Harmful Algae 69 (2017) 18–27

Contents lists available at ScienceDirect

Harmful Algae

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

Research paper

Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling

Friedrich Recknagel[a,*], Philip T. Orr[b,c], Michael Bartkow[b], Annelie Swanepoel[d], Hongqing Cao[a]

[a] School of Biological Sciences, University of Adelaide, Adelaide 5005, Australia
[b] Seqwater, Ipswich 4305, Australia
[c] Australian Rivers Institute, Griffith University, Nathan 4111, Australia
[d] Rand Water, Vereeniging 1930, South Africa

ARTICLE INFO

Article history:
Received 17 March 2017
Received in revised form 6 September 2017
Accepted 17 September 2017
Available online xxx

Keywords:
Early warning
Dolicospermum
Cylindrospermopsis
Microcystins
In situ sensor monitoring
Inferential models
HEA

ABSTRACT

An early warning scheme is proposed that runs ensembles of inferential models for predicting the cyanobacterial population dynamics and cyanotoxin concentrations in drinking water reservoirs on a diel basis driven by *in situ* sonde water quality data. When the 10- to 30-day-ahead predicted concentrations of cyanobacteria cells or cyanotoxins exceed pre-defined limit values, an early warning automatically activates an action plan considering in-lake control, *e.g.* intermittent mixing and ad hoc water treatment in water works, respectively. Case studies of the sub-tropical Lake Wivenhoe (Australia) and the Mediterranean Vaal Reservoir (South Africa) demonstrate that ensembles of inferential models developed by the hybrid evolutionary algorithm HEA are capable of up to 30 days forecasts of cyanobacteria and cyanotoxins using data collected *in situ*. The resulting models for *Dolicospermum circinale* displayed validity for up to 10 days ahead, whilst concentrations of *Cylindrospermopsis raciborskii* and microcystins were successfully predicted up to 30 days ahead. Implementing the proposed scheme for drinking water reservoirs enhances current water quality monitoring practices by solely utilising *in situ* monitoring data, in addition to cyanobacteria and cyanotoxin measurements. Access to routinely measured cyanotoxin data allows for development of models that predict explicitly cyanotoxin concentrations that avoid to inadvertently model and predict non-toxic cyanobacterial strains.

1. Introduction

Global eutrophication and warming are suggested to cause shifts in plankton communities towards cyanobacteria which are favoured by high concentrations of nutrients, increased atmospheric CO_2 supplies, high temperatures, enhanced thermal stratification, as well as are tolerant to UV radiation (*e.g.* Wagner and Adrian, 2009; Huber et al., 2012; Paerl and Otten, 2013; Michalak et al., 2013). Cyanobacteria species such as *Dolicospermum circinale* Rabenhorst ex Bornet & Flahault (basionym Anabaena circinalis) and *Cylindrospermopsis raciborskii* (Wołoszyńska) Seenayya & Subba Raju may be capable of atmospheric nitrogen fixation, possibly giving them a competitive advantage over other phytoplankton (*e.g.* Paerl, 1990; Burford et al., 2016). Like species from the non-N_2-fixing genus *Microcystis*, *D. circinalie* and *C. raciborskii* may also produce cyanotoxins (saxitoxins and cylindrospermopsins) that have been postulated to have a role in reducing grazing losses by zooplankton (Jang et al., 2007). These cyanotoxins are both neurotoxic and hepatotoxic, and have been shown to cause serious liver, digestive and skin diseases, neurological impairment, and death (Carmichael, 2001). All cyanotoxins pose a risk to human health. High cell concentrations of *D. circinale*, *C. raciborskii* and *Microcystis* sp. in freshwaters, commensurate with cyanotoxin concentrations that exceed guideline or health alert levels, impair water quality and aquatic biodiversity resulting in economic, social and ecological costs worldwide (*e.g.* Smith, 2003).

Increasingly frequent cyanobacteria bloom events in rivers and drinking water supplies – possibly driven by increasing atmospheric and dissolved inorganic carbon concentrations and increased temperatures, challenge water authorities to improve operational and strategic control of cyanobacterial blooms and their toxins. Whilst strategic control is traditionally being supported by scenario analyses using process-based lake models (*e.g.* Recknagel et al., 1995; Chen et al., 2014; Nguyen et al., 2017),

---

[*] Corresponding author.
E-mail address: friedrich.recknagel@adelaide.edu.au (F. Recknagel).

https://doi.org/10.1016/j.hal.2017.09.003
1568-9883/

remote sensing (*e.g.* Lunetta et al., 2015; Matthews and Odermatt, 2015) and inferential models (*e.g.* Recknagel et al., 2014; Ye et al., 2014) have been proven to be suitable for operational forecasting and control. Fig. 1 illustrates the use of process-based and inferential models for different forecasting horizons for cyanobacterial blooms.

This study proposes an early warning scheme (Fig. 2) that runs ensembles of inferential models developed by the hybrid evolutionary algorithm HEA that have been shown to accurately predict the development of cyanobacterial cell and cyanotoxin concentrations in freshwater bodies, on a diel basis driven by present-day *in situ* water quality data from on-line monitoring stations. When the 10- to 30-day-ahead predicted cell concentrations and cell division rates of cyanobacteria, or cyanotoxin concentrations and proliferation rates exceed limit values defined by water authorities in accordance with national drinking water guidelines, operational in-lake control such as intermittent mixing, as well as enhanced monitoring and treatment of cell and toxin concentrations can be implemented.

The scheme proposed in Fig. 2 is applied to Lake Wivenhoe (Australia) and the Vaal Reservoir (South Africa). The two case studies aim to demonstrate that ensembles of inferential models developed by HEA based on high-frequency sensor monitoring data and weekly to biweekly cyanobacteria cell counts or microcystins concentrations can forecast population dynamics of cyanobacteria and total microcystin concentrations up to 30 days ahead. Forecasting horizons of 10 to 30 days would enable water authorities to implement ad hoc operational control measures as suggested in Fig. 1.

2. Materials and methods

2.1. Lake Wivenhoe

Wivenhoe Dam (27° 24′S, 152° 36′E) is located about 45 km to the west of Brisbane in subtropical southeast of Queensland, Australia. Lake Wivenhoe is a warm-monomictic and mesotrophic reservoir about 30 km long with a catchment area of 7020 km^2. It has a maximum depth in excess of 45 m near the dam wall, an average depth of 11 m and a surface area of 108 km^2. It has a drinking water storage capacity of 1.165 GL and additional temporary flood storage capacity 1.4 GL.

Blooms of the filamentous cyanobacteria *C. raciborskii* and *D. circinale* occur regularly in Lake Wivenhoe (Orr et al., 2010). *C. raciborskii* is capable of producing hepatotoxic cylindrospermopsins, and strains in Lake Wivenhoe are toxic. *D. circinale* can produce neurotoxic saxitoxins although to date, strains in Lake Wivenhoe have consistently been shown by HPLC analysis to be non-toxic. Both cyanotoxins present risks to human health (*e.g.* Hawkins et al., 1985) and when present, cyanotoxin concentrations must be reduced to levels below the Health Alert Levels presented in the Australian Drinking Water Guidelines (NHMRC and NRMMC, 2011) during water treatment. Controlling the development of *C. raciborskii* and *D. circinale* within the reservoir is a key goal of Seqwater (www.seqwater.com.au), the water authority responsible for the management of the reservoir. However, these cyanobacteria (especially *C. raciborskii*) are ecologically adaptable (Burford et al., 2016) and can bloom under a range of light, temperature and nutrient regimes. Both species are capable of fixing atmospheric nitrogen (N_2) if necessary (Reynolds, 1984; Bouvy et al., 2000; Moisander et al., 2008).

Daily *in situ* sensor data from Seqwater's routine monitoring program from 2007 until 2015 (see Table 1) were used to develop models for predicting *C. raciborskii* and *D. circinale* cell concentrations and cell division rates for 10, 20 and 30 days ahead through the use of a hybrid evolutionary algorithm HEA. The models were tested against measured cell counts – also from Seqwater's routine monitoring program – to test their accuracy and efficacy in predicting cell concentrations, and to eventually allow them to be incorporated into Seqwater's monitoring program to provide longer range warnings of cell concentrations of *C. raciborskii* and/or *D. circinale* in source water that might exceed operational trigger levels.

2.2. Vaal Reservoir

The Vaal Reservoir (26° 56′S, 28° 7′E) is located approximately 75 km south of Johannesburg and South Africa's largest drinking water reservoir. It is a warm-monomictic and mesotrophic reservoir about 45 km long with a catchment area of 38,500 km^2, a maximum depth of 47 m, a surface of 320 km^2 and a maximum volume of 2.61 GL.

Favourable climate and water quality conditions can lead to blooms of microcystin-producing strains of *Microcystis* sp. in the Vaal Reservoir (Conradie and Barnard, 2012). Therefore, monitoring and control of cyanobacterial blooms is a high priority for Rand

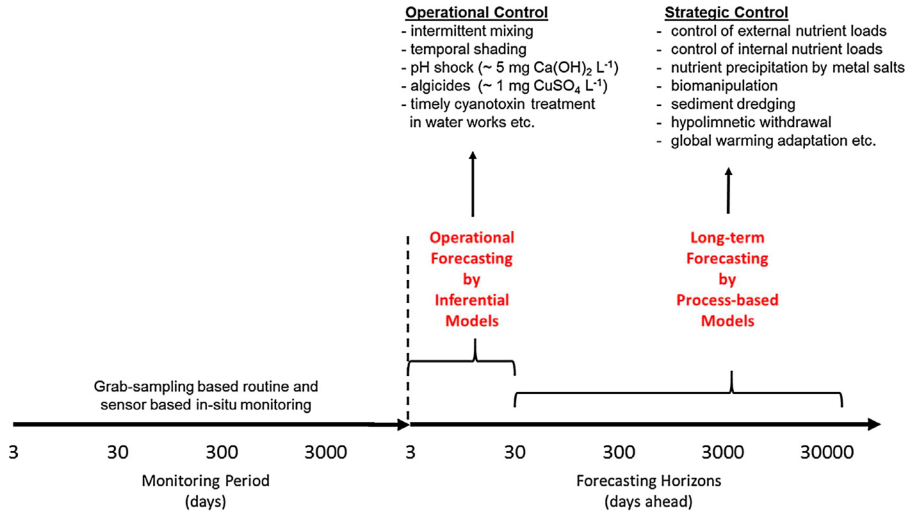

Fig. 1. Suitability of inferential and process-based models for different forecasting horizons of cyanobacteria blooms.

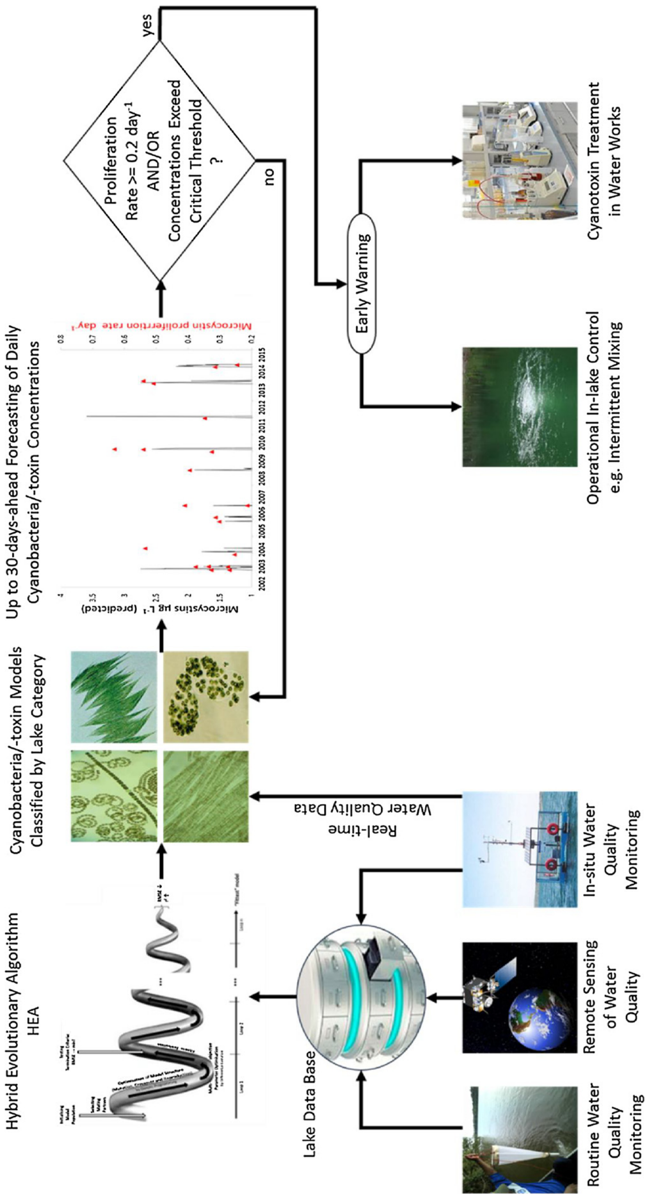

Fig. 2. Conceptual scheme of early warning of cyanobacteria blooms by inferential models based on HEA.

Table 1
Limnological data for Lake Wivenhoe.

Water Quality Variables	Units	2007–2015 Mean/Min/Max
Physical-chemical parameters		
WT (Temperature)	°C	22.59/14.83/29.5
TURB (Turbidity)	NTU	10.08/0.3/288.1
pH		8.1/6.5/9.7
DO (dissolved oxygen)	$mg\,L^{-1}$	7.96/4.85/11.16
EC (electrical conductivity)	$\mu S\,cm^{-1}$	335/140/496.5
Biological parameters		
Dolicospermum circinalie	cells mL^{-1}	1277/0/27200
Cylindrospermopsis raciborskii	cells mL^{-1}	7126/0/74700

Water (www.randwater.co.za), the water authority responsible for the management of Vaal Reservoir.

Both *in situ* water quality data measured by YSI sondes and concentrations of the total microcystin determined by the Enzyme Linked Immuno Sorbent Assay (ELISA) of EnviroLogix^{TM} from the Rand Water's routine monitoring program were available for this study from 2002 until 2015, and are summarised in Table 2. The provisional guideline concentration for microcystins in drinking water that has been proposed by the WHO (2003) is 1 µg (microcystin-LR toxicity equivalents) L^{-1}.

2.3. Hybrid evolutionary algorithm HEA

Evolutionary computation infers models from data using principles of natural selection and evolution (Holland, 1975; Holland et al., 1986). The hybrid evolutionary algorithm HEA (Cao et al., 2014, 2016) has been designed to evolve 'fittest' IF-THEN-ELSE rules from ecological data by integrating genetic programming (GP) and differential evolution (DE) (Fig. 3a). It applies GP according to Koza (1992) to evolve the optimum structure of the rule model, and DE according to Storn and Price (1997) to optimise the parameters of the rule model. Since GP typically operates on parse trees rather than on bit strings, it is well suited to evolve IF-THEN-ELSE rules for multivariate relationships. GP uses the logic functions FL = {AND, OR}, comparison functions $FC = \{>, <, \geq, \leq\}$, and arithmetic functions $FA = \{+, -, *, /, \exp, \ln\}$ to represent IF-THEN-ELSE rules as vectors of multiple trees. Tree 1 denotes the IF condition with the function set $F_{tree1} = FL \cup FC \cup FA$, Tree 2 and Tree 3 respectively denote the THEN and ELSE branches with the function set $F_{tree2/tree3} = FA$.

Fig. 4 illustrates one crossover step by GP for the optimisation of the IF trees of two parent models for a 10-day-ahead prediction of cell concentrations of the cyanobacterium *D. circinale* as cells mL^{-1} in Lake Wivenhoe, Australia. Fig. 4a and e represent two parent models. Fig. 4b, c, f and g illustrate the selection of crossover points

Table 2
Limnological data of Vaal Reservoir.

Water Quality Variables	Units	2002–2015 Mean/Min/Max
Physical-chemical parameters		
WT (Temperature)	°C	17.5/8.8/26
TURB (Turbidity)	NTU	57.1/8.6/141
pH		7.7/5.8/10.7
DO (dissolved oxygen)	$mg\,L^{-1}$	7.4/3.2/11.1
EC (electrical conductivity)	$mS\,cm^{-1}$	19.6/13.9/55
Biological parameters		
Chlorophyll-a	$\mu g\,L^{-1}$	11.9/0.67/101
Total microcystin	$\mu g\,L^{-1}$	0.42/0.1/5

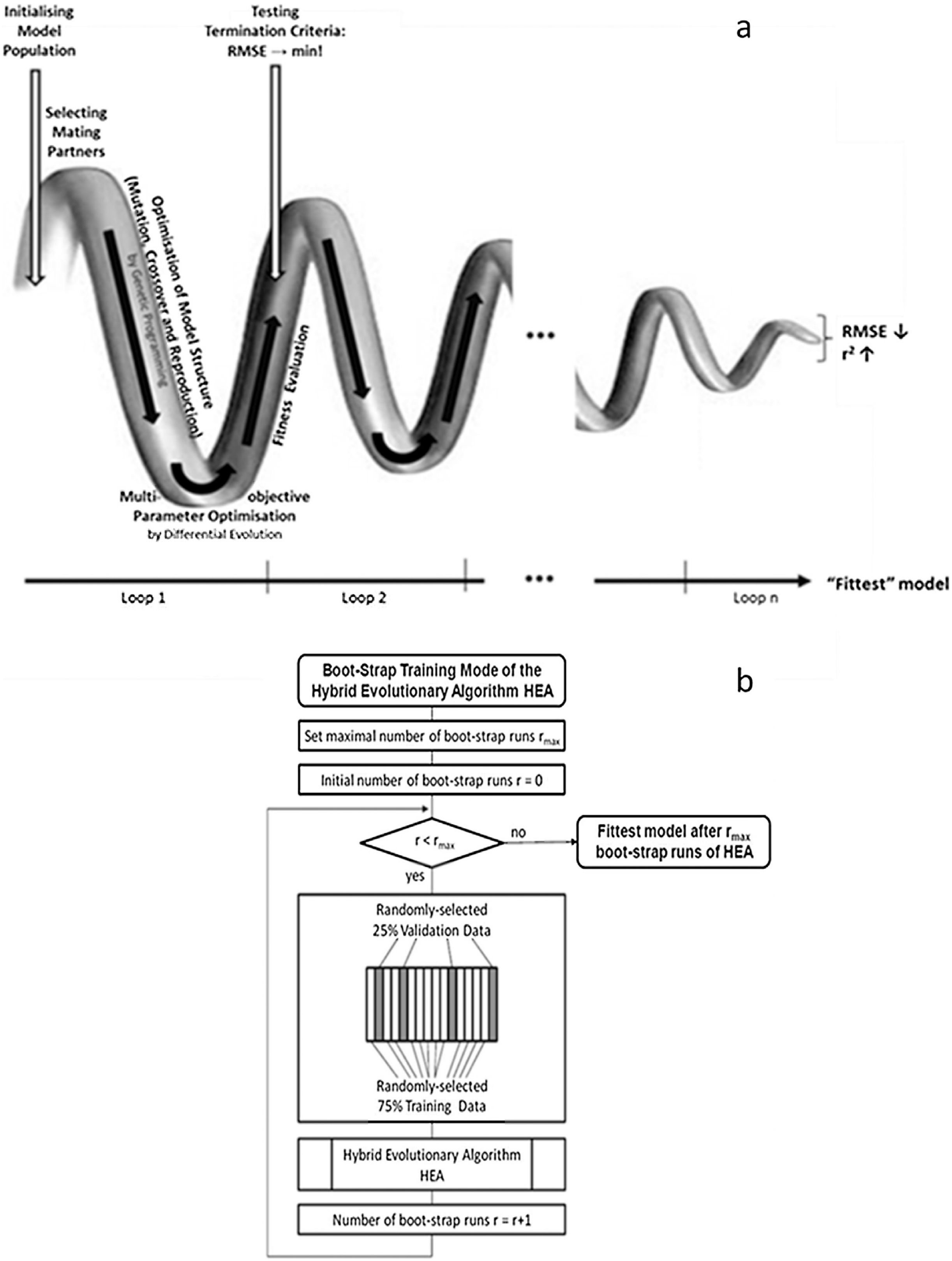

Fig. 3. Hybrid evolutionary algorithm HEA. (a) Design and functioning; (b) Boot-strap scheme.

and the crossover between the IF trees of the parent models. Fig. 4d and h represents the offspring models after the crossover.

Differential evolution DE extracts information on distance and direction of the current population of solutions towards a global optimum to guide the search for optimal parameters in the IF-THEN-ELSE rules. Since DE doesn't require separate probability distributions, the scheme becomes completely self-organizing. DE has been implemented in HEA for multi-objective parameter optimization as described by Cao et al. (2014).

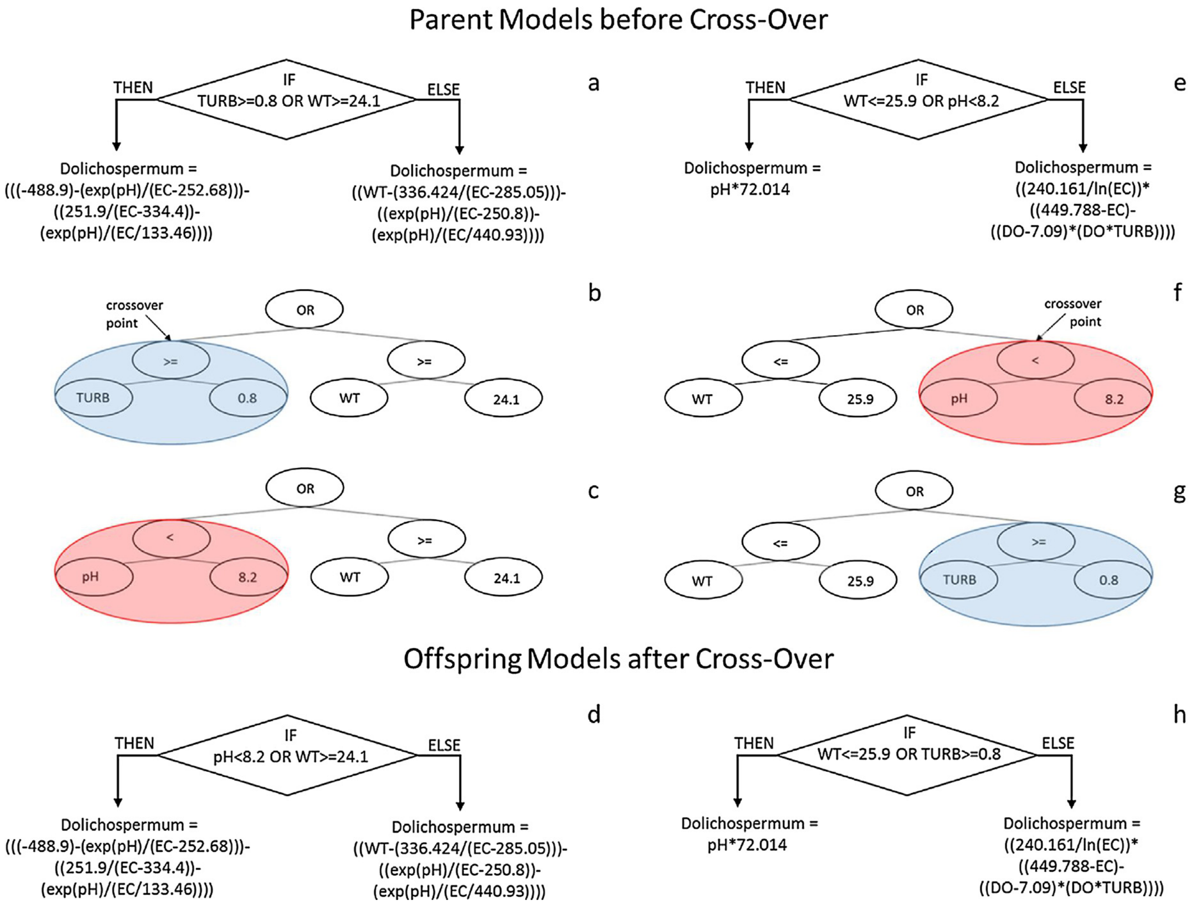

Fig. 4. Genetic programming creates two offspring models by cross-over between the IF-trees of two parent models. (a) and (e): parent models; (b) and (f): selection of crossover points of IF-trees; (c) and (g): IF-trees after crossover; (d) and (h): offspring models. (WT = water temperature °C, TURB = turbidity NTU, EC = electrical conductivity μS cm^{-1}, DO = dissolved oxygen mg L^{-1}).

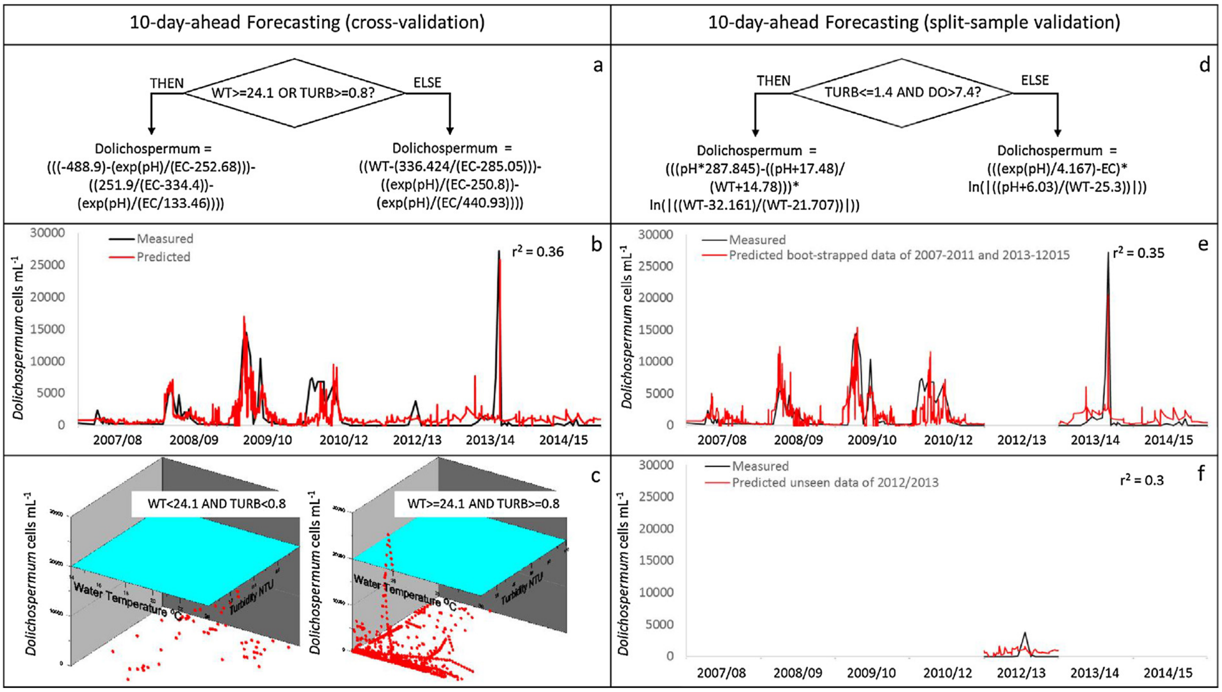

Fig. 5. 10-day-ahead forecasting models for *D. circinale* in Lake Wivenhoe from 2007 to 2015 (flood year 2011 not included). Cross-validation: IF-THEN-ELSE model (a), validation result (b), separation of high and low cell numbers (red dots) by IF-conditions; the light-blue horizontal plane marks the limit value 2×10^4 cells mL^{-1} (c) Split-sample validation: IF-THEN-ELSE model (d), validation results (e), (f). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

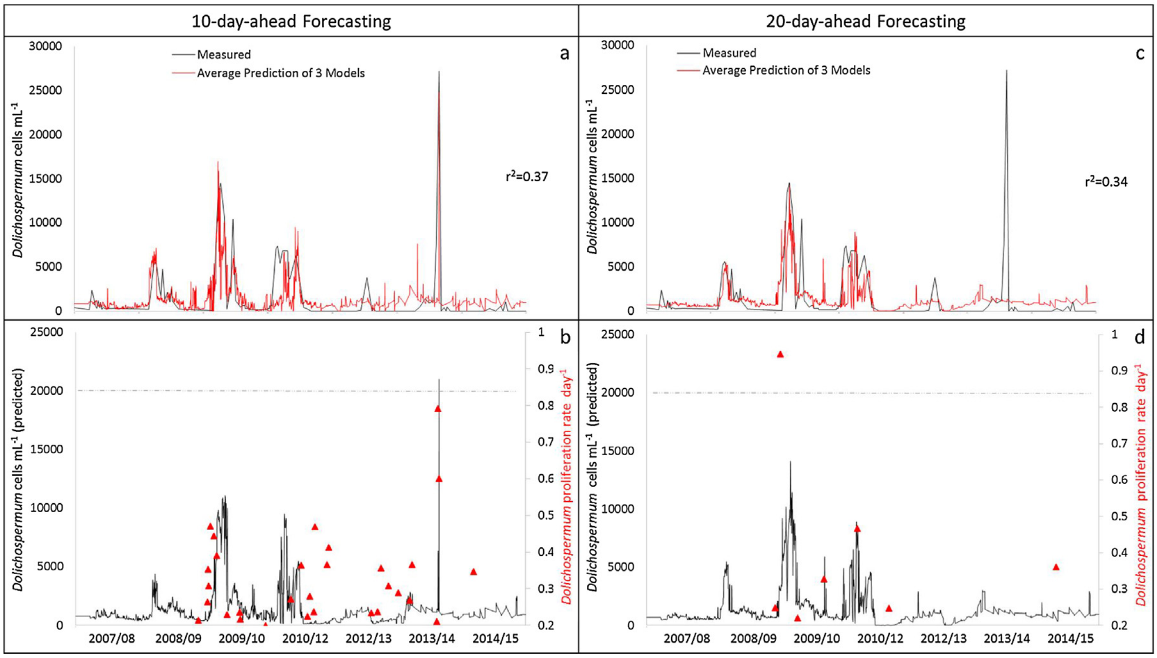

Fig. 6. Forecasting of *D. circinale* in Lake Wivenhoe from 2007 to 2015 (flood year 2011 not included). 10-day-ahead forecasting: Cross-validation of the average of 3 best models (a), Cell concentrations and cell division rates above 0.2 day^{-1} predicted by the average of 3 best models (dotted line marks limit cell concentration of 2×10^4 cells mL^{-1}) (b) 20-day-ahead forecasting: Cross-validation of the average of 3 best models (c), Cell concentrations and cell division rates above 0.2 day^{-1} predicted by the average of 3 best models (dotted line marks limit cell concentration of 2×10^4 cells mL^{-1}) (d).

To take full advantage of the information content of the data, a cyclic boot-strap scheme was applied that randomly selected different data-subsets for training (75% of all data points) and testing (25% of all data points) (see Fig. 3b) for each of 80 generations of models. After 100 boot-strap runs it determined the overall "fittest model" of all generations by cross-validation for the 19 years of data assuming that there is no bias by the 25% of data points randomly selected for testing during the boot-strap runs.

The fitness of each model was evaluated by the root mean squared error (RMSE) between the measured training data $\hat{y}_i$ and the predicted data y_i defined in Eq. (1).

$$\text{Fitness} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\hat{y}_i - y_i)^2}$$

The software HEA automatically carries out sensitivity analyses for the input variables of each model. For this purpose it calculates output trajectories separately for each input range (mean ± SD) by keeping remaining input variables constant at mean values. Resulting sensitivity curves visualise the output trajectories in percentage terms (0–100%) within their range of each input.

Since HEA induces models from long-term data patterns, it appears that the more event-related patterns the historical data contains, the more generic the models tend to become, and the more likely the model's predictive validity reaches beyond the data limits. Ongoing ecosystem evolution requires that models become regularly upgraded by the most recent monitoring data.

3. Results

3.1. D. circinale *and* C. raciborskii *in Lake Wivenhoe*

The best performing model for 10-day-ahead prediction of *D. circinale* cell concentrations in Fig. 5a suggests that increases are driven either by water temperatures greater than 24.1 °C or by turbidity greater than 0.8 NTU. The cross-validation in Fig. 5b revealed that the model failed to predict the correct onset of log-phase growth in 2010 but accurately predicted both the timing and peak cell concentrations of the major bloom in the summer of 2013/2014. Cell concentrations during that bloom exceeded 2×10^4 cells mL^{-1} (Fig. 5b), the cell concentration that corresponded to the 3 μg (saxitoxin-toxicity-equivalents) L^{-1} Health Alert Level suggested in the Australian Drinking Water Guidelines (NHMRC and NRMMC, 2011). Fig. 5c illustrates how the IF-conditions of the model successfully separated low and high cell concentrations of *D. circinale* as a prerequisite for accurate prediction of cell concentrations during bloom events.

The split-sample validation of the model documented in Fig. 5d–f used unseen data from July 2012 until June 2013 for testing and data from the remaining 6 years for evolving the model by HEA. The resulting model successfully predicted the population dynamics of *D. circinale* during the 6 years of training and predicted the major bloom event in early 2014 when cell concentrations exceeded 2×10^4 cells mL^{-1} (Fig. 5e). The model correctly predicted cell concentrations well below the threshold for bloom events for the testing period 2012/2013 but failed to predict the smaller summer peak in early 2013 where cell concentrations were 3×10^3 cells mL^{-1} (Fig. 5f) which is below the ADWG (NHMRC and NRMMC, 2011) Health Alert Level.

Results from both, cross- and split-sample validation suggest that models for 10-day-ahead prediction are suitable for prediction of *D. circinale* blooms in Lake Wivenhoe. Fig. 6 shows validation results that were averaged from ensembles of the 3 best models designed for predicting population dynamics of *D. circinale* for 10- and 20-days-ahead respectively. The averages of the two ensembles in Fig. 6a and c satisfactorily predicted both the timing and the peak cell concentrations of *D. circinale* between 2007 and 2010, but only the 10-day-ahead models in Fig. 6a predicted the major bloom event in 2013/2014. Fig. 6b and d illustrate the cell division rates in conjunction with the population dynamics predicted by the two models. The 10-day-ahead model ensemble correctly predicted the first order specific cell division rates (μ_c) of greater than 0.2 day^{-1} for the years 2008 to 2010 as well as 2013/2014 (Fig. 6b), which, according to Fig. 2 would have provided

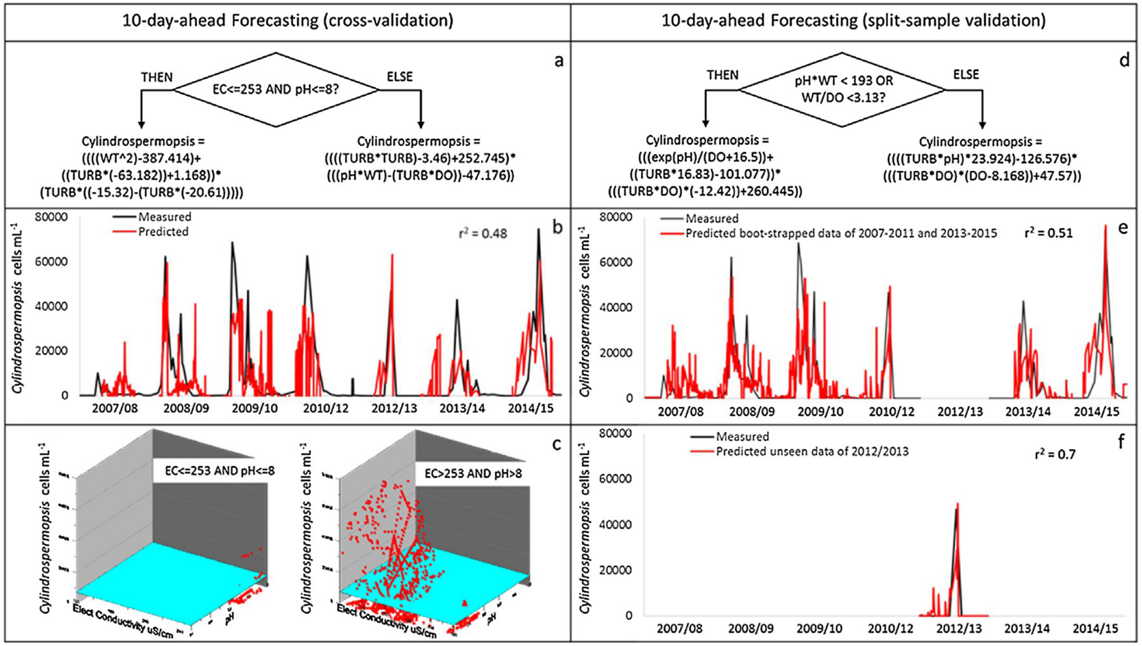

Fig. 7. 10-day-ahead forecasting models for *C. raciborskii* in Lake Wivenhoe from 2007 to 2015 (flood year 2011 not included). Cross-validation: IF-THEN-ELSE model (a), validation result (b), separation of high and low cell numbers (red dots) by IF-conditions; the light-blue horizontal plane marks the limit value 8×10^3 cells mL^{-1} (c) Split-sample validation: IF-THEN-ELSE model (d), validation results (e), (f). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

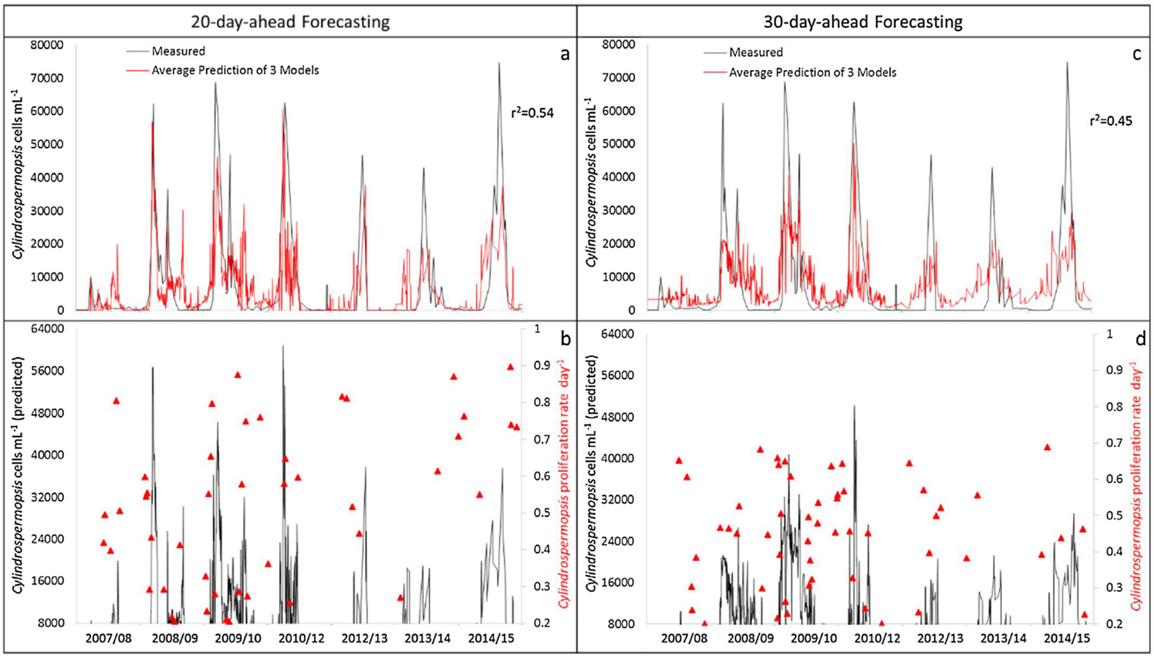

Fig. 8. Forecasting of *C. raciborskii* in Lake Wivenhoe from 2007 to 2015 (flood year 2011 not included). 20-day-ahead forecasting: Cross-validation of the average of 3 best models (a), Cell concentrations above the limit value of 8×10^3 cells mL^{-1} and cell division rates above 0.2 day^{-1} predicted by the average of 3 best models (b) 30-day-ahead forecasting: Cross-validation of the average of 3 best models (c), Cell concentrations above the limit value of 8×10^3 cells mL^{-1} and cell division rates above 0.2 day^{-1} predicted by the average of 3 best models (d).

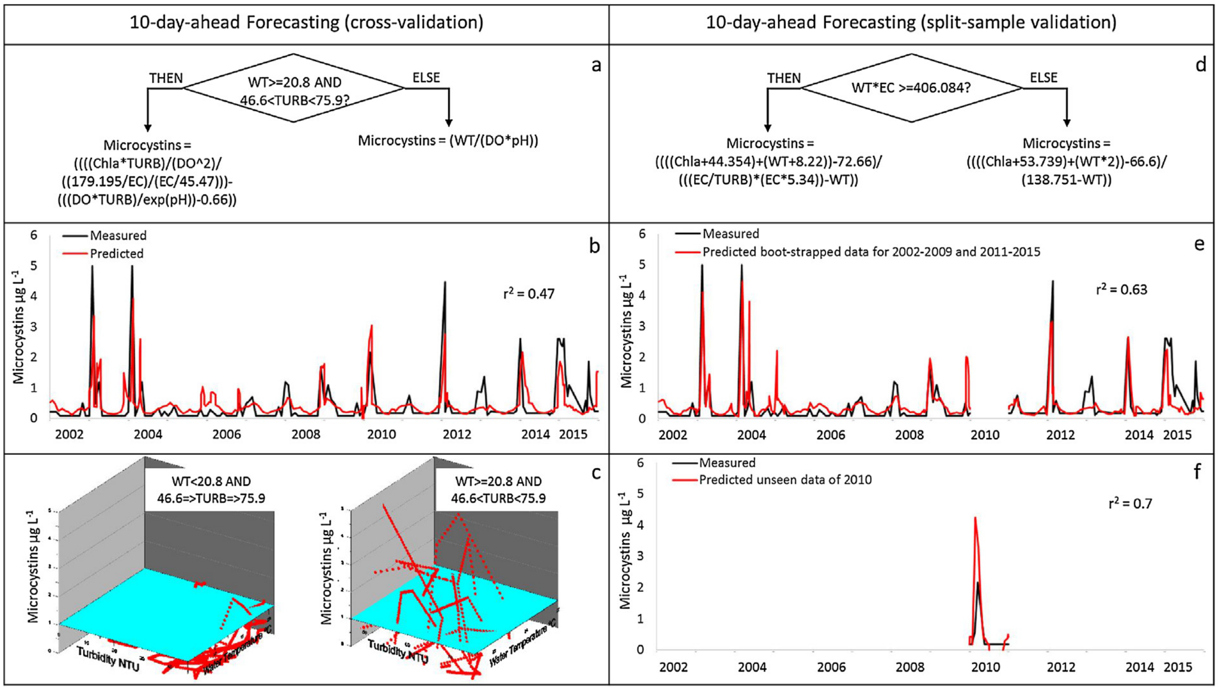

Fig. 9. 10-day-ahead forecasting models for total microcystins in Vaal Reservoir from 2002 to 2005. Crossvalidation: IF-THEN-ELSE model (a), validation result (b), separation of high and low concentrations (red dots) by IF-conditions: the light-blue horizontal plane marks the limit value 1 μg L^{-1}. (c) Split-sample validation: IF-THEN-ELSE model (d), validation results (e), (f). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

early warning for the 2008 to 2010 blooms and predicted the trigger level for the bloom of 2013/2014. The 20-day-ahead model ensemble would have correctly predicted the bloom events in 2008 to 2010 when μ_c exceeded 0.2 day^{-1} (Fig. 6d), but failed to predict the bloom event in 2013/2014 when peak cell concentrations exceeded 2×10^4 cells mL^{-1}.

Both cross- and split-sample-validation displayed good correlation between observed and 10-day-ahead predictions of the population dynamics of *C. raciborskii* in Lake Wivenhoe for 2007 until 2015 (Fig. 7b, e and f). The best-performing model shown in Fig. 7a achieved a correlation coefficient of determination (r^2) of 0.48 at $p < 0.05$, and its IF-condition suggested that distinct ranges of pH and electrical conductivity separate high and low cell concentrations (see Fig. 7c). This is based on the assumption that a cell concentration of 8×10^3 cells mL^{-1} of *C. raciborskii* corresponds to the proposed 1 μg L^{-1} guideline

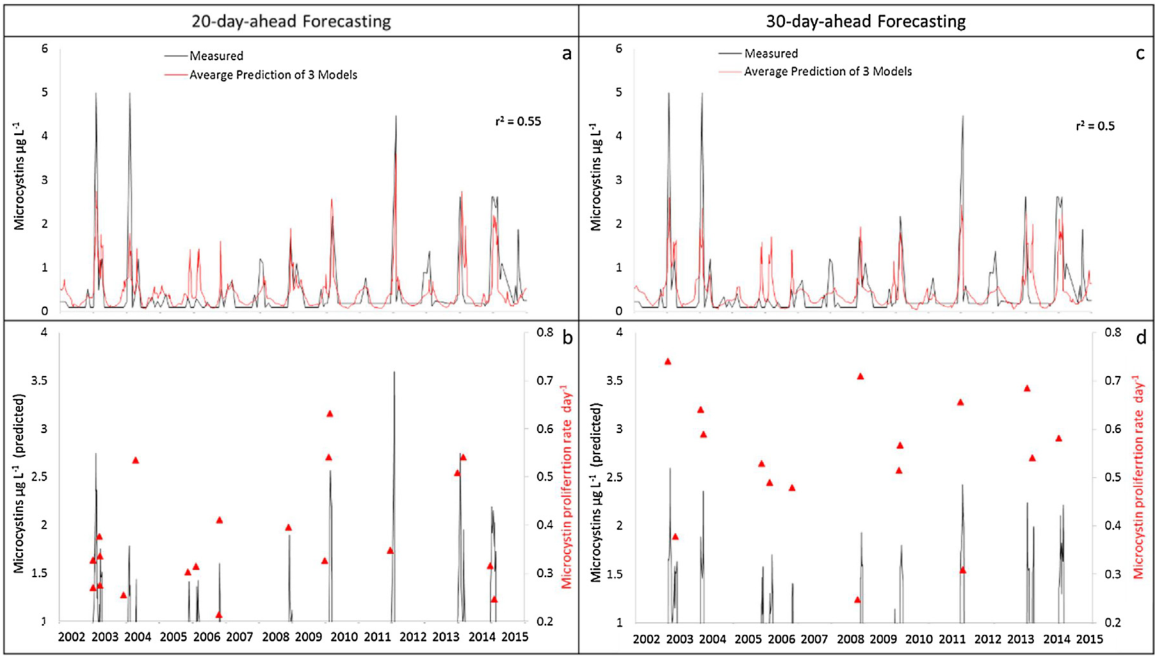

Fig. 10. Forecasting models for total microcystins in Vaal Reservoir from 2002 to 2015. 20-day-ahead forecasting: cross-validation (a), forecasted daily concentrations above the limit concentration of 1 μg L^{-1} and proliferation rates above 0.2 day^{-1}. (b) 30-day-ahead forecasting: cross-validation (c), forecasted daily concentrations above the limit concentration of 1 μg L^{-1} and proliferation rates above 0.2 day^{-1} (d).

concentration for cylindrospermopsins in drinking water currently being considered by the WHO (2003). The timing of blooms between 2008 and 2015 was successfully predicted, except for a smaller bloom event in the summer of 2007/2008 (Fig. 7b). Even though predicted peak cell concentrations in 2009, 2010 and 2013 where lower than the observed data, the model still correctly predicted observed bloom events when peak cell concentrations exceeded 8×10^3 cells mL^{-1}.

Fig. 8 shows validation results for the 20- and 30-day-ahead predictions of *C. raciborskii* cell concentrations that were averaged from ensembles of the 3 best models for each case. The averaged forecasts of the model ensembles matched very well with the timing of observed bloom events between 2008 and 2015 but again underestimated peak cell concentrations for most of the bloom events (Fig. 8a and c). However, both the 20- and 30-day-ahead models correctly predict all μ_c greater than 0.2 d^{-1} and all peak cell concentrations higher than 8×10^3 cells mL^{-1} that according to Fig. 2 would constitute a cyanobacterial bloom alert.

3.2. Microcystins in Vaal Reservoir (South Africa)

In-situ water quality data and total microcystins concentrations measured from 2002 until 2015 (see Table 2) were used to develop predictive models for total microcystins for 10 to 30 days ahead by means of the HEA. The models were tested for their suitability to be used for early warning of microcystins concentrations that exceed 1 μg L^{-1} in Vaal Reservoir. Such models would directly target the WHO provisional guideline level for microcystins rather than predicting cell concentrations of potentially non-toxic *Microcystis aeruginosa* strains.

Fig. 9 documents cross- and split-sample validation for best-performing 10-day-ahead prediction models of microcystins. Both models failed to predict a minor peak event in 2013 but correlated well with the timing of major peaks in 2003, 2004, 2010, 2012, 2014 and 2015 when concentrations of microcystins exceeded 1 μg L^{-1} (Fig. 9b, d, e). The threshold conditions of the model in Fig. 9a suggest that in Vaal Reservoir, water temperatures greater than 20.8 °C and turbidity in the range between 46 and 75 NTU were indicative of increased concentrations of microcystins. This finding reflects the fact that highest dissolved microcystins concentrations can be expected during the collapse of a *Microcystis* bloom typically occurring at warmer water temperatures. Fig. 9c illustrates how these thresholds separate microcystins concentrations above and below 1 μg L^{-1} as a prerequisite for the model's forecasting performance.

Fig. 10 shows averaged 20- and 30-day-ahead predictions from ensembles of the 3 best models for microcystins concentrations in Vaal Reservoir. The model ensembles accurately predicted the major peak microcystins events in 2003, 2004 and 2011 when microcystins concentrations exceeded 1 μg L^{-1} (Fig. 10a and c). Fig. 10b and d shows daily proliferation rates greater than 0.2 day^{-1} before and during events of significantly increased microcystin concentrations.

4. Discussion and conclusions

Even though there is a highly complex synergism between environmental and climatic factors and the development of cyanobacterial blooms and cyanotoxins in freshwaters, this study has shown that HEA can successfully predict short-term temporal dynamics of cyanobacteria and microcystins by means of inferential models of *in situ* water quality data. These results open up the possibility of accurately predicting cyanobacteria cell concentrations in real-time using only online water quality data monitored by sensors, which can replace or complement current water quality monitoring practices.

The two case studies allow to draw following conclusions:

(1) The models for *D. circinale* proved to be suitable only for 10-day-ahead predictions and early warning of bloom events, but failed 20- and 30-days-ahead forecasts of the major bloom event in 2013/2014.
(2) The models for *C. raciborskii* forecasted satisfactorily all bloom events that exceeded proposed guideline levels up to 30 days

ahead, even though peak cell concentrations were sometimes underestimated.

(3) IF-conditions of the models for Lake Wivenhoe suggested that *D. circinale* has the fastest cell division rates at water temperatures greater than 24.1 °C which corresponds with findings by Reynolds (1984), and that *C. raciborskii* grows fastest at water temperatures greater than 26.1 °C which correlates well findings by Briand et al. (2002). The thresholds also suggested that *C. raciborskii* is tolerant to electrical conductivity higher than 253 $\mu S\,m^{-1}$ as previously suggested by Briand et al. (2002) and Moisander et al. (2008).

(4) 10- to 30-day-ahead forecasts of total microcystin concentrations in the Vaal Reservoir proved to be valid as an early warning of events that exceeded 1 $\mu g\,L^{-1}$.

(5) IF-conditions of the models suggested that water temperatures greater than 20.8 °C and turbidity in the range between 46 and 75 NTU may be indicative for increasing concentrations of microcystins in the Vaal Reservoir.

(6) Forecasting models for microcystins in Vaal Reservoir required neither measured cyanobacterial cell counts, nor measured nutrient concentrations but were solely derived using *in situ* sensor water quality data and laboratory measurements of total microcystins.

(7) Inferential models predicting explicitly cyanotoxin concentrations avoid to inadvertently model and predict non-toxic cyanobacterial strains.

To the best of the authors' knowledge, there are no published examples of models solely driven by *in situ* sensor monitoring data that have the capacity to reasonably accurate forecast daily concentrations of cyanobacteria and cyanotoxin species in lakes for up to 30 days ahead, as demonstrated by this research. Future research will test the modelling approach for different cyanobacteria and cyanotoxin species in a variety of lakes and reservoirs that will allow to generalise species specific models for lakes categorised by morphometry, circulation type and trophic state, as suggested by Recknagel et al. (2008). Resulting model libraries will facilitate the sharing of species-specific models across lakes of the same category. Further improvements to the predictive capability of the hybrid evolutionary algorithm HEA can be expected by targeting the accuracy of timing and peak concentrations of cyanobacteria bloom events in a more differentiated way through testing alternative fitness criteria such as Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) and percent bias (PBIAS) (Gupta et al., 1998).

Acknowledgments

The authors acknowledge funding of this research by the Australian Research Council (LP0990453). We also appreciate valuable comments of the anonymous reviewer that improved significantly the clarity of the manuscript.[CG]

References

Bouvy, M., Falcao, D., Marinho, M., Pagano, M., Moura, A., 2000. Occurrence of *Cylindrospermopsis* (Cyanobacteria) in 39 Brazilian tropical reservoirs during the 1998 drought. Aquat. Microb. Ecol. 23, 13–27.

Briand, J.F., Robillot, C., Quiblier-Lloberas, C., Humbert, J.F., Coute, A., Bernard, C., 2002. Environmental context of *Cylindrospermopsis raciborskii* (Cyanobacteria) blooms in a shallow pond in France. Water Res. 36, 3183–3192.

Burford, M.A., Beardall, J., Willis, A., Orr, P.T., Magalhaes, V.F., Rangel, L.M., Azevedo, S.M.F.O.E., Neilan, B.A., 2016. Understanding the winning strategies used by the bloom forming cyanobacterium *Cylindrospermopsis raciborskii*. Harmful Algae 54, 44–53.

Cao, H., Recknagel, F., Orr, P.T., 2014. Parameter optimisation algorithms for evolving rule models applied to freshwater ecosystem. IEEE Trans. Evol. Comput. 18 (6), 793–806.

Cao, H., Recknagel, F., Bartkow, M., 2016. Spatially-explicit forecasting of cyanobacteria assemblages in freshwater lakes by multi-objective hybrid evolutionary algorithms. Ecol. Model. 342, 97–112.

Carmichael, W.W., 2001. Health effects of toxin producing cyanobacteria: the CyanoHABS'. Hum. Ecol. Risk Assess. 7, 1393–1407.

Chen, Q., Zhang, Z., Recknagel, F., Guo, J., Blanckaert, K., 2014. Adaptation and multiple parameter optimization of the simulation model SALMO as prerequisite for scenario analysis on a shallow eutrophic lake. Ecol. Modell. 273, 109–116.

Conradie, R.C., Barnard, S., 2012. The dynamics of toxic *Microcystis* strains and microcystin production in two hypertrofic South African reservoirs. Harmful Algae 20, 1–10.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour. Res. 34 (4), 751–763.

Hawkins, P.R., Runnegar, M.T.C., Jackson, A.R.B., Falconer, I., 1985. Severe hepatotoxicity caused by the tropical cyanobacterium (blue-green alga) *Cylindrospermopsis raciborskii* (Woloszynska) Seenaya and Subba Raju isolated from a domestic supply reservoir. Appl. Environ. Microbiol. 50, 1292–1295.

Holland, J.H., et al., 1986. Induction. Process of Inference, Learning and Discovery. MIT Press, Cambridge.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbour, MI.

Huber, V., Wagner, C., Gerten, D., Adrian, R., 2012. To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. Oecologia 169, 245–256.

Jang, M.H., Jung, J.M., Takamura, N., 2007. Changes in microcystin production in cyanobacteria exposed to zooplankton at different population densities and infochemical concentrations. Limnol. Oceanogr. 52 (4), 1454–1466.

Koza, J.R., 1992. Genetic Programming On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA.

Lunetta, R.S., et al., 2015. Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. Remote Sens. Environ. 157, 24–34.

Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sens. Environ. 156, 374–382.

Michalak, A.M., et al., 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. Natl. Acad. Sci. U. S. A. 110 (16), 6448–6452.

Moisander, P.H., Paerl, H.W., Zehr, J.P., 2008. Effects of inorganic nitrogen on taxa specific cyanobacterial growth and nifH expression in a subtropical estuary. Limnol. Oceanogr. 53, 2519–2532.

NHMRC, NRMMC, 2011. Australian Drinking Water Guidelines Paper 6 National Water Quality Management Strategy. National Health and Medical Research Council, National Resource Management Ministerial Council, Commonwealth of Australia, Canberra.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, part I—a discussion of principles. J. Hydrol. 10, 282–290.

Nguyen, H.H., Recknagel, F., Meyer, W., Frizenschaf, J., Shrestha, M., 2017. Modelling the impacts of altered management practices, land use and climate changes on the water quality of the Millbrook catchment reservoir system in South Australia. J. Environ. Manage. 202 (Part 1), 1–11 (in press).

Orr, P.T., Rasmussen, P., Burford, M.A., Eaglesham, G.K., Lennox, S.M., 2010. Evaluation of quantitative real-time PCR to characterise spatial and temporal variations in cyanobacteria, *Cylindrospermopsis raciborskii* (Woloszynska) Seenaya et Subba Raju and cylindrospermopsin concentrations in three subtropical Australian reservoirs. Harmful Algae 9, 243–254.

Paerl, H.W., Otten, T.G., 2013. Blooms bite the hand that feeds them. Science 342, 433–434.

Paerl, H.W., 1990. Physiological ecology and regulation of N_2 fixation in natural waters. Adv. Microb. Ecol. 11, 305–344.

Recknagel, F., Hosomi, M., Fukushima, T., Kong, D.-S., 1995. Short- and long- term control of external and internal phosphorus loads in lakes: a scenario analysis. Water Res. 29 (7), 1767–1779.

Recknagel, F., Cao, H., van Ginkel, C., van der Molen, D., Park, H., Takamura, N., 2008. Adaptive agents for forecasting seasonal outbreaks of blue-green algal populations in lakes categorised by circulation type and trophic state. Verh. Int. Verein Limnol. 30 (2), 191–197.

Recknagel, F., Orr, P., Cao, H., 2014. Inductive reasoning and forecasting of population dynamics of *Cylindrospermopsis raciborskii* in three sub-tropical reservoirs by evolutionary computation. Harmful Algae 31, 26–34.

Reynolds, C., 1984. The Ecology of Freshwater Phytoplankton. Cambridge University Press, Cambridge.

Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems. A global problem. Environ. Sci. Pollut. Res. 10 (2), 126–139.

Storn, R., Price, K., 1997. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. 11, 341–359.

World Health Organisation (WHO), 2003. Guidelines for Safe Recreational Waters. Volume1: Coastal and Fresh Waters. WHO Publishing, Geneva, Switzerland.

Wagner, C., Adrian, R., 2009. Cyanobacteria dominance: quantifying the effects of climate change. Limnol. Oceanogr. 54, 2460–2468.

Ye, L., Cai, Q., Zhang, M., Tan, L., 2014. Real-time observations, early warning and forecasting phytoplankton blooms by integrating in situ observations, online sondes and hybrid evolutionary algorithms. Ecol. Inf. 22, 44–51.