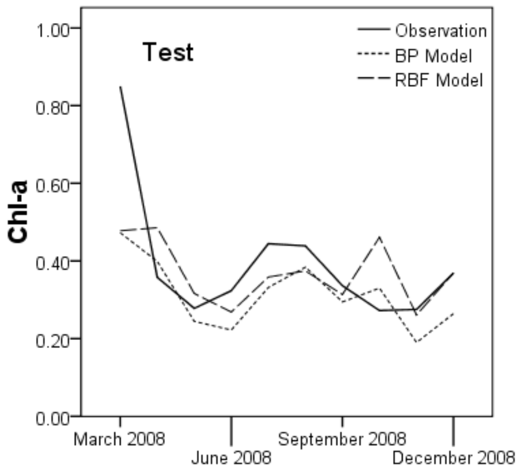



**Figure 10** Predicted and observed Chl-a for testing set of BP and RBF models




Furthermore, the correlation coefficients of the two datasets for the BP-ANN model were superior to those of the RBF-ANN model at 0.90 and 0.81 for the training set and 0.76 and 0.49 for the testing set, respectively. The RMSEs of the training data for the BP-ANN were lower than that of the RBF-ANN, at 0.07 and 0.09, respectively. This demonstrates that although both models provide good results, the BP-ANN model predictions are better correlated with the observed Chl-a values for both the training and testing sets.

From the above results obtained by the ANN models for TN, SD, DO and Chl-a, it is evident that the DO and Chl-a prediction results were much better than the other two variables. This may be related to the number of input variables. For example, the input variables for the TN model were only $NH_3$-N and SD; other parameters that are closely related to the concentration of TN, such as nitrate and nitrite, did not appear in the input layer owing to a lack of relevant data. In addition to Chl-a and month, it is likely that SD is also affected by other abiotic factors such as suspended particles and colour. However, these factors are insufficiently represented in the collected data. Thus, to establish a superior prediction system for eutrophication in Lake Fuxian, a continuing monitoring network is needed.

Generally, the prediction performances of the BP-ANN models were superior to those of the RBF-ANN models for the lake under investigation. This is primarily because the principles of the two algorithms are different. In the BP algorithm, a set of inputs and outputs is selected from the training data and the network calculates the output based on the inputs. This output is subtracted from the actual output to find the output-layer error. The error is back-propagated through the network and the weights are suitably adjusted. This process continues for a number of prescribed sweeps or until a predetermined error tolerance is reached. In contrast, for a given training error, the RBF network can automatically add/remove nodes to/from the hidden layer as required until the actual error is below the given error. The training process then stops and the size of the network may be larger than that of the best model, which is achieved with fewer free parameters requiring estimation (Lu et al., 2004).

TN, SD, DO and Chl-a are the main indicators for lake eutrophication. The proposed ANN models were able to predict TN, SD, DO and Chl-a concentrations fairly well using input parameters, even though there were limited datasets and some unknown factors that affect water quality. The results show that the complex behaviour in the eutrophication process could be modelled using the ANN technique and we successfully estimated some extreme values from the training dataset that was used to train the neural network. The proposed model can be used to:

1 estimate the values of TN, SD, DO and Chl-a when real values cannot be obtained

2 estimate interpolated data between two consecutives samples

3 simulate different water quality scenarios for extreme ranges of input and output parameters.

This eutrophication modelling approach is helpful for predicting TN, SD, DO and Chl-a levels at any location or time in the domain of interest where there are training stations. The ANN was able to satisfactorily model TN, SD, DO and Chl-a levels in Lake Fuxian in very short computational time. The TN and SD models will be further tuned for higher accuracy by including data on nitrate, nitrite, suspended particles and colour as input parameters. ANN modelling can help optimise the monitoring network and might help to analyse the factors that control the occurrence and development of harmful algae blooms, as well as providing fast predictions for forecasting.

## 5 Conclusions

Eight ANN models based on BP and RBF algorithms were developed to predict the major constituent concentrations of water quality of Lake Fuxian. Despite largely unknown factors that control eutrophication and the limited dataset size, relatively good correlations were obtained between the observed and predicted values. The simulation results have shown that the ANN models can preserve nonlinear characteristics between input and output variables for the lake under investigation. The RMSE between the predicted values and the observed data was less than 0.2 for all eight models. The correlation coefficients between the predicted values and the observed data were all above 0.7 for the four BP-ANN models. In this study, the performance of the BP-ANN model was better than the RBF-ANN model, indicating that the complex mechanisms in Lake Fuxian can be better quantified and expressed by the BP-ANN models. The BP-ANN model was proved to be a robust and highly accurate intelligence technique to predict TN, SD, DO and Chl-a values. The modelling approach described in this study for the analysis of the eutrophication problem in 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

Adeloye, A. (2009) 'The relative utility of regression and artificial neural networks models for rapidly predicting the capacity of water supply reservoirs', *Environmental Modelling & Software*, Vol. 24, No. 10, pp.1233–1240.

Bagheri, S. and Yu, T. (2008) 'Hyperspectral sensing for assessing nearshore water quality conditions of Hudson/Raritan Estuary', *Journal of Environmental Informatics*, Vol. 11, No. 2, pp.123–130.

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', *Ecological Modelling*, Vol. 120, Nos. 2–3, pp.199–211.

Cui, Y.D., Liu, X.Q. and Wang, H.Z. (2008) 'Macrozoobenthic community of Fuxian Lake, the deepest lake of southwest China', *Limnologica*, Vol. 38, No. 2, pp.116–125

French, M. and Recknagel, F. (1994) 'Modeling algal blooms in freshwaters using artificial neural networks', in Zanetti, P. (Ed.): *Computer Techniques in Environmental Studies V*, Vol. 2, pp.87–94, Environment Systems, Computational Mechanics Publications, Boston.

Gao, J.Q., Xiong, Z.T., Zhang, J.D. and Zhang, W.H. (2009) 'Phosphorus removal from water of eutrophic Lake Donghu by five submerged macrophytes', *Desalination*, Vol. 242, Nos. 1–3, pp.193–204.

Gardner, M.W. and Dorling, S.R. (1998) 'Artificial neural network: the multilayer perceptron: a review of applications in atmospheric sciences', *Atmospheric Environment*, Vol. 32, No. 14, pp.2627–2636.

Haykin, S. (1998) *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.

Huang, Y.T. and Liu, L. (2008) 'A hybrid perturbation and morris approach for identifying sensitive parameters in surface water quality models', *Journal of Environmental Informatics*, Vol. 12, No. 2, pp.150–159.

Jin, J.L. and Wei, Y.M. (2008) *Generalized Intelligent Assessment Methods for Complex Systems and Applications*, Science Press, Beijing (in Chinese).

Kunwar, P.S., Ankita, B., Amrita, M. and Gunja, J. (2009) 'Artificial neural network modeling of the river water quality – a case study', *Ecological Modelling*, Vol. 22, No. 6, pp.888–895.

Kuo, J.T., Hsieh, M.H., Lung, W.S. and She, N. (2007) 'Using artificial neural network for reservoir eutrophication prediction', *Ecological Modelling*, Vol. 200, Nos. 1–2, pp.171–177.

Kurunça, A., Yüreklia, K. and Çevikb, O. (2005) 'Performance of two stochastic approaches for forecasting water quality and streamflow data from Yeşilırmak River, Turkey', *Environmental Modelling & Software*, Vol. 20, No. 9, pp.1195–1200.

Lee, G.F., Rast, W. and Jones, R.A. (1978) 'Water report: eutrophication of water bodies: insights for an age old problem', *Environmental Science & Technology*, Vol. 12, No. 8, pp.900–908.

Lee, J.H.W., Yan, H., Dickman, M. and Jayawardena, A.W. (2003) 'Neural network modelling of coastal algal blooms', *Ecological Modelling*, Vol. 159, Nos. 2–3, pp.179–201.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S. (1996) 'Application of neural networks to modeling nonlinear relationships in ecology', *Ecological Modelling*, Vol. 90, No. 1, pp.39–52.

Lu, W.Z., Wang, W.J., Wang, X.K., Yan, S.H. and Lam, J.C. (2004) 'Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong', *Environmental Research*, Vol. 96, No. 1, pp.79–87.

Maier, H.R., Dandy, G.C. and Burch, M.D. (1998) 'Use of artificial neural networks for modeling cyanobacteria Anabaena spp. in the River Murray, South Australia', *Ecological Modelling*, Vol. 105, No. 1, pp.257–272.

Maier, H.R., Morgan, N. and Chow, C.W.K. (2004) 'Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters', *Environmental Modelling & Software*, Vol. 19, No. 3, pp.485–494.

Palani, S., Liong, S.Y. and Tkalich, P. (2008) 'An ANN application for water quality forecasting', *Marine Pollution Bulletin*, Vol. 56, No. 9, pp.1586–1597.

Portielje, R. and van der Molen, D.T. (1999) 'Relationships between eutrophication variables: from nutrient loading to transparency', *Hydrobiologia*, Vols. 408–409, No. 0, pp.375–387.

Schalkoff, R. (1992) *Pattern Recognition: Statistical, Structural and Neural Approaches*, Wiley, NewYork.

Shuhaibar, B.N. and Riffat, R. (2008) 'a process for harmful algal bloom location prediction using GIS and trend analysis for the terrestrial waters of Kuwait', *Journal of Environmental Informatics*, Vol. 12, No. 2, pp.160–173.

Singh, K.P., Basant, A., Malik, A. and Jain, G. (2009) 'Artificial neural network modeling of the river waterquality – a case study', *Ecological Modelling*, Vol. 220, No. 6, pp.888–895.

Wieland, R. and Wilfried, M. (2008) 'Adaptive fuzzy modeling versus artificial neural networks', *Environmental Modelling & Software*, Vol. 23, No. 2, pp.215–224.

Young, 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', *Environmental Modelling & Software*, Vol. 26, No. 10, pp.1199–1210.

Zeng, W.H., Song, Q.L., Liu, H.C. and Wang, T. (2010) 'Research on ANN-based pre-warning water bloom model of LiuHai Lake in Beijing', *Procedia Environmental Sciences*, Vol. 2, No. 1, pp.625–635.

Zhang, Q. and Stanley, S.J. (1997) 'Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling', *Water Research*, Vol. 31, No. 9, pp.2340–2350.

Zheng, C., Yang, W. and Yang, Z.F. (2011) 'Strategies for managing environmental flows based on the spatial distribution of water quality: a case study of Baiyangdian Lake, China', *Journal of Environmental Informatics*, Vol. 18, No. 2, pp.84–90.