Ecological Modelling 364 (2017) 42–52

Contents lists available at ScienceDirect

Ecological Modelling

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

An optimization of artificial neural network model for predicting chlorophyll dynamics

Wenchong Tian[a], Zhenliang Liao[a,*], Jin Zhang[b,*]

[a] Key Laboratory of Yangtze River Water Environment, Ministry of Education, College of Environmental Science and Engineering, Tongji University, 200092 Shanghai, China

[b] Institute of Urban Water Management, Technische Universität Dresden, 01062 Dresden, Germany

ARTICLE INFO

Article history:
Received 3 February 2017
Received in revised form 16 August 2017
Accepted 12 September 2017

Keywords:
Artificial neural networks
Chlorophyll dynamics
Algal bloom
Non-stationary time series

ABSTRACT

As one of the factors to represent some species of algae, chlorophyll dynamics model has been regarded as one of the early-warning proactive approaches to prevent or mitigate the occurrence of some algal blooms. To decrease the cost of aquatic environmental in-situ monitoring and increase the accuracy of bloom forecasting, a traditional artificial neural network (ANN) based chlorophyll dynamics prediction model had been optimized. This optimization approach was conducted by presenting the change of chlorophyll value rather than the base value of chlorophyll as the output variable of the network. Both of the optimized and traditional networks had been applied to a case study. The results of model performance indices show that the optimized network predicts better than the traditional network. Furthermore, the non-stationary time series was employed to explain this phenomenon from a theoretical aspect. The proposed approach for chlorophyll dynamics ANN model optimization could assist the essential proactive strategy for algal bloom control.

1. Introduction

It has been widely reported that lakes and reservoirs are commonly susceptible to eutrophication, and the subsequent algal blooms result in an adverse effect on the drink water security (Gu et al., 2017). Due to the adverse effect on the water quality, the eutrophication induced algal blooms can disrupt water supply to the surrounding cities (Zhang et al., 2014). Traditionally, algal blooms in-situ management programs require routine monitoring and/or reactive monitoring whereby blooms are monitored episodically and at a greater frequency only when a problematic algal species are detected, or a bloom is visually observed (Coad et al., 2014). These programs have limited capacity for environmental managers to adequately monitors and respond to algal blooms due to constraints such as (i) the expense of field monitoring, (ii) staffs availability and resources, (iii) field safety issues, and (iv) large time intervals between data collection, reporting and public notification. Therefore, to decrease the cost of aquatic environmental in-situ monitoring and increase the accuracy of blooms forecasting, an early-warning proactive approach of the algae blooms forecasting model is essential to prevent or mitigate the occurrence of algal blooms, and eventually facilitate the minimization of the adverse effect of algal blooms on the water bodies (Oh et al., 2007).

In terms of the model development, there are typically two kinds of forecasting models: deductive models and inductive models. The deductive models are developed based on the existing theories and knowledge which enable users to simulate the systems' behavior (Recknagel, 1997). Many deductive models have been proposed to predict the algal blooms. For example, Grover (1991) built a mechanistic deductive model by considering the theoretical elements needed for algae growth. Wei et al. (2014) used a coupled hydrodynamic-algal biomass model to forecast the short-term cyanobacteria blooms in Taihu Lake, China. Except for the wide application, however, deductive models require detailed descriptions of physical, chemical and biological processes, and usually contain a lot of parameters for calibration (Zhang et al., 2014). To a great extent, therefore, the prediction accuracy was restricted by the lacking of knowledge about the algae-growth mechanism and chlorophyll dynamics.

In terms of the inductive models, they produce holistic information extracted from the empirical data patterns by statistic, correlation and machine learning methods which enable users to predict rather than to explain the systems' behavior (Recknagel,

* Corresponding authors.
E-mail addresses: zl_liao@tongji.edu.cn (Z. Liao), jin.zhang@hotmail.com (J. Zhang).

http://dx.doi.org/10.1016/j.ecolmodel.2017.09.013

1997; Zhao et al., 2016). Jeong et al. (2003) modeled microcystins aeruginosin bloom dynamics in Nakdong River by means of evolutionary computation and statistical approach. Cha et al. (2014) reported a Bayesian Poisson model to probabilistically predict the *Cyanobacteria* abundance in a Korean reservoir.

Among these inductive models, artificial neural network (ANN) is a kind of model which could deal with the complex information among the data by using machine learning theories. For normal forward neural networks, their structures include input-layer, hidden-layer, and output-layer. There are multiple neurons in each layer, and each of them is assigned with two parameters, namely weight value and the threshold value. By training ANN with enough samples obtained from the history of a natural system, these parameters can be adjusted to reproduce the behavior of this system (Ethem, 2010). ANN has been applied to the algae blooms and chlorophyll dynamics forecasting in the last two decades. For example, ANN was used for modeling and prediction of algal blooms (Recknagel et al., 1997) and also for modeling and prediction of zooplankton dynamics in Lake Kasumigaura (Recknagel et al., 1998). In Huelva, Western Andalucia, Spain, ANN was adapted for one-step weekly prediction of *Dinophysis Acuminata* blooms (Velo-Suárez and Gutiérrez-Estrada, 2007).

The ANN model was also applied on chlorophyll dynamics, as it is one of the factors to represent some species of algae and has been regarded as one of the early-warning proactive approaches to prevent or mitigate the occurrence of some algal blooms. To improve the understanding of chlorophyll dynamics, Coad et al. (2014) used ANN to predict daily *Chlorophyll-a* concentration. In a case study on the Yuqiao Reservoir in North China, an ANN was employed for the eutrophication forecasting and management (Zhang et al., 2015). Despite its successful application, it is stressed that the optimal ANN is generally problem dependent (Maier and Dandy, 2000; Recknagel, 2001). For this reason, it is necessary to develop and optimize the ANN for different problems to obtain the best model configurations that have a lower error with short training time and higher accuracy. Traditionally, an optimal ANN model was found by trial and error with adjusting of its structures and parameters (Maier and Dandy, 2001; Dedecker et al., 2004). However, it is difficult to find the optimal set of the possible structures and parameters. Therefore, a new method for developing and optimizing ANN models with the easier operation to predict chlorophyll dynamics is needed.

In terms of the chlorophyll dynamics, it has been widely recognized that the water quality, hydrology, and climate condition are the main influencing factors on the chlorophyll dynamics (Coad et al., 2014; Seitzinger, 1991). Compared with the base value of chlorophyll (refers to the value at the beginning of a period used as a reference or starting point for the estimation process), the value change of chlorophyll (refers to the difference between the size of the value to the end and the beginning of a period) is more sensitive to these influencing factors. In other words, the influencing factors will first determine the value change of chlorophyll, and then influence the base value of chlorophyll within a given time period. Thus, the primary hypothesis of this study is that the ANN based chlorophyll dynamics prediction model could be optimized by computing the correlations between the value change of chlorophyll and its influencing factors rather than the correlations between the base value and the influencing factors.

Consequently, the objectives of this study were to (i) explore a method to optimize ANN based chlorophyll dynamics prediction model, (ii) apply this optimized model and forecast daily *Chlorophyll-a* concentrations in a water body, and (iii) compare the accuracy of the optimized model with a traditional ANN based chlorophyll dynamics prediction model.

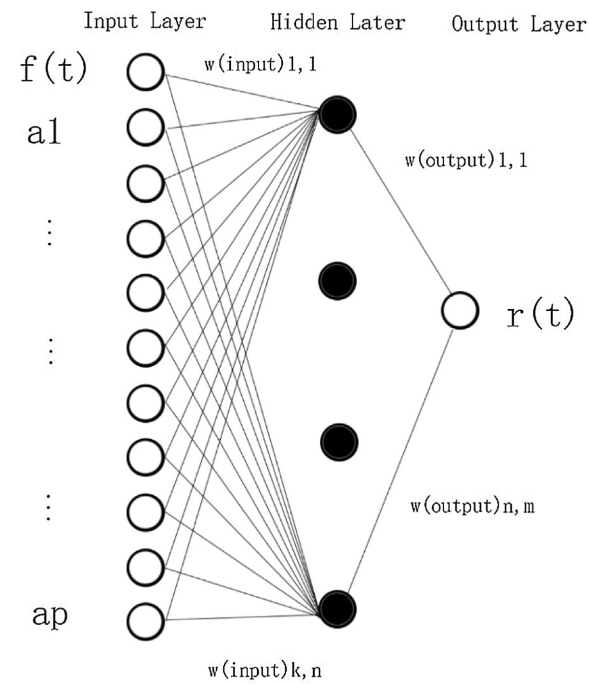

Fig. 1. Structure of an ANN model.

2. Materials and methods

2.1. Theory of ANN model

ANN applied in this study consists of an input layer with *p+1* nodes, a hidden layer with N nodes, and an output layer with one node as given in Fig. 1. *p* is the number of variables which depends on different the model designing. The number N is subjected by Eq. (1) as follows.

$$N \leq \frac{N^{TR}}{\left(N^{I}+1\right)} \tag{1}$$

where, N is the nodes number in the hidden layer; N^{TR} is the number of training samples; and N^{I} is the number of inputs. There are mainly two kinds of transfer functions for each node, i.e.: the log sigmoid as shown in Eq. (2) and the tangent sigmoid as shown in Eq. (3). In this study, the tangent sigmoid was chosen as the transfer function for the hidden layer nodes and the output layer nodes.

$$\varphi_0 = \log sig\,(x) = \frac{1}{1+exp\,(-x)} \tag{2}$$

$$\varphi_h = \tan sig\,(x) = \frac{2}{1+exp\,(-2x)} - 1 \tag{3}$$

The connections among nodes in each layer are represented by the weights ($W\,(input)$ and $W\,(output)$) and thresholds (b_1 and b_2). The initial values of weights are determined by a random starting, and randomly set between −1 and 1. Thresholds are corresponding with inputs which are set randomly between −1 and 1 at the beginning. Then the weights and thresholds are adjusted by training with samples. If let functions $a_i\,(t)\,,\, i = 1, \ldots, p$ as the inputs of ANN, and let function $f\,(t)$ as the output of ANN, t is time variable. Then the model is represented by the following equations:

$$f(t)_N = \varphi_h\{\sum_{j=1}^{N} W(output)_j \cdot \varphi_h[\sum_{i=1}^{p} W(input)_{ij} \cdot I_{iN}\,(t) + b_{1j}] + b_2\} \tag{4}$$