Water Research xxx (2016) 1–10

Contents lists available at ScienceDirect

Water Research

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

A novel single-parameter approach for forecasting algal blooms

Xi Xiao [a, b, c], Junyu He [a, b, c], Haomin Huang [b, c, d], Todd R. Miller [e], George Christakos [a], Elke S. Reichwaldt [f], Anas Ghadouani [f], Shengpan Lin [g], Xinhua Xu [b], Jiyan Shi [b, c, *]

[a] Ocean College, Zhejiang University, Zhoushan, PR China
[b] College of Environmental & Resource Sciences, Zhejiang University, Hangzhou, PR China
[c] Key Laboratory for Water Pollution Control and Environmental Safety, Zhejiang Province, PR China
[d] School of Environment and Energy, South China University of Technology, Guangzhou, PR China
[e] Joseph J. Zilber School of Public Health, University of Wisconsin - Milwaukee, USA
[f] Aquatic Ecology and Ecosystem Studies, School of Civil, Environmental and Mining Engineering, The University of Western Australia, Crawley, WA, Australia
[g] Department of Integrative Biology, Michigan State University, East Lansing, MI, USA

A R T I C L E I N F O

Article history:
Received 20 April 2016
Received in revised form
29 September 2016
Accepted 30 October 2016
Available online xxx

Keywords:
Artificial neural network
Forecasting
Harmful algal bloom
Phytoplankton
Single-parameter
Wavelet analysis

A B S T R A C T

Harmful algal blooms frequently occur globally, and forecasting could constitute an essential proactive strategy for bloom control. To decrease the cost of aquatic environmental monitoring and increase the accuracy of bloom forecasting, a novel single-parameter approach combining wavelet analysis with artificial neural networks (WNN) was developed and verified based on daily online monitoring datasets of algal density in the Siling Reservoir, China and Lake Winnebago, U.S.A. Firstly, a detailed modeling process was illustrated using the forecasting of cyanobacterial cell density in the Chinese reservoir as an example. Three WNN models occupying various prediction time intervals were optimized through model training using an early stopped training approach. All models performed well in fitting historical data and predicting the dynamics of cyanobacterial cell density, with the best model predicting cyanobacteria density one-day ahead (r = 0.986 and mean absolute error = 0.103 $\times$ 10^4 cells mL^{-1}). Secondly, the potential of this novel approach was further confirmed by the precise predictions of algal biomass dynamics measured as *chl a* in both study sites, demonstrating its high performance in forecasting algal blooms, including cyanobacteria as well as other blooming species. Thirdly, the WNN model was compared to current algal forecasting methods (i.e. artificial neural networks, autoregressive integrated moving average model), and was found to be more accurate. In addition, the application of this novel single-parameter approach is cost effective as it requires only a buoy-mounted fluorescent probe, which is merely a fraction (~15%) of the cost of a typical auto-monitoring system. As such, the newly developed approach presents a promising and cost-effective tool for the future prediction and management of harmful algal blooms.

1. Introduction

Algal blooms, one of the most serious aquatic environmental problems globally, cause water quality problems and ecosystem damage (Paerl and Otten, 2013). Accumulated algal biomass is respired by heterotrophic bacteria causing reductions in dissolved oxygen levels. Furthermore, toxic metabolites produced by both eukaryotic algae and cyanobacteria pose threats to ecosystem and human health (Conley et al., 2009). In situ monitoring and forecasting has the potential to be a proactive strategy for algal bloom control (Coad et al., 2014). However, previously published strategies require monitoring multiple parameters with expensive sensor packages that are cost prohibitive for a wider application (Storey et al., 2011; Ye et al., 2014).

An important scientific question is "whether these datasets are disproportionate for harmful algal bloom (HAB) forecasting and whether a more economical and practical monitoring protocol could be developed to achieve the same goal". Recknagel et al. (2014a) developed two models to forecast cyanobacteria (i.e. *Cylindrospermopsis raciborskii*) blooms: one of which used all

* Corresponding author. College of Environmental & Resource Science, Zhejiang University, 310058 Hangzhou, PR China.
E-mail address: shijiyan@zju.edu.cn (J. Shi).

http://dx.doi.org/10.1016/j.watres.2016.10.076
0043-1354/

measured parameters (9 parameters), while the other one used only electrically measurable parameters including water temperature, electrical conductivity, pH, DO, and turbidity. The model using only the electrically measurable parameters was found to be as accurate as the model using all parameters (Recknagel et al., 2014a). HAB forecasting in Three Gorges Reservoir has also shown that nutrients are not needed as inputs of the early-warning system, and acceptable accuracies can be achieved in predictions of phytoplankton blooms using only a small number of parameters (Ye et al., 2014). Using a simplified data input not only saves costs in labor and sensor instruments, thereby making HAB monitoring and prediction more efficient, but also makes the running of the numerical model simpler. The most relevant indicator for algal growth quantification is its direct indicator - the algal biomass itself. Consequently, algal biomass (either algal cell density or *chl a* concentration) could be the most meaningful and necessary input for HAB forecasting models.

Algal biomass dynamics are non-linear and non-stationary due to the complex interaction of physical, chemical and biological parameters affecting growth and accumulation of biomass. Therefore, traditional statistical models (e.g. multiple linear regression, autoregressive moving average models) are problematic and difficult (Recknagel et al., 2013). Recently, a number of advanced methods have been developed to predict HABs, including evolutionary computation (Recknagel et al., 2014a, 2014b; Ye et al., 2014), autoregressive fuzzy models (Kim et al., 2014), and artificial neural network (ANN) models (Kuo et al., 2007). As a black box model, ANN models are well suited to simulating dynamic non-linear systems, and have the advantages of self-learning, self-organizing and self-adaption (Lek et al., 1996; Feng and Hong, 2008). However, ANNs and other linear and non-linear models often have limitations when dealing with non-stationary data (Cannas et al., 2006; Wang et al., 2011). To take advantage of the ANN algorithm, preprocessing of input data into several stationary variables for algal bloom forecasting is therefore needed, for which Fourier Transform and serial-correlation analysis are popular choices (Cannas et al., 2006; Tiwari and Chatterjee, 2010). Wavelet analysis (WA) appears to be a more effective tool for preprocessing of input data due to its ability to reveal both spectral and temporal information simultaneously, within one signal (Nourani et al., 2009; Li et al., 2015). We hypothesized that ANNs coupled with WA would effectively and precisely handle non-stationary data such as algal bloom signals, and that the method would be more practical for HAB predictions than traditional statistical models.

The objective of this study was to explore modeling HABs based on a single-parameter and to verify the appropriateness of algal biomass as the key input of such a model. To achieve this, a novel single-parameter WA and ANN hybrid forecasting approach was developed and validated by: (1) simulating the daily cell density of cyanobacteria in the Siling Reservoir, China; (2) confirming its robustness to forecast daily algal biomass measured as *chl a* concentrations in two different water bodies (Lake Winnebago, U.S.A. and Siling Reservoir, China); (3) comparing the accuracy of prediction to up-to-date HAB forecasting approaches, i.e. ANN and autoregressive integrated moving average models (ARIMA). The developed approach may be useful in decreasing the essential number of input parameters necessary to achieve a robust and cost-effective HAB forecasting system.

2. Materials and methods

2.1. Real-time online monitoring datasets for the Siling Reservoir and Lake Winnebago

The Siling Reservoir (30°20'N, 119°43'E) is located in the northeast rural area of Hangzhou, Zhejiang province, China, and Lake Winnebago (44°01'N, 88°25'W) is located in eastern Wisconsin, U.S.A. (Fig. 1). Both water-bodies suffer from consistent and intense harmful algal blooms.

The storage capacity and total catchment area of the Siling Reservoir are 27.8 million m^3 and 71.6 km^2, respectively. As a medium sized reservoir, it provides a drinking water source for the

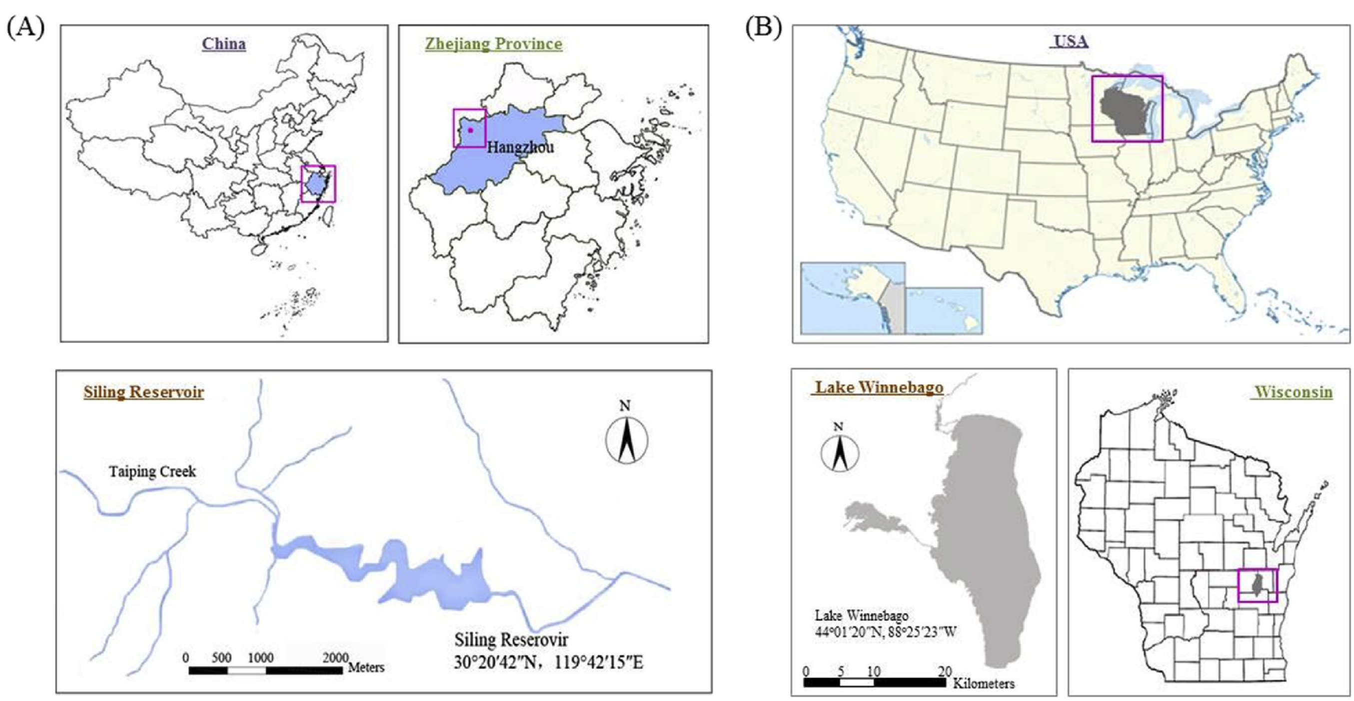

Fig. 1. Sketches of the Siling Reservoir, Zhejiang province, China, and the Lake Winnebago, Wisconsin, U.S.A.

Please cite this article in press as: Xiao, X., et al., A novel single-parameter approach for forecasting algal blooms, Water Research (2016), http://dx.doi.org/10.1016/j.watres.2016.10.076