



Spatially-explicit forecasting of cyanobacteria assemblages in freshwater lakes by multi-objective hybrid evolutionary algorithms



Hongqing Cao^{a,*}, Friedrich Recknagel^a, Michael Bartkow^b

^a School of Biological Sciences, University of Adelaide, 5005, Australia

^b Seqwater, PO Box 16146, City East, QLD 4002, Australia

ARTICLE INFO

Article history:

Received 2 January 2016

Received in revised form 1 September 2016

Accepted 26 September 2016

Keywords:

Multi-objective optimization

Genetic programming

Hill climbing

Multi-output rule models

Cyanobacteria blooms

ABSTRACT

This paper proposes a novel multi-objective hybrid evolutionary algorithm (MOHEA) that allows spatially-explicit modelling of local outbreaks and dispersal of population density. The MOHEA was tested for modelling at once two cyanobacteria populations at one lake site, same population in two different lakes and same population at three different sites of one lake. All experiments with MOHEA utilized water quality time-series and abundances of *Anabaena* and *Cylindrospermopsis* monitored in the sub-tropical Lakes Wivenhoe and Somerset in Queensland (Australia) from 1999 to 2010. Results have demonstrated the capacity of MOHEA to determine generic rules that: (1) reveal crucial thresholds for outbreaks of cyanobacteria blooms, and (2) perform spatially-explicit forecasting of timing and magnitudes 7-day-ahead of bloom events.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The current economic development in Australia and world-wide goes side by side with the global problems of eutrophication and climate change. There is evidence that high nutrient loads, rising temperatures, enhanced stratification, increased residence time and salinisation of drinking water reservoirs and lakes favor the dominance of cyanobacteria (Paerl and Huisman, 2008). Therefore water industries have to consider coinciding effects of eutrophication and climate change in their strategies to manage cyanobacterial blooms. However our ability to predict the occurrence and composition of cyanobacteria blooms has lagged well behind our ability to control total algal biomass. We urgently need advances in our ability to predict and prevent the growth of undesirable algae and other nuisance-forming organisms (Smith and Schindler, 2009). To develop comprehensive lake-based monitoring and early warning systems for water quality and cyanobacteria is therefore the right step forward (Schindler, 2009). Frequent population outbreaks of toxic cyanobacteria in drinking water reservoirs and lakes will have detrimental effects on raw water quality and aquatic biodiversity, and costly technology will be required to sustain safe human water supplies (e.g. Dodds et al.,

2009). In order to assist water industries in making informed decisions and timely adaptations of measures for preventing and controlling effects of cyanobacteria, more adequate computer models are required (Jackson et al., 2001).

Traditionally, process-based models which allow simulations of food web dynamics and nutrient cycles over time by using ordinary differential equations (ODEs) (Pei and Ma, 2002; Arhonditsis and Brett, 2005; Chen et al., 2014) are widely used. However, there are some shortcomings to use these process-based models. Firstly, process-based models may hardly comprehend the causal complexity of the phytoplankton community in order to make accurate daily forecasts of population dynamics of algal species. Secondly, process-based models are calibrated for a limited number of years with annual data that constrains their validity to those years. Thirdly, the data demand of process-based models by far exceeds operationally-available data of a lake or a lake site at a certain point in time. Therefore it is unlikely that process-based models may ever been applicable as operational forecasting tools for early warning.

With rapidly growing amounts of ecological data and progress in computing technology, powerful tools for inductive reasoning and forecasting from complex data become available. Artificial neural networks (Hornik et al., 1989) approximate complex data with high accuracy by multivariate nonlinear models (Recknagel et al., 1997; Wei et al., 2001; Jeong et al., 2001), but lack the explicit representation of models extracted from data. In recent years, the use of evolutionary algorithms (EAs) (Holland, 1975) has gained wide

* Corresponding author.

E-mail addresses: hongqing.cao@adelaide.edu.au (H. Cao), friedrich.recknagel@adelaide.edu.au (F. Recknagel), michael.bartkow@seqwater.com.au (M. Bartkow).

popularity in domains, such as machine learning, pattern recognition, economic prediction and so on, due to their characteristics of self-adaptation, self-organization, self-learning and generality (Bäck et al., 1997). Since EA applications for ecological modelling have been pioneered by Bobbin and Recknagel (2001), Cao et al. (2006) developed the hybrid evolutionary algorithm (HEA) that is now worldwide applied for non-spatially-explicit modelling of cyanobacteria blooms in lakes and rivers (e.g. Kim et al., 2007; Chan et al., 2007; Recknagel et al., 2014a) as well as for knowledge discovery (Recknagel et al., 2014b, 2016). Since the HEA was designed to develop non-spatially-explicit models, resulting typically single output rule models did not represent spatial or multi-species relationships. However plankton communities in lakes vary seasonally and spatially by abiotic factors like advection, thermal stratification, nutrient loads as well as by biotic factors like competition, grazing, and predation. Therefore there is a demand for models allowing spatially-explicit forecasting that can identify local hotspots for seasonal outbreaks of cyanobacteria blooms.

It is well known that multi-objective optimization (MOO) techniques (Marler and Arora, 2004; Miettinen, 1999; Deb, 2001; Hanne, 2000) have been widely applied in many fields. The multi-objective hybrid evolutionary algorithm (MOHEA) proposed in this study allows to develop IF-THEN-ELSE rules with multiple outputs whereby fitting errors of all outputs are minimized by MOO. Resulting IF-THEN-ELSE rules with multiple outputs provide the benefit of: (1) revealing threshold conditions (IF-condition) that trigger population outbreaks being generic for all outputs, and (2) forecasting multiple species at a single site and single species at multiple sites (see Fig. 1). The functionality of MOHEA is tested for 7-day-ahead forecasting of the cyanobacteria *Anabaena* and *Cylindrospermopsis* in the Lakes Wivenhoe and Somerset, Queensland (Australia) based on physical and chemical water quality data monitored from 1999 to 2010. The paper validates forecasting results of different types of multi-output models and discusses ecological relationships revealed by input sensitivity analyses of the models.

2. Materials and methods

2.1. Study sites and data

Different data were utilized for developing the three types of multi-output rule models. Eleven years of water quality data from 1999 to 2010 from Lake Wivenhoe in Queensland, Australia were used to develop single-site multi-species and multi-site single-species models. Measured data from Site30001 of Lake Wivenhoe (see Fig. 2) were used for developing single-site multi-species models and the measured data from sites 30015, 30016 and 30017 were used for developing multi-site single-species models. Tables 1 and 2 show the water quality variables of the 11-year period utilized for developing single-site multi-species models and multi-site single-species models respectively.

To develop multi-lake single-species models, water quality data of the same time period from Lake Wivenhoe (Site30001) and Lake Somerset (Site20001) were utilized for the modelling. Fig. 3 shows the map of Lake Somerset and all sampling sites. Table 3 shows the measured water quality data utilized for developing multi-lake single-species models.

It is notable that the monthly measured data from the two lakes were averaged over the euphotic depth of 3 m. Since the measurement intervals of the raw data were highly irregular and sampling dates differed between physical, chemical and biological variables, the data were linearly interpolated to suit daily time steps for forecasting. In order to develop 7-days-ahead forecasting models the daily interpolated input data have been shifted by 7 days against the daily output data.

2.2. Multi-objective hybrid evolutionary algorithm

The hybrid evolutionary algorithm (HEA) developed by Cao et al. (2006) evolves single rule models by genetic programming and optimizes model parameters by a genetic algorithm. Our recent study (Cao et al., 2014) intensively investigated the performance of six different parameter optimization algorithms, which include hill climbing (HC), simulated annealing, genetic algorithm, differential evolution, covariance matrix adaptation evolution strategy and estimation of distribution algorithm, when developing single rule models for chlorophyll-*a* and *Cylindrospermopsis*. The experimental results demonstrated that HC always performed best. Hence this study used HC to optimize constant parameters of the rule models and extended the previous research by improving the rule model types from single output to multiple outputs. The multi-objective hybrid evolutionary algorithm (MOHEA) has been developed by integrating multi-objective optimization into HEA. The conceptual diagram in Fig. 4 shows (of) the structure and functioning of MOHEA using water quality data as inputs and *Anabaena* data as outputs. It shows that genetic programming (GP) identifies the optimal model structure by applying the genetic operators: reproduction, crossover and mutation. Simultaneously the HC algorithm is utilized to optimize the model parameters. The overall “fittest” model is determined by the fitness function that applies to multiple outputs by multi-objective optimization. The details of GP and HC were briefly described in what follows.

2.2.1. Structure optimization of multi-output rule models

1) Model representation

We used GP (Koza, 1992, 1994) as the main technique to evolve the rule model structure. Because GP typically operates on parse trees instead of bit strings as traditional GA does, it is well suited to evolve an equation or formula relating the output and input variables. Firstly we defined the following three function sets as:

- Logic function set: $F_L = \{\text{AND, OR}\}$;
- Comparison function set: $F_C = \{>, <, \geq, \leq\}$;
- Arithmetic function set: $F_A = \{+, -, *, /, \exp, \ln\}$.

In our case, a multi-output rule model with an IF-THEN-ELSE structure is represented as a vector of multiple trees in GP with the form of (IF Tree, THEN Tree 1, THEN Tree 2, ..., THEN TreeN, ELSE Tree 1, ELSE Tree 2, ..., ELSE TreeN). N is the total number of outputs. IF Tree denotes the IF condition branch, THEN Tree and ELSE Tree denote the result branches of the THEN and ELSE branch respectively. Their function sets are:

$$F_{\text{IFTree}} = F_L \cup F_C \cup F_A \text{ and } F_{\text{THENTree/ELSETree}} = F_A.$$

They have the same terminal set as:

$$T = \{x_1, \dots, x_n, c\}$$

where n is the number of input variables and c is a random constant.

Fig. 5 shows an example of a multi-output rule model for predicting *Cylindrospermopsis* (*Cylind*) abundance in different sites from Site1 to Site3 (Fig. 5(a)) and its multi-tree representation in GP (Fig. 5(b)–(h)). The rule model shows the relationship between the abundance of *Cylind* in different sites and some selected water quality parameters which include water temperature (WT), conductivity (Cond), dissolved oxygen (DO), pH, total phosphorus (TP), and total nitrogen (TP). Note that we provided this example for the purpose of showing the rule structure only which may not have any biological meaning.

2) Genetic operators

Download English Version:

<https://daneshyari.com/en/article/4375496>

Download Persian Version:

<https://daneshyari.com/article/4375496>

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