An ensemble simulation approach for artificial neural network: An example from chlorophyll a simulation in Lake Poyang, China

Jiacong Huang, Junfeng Gao *

Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, 73 East Beijing Road, Nanjing 210008, China

ARTICLE INFO

Article history:
Received 11 October 2016
Received in revised form 29 November 2016
Accepted 30 November 2016
Available online 02 December 2016

Keywords:
Artificial neural network
Ensemble
Stability
Model fit
Poyang

ABSTRACT

Artificial neural network (ANN) models have been widely used in environmental modeling with considerable success. To improve the reliability of ANN models, ensemble simulations were applied in this study to develop four ANN ensemble models for chlorophyll a simulation in the largest freshwater lake (Lake Poyang) in China. Reliability (evaluated by model fit and stability) of these ANN ensemble models was compared with that of single ANN models from ensemble members. The model fit of these single ANN models varied significantly over repeated runs, indicating the unstable performance of the single ANN models. Comparing with the single ANN models, the ANN ensemble models showed a better model fit and stability, implying the potential of ensemble simulation in achieving a more reliable model. An ensemble size of 30 was adequate for the ANN ensemble models to achieve a good model fit, while an ensemble size of 50 was adequate to achieve good stability. This case study highlighted both the necessity and potential of the ensemble simulation approach to achieve a reliable ANN model with good model fit and stability.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Artificial neural network (ANN) has been increasingly used in environmental modeling due to its high potential in forecasting complex relationships (Lohani et al., 2011; Sudheer et al., 2002; Sudheer and Jain, 2004; Wu et al., 2014a,b). Its model fit was widely demonstrated to be better than other models, such as regression and mechanistic models (Amiri and Nakane, 2009; Huang et al., 2014). However, different from mechanistic models, some processes in developing ANN models, such as data division, were implemented based on a random basis (Maier et al., 2010). This means that an ANN model may perform differently over repeated runs (simulations). For example, a simulation with a good model fit may be occasionally obtained by a modeler, however, may not be easy to reproduce by other modelers (Abrahart et al., 2008; Elshorbagy et al., 2010). This weakness constrains us to get consistent findings as previous studies using ANN models. From this perspective, a reliable ANN model requires not only an acceptable model fit, but also an acceptable stability (similar results over repeated runs). Strategies are needed to improve the repeatability of ANN models.

Ensemble simulations have been increasingly used for ANN models to improve their model fits in environmental modeling (Jeong and Kim, 2005; Kasiviswanathan et al., 2013; Khalil et al., 2011; McIntyre et al., 2005; Zaier et al., 2010). To implement the ensemble simulations, many ANN models were first produced based on different training data sets. These ANN models were then combined together, and could generally achieve a better model fit rather than a single ANN model (Shu and Burn, 2004). Many ensemble techniques can support the implementation of ensemble simulations, such as bagging and boosting for creating ensemble members, and averaging and stacking for combining ensemble members (Breiman, 1996a; Jeong and Kim, 2009; Shu and Burn, 2004; Wolpert, 1992; Zhou, 2012). Ensemble simulations for ANN models have been highly emphasized for its necessity (Hansen and Salamon, 1990; Shu and Ouarda, 2007; Zhou, 2012). However, statistics on 210 ANN applications reviewed by Maier et al. (2010) showed that single ANN model was far more widely used than ANN ensemble model.

The main objective of this paper is to develop a stable ANN model for a case of chlorophyll a simulation in Lake Poyang, China. A strategy of ensemble simulations was coupled with ANN models for this purpose. The potential of the ensemble simulations in improving model fitting and stability of ANN models was discussed based on the case study.

2. Material and methods

2.1. Study area and data

Lake Poyang, the largest freshwater lake in China, has a surface area of 3283 km² and a mean depth of 5.1 m. Five major inflows include Ganjiang, Fuhe, Xinjiang, Raohe and Xiushui rivers (Fig. 1). The water of Lake Poyang flows into the Yangtze River through a narrow outlet with an annual mean outflow of 5067 m³/s. Due to the high discharge of the connected rivers, the mean hydraulic retention time is as short...
During the last decade, a large amount of pollutants was loaded into the lake due to human activities (Lu et al., 2012). Therefore, eutrophication of Lake Poyang increasingly concerned (Wu et al., 2013; Wu et al., 2014a,b).

Data that are highly related to phytoplankton dynamics were collected, including meteorological, hydrological and water quality data. The daily meteorological data including precipitation ($P_r$, 0–48 mm) and sunshine hours ($PAR$, 0–12.5 h) were collected from three weather stations of the China Meteorological Administration (Fig. 1). $P_r$ data is included because the chlorophyll $a$ in Lake Poyang might be diluted by the rainfall. The daily hydrological data include discharge ($Q$, $-$1560 to 19,100 m$^3$/s) and water level ($WL$, 7.51–19.77 m). These data were recorded at the hydrological station of Site 1 (Fig. 1) by the Bureau of Hydrology, Changjiang Water Resources Commission. The water quality data, consisting of the chlorophyll $a$ concentration ($Chl_a$, 0.58–14.90 μg/L), water temperature ($WT$, 7.0–32.0 °C), total phosphorus ($TP$, 0.01–0.28 mg/L) and nitrogen ($TN$, 0.42–1.84 mg/L), were obtained from the Ministry of Environmental Protection of the People's Republic of China. Water samples at Site 2 were collected seasonally during 2005–2009, and were collected monthly during 2009–2011. In 2011, monthly sampling was carried out at another six sampling sites (Fig. 1). During this 7-year period, 166 instances (samples) were collected for model development.

2.2. Artificial neural network ensembles

Taking chlorophyll $a$ simulation of Lake Poyang as an example, a single ANN model and four ANN ensemble models were developed using different approaches for creating and combining ANN ensemble members. Based on these models implemented in Matlab software, a number of simulations were carried out and compared to evaluate their model fits and stabilities. Workflow is given in Fig. 2 with a step-by-step description given as follows (Sections 2.2.2–2.2.4).

2.2.1. Configurations for artificial neural networks

(1) Data rescale and division

All the variables were normalized to 0.1–0.9 range using a linear transformation to ensure an equal attention to each variable. Available data were randomized and divided into three datasets for model training, testing and validation, respectively. The proportion for training, testing and validation datasets is about 0.75:0.1:0.15. Thus, 124, 17 and 25 instances were used for model training, testing and validation, respectively.

(2) Model structure

The ANN structure used in this study includes an input layer, a hidden layer and an output layer (Fig. 2). The input layer includes seven input variables with a linear transfer function, and the output layer includes a single node of $Chl_a$ with a typical transfer function of sigmoid (Fig. 2). This sigmoid transfer function has been commonly used in ANN modeling (Alizadeh et al., 2012; Mohanraj et al., 2012; Piotrowski et al., 2014). The input and output variables were linked by the nodes of hidden layer. The node number of the hidden layer generally varied from 2 to 20. In this case, eight nodes for the hidden layer were found to be optimal using a trial-and-error approach.

(3) Model training

ANN models were trained using the Levenberg-Marquardt backpropagation algorithm. Selection of this training function was based on testing the commonly used functions, such as Levenberg-Marquardt backpropagation, gradient descent with momentum backpropagation and gradient descent with momentum and adaptive learning rate backpropagation (Maier et al., 2010). The training process was stopped when the training error (root mean square error) for the testing dataset...
began to rise, or the training error was lower than the goal value (0.1), or a maximum of 500 iterations were reached. This early stopping rule aimed to prevent ANN model from overfitting.

### 2.2.2. Creating ensemble members of artificial neural network

For creating ANN ensemble members, bagging and boosting approaches have attracted much attention in recent years (Zhou, 2012). Both bagging and boosting approaches are based on resampling techniques to alter the training sets for training each ANN ensemble member. In this case study, the training dataset \( D \) consisted of 124 instances \( \{XY_1, XY_2, ..., XY_i, ..., XY_{124}\} \), where \( XY_i \) is the \( ith \) instance, \( x_i \) and \( y_i \) are the input variables and output variable (chlorophyll \( a \) of the \( ith \) instance (Fig. 2).

The bagging approach aimed to generate a training data set \( D_t \) for each ANN ensemble member. Each instance in \( D \) was assigned a probability of 1/124 to be selected into the training dataset of each ANN ensemble member. The training set of a new ANN ensemble member was generated by randomly sampling 124 times from the training dataset of \( D \). Thus, some instances in \( D \) may be selected several times into \( D_t \), while others may be left out.

The main idea of the boosting approach was to generate a sequence of ANN ensemble members. The training set of a new ANN ensemble member was obtained based on the performance of the previous member. In other words, the instances that are not well predicted by the previous member have a higher probability to be selected in the next ensemble member. An improved algorithm of the boosting approach (named AdaBoost.R2) was employed in this case study (Sharkey, 1999). The AdaBoost.R2 algorithm updated the rule to determine the training dataset \( D_t \) continually, and was implemented by following the steps.

1. At the first step, each instance in \( D \) had a same probability to be included in the training set of \( D_t \). A probability distribution \( P_t(i), i = 1, 2, ..., n \) was used to select instance from \( D \) to \( D_t \) at the time step of \( t \). \( P_t(i) \) represented the probability of the \( ith \) instance to be included in the training set of \( D_t \), \( n \) is the instance number \( n = 124 \). Thus, for \( t = 1 \),

\[
P_{t=1}(i) = 1/n, \quad i = 1, 2, ..., n
\]

(1)

2. Generate a new training dataset \( D_t \) from the training dataset \( D \) based on the probability distribution of \( P_t(i) \).

3. Generate an ANN model (i.e., an ANN ensemble member) using the training set of \( D_t \).

4. Calculate the maximum loss \( L_{\text{max}} \) between the measured \( y_i \) and simulation \( \hat{y}_i \) values over all the \( n \) training instances,

\[
L_{\text{max}} = \max \left( \left| y_i - \hat{y}_i \right| \right) \quad i = 1, 2, ..., n
\]

(2)

5. Calculate the loss \( L_i \) for each instance in the training dataset of \( D_t \),

\[
L_i = 1 - \exp \left[ -\frac{\left| y_i - \hat{y}_i \right|}{L_{\text{max}}} \right]
\]

(3)

6. Calculate the weight-averaged loss \( \bar{L} \),

\[
\bar{L} = \frac{\sum_{k=1}^{m} L_k P_t(i)}{\sum_{k=1}^{m} P_t(i)}
\]

(4)

7. Update the probability distribution of \( P_t(i) \),

\[
P_{t+1}(i) = \frac{P_t(i) \left( \frac{\sum_{k=1}^{m} L_k}{\bar{L}} \right)^{1-L_i}}{Z_t}
\]

(5)

where \( Z_t \) is a normalization factor.

### 2.2.3. Combining ensemble members of artificial neural network

It is widely recognized that combination of many ANN ensemble members can potentially achieve a better model fit than one (Hansen and Salamon, 1990; Zhou et al., 2002). Many approaches have been proposed to combine these ANN ensemble members. Two most widely used methods, i.e., averaging and stacking (Sharkey, 1999), were used in this study.

In the averaging method, the ANN ensemble output \( \hat{y}_\text{Ave} \) is obtained by computing the mean of these single ANN models \( \hat{y}_k \).

\[
\hat{y}_\text{Ave} = \frac{1}{m} \sum_{k=1}^{m} \hat{y}_k
\]

(6)

where \( m \) is the ensemble size.

The stacking approach combines the networks with weights varying over the feature space (Wolpert, 1992). This method has been widely used to improve the generalization capability of ANN models (Breiman, 1996b; Shu and Burn, 2004; Ting and Witten, 1999; Zailer et al., 2010; Zhou et al., 2002). Shu and Burn (2004) suggested to estimate the weights by minimizing the objective function of \( G \),

\[
G = \sum_{i=1}^{n} \left( y_i - \frac{\sum_{k=1}^{m} c_k \hat{y}_k}{\hat{y}_i} \right)^2
\]

(7)

Based on above equation, the weight coefficient of \( c_k \) \( k = 1, 2, ..., m \) for the \( kth \) ANN ensemble member was estimated by minimizing the squared absolute differences between measurement \( y_i \) and simulation \( \hat{y}_k \) (Shu and Burn, 2004). The estimated weight coefficient \( c_k \) was then used to calculate the ANN ensemble output \( \hat{y}_\text{Stack} \).

\[
\hat{y}_\text{Stack} = \sum_{k=1}^{m} c_k \hat{y}_k
\]

(8)

where \( m \) is the ensemble size.

### 2.2.4. Model fit and stability evaluation

By using different approaches for creating and combining ANN ensemble members (named ANN_Single), four ANN ensemble models were developed named ANN_Bag_Ave, ANN_Boost_Ave, ANN_Bag_Stack and ANN_Boost_Stack, respectively. ANN_Bag_Ave and ANN_Bag_Stack used the bagging approach to create ANN ensemble members, and used the averaging and stacking approaches to combine them, respectively. ANN_Boost_Ave and ANN_Boost_Stack used the boosting approach to create ANN ensemble members, and used the averaging and stacking approaches to combine them, respectively.

Ten repeated simulations (runs) for the ANN ensemble model and single ANN model were carried out, and were compared to evaluate the stability of ANN ensemble models. Our model testing experiment revealed that ten repeated simulations were adequate for stability investigation of ANN models, because the probability was very low for an unstable model to achieve similar results in all these ten repeated simulations. More similar simulation results of these ten simulations implied a higher stability. An ensemble size of 30 was chosen for these ten simulations based on the response investigation of the model performance to ensemble size. In order to evaluate model fitting of the ANN ensemble models, these ten simulations were also compared with the simulations using the single ANN model (ANN_Single). Considering only 166 samples (Section 2.1) were available for ANN
development, a k-fold cross-validation was used for model validation. These 166 samples were divided into seven sub-sets. These seven subsets were used in turn for model validation. More implementation details on the cross-validation approach can be found in Huang et al. (2016).

To determine a proper ensemble size for the ANN Ensemble model, another 96 simulations with the ensemble size varying from 5 to 100 were carried out for each ANN ensemble model. Their simulation outputs were compared to test the influence of ensemble size on model fit and stability of these ANN ensemble models. The above-mentioned simulations were compared based on five widely-used indicators of root mean square error (RMSE, μg/L), coefficient of determination (R²), mean absolute percent error (MAPE, %), mean absolute error (MAE, μg/L) and standard deviation. Computation of these model fit indicators can be found in Harmel et al. (2015) and Huang et al. (2013).

3. Results

3.1. Model fits of artificial neural network ensembles

Model fitting (evaluated by RMSE, R², MAPE and MAE) of four ANN ensemble models during validation period were compared with that of the single ANN model (ANN_Single). The cell (RMSE) values of the first row varied significantly among ten repeated simulations in Simulations 1–7 (Fig. 3), indicating that ANN_Single had significantly different model fits in these ten repeated simulations. There was no significant difference of RMSE values among Simulations 1–7 with different validation datasets (Fig. 3). The cell (RMSE) values of the first row were mostly higher than other rows (Fig. 3), indicating that the ANN ensemble models performed better than ANN_Single during the validation period. These four ANN ensemble models had an average RMSE value of 2.34, while ANN_Single had a RMSE value of 2.72 (Table 1). Model fitting did not show significant differences among these four ANN ensemble models, with their RMSE values ranging from 2.31 to 2.37 during the validation period (Table 1 and Fig. 3).

3.2. Stability of artificial neural network ensembles

Among these ten repeated simulations, both single ANN model and ANN ensemble models showed a large Chl a variation during the Chl a peak period (Fig. 4). Chl a variation among these ten repeated simulations was relatively low during the low-Chl a period. Compared with the single ANN model (ANN_Single), all these four ANN ensemble models (ANN_Bag_Ave, ANN_Boost_Ave, ANN_Bag_Stack and ANN_Boost_Stack) showed a significantly better stability (Fig. 4). Chl a variation among these ten repeated simulations was not significantly different among these four ANN ensemble models (Fig. 4).

3.3. Ensemble size

The four ANN ensemble models performed differently using different ensemble size. Model fitting of these four ANN ensemble models were significantly improved when ensemble size increases from 5 to 20, and were generally stable with the ensemble size higher than 30 (Fig. 5).

The standard deviations of these four ANN ensemble models showed a clear decreasing trend in ten repeated simulations when the ensemble size increased from 5 to 30. The models had relatively low fluctuation in RMSE values with the ensemble size larger than 60, implying that they were stable. ANN_Bag_Ave and ANN_Boost_Ave using the averaging approach had slightly higher stability than ANN_Bag_Stack and ANN_Boost_Stack using the stacking approach (Fig. 6).

4. Discussion

4.1. Performance of ANN ensemble models

Chl a in lake ecosystems has been widely simulated using various models, such as mechanistic and statistical models. Compared with 124 Chl a modeling cases using mechanistic models reviewed by Shimoda and Arhonditsis (2016), the MAPE value in the validation period (Table 1) implied that the ANN ensemble models performed better than 75% of the case studies. Moreover, the RMSE values of the ANN ensemble models among Simulations 1–7 with different validation datasets did not show significant differences (Fig. 3), indicating that the model validation was reliable. The acceptable model fit of the ANN ensemble models was mainly attributed to the high potential of ANN and ensemble simulation techniques in describing Chl a dynamics. Compared with the single ANN models, the ANN ensemble models have potential in reducing model error. For example, some single ANN models may over-estimate Chl a, while other ANN models may underestimate Chl a. Combining all these single ANN models by simple averaging method would result in a better prediction of Chl a. The model fits of the ANN ensemble models (Table 1) were not as good as the model fits in some other ANN case studies (e.g., Panda et al., 2010; Singh et al., 2009). This is because Lake Poyang is a very complex aquatic system with a large annual Chl a variation. Moreover, simulating Chl a dynamics in a lake are generally a more challenging task comparing to simulating many other indicators (e.g., water level).

4.2. Implications for environmental modeling using artificial neural network

For environmental modeling studies, it is clear that only case studies with a repeatable methodology and reproducible results can well support their conclusions (Cunningham et al., 2000). The large differences of the repeated ANN_Single simulations (Fig. 4) suggested that good model fit may be occasionally obtained by a modeler, however, may be difficult to reproduce by other modelers. Conclusions, drawn based on the simulation results of this unstable model, would show high uncertainty. It is thus worth developing a stable ANN model, that would perform consistently among different repeated runs.

Compared with the single ANN models, the ANN ensemble models obtained better model fit (Fig. 3), indicating the potential of ensemble simulation in improving model fit. Such conclusion is consistent with previous studies (Shu and Burn, 2004; Zaier et al., 2010). The similar performances of ANN ensemble models in different runs (Fig. 4) indicated that the ANN ensemble models were more stable than the single ANN models. The stability of ANN ensemble models implied that other modelers can reproduce the simulation results in this study, and achieve a consensus finding using an identical dataset.

Fig. 3. The root mean square error (RMSE) of the single ANN model (ANN_Single) and four ANN ensemble models (ANN_Bag_Ave, ANN_Boost_Ave, ANN_Bag_Stack and ANN_Boost_Stack) during the validation period. Simulations 1–7 represented the model validated using different dataset based on k-fold cross-validation method.
The potential of ensemble simulations in improving model fit and stability was not adequately investigated in the context of environmental modeling (Trolle et al., 2014). This study improved ANN models based on the advances in machine learning during the last few decades (Granitto et al., 2005; Hansen and Salamon, 1990; Zhou, 2012; Zhou et al., 2002), and succeeded in combining ANN and ensemble simulation techniques for Chl $\alpha$ simulation. Considering the potential of ensemble simulations in improving model fit and stability, ANN ensemble models were encouraged to use in environmental and ecological modeling as an alternative of single ANN models.

In this study, only 166 samples collected from Northern Lake Poyang were used for ANN model development. Data from Southern Lake Poyang were available, however, were not used in this study. This is because Lake Poyang is a large lake with high spatial heterogeneity of environmental factors. The response of Chl $\alpha$ to environmental factors in different areas can vary significantly, and was thus not reliable to be described in an ANN model (Huang et al., 2015). The weakness of limited Chl $\alpha$ data can be partly overcome by the increasing use of satellite images for Chl $\alpha$ estimation. Because Chl $\alpha$ estimation using satellite images showed great advance in both methods and applications during the past decade (Awad, 2014; Werdell et al., 2009; Zhou et al., 2015), and can provide Chl $\alpha$ data with high spatial and temporal resolutions for ANN model development.

### 4.3. Methodological issues for ensemble simulations

In this study, the combination of ANN models and ensemble simulation showed the advantage of improving model fit and stability. Such

<table>
<thead>
<tr>
<th>Period</th>
<th>Indicator</th>
<th>ANN_Single</th>
<th>ANN_Bag_Ave</th>
<th>ANN_Boost_Ave</th>
<th>ANN_Bag_Stack</th>
<th>ANN_Boost_Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>RMSE</td>
<td>2.46</td>
<td>1.69</td>
<td>1.70</td>
<td>1.68</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.25</td>
<td>0.54</td>
<td>0.53</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
<td>31.8%</td>
<td>24.7%</td>
<td>25.1%</td>
<td>24.2%</td>
<td>25.0%</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>1.48</td>
<td>1.08</td>
<td>1.08</td>
<td>1.11</td>
<td>1.12</td>
</tr>
<tr>
<td>Testing</td>
<td>RMSE</td>
<td>1.73</td>
<td>1.16</td>
<td>1.12</td>
<td>1.33</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.10</td>
<td>0.26</td>
<td>0.25</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
<td>33.3%</td>
<td>19.7%</td>
<td>19.2%</td>
<td>21.9%</td>
<td>20.8%</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>1.28</td>
<td>0.86</td>
<td>0.82</td>
<td>0.99</td>
<td>0.94</td>
</tr>
<tr>
<td>Validation</td>
<td>RMSE</td>
<td>2.72</td>
<td>2.35</td>
<td>2.31</td>
<td>2.37</td>
<td>2.32</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.21</td>
<td>0.31</td>
<td>0.33</td>
<td>0.28</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>MAPE</td>
<td>41.2%</td>
<td>29.4%</td>
<td>28.7%</td>
<td>30.5%</td>
<td>29.1%</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>1.67</td>
<td>1.31</td>
<td>1.29</td>
<td>1.37</td>
<td>1.33</td>
</tr>
</tbody>
</table>

Note: RMSE, root mean square error ($\mu$g/L); $R^2$, coefficient of determination; MAPE, mean absolute percent error (%); MAE, mean absolute error ($\mu$g/L).

Fig. 4. Measured and simulation chlorophyll $\alpha$ of four ANN ensemble models (ANN_Bag_Ave, ANN_Boost_Ave, ANN_Bag_Stack and ANN_Boost_Stack) and the single ANN model (ANN_Single) in ten repeated simulations.
strategy can be potentially used in other ANN models from methodological perspective. However, it is important to keep in mind that the success of ANN models is highly depended on the training data. In other word, the training data should well reflect the relationships between the target variable (Chl-a) and environmental factors. We can never expect to develop a successful ANN ensemble model without any successful single ANN models. To develop a successful ANN model, modelers should pay adequate attentions to the methodological issues, such as data division, and selection of an appropriate transfer function and backpropagation network training function.

Some other methodological issues should be properly addressed to ward good practice of developing ANN ensemble models. One important methodological issue is the approach selection for ANN ensemble creation and combination. These four ANN ensemble models showed very similar performances (Fig. 3), implying that the averaging and stack approaches had similar potential to improve model fit in this case study. This finding was different from previous studies, where the stacking approach generated better model fit than the averaging approach (Shu and Burn, 2004; Zaier et al., 2010). The relatively low standard deviations of ANN_Bag_Ave and ANN_Boost_Ave implied that averaging approach was better than stacking approach to obtain a stable model for this case study. This meant that good performance of stacking approach was highly relied on a proper estimation of ANN ensemble weight.

Another important methodological issue is the ensemble size determination. The simulations with different ensemble sizes (Figs. 5 and 6) revealed that increasing in ensemble size improved both the model fit and stability of ANN ensemble model significantly, however, resulted in higher computation cost. To balance between model stability and computation cost, a minimum ensemble size to achieve a reliable model is ideal. In this case study, an ensemble size of 30 was adequate to achieve a good model fit, while the ideal ensemble size in previous case studies varied from 10 to 30 (Agrafiotis et al., 2002; Hansen and Salamon, 1990; Opitz and Maclin, 1999; Zaier et al., 2010). An ensemble size of 50 was adequate to achieve good stability in this study, suggesting that good model fit and stability may require different ensemble sizes to achieve. This variation of ideal ensemble size from case to case revealed that ideal ensemble size should be determined for a specific case based on its range from previous case studies.

5. Conclusions

A strategy of ensemble simulation was coupled with ANN models for chlorophyll a simulation in Lake Poyang, China. The simulation results in this case study revealed that ensemble simulations had potential in improving both model fit and stability of the developed ANN models. Determining a proper ensemble size was important for ensemble simulations. For chlorophyll a simulation in Lake Poyang, an ensemble size of 30 was adequate to achieve a good model fit, while an ensemble size of 50 was adequate to achieve good stability. This study demonstrated the success of the ensemble simulation approach for ANN models. This ensemble simulation approach can be potentially used in other ANN models.

Acknowledgments

The project was financially supported by Major Water Resources Science and Technology Program of Jiangxi Water Resources Department (KT201406) and Model Development Project for Aquatic Ecology of Lake Poyang. The authors would like to thank the Ministry of Environmental Protection of the People’s Republic of China, the Changjiang Water Resources Commission, and the National Meteorological Information Center for providing the measured data for model development. We would like to thank two anonymous reviewers for their comments on the manuscript.