Measurement 53 (2014) 224–233

Contents lists available at ScienceDirect

Measurement

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

Factors affecting algal blooms in a man-made lake and prediction using an artificial neural network

Sohyun Cho [a], Byungjin Lim [a], Jaewoon Jung [a,b], Sangdon Kim [a], Hyunmi Chae [a], Jonghwan Park [a], Seoksoon Park [c], Jae K. Park [d,*]

[a] Yeongsan River Environment Research Laboratory, National Institute of Environmental Research, Gwangju 500-480, Republic of Korea
[b] Department of Rural & Biosystems Engineering, Chonnam National University, Gwangju 500-757, Republic of Korea
[c] Department of Environmental Science and Engineering, University of Ewha Womans, Seoul 120-75, Republic of Korea
[d] Department of Civil and Environmental Engineering, University of Wisconsin – Madison, 1415 Engineering Drive, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 20 December 2013
Received in revised form 18 March 2014
Accepted 31 March 2014
Available online 13 April 2014

Keywords:
Algae
Artificial neural network (ANN)
Chlorophyll-a (Chl-a)
Man-made lake
Prediction
Total organic carbon (TOC)

ABSTRACT

It is difficult to predict when, where, and how long algal blooms will occur in a water body. The objectives of this study were to determine the factors affecting algal bloom and predict chlorophyll-a (Chl-a) levels in the reservoir formed by damming a river using an artificial neural network (ANN). The automatic water quality monitoring data [water temperature, pH, dissolved oxygen (DO), electric conductivity, total organic carbon (TOC), Chl-a, total nitrogen (T-N), and total phosphorus (T-P)], weather data (precipitation, temperature, insolation, and duration of sunshine) and hydrologic data (water level, discharges, and inflows) in the man-made Lake Juam during 2008–2010 were used to develop a model to predict Chl-a as an indirect measure of the abundance of algae. The ANN was trained using the collected data during 2008–2010 and the accuracy of the model was verified using the data collected in 2011. It was found that Chl-a concentration, TOC, pH and atmospheric and water temperatures were the most important parameters in predicting Chl-a concentrations. The Chl-a prediction was most influenced by the parameters showing the algal activities such as Chl-a, TOC and pH. Due to the relatively long hydraulic retention time of ~131 days, the inflow and outflow did not affect the prediction much. Likewise, atmospheric and water temperatures did not respond to the change of the Chl-a concentration due to the lake's relatively slow response to the temperature. Most importantly, T-N and T-P were not the major factors in predicting Chl-a levels at Lake Juam. The ANN trained with the time series data successfully predicted the Chl-a concentration and provided information regarding the principal factors affecting algal bloom at Lake Juam.

1. Introduction

Lake Juam is the major water supply source serving about 2.5 million people in the City of Gwangju and part of Jeonnam province. There has been a growing concern on algal blooms and deterioration of water quality. Although the water quality of Lake Juam has maintained 'Good (II)' based on chemical oxygen demand (COD) ranging from 1 to 3 mg/L according to 'Korea Water Quality Standard for Lakes'. For the past 10 years, algal blooms have frequently occurred. Algal blooms may cause the taste-and-odor problem in drinking water, discoloration of water or scum formation. It is also necessary to prevent algal blooms because it causes various problems such as aquatic ecosystem destruction, filter clogging in water purification plants, and toxicity by blue–green algae after the rainy season [1].

* Corresponding author: Tel.: +1 608 262 7247; fax: +1 608 262 5199.
E-mail address: jkpark@wisc.edu (J.K. Park).

http://dx.doi.org/10.1016/j.measurement.2014.03.044
0263-2241/

It is necessary to investigate the conditions causing algal blooms for prevention [2,3]. The behaviors of algae in water bodies have been investigated through mathematical modeling [4–8] and other various statistical methods [9–13]. However, a modeling method has a few disadvantages: an enormous input data requirement and calibration/validation of the model by adjusting the parameters used in the model, requiring a lot of time and effort. The statistical methods are simple to predict Chl-a concentration but exhibit a limitation in its accuracy and long-term prediction. Thus, the artificial neural network (ANN), which has an easy input data entry but high accuracy, has been used in predicting changes in complex and nonlinear algal movement [14–16].

Yeon et al. [17] investigated effective factors and proper model structures of the ANN in predicting Chl-a by applying the water quality and hydrologic data to predict daily Chl-a in Lake Daecheong, Korea. Maier and Dandy [18] predicted the *Anabaena* species of cyanobacteria using weekly water quality data and hydrologic data. Recknagel et al. [19] compared the dominant species of algae in different lakes of Japan, Finland, and Australia using ANNs. Whitehead et al. [4] predicted the algal concentrations in the River Thames using time series analysis, dynamic mass balance and growth equations and ANN approaches. Lee et al. [14] also predicted the *Skeletonema* species that is one of phytoplanktons in the Tolo Harbour in Hongkong and Lamma Island using the ANN method. Yao et al. [20] predicted blue–green algae using a distributed genetic algorithm for configuring the ANN. Oh et al. [21] predicted the behavior of algae using the ANN method after analyzing the pattern of factors affecting algae population using a self-organizing method.

Although various studies on algae have been conducted using ANNs in different areas, the prediction of algae in Lake Juam using ANN are still insufficient. The objectives of this study were to investigate the principal factors affecting Chl-a using the water quality and hydrologic data collected at Lake Juam and to evaluate the applicability of ANN to Lake Juam.

2. Materials and methods

2.1. Field and monitoring site

Lake Juam is an artificial reservoir constructed in the midstream of the Bosung River (35°3′N, 127°14′E) that is the primary branch of the Seomjin River, which is one of the five major rivers in Korea, and connected to Lake Sangsa through a waterway tunnel. Lake Sangsa is designed to control the discharge to Isa stream. The surface area of Lake Juam is 33 km^2, the maximum depth is 40 m, the hydrologic retention time is approximately 131 days, and the maximum water volume is 4.5×10^8 m^3. The lake water is used for drinking water, river maintenance water, and industrial water. The water is also used for hydroelectricity generation and flood/drought control while providing recreational space. The study site is shown in Fig. 1.

2.2. Dominant algal species in Lake Juam

Lake Juam has been a relatively pristine water body and thus had few algal bloom problems until recently. However, due to increased nonpoint sources, Lake Juam had the first algal bloom warning since 1998 on September 8, 2009. Korea introduced the algal bloom warning system in 1998 [22]. The criteria for issuing an algal bloom warning are shown in Table 1. Excess growth of green algae was caused by high-level nutrient and algae input from forest areas after heavy rain and exceptionally high temperatures [23]. The dominant algal species reported by Lee et al. [24] are summarized in Table 2.

2.3. Construction of a neural network

The ANN is a feed forward network that has been used in many disciplines mainly because of its aptitude to approximate the transfer function between a given set of input and the equivalent output data [25]. A typical ANN configuration consists of an input layer, hidden layers and an output layer. Among many variations of ANN, Levenberg–Marquardt algorithm was selected because it was reported to have the fastest method for training moderate-sized feed-forward neural networks [26]. The basic calculation process used in this case is presented as follows:

$$h_j = \sum_{i=0}^{\infty} w_{ij} r_i (j = 1, \ldots, n) \tag{1}$$

where n is the total number of nodal points in the hidden layer, w is the weighted value applied to the paths from i to j, r_i is the input value from the nodal point of i, and h is the value at the nodal point of j in the hidden layer. The effective reception signal at the nodal point j is the sum of all input signals. Then, the combined signal is transformed to generate output signals using a conversion function as follows:

$$o_k = \sum_{j=1}^{n} w_{kj} f(h_j) = \sum_{j=1}^{n} w_{kj} f\left(\sum_{i=1}^{m} w_{ji} r_i\right)(k = 1, \ldots, l) \tag{2}$$

where j is the selected conversion function, w is the weighted value applied to the paths from j to k, o_k is the output value, and l is the total number of nodal points in the output layer.

The nonlinear conversion function makes it possible to consider a nonlinear relationship between input and output data in a neural network using the following equation:

$$f(h) = \frac{2}{1 + \alpha e^{-h}} - 1 \tag{3}$$

where h is the input value of a nodal point, $f(h)$ is the output value of a nodal point, and α is an amplification coefficient for considering the nonlinear behavior of the input data.

The training process can be configured by determining a set of new weighted values that minimize the mean square error in the output as follows:

$$E = \sum_{i=0}^{\infty} (t_{kj} - o_k)^2 \tag{4}$$

where t_k is the target output at the nodal point of k.