



Spatially-explicit modelling and forecasting of cyanobacteria growth in Lake Taihu by evolutionary computation



Xiaoqing Zhang^a, Friedrich Recknagel^c, Qiuwen Chen^{a,b,*}, Hongqing Cao^c, Ruonan Li^a

^a RCEES, Chinese Academy of Sciences, Beijing 100085, China

^b CEER, Nanjing Hydraulic Research Institute, Nanjing 210029, China

^c School of Earth & Environmental Sciences, University of Adelaide, 5005 Adelaide, SA, Australia

ARTICLE INFO

Article history:

Available online 18 June 2014

Keywords:

Cyanobacteria blooms
Hybrid evolutionary algorithm
Forecasting
Thresholds
Sensitivity analysis

ABSTRACT

Models are proved to be effective instrument for algae bloom prediction and management. The commonly available prediction models are physically based numerical approach or data-driven approach. However, these models are sometimes restricted by the lack of an explicit representation function or by insufficient data. The present research aimed to develop forecasting models that provide early warning on cyanobacteria outbreaks, as well as understand the ecological thresholds and relationships that determine such events, by means of evolutionary computation. The Lake Taihu, which has been suffering from severe cyanobacteria blooms over the last decades due to eutrophication, was taken as study case. Two modelling approaches were used based on water quality data collected from 31 monitoring sites from 2008 to 2012. First, eight sampling sites representing spatially different environmental conditions across Lake Taihu were selected to develop 2-day ahead forecasting models. The resulting models well-matched the timing and magnitude of the observed cyanobacteria dynamics for all eight sites, which was reflected by coefficients of determination (r^2) of 0.62 for eastern site 24 being least favourable to cyanobacteria growth and 0.83 for north-western site 6 being most favourable. The sensitivity analyses revealed inhibitory relationships with nitrate at water temperatures greater than 18 °C and excitatory relationships with phosphate at lower water temperatures for most sites, which suggested N-limitation of the lake existed locally in summer and autumn. Second, the aggregated data from all 31 sites were used to develop a generic 2-day ahead forecasting model. When compared with the observed cyanobacteria data of the eight selected sampling sites, the generic model achieved slightly lower coefficients of determination than the site-specific models, with the lowest r^2 value for site 24 (0.36) and the highest r^2 value for site 6 (0.77). The sensitivity analysis for the generic model revealed a much lower water temperature threshold of 13.01 °C, above which N-limitation for cyanobacteria growth was indicated. Overall, both the spatially-explicit models and the generic model were suitable for early warning of cyanobacteria blooms at most sampling sites, and specified understanding on the environmental conditions that favour cyanobacteria growth across Lake Taihu.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Blooms of cyanobacteria are a global environmental concern for drinking water security and aquatic biodiversity conservation (Dodds et al., 2008; Yang et al., 2008). Lake Taihu is a shallow, polymictic lake located in the Yangtze River delta, which is the most rapidly developing region in China. The lake is a key drinking, fishing and tourism resource in the region, and also serves as a

waste repository for urban, agricultural and industrial sectors (Guo, 2007; Qin et al., 2007). In recent years, algal blooms occurred frequently throughout much of the Lake. The main species was cyanobacteria, whose biomass accounted averagely for 38.3% of total phytoplankton biomass in Meiliang Bay between 1992 and 2002. Till 2009, cyanobacteria population accounted for 93–100% of the total phytoplankton cells (Wu et al., 2013). Microcystis was dominant among the cyanobacteria, which accounted for 85.7% of the total cyanobacteria biomass (Chen et al., 2003a; Liu et al., 2011).

Great efforts are being undertaken to better understand the physical and chemical driving forces that cause fast cyanobacteria growth events in Lake Taihu (Xu et al., 2010; Qin et al., 2010; Chen et al., 2014a) as well as how to predict such events by applying

* Corresponding author at: Nanjing Hydraulic Research Institute, Guangzhou Road 223, Nanjing, China. Tel.: +86 25 85829765; fax: +86 25 85829765.
E-mail address: qchen@rcees.ac.cn (Q. Chen).

different modelling approaches. The models are in general categorized into either physically based approach or data-mining approach. The physically based models describe the underline processes of algae growth and outbreak and have achieved remarkable success in many lakes (Hamrick, 1992; Cole and Buchak, 1995; Moll and Radach, 2003; Chau, 2007; Mao et al., 2008; Gal et al., 2009; Zhang et al., 2013; Chen et al., 2014b). However, these models are based on the detailed description of physical, chemical and biological processes, and usually contain a large number of parameters for calibration. The models are sometimes restricted by the lack of explicit representation functions or by the high uncertainty from parameters calibration. Being an alternative, data driven models based on such as artificial neural networks, genetic programming and fuzzy logic have been rapidly developed and applied to ecological study (Maier et al., 1998; Whigham and Recknagel, 2001; Chen and Mynett, 2003; Zhao et al., 2006; Chan et al., 2007). Despite artificial neural networks are powerful in predicting data by minimizing the root mean square error of approximators, the method lacks of a numerical expression to represent the modelled system (Recknagel et al., 2002; Muttill and Chau, 2006, 2007; Xie et al., 2006). Although the fuzzy logic models showed great advantages in encapsulating empirical knowledge from experts, they usually encounter difficulties in defining membership functions and inference rules (Chen and Mynett, 2003; Chen et al., 2014b).

Current progress in sensor technology makes online surveillance possible and provides instantaneous data for early warning through applicable lake forecasting models. Inductive reasoning and prediction by evolutionary computation is superior for developing operational models, as demonstrated by 7-day ahead forecasting models for *Cylindrospermopsis* sp. developed from four years of online water quality data of the Wivenhoe Reservoir in Australia using the hybrid evolutionary algorithm (HEA) (Recknagel et al., 2013b). The HEA can generate the optimized rule sets to represent complex relationship between multiple variables and provide a discovery of explanatory rules. Moreover the most relevant input variables can be identified by the statistics of numerous random runs of the HEA without a priori knowledge (Cao et al., 2006). The major shortcoming of previous studies using the HEA method is that the predictive models focused on each single site of particular lakes or rivers with different eutrophication and circulation patterns and could not capture the generic feature of the whole water bodies. The resulted site-specific models were not only redundant, but also were limited to predict cyanobacteria blooms for narrow environmental conditions. Therefore, there is a demand to develop generic HEA models for a certain category of lakes to determine the most significant explanatory variables and describe the outbreak of cyanobacteria in various conditions.

The present research applied HEA to develop eight spatially-explicit models and a generic model for 2-day ahead forecasting of cyanobacteria dynamics in Lake Taihu based on data monitored at 31 sampling sites from 2008 to 2012. The resulting models aimed to be applicable for early warning of cyanobacteria blooms. Through analysing the site-specific models and comparing to the generic model, the results are expected to improve our understanding of the environmental conditions that favour outbreaks of bloom events in Lake Taihu.

2. Material and methods

2.1. Study sites and data collection

Lake Taihu (E30°56'–31°33', N119°54'–120°36') is a shallow water body, with a mean depth of 1.9 m, a surface area of 2338 km² and a volume of 4.4 billion m³ (Qin et al., 2007, 2010). The lake basin has a drainage area of 36,500 km², and more than 30 canals and

rivers discharge water into the lake. Generally, rivers in the west of the lake are defined as upstream and the main outflow rivers are located in the southeast. The Meiliang Bay, Zhushan Bay and Western Lake are the most eutrophic areas of Lake Taihu. There are two main rivers, the Liangxi River and the Zhihu Gang, connected to the Meiliang bay (Fig. 1). They discharge the effluents from the cities of Wuxi and Changzhou into the lake. The Yincun Gang is the largest inflow in the northwest, which is connected to the Zhushan Bay. The Dapu Gang is the main river connected to the Western Lake and the discharge account for 50% of the total inflow of the area. The Wangyu River, which is connected to the Gonghu Bay delivers Yangtze River water to the lake. The Taihu River is the main river discharging water out of the lake.

The Taihu Lake basin occupies only 0.4% of China's land area and 3% of China's population, but it accounts for 11% of the Gross Domestic Product (Qin et al., 2007) of China. Influenced by the economic development of the surrounding region over the past three decades, the lake has changed from a mesotrophic, diatom-dominated lake to a hyper-eutrophic, cyanobacteria-dominated system (Chen et al., 2003a,b). In recent years, severe cyanobacteria dominated by *Microcystis* blooms occurred throughout much of the lake from May to October.

Monthly surface (at depth of 0.5 m) water samples were taken at 31 sites across the lake (Fig. 1) from 2008 to 2012. The water temperature (WT), Secchi depth (SD) and pH were measured in situ by using YSI 6600V2. The nitrate-N (NO₃-N), ammonia-N (NH₃-N) and phosphate-P (PO₄-P) concentrations were measured in laboratory, following the national standard methods (National Environmental Protection Burcau (NEPB), 2002). The cyanobacteria concentrations were analysed by microscope in laboratory. Table 1 summarises the basic statistics of the collected data. It is easily understood that high-resolution data such as daily data is better for model development. However, due to cost concern, there is a lack of long-term daily data. Therefore, before model development, the monthly data were linearly interpolated (Bobbin and Recknagel, 2001; Chan et al., 2007; Recknagel et al., 2013a) for each site to obtain daily values.

2.2. Hybrid evolutionary algorithms (HEA)

The research was conducted using the hybrid evolutionary algorithm (Cao et al., 2013), which has been applied previously for the predictive modelling of cyanobacteria growth in a variety of lakes and rivers worldwide (e.g. Chan et al., 2007; Kim et al., 2007; Recknagel et al., 2008, 2013a,b). This algorithm is designed to evolve over time to improve the fitness between model results and observations by combining genetic programming (GP) for optimizing the model structure and differential evolution (DE) for optimizing model parameters (see Fig. 2). GP is an evolutionary algorithm, in which the genetic population consists of computer programs of different sizes and shapes (Koza, 1992, 1994). Parse trees represent computer programs. These programs are subsequently evaluated by means of "fitness" function. Fitter programs are selected for recombination by using arithmetic and logic operators (such as crossover, mutation and reproduction) in order to create the next generation. This step is iterated for continual generations until the termination criterion has been satisfied (Cao et al., 2006). Differential evolution is an evolutionary algorithm designed for parameter optimization (Storn and Price, 1997), which extracts differential information (i.e., distance and direction to global optimum) from the current population of solutions to guide the search for the global optimum. It can outcompete the other optimization methods in convergence speed and robustness. By comparison with conventional optimization algorithms, the DE could make it a self-organizing scheme, and does not depend on additional probability distributions (Cao et al., 2006).

Download English Version:

<https://daneshyari.com/en/article/4375726>

Download Persian Version:

<https://daneshyari.com/article/4375726>

[Daneshyari.com](https://daneshyari.com)