Ecological Engineering 95 (2016) 422–429

Contents lists available at ScienceDirect

Ecological Engineering

journal homepage: www.elsevier.com/locate/ecoleng

Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin

Fang Lu[a], Zhi Chen[a], Wenquan Liu[b,*], Hongbo Shao[c,*]

[a] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada
[b] Key Laboratory of Marine Sedimentology and Environmental Geology, First Institute of Oceanography, SOA, Qingdao 266061, China
[c] Institute of Agro-biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing210014, China

ARTICLE INFO

Article history:
Received 2 April 2016
Received in revised form 2 June 2016
Accepted 16 June 2016
Available online 4 July 2016

Keywords:
Back-propagation artificial neural network
Chlorophyll-a
Prediction
Lake water quality
Eco-restoring

ABSTRACT

A back-propagation artificial neural network (BPANN) model was developed in this study for the prediction of chlorophyll-a concentration in Lake Champlain. 21 years of monitoring data (1992–2012) of water quality parameters was used to train, validate and test the BPANN models. The optimal input parameters of the model were selected on the basis of the performance of models built with different combinations of input variables. To verify the model performances, the trained models were applied to field monitoring data from Lake Champlain. Prediction accuracy was measured by using the coefficient of determination (R^2) and RMSE-observations standard deviation ratio (RSR). The R^2 values of the best-performed model in the training set, validation set, testing set, and all-year data were 0.82, 0.93, 0.81, and 0.87, respectively. The corresponding RSR values of the three data sets and all-year were 0.62, 0.38, 0.53, and 0.48, respectively. Results indicated that the developed BPANN model can predict chlorophyll-a concentrations in Lake Champlain with high accuracy and provide a quick assessment of chlorophyll-a variation for lake management and eco-restoration.

1. Introduction

Lakes are significant sources of water for recreation, fishing and drinking water supply. The lake water system is affected by the municipal, agricultural and industrial waste disposal directly or indirectly discharged into lakes. Chlorophyll-a is a widely used environmental indicator of algae biomass and the eutrophication condition in lakes (Latif et al., 2003). Elevated levels of chlorophyll-a are interpreted as the possible presence of algae blooms, and the amount of algae in a lake has significant effects on the lake's physical, chemical, and biological processes (United States Environmental Protection Agency, 2014). Algal blooms could also decline lake water quality. Cyanotoxins generated by cyanobacteria in lake water could introduce human health risk during recreational activities or by drinking water (Kalaji et al., 2016; McQuaid et al., 2011; Watzin et al., 2006). In the case that immediately concentration of cyanotoxins is unavailable, chlorophyll-a is also widely accepted as a surrogate measure of cyanobacterial density and of the potential public health risk posed by cyanobacterial blooms (Wheeler et al., 2012). Therefore, it's essential to predict the chlorophyll-a concentration and, in turn, to provide advance information for water quality management and facilitate public health risk assessment.

Several studies have recently been conducted on water quality prediction models (Chibole, 2013; Wu and Xu, 2011; Zhao et al., 2012). Physical-based water quality modeling approaches are capable of simulating the internal physical processes of the aquatic system, but require extensive information that is not easily accessible (Dogan et al., 2009). Moreover, many physical-based water quality models are time-consuming (Singh et al., 2009). Artificial neural networks have several advantages over the physical-based models, they avoid the need for a hydraulic model and the specialized knowledge of the physical processes governing the fate and transport of pollutants (Bowden et al., 2006). Since a large number of factors that affect the water quality have a complicated non-linear relationship, an artificial neural network (ANN) tries to simulate the learning processes of human being; through training and calibration of the network, ANN has the ability to reflect the linear or non-linear relationships among the data (Dogan et al., 2009). Once the network is trained satisfactorily, the ANN model should be able to obtain output for a new data set. Therefore, there has been a significant increase in their application in different scientific areas, such as modeling process, pattern recognition, and time series analysis (Li et al., 2007; Patil et al., 2008).

The applications of ANN models in the prediction and forecasting of water resource variables were reviewed by Maier and Dandy

* Corresponding authors.
E-mail addresses: liuwq05@163.com (W. Liu), shaohongbochu@126.com (H. Shao).

http://dx.doi.org/10.1016/j.ecoleng.2016.06.072
0925-8574/

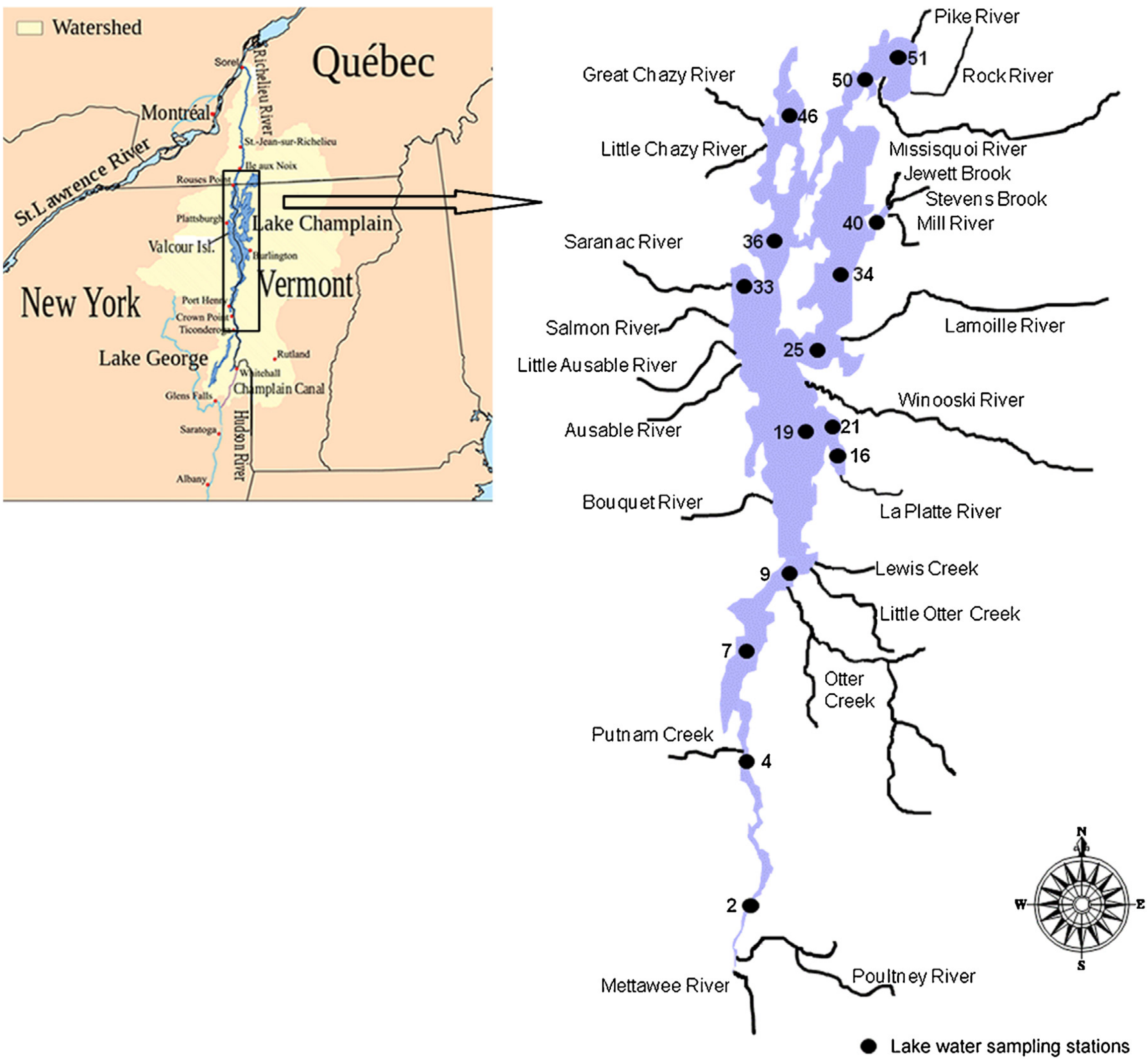

Fig. 1. Study area and sampling sites (Vermont Department of Environmental Conservation Water Quality Division and New York State Department of Environmental Conservation, 2012).

(2000), and the majority of those models used back-propagation artificial neural networks (BPANN). The BPANN model showed great potential to simulate water quality parameters, such as water temperature, salinity, DO, and Chl-a in Singapore coastal water (Palani et al., 2008). In Sahoo's study, a genetic algorithm-optimized BPANN model, multiple regression analysis, and chaotic non-linear dynamics algorithms were used to predict the stream water temperature, the results indicated that the performance efficiency of the optimized BPANN was the highest among all considered algorithms (Sahoo et al., 2009). By comparing the results of the BPANN model and the MIKE 11 physically based hydrodynamic model in the simulation of water levels in a river, the results obtained from the BPANN model were much better than that of the MIKE 11, as indicated by the Nash-Sutcliffe index and root mean square error (Panda et al., 2010). Besides the vast references of BPANN applications, the methods used for determining model inputs, dividing data sets, determining the best model structure, and the comparison between different training algorithms were also reviewed (Maier and Dandy, 2000; Maier et al., 2010; May et al., 2008; Piotrowski et al., 2014).

In Lake Champlain area, physical-based models have been applied for lake water quality modeling. A hydrodynamic model in conjunction with a water quality model was applied to Lake Champlain to simulate the current, transport and mixing for the main lake basin, as well as the kinetics, distribution and concentration of phosphorus in the lake (Mendelsohn et al., 1996). By estimating the mean annual tributary loadings using the FLUX program, a mass balance model was developed and calibrated using the BATHTUB program to simulate the total phosphorus concentration in the thirteen lake segments of Lake Champlain (Smeltzer and Quinn, 1996; Vermont DEC and New York State DEC, 1997). As illustrated previously, these models require numerous input data that are related to the modeling parameters, e.g., water quality sampling data in the lake and tributaries, tributary flow data, wastewater treatment facilities sampling data, meteorology data (including wind, precipitation and evaporation), biological data (including phytoplankton and zooplankton data), and sediment sampling data. Some of these data may be difficult to obtain, and the simulation is time-consuming if a large quantity of numerical calculation is included.

This study presents a first attempt to predict the chlorophyll-a concentration in Lake Champlain by using the back propagation artificial neural network model (BPANN). The optimal input variables of the model will be selected based on linear correlation analysis and domain knowledge. The results of the developed ANN model will be validated with the observed chlorophyll-a concentration, and the model performance will be evaluated using statistical