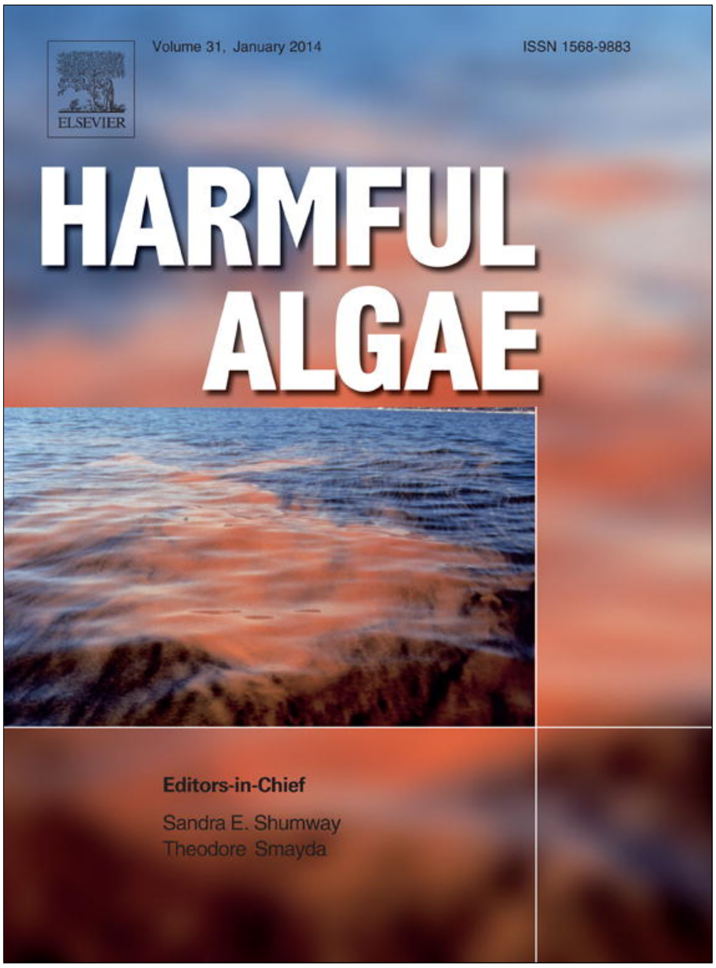
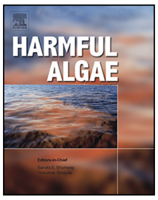

Volume 31, January 2014
ISSN 1568-9883
ELSEVIER
HARMFUL
ALGAE
Editors-in-Chief
Sandra E. Shumway
Theodore Smayda

Harmful Algae 31 (2014) 26–34

Contents lists available at ScienceDirect

Harmful Algae

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



Inductive reasoning and forecasting of population dynamics of *Cylindrospermopsis raciborskii* in three sub-tropical reservoirs by evolutionary computation

Friedrich Recknagel [a,*], Philip T. Orr [b], Hongqing Cao [a]

[a] *School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia*
[b] *Seqwater, PO Box 16146, City East, Qld 4002, Australia*

A R T I C L E I N F O

Article history:
Received 15 February 2013
Received in revised form 10 September 2013
Accepted 10 September 2013

Keywords:
Cylindrospermopsis raciborskii
Cyanobacteria blooms
Evolutionary computation
Ecological thresholds
Ecological relationships
Forecasting

A B S T R A C T

Seven-day-ahead forecasting models of *Cylindrospermopsis raciborskii* in three warm-monomictic and mesotrophic reservoirs in south-east Queensland have been developed by means of water quality data from 1999 to 2010 and the hybrid evolutionary algorithm HEA. Resulting models using all measured variables as inputs as well as models using electronically measurable variables only as inputs forecasted accurately timing of overgrowth of *C. raciborskii* and matched well high and low magnitudes of observed bloom events with $0.45 \le r^2 > 0.61$ and $0.4 \le r^2 > 0.57$, respectively. The models also revealed relationships and thresholds triggering bloom events that provide valuable information on synergism between water quality conditions and population dynamics of *C. raciborskii*. Best performing models based on using all measured variables as inputs indicated electrical conductivity (EC) within the range of 206–280 mS m^{-1} as threshold above which fast growth and high abundances of *C. raciborskii* have been observed for the three lakes. Best models based on electronically measurable variables for the Lakes Wivenhoe and Somerset indicated a water temperature (WT) range of 25.5–32.7 °C within which fast growth and high abundances of *C. raciborskii* can be expected. By contrast the model for Lake Samsonvale highlighted a turbidity (TURB) level of 4.8 NTU as indicator for mass developments of *C. raciborskii*.

Experiments with online measured water quality data of the Lake Wivenhoe from 2007 to 2010 resulted in predictive models with $0.61 \le r^2 > 0.65$ whereby again similar levels of EC and WT have been discovered as thresholds for outgrowth of *C. raciborskii*. The highest validity of $r^2 = 0.75$ for an in situ data-based model has been achieved after considering time lags for EC by 7 days and dissolved oxygen by 1 day. These time lags have been discovered by a systematic screening of all possible combinations of time lags between 0 and 10 days for all electronically measurable variables. The so-developed model performs seven-day-ahead forecasts and is currently implemented and tested for early warning of *C. raciborskii* blooms in the Wivenhoe reservoir.

1. Introduction

Lakes Wivenhoe, Somerset and Samsonvale are subtropical, warm-monomictic and mesotrophic reservoirs located near Brisbane in the subtropical southeast of Queensland, Australia. Each of these reservoirs have annual recurring blooms of the potentially toxic cyanobacterium *Cylindrospermopsis raciborskii* (Woloszynska) Seenaya et Subba Raju (Orr et al., 2010). *C. raciborskii* produces cylindrospermopsins (CYN's) which are hepatotoxic alkaloid cyanotoxins that present a risk to human health (e.g. Hawkins et al., 1985) and which must be removed from raw water during water treatment. Controlling the development of *C. raciborskii* within these reservoirs is a key goal of Seqwater, the water authority responsible for the management of the reservoirs. However, *C. raciborskii* is ecologically adaptable and can form blooms under a range of light, temperature and nutrient regimes (Isvanovics et al., 2000; Sprober et al., 2003; Briand et al., 2002). It also tolerates a wide range of other environmental conditions including oligohaline waters (Caldwell, 2001) and nitrogen depleted waters through its ability to fix atmospheric nitrogen (N_2) (Bouvy et al., 2000; Moisander et al., 2008). Although *C. raciborskii* is being considered a tropical or sub-tropical species and recorded in tropical countries such as Brazil (e.g. Bouvy et al., 2000; Branco and Senna, 1994; DeSouza et al., 1998) it is now increasingly detected in temperate regions of Europe (Dokulil and Mayer, 1996; Padisak, 1997; Briand et al., 2002; Fastner et al.,

---

[*] Corresponding author. Tel.: +61 8 8313 3953; fax: +61 8 8303 6222.
E-mail address: friedrich.recknagel@adelaide.edu.au (F. Recknagel).

1568-9883/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.hal.2013.09.004

Table 1
Summary of water quality data of the three reservoirs used in this study.

	Wivenhoe		Somerset	Samsonvale
	Historical data 1999–2009	On-line data 2007–2010	Historical data 1999–2009	Historical data 1999–2009
	Mean ± SD			
Electrical conductivity (EC) μS cm^{-1}	354.90 ± 72.64	348.75 ± 72.1	235.75 ± 45.35	226.94 ± 41.62
Turbidity (TURB) NTU	12.92 ± 56.76	2.71 ± 3.59	2.32 ± 2.25	2.72 ± 1.4
Water temperature (WT) °C	22.66 ± 3.70	22.51 ± 3.61	22.69 ± 3.98	22.47 ± 3.64
Dissolved oxygen (DO) mg L^{-1}	8.67 ± 1.25	7.81 ± 1.61	7.55 ± 1.82	7.96 ± 1.4
pH	8.21 ± 0.32	8.15 ± 0.54	7.97 ± 0.46	7.95 ± 0.34
Silica (SiO_2) mg L^{-1}	3.24 ± 2.69		4.25 ± 2.13	2.27 ± 1.78
Total nitrogen (TN) mg L^{-1}	0.48 ± 0.078		0.59 ± 0.15	0.56 ± 0.09
Total phosphorus (TP) mg L^{-1}	0.019 ± 0.012		0.028 ± 0.017	0.018 ± 0.007
Chlorophyll_a (Chl_*a*) mg L^{-1}	7.82 ± 4.54	6.41 ± 3.58	10.66 ± 7.49	11.6 ± 5.23
Cylindrospermopsis cells/mL	10,108 ± 20,698	6098 ± 12,466	10,259 ± 250,489	14,853 ± 19,110

2007), New Zealand (Wood and Stirling, 2003), Canada (Hamilton et al., 2005) and the U.S. (Chapman and Schelske, 1997; Calandrino and Perl, 2011).

The drivers of *C. raciborskii* blooms are highly complex and the synergism between *C. raciborskii* bloom development and physical-chemical water quality conditions is still poorly understood. Novel techniques for inductive reasoning and forecasting from historical data can reveal environmental conditions and thresholds that have triggered blooms of *C. raciborskii* in the past, and synthesise short-term forecasting models for up to 14 days ahead. These models can give early warning for timely operational control and risk management of cyanobacteria bloom events.

Artificial neural networks (ANN) and evolutionary computation (EC) have proved to be powerful techniques for inductive reasoning and forecasting of highly complex limnological data (e.g. Recknagel et al., 1997, 2013). Even though both ANN and EC achieve similar good multivariate forecasting, only EC performs mathematically explicit syntheses and representation of underlying models. The hybrid evolutionary algorithm HEA (Cao et al., 2006, 2013) has been specifically designed and customised for inducing predictive and explanatory models from complex ecological data.

This study applied HEA to 11 years of water quality data collected from sites close to the dam wall of the three reservoirs and developed forecasting models to predict *C. raciborskii* cell concentrations up to 2 weeks in advance. The models also reveal underlying ecological relationships that trigger mass developments of *C. raciborskii*. Since the three lakes are similar in climate, circulation patterns and eutrophication levels, collective properties of *C. raciborskii* in the three lakes have been identified.

2. Materials and methods

2.1. The Lakes Wivenhoe, Somerset and Samsonvale

Eleven years of water quality data collected from Lakes Wivenhoe, Somerset and Samsonvale in south-east Queensland, Australia between 1999 and 2010 (see Table 1) were used for this study. The reservoirs are located to the north and west of Brisbane (Fig. 1) and supply drinking water for about 1.5 million people. Lake Samsonvale (also known as North Pine Dam) (27°15′S, 152°55′E) is the smallest reservoir with a surface area of 21.8 km^2 at full drinking water supply volume of 215,000 mL and has a bubble plume destratifier which is activated during the summer months to reduce stratification and help control *Cylindrospermopsis raciborskii* cell concentrations. Lake Somerset (27°7′S, 152° 33′E) is smaller with a surface area of 42.1 km^2 at full supply volume of 380,000 mL and is located upstream of the third, largest and most spatially diverse reservoir, Lake Wivenhoe (27°24′S, 152°36′E) (Orr et al., 2010). Lake Wivenhoe has a surface area of 107.5 km^2 and full supply volume of 1,165,000 mL. The catchments for these reservoirs are typically unprotected, with more than 50% dominated by cattle grazing pasture, and approximately 20% being natural vegetation (Orr et al., 2010).

Table 1 summarises the water quality data measured in the three reservoirs over an 11 year periods that have been utilised for modelling. Since the intervals for measuring the historical data ranged between weekly and biweekly, and sampling dates differed for physical, chemical and biological variables, the data have been interpolated to suit daily time steps for forecasting. For simplicity, we used linear interpolation and in order to

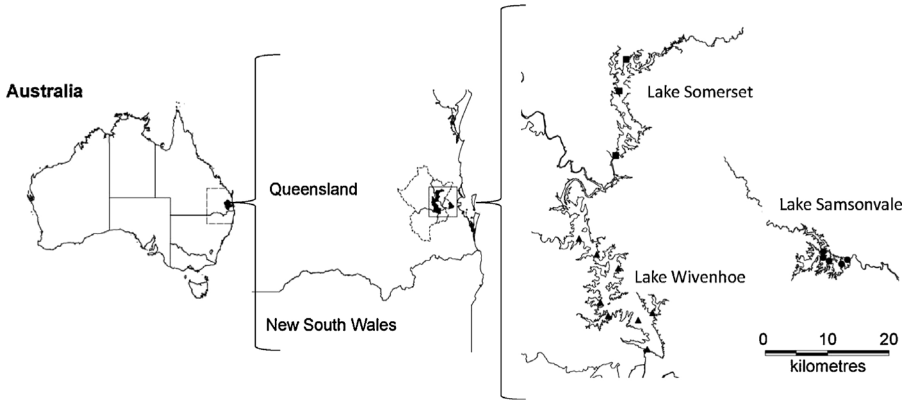


Fig. 1. The locations of the three reservoirs within Australia.

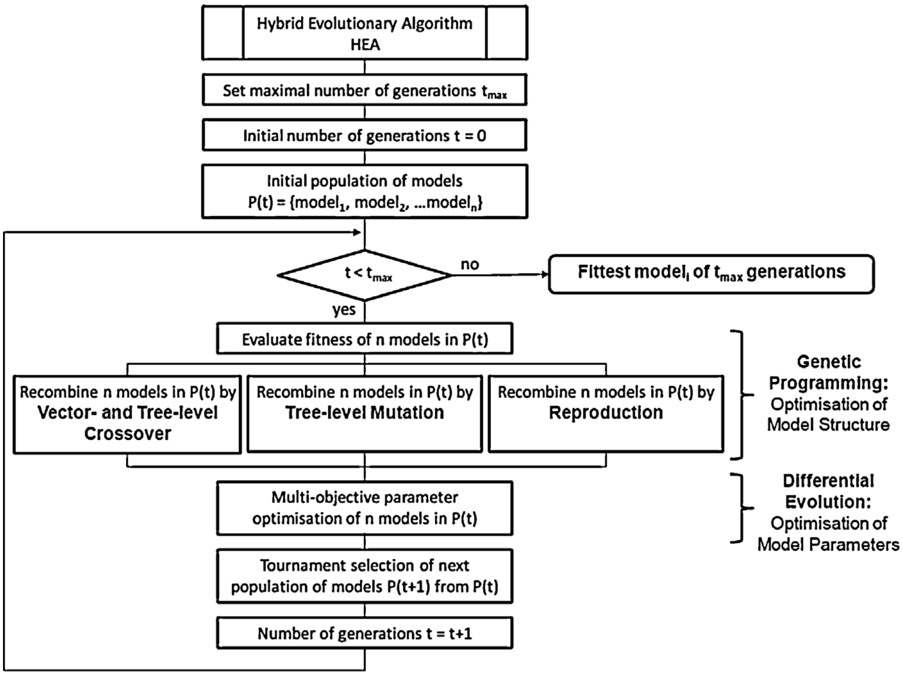


Fig. 2. Design and functioning of the hybrid evolutionary algorithm HEA (Recknagel et al., 2013).

develop models for 7-day-ahead forecasting the interpolated input data have been shifted by 7 days against the daily output data. On-line measured water quality data were available for Lake Wivenhoe from 2007 to 2010.

2.2. Hybrid evolutionary algorithm HEA

This study was conducted using the hybrid evolutionary algorithm HEA developed by (Cao et al., 2006, 2013) and which had been successfully applied for 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). The design of HEA combines genetic programming (GP) for optimising the model structure and differential evolution (DE) for optimisation of model parameters (see Fig. 2). We use GP according to Koza (1992) to evolve the rule model structure. Since GP typically operates on parse trees rather than traditionally used bit strings, it suits well to evolve equations or formula for input-output relationships. Based on the three function sets: logic functions ($\mathrm{FL} = \{\mathrm{AND}, \mathrm{OR}\}$), comparison functions ($\mathrm{FC} = \{>, <, \geq, \leq\}$) and arithmetic functions ($\mathrm{FA} = \{+, -, *, /, \exp, \ln\}$), GP represents IF-THEN-ELSE rule models as a vector of multiple trees (Tree 1, Tree 2, Tree 3) where Tree 1 denotes the IF condition branch, Tree 2 and Tree 3 denote the THEN branch and ELSE branch respectively. Hence their function sets are: $\mathrm{FTree1} = \mathrm{FL} \cup \mathrm{FC} \cup \mathrm{FA}$ and $\mathrm{FTree2/Tree3} = \mathrm{FA}$. Cao et al. (2006) illustrate in great detail the design of IF-THEN-ELSE model by GP.

Differential evolution (DE) is one of the most recent evolutionary algorithms for solving real-parameter optimisation problems proposed by Storn and Price (1997). It is an effective global optimisation algorithm which extracts the differential information (i.e., information on distance and direction towards global optimum) from the current population of solutions to guide its further search. It doesn't require separate probability distribution that makes the scheme completely self-organizing. More details on how DE is implemented in HEA are provided in Cao et al. (2013).

Since HEA induces models from data patterns, it requires cleansed and formatted cross-sectional or time-series data that are representative of the system to be modelled for both the number of observations and their relevance. The more event-related patterns the historical data contains, the more generic models tend to become, and the more likely the model's predictive validity reaches beyond the data limits. Ecosystem evolution requires that models become regularly upgraded by updated data.

Time series between 1999 and 2010 from the three reservoirs used in this study provided seasonal and interannual patterns of physical and chemical water quality data as well as population dynamics of *C. raciborskii*. In order to utilise the information content of available data the boot-strap scheme has been implemented for the training of HEA as illustrated in Fig. 3. It

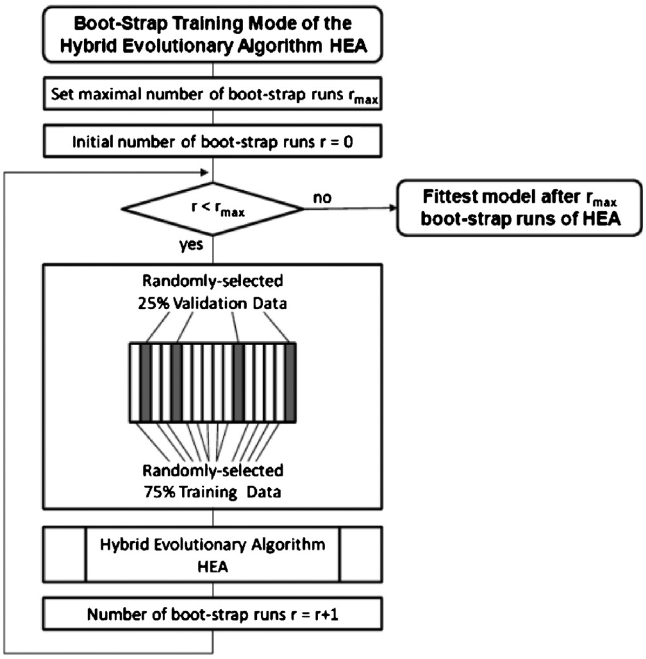


Fig. 3. Scheme of boot-strap training applied to the hybrid evolutionary algorithm HEA (Recknagel et al., 2013).

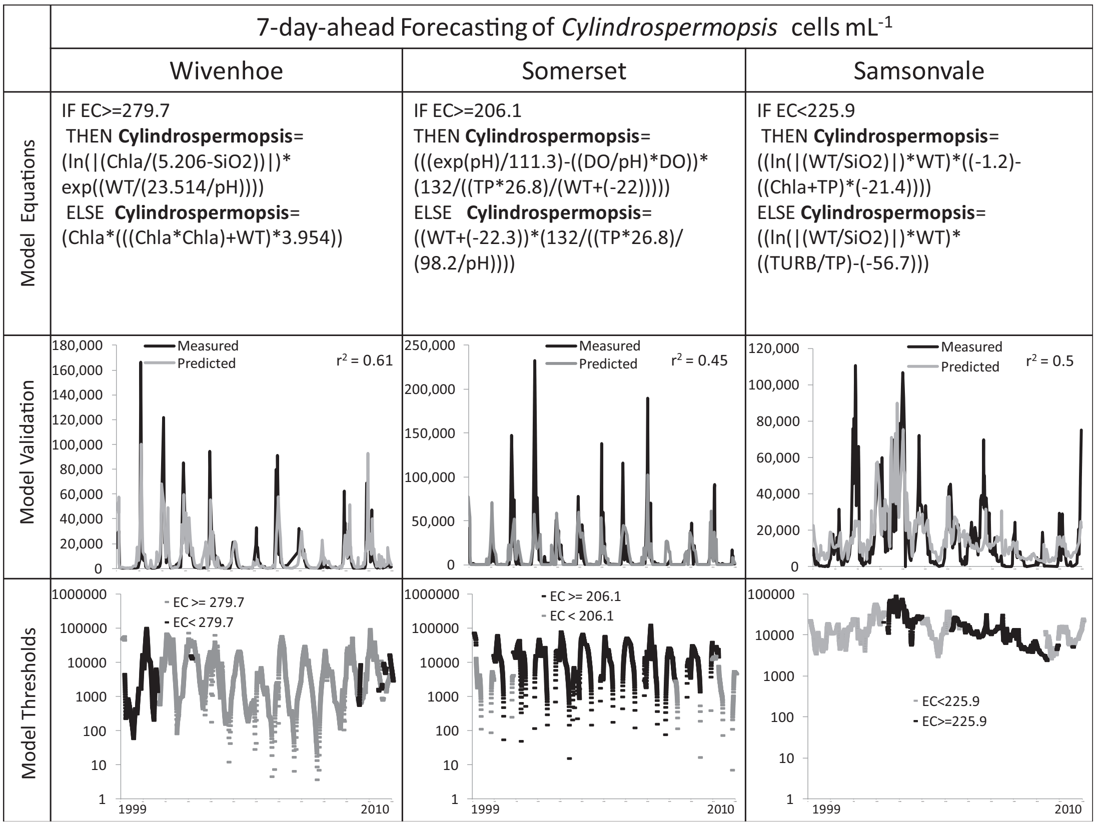


Fig. 4. 7-day-ahead forecasting of *C. raciborskii* utilising all water quality data of the Lakes Wivenhoe (left column), Somerset (middle column) and Samsonvale (right column) documented by the model equations (upper row), model validation (middle row) and thresholds (bottom row).

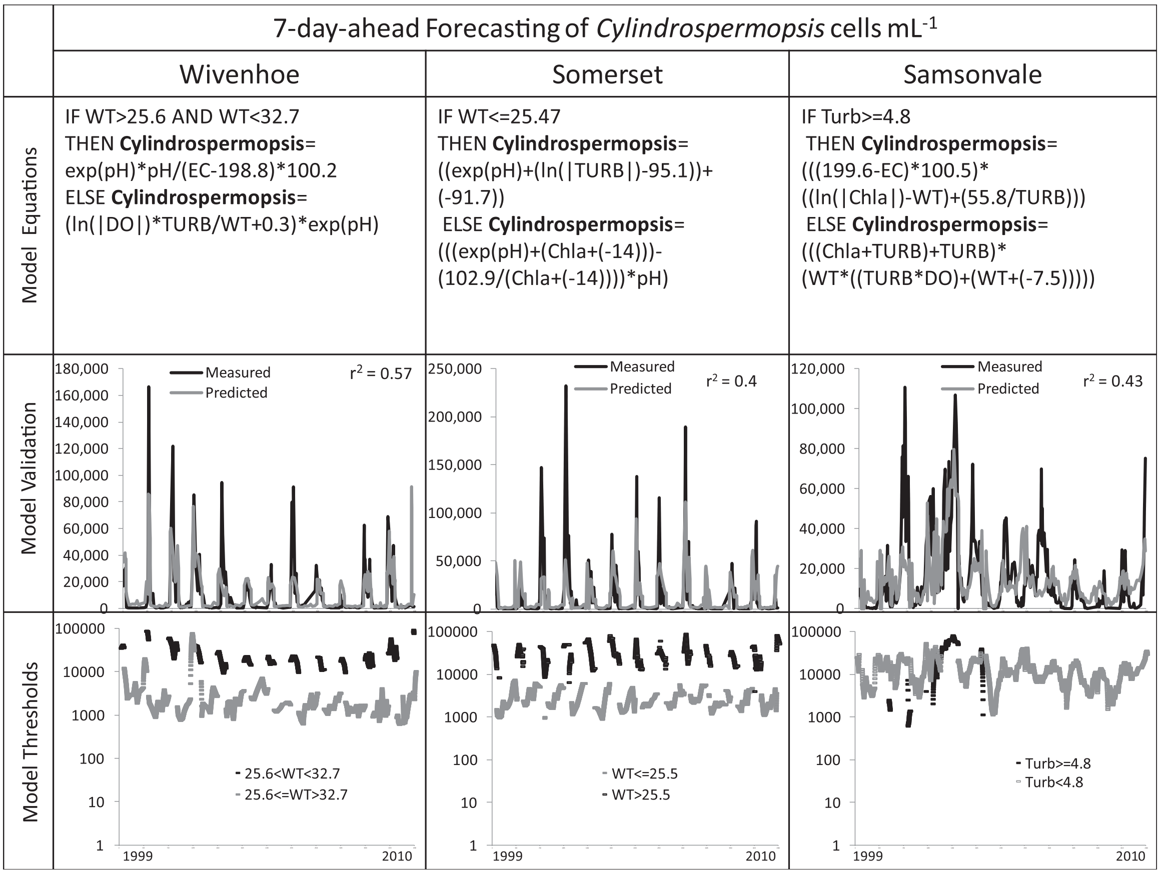


Fig. 5. 7-day-ahead forecasting of *C. raciborskii* utilising electronically measurable water quality data of the Lakes Wivenhoe (left column), Somerset (middle column) and Samsonvale (right column) documented by the model equations (upper row), model validation (middle row) and thresholds (bottom row).

randomly selects r_{max} data-subsets for training and testing for each of which t_{max} generations of models are evolved. After the completion of r_{max} boot-strap runs it determines the overall "fittest model" of all the generations evolved through the principles of genetic programming and differential evolution. In this study a value of $r_{max} = 100$ was applied to all modelling experiments while $t_{max} = 80$ was used for modelling historical data from the three reservoirs and $t_{max} = 100$ was used for modelling on-line data from Lake Wivenhoe.

3. Results

3.1. *Forecasting models of* C. raciborskii *based on historical water quality data from 1999 to 2010*

3.1.1. *Model validation and thresholds*

Modelling results for the three reservoirs based on 11 years of water quality data summarised in Table 1 are shown in Fig. 4. The best performing models match the timing of *Cylindrospermopsis raciborskii* bloom development for the 11 consecutive years but sometimes underestimate the magnitudes of bloom events and this is reflected by r^2 values of 0.61, 0.45 and 0.5 respectively. The three models conform by suggesting low levels of electrical conductivity EC as thresholds for triggering different abundances of *C. raciborskii* as shown in the bottom row of Fig. 4.

The three best performing models for the three lakes developed only from electronically measurable input variables including WT, EC, pH, DO, TURB and Chl_*a* fluorescence are documented in Fig. 5. Whilst r^2 values of 0.57 for Wivenhoe, 0.4 for Somerset and 0.43 for Samsonvale are slightly lower compared with the models where chemical and biological data are included (Fig. 4), the annual timing of *C. raciborskii* bloom development is still accurately predicted, and magnitudes of peak abundances well approximated. The models also indicate that bloom development in Wivenhoe and Somerset may be triggered by WT in contrast to Samsonvale where TURB has been indicated as a criterion for excessive or diminished abundances of *C. raciborskii.* The derived water temperature range of $25.6 < WT < 32.7$ °C for Wivenhoe (Fig. 5, left column) correlates well with the known temperature optimum of *C. raciborskii* (Briand et al., 2002). A similar threshold value of 25.5 °C for WT triggering *C. raciborskii* blooms in Somerset (Fig. 5, middle column) also correlates with the published temperature optimum of this species.

The turbidity level of 4.8 NTU discovered as threshold for high abundances of *C. raciborskii* in Samsonvale may reflect inter-relationships between increasing water cloudiness by growing cell concentrations and underwater light conditions. Even though one peak in 2003 seems to be related to $TURB >= 4.8$ NTU the remaining peak events correspond with $TURB < 4.8$ NTU and therefore higher water transparency.

3.1.2. *Ecological relationships*

Fig. 6 shows the input selection frequencies for the 20 best models of each of the three reservoirs. Water temperature was the most frequently selected parameter by more than 90% of the models while electrical conductivity was shown to be least frequently selected. The biggest difference in the selection frequency between the three reservoirs was observed for TURB that ranked second after WT for Wivenhoe and Samsonvale but ranked last for Somerset.

Fig. 7 summarises the relationship between *C. raciborskii* cell concentrations, WT and TURB that have been discovered by

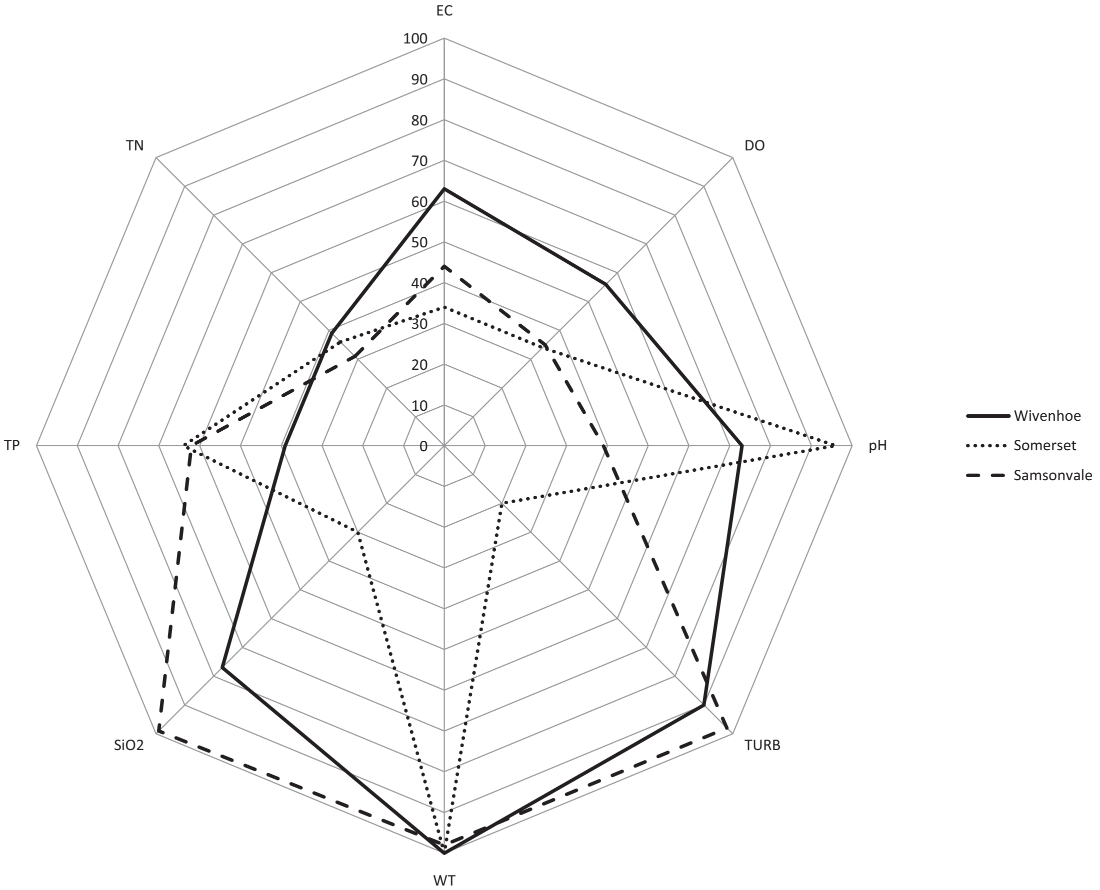


Fig. 6. Input selection frequencies of the twenty best performing models for the three Lakes Wivenhoe, Somerset and Samsonvale.

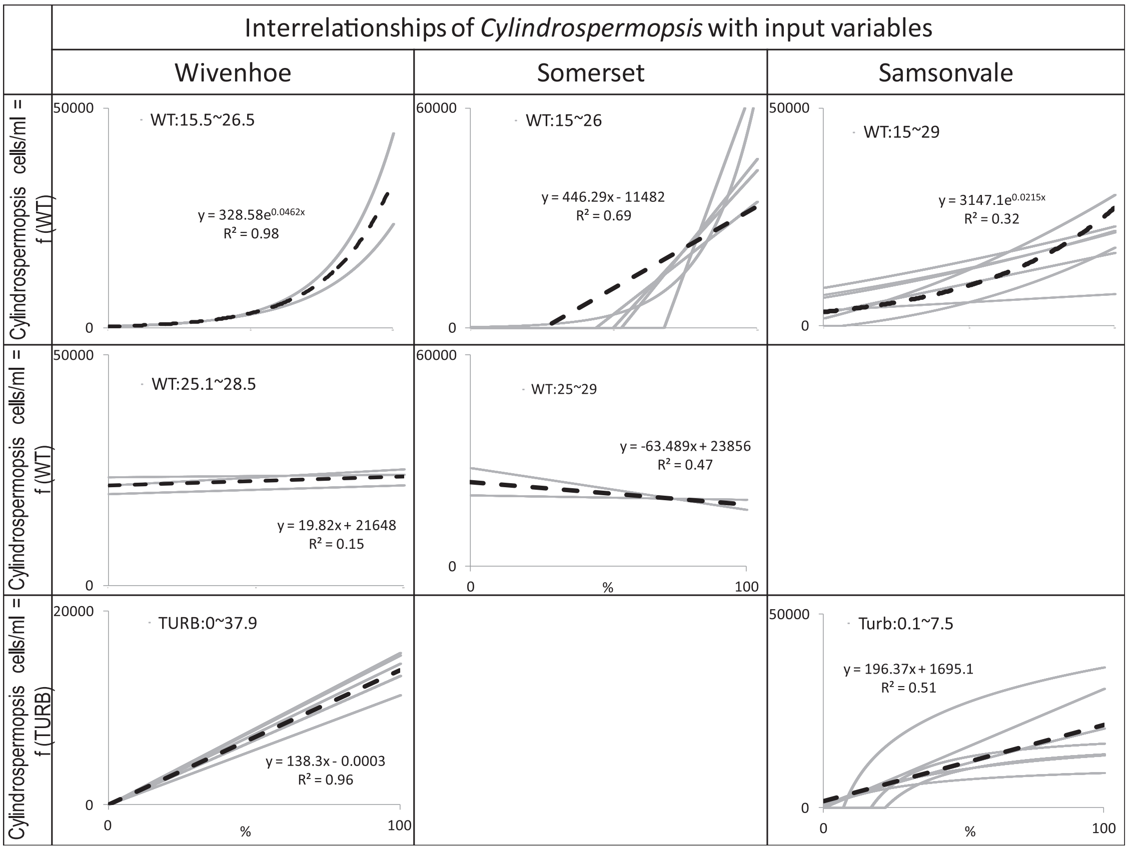


Fig. 7. Interrelationships between *C. raciborskii* and water temperature (top and middle row) and turbidity TURB (bottom row) extracted from sensitivity analysis of 10 best performing models of the Lakes Wivenhoe (left column), Somerset (middle column) and Samsonvale (right column). Dashed lines indicate trend lines.

sensitivity analyses of the 10 best models for the three lakes. It suggests that *C. raciborskii* starts to grow at WT higher than 20 °C and reaches its fastest growth at approximately 23 °C (Fig. 7, top row). Results from Wivenhoe and Somerset indicate a WT range of between 25 and 29 °C are required for sustained cell concentrations in excess of 2×10^4 cells mL^{-1} (Fig. 7, middle row). Rising cell concentrations coincide with increasing turbidity levels in Lakes Wivenhoe and Samsonvale (Fig. 7, bottom row).

The relationships between *C. raciborskii*, EC and pH are shown in Fig. 8. The results for Wivenhoe suggest that cell concentrations of up to 5×10^5 cells mL^{-1} are exponentially declining within the EC range from 203 to 504 $\mu\text{S cm}^{-1}$. Declining effects on cell

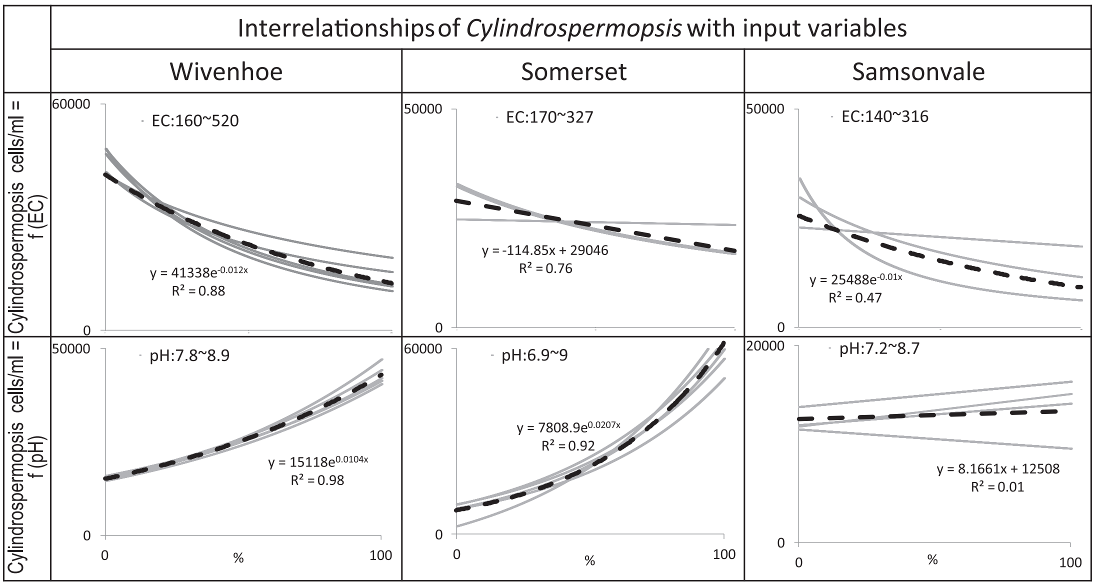


Fig. 8. Interrelationships between *C. raciborskii* and electrical conductivity (top row) and pH (bottom row) extracted from sensitivity analysis of 10 best performing models of the Lakes Wivenhoe (left column), Somerset (middle column) and Samsonvale (right column). Dashed lines indicate trend lines.

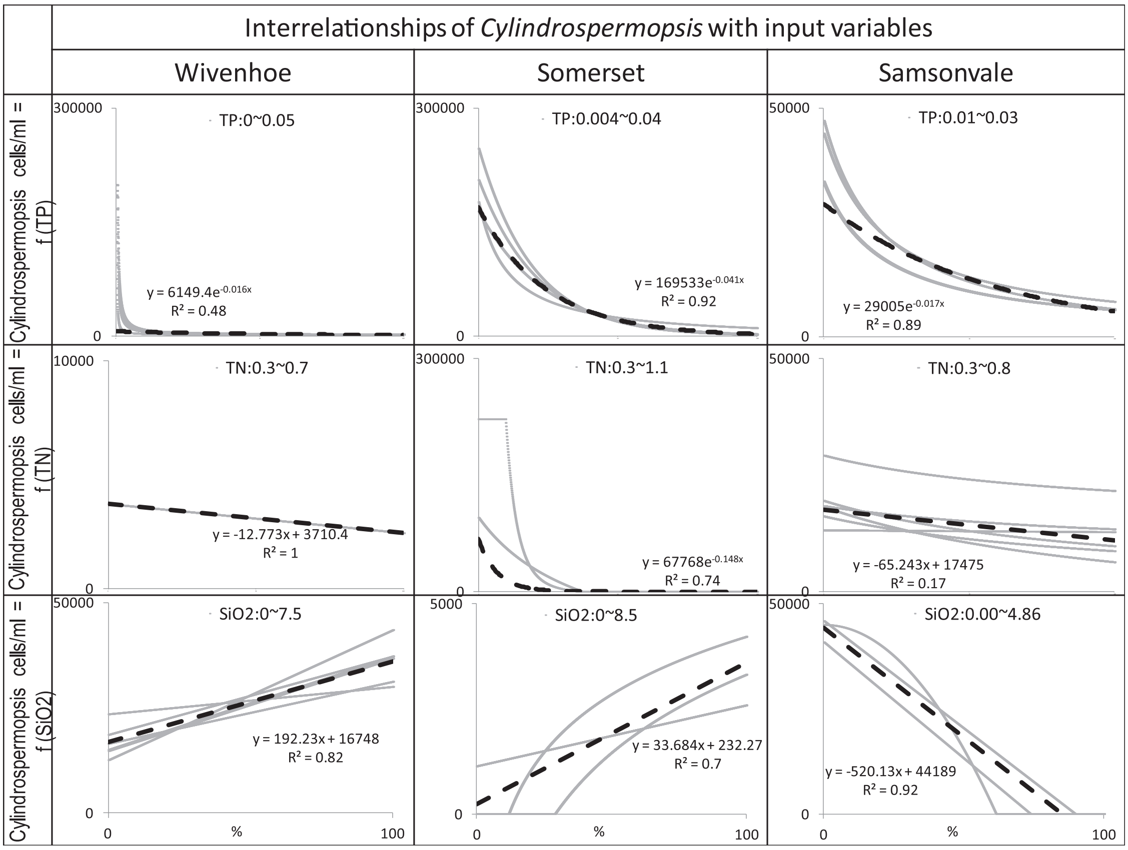


Fig. 9. Interrelationships between *C. raciborskii* and total phosphorus (TP) (top row), total nitrogen (TN) (middle row) and SiO_2 concentrations (bottom row) extracted from sensitivity analysis of 10 best performing models of the Lakes Wivenhoe (left column), Somerset (middle column) and Samsonvale (right column). Dashed lines indicate trend lines.

concentrations along somewhat lower EC ranges have also been discovered for Somerset and Samsonvale. Rising cell concentrations of *C. raciborskii* correspond with increasing pH in Wivenhoe and Somerset, whilst cell concentrations in Samsonvale stagnate within a comparable pH range (Fig. 8, bottom row).

Comparing responses of *C. raciborskii* to changing concentrations of TP and TN in Fig. 9 suggests that all three lakes experience highest cell concentrations when TP and TN concentrations are lowest, and lowest cell concentrations when nutrient concentrations are highest (Fig. 9, top and middle row). Since SiO_2 concentrations of Samsonvale were noticeably lower compared to the other two lakes (Table 1), Wivenhoe and Somerset experienced stimulating effects on cell concentrations with changing SiO_2 concentrations along the range between 2 and 8.5 mg L^{-1} but cell concentrations at Samsonvale showed declining effects along the range from 0 to 4.8 mg L^{-1} (Fig. 9, bottom row).

3.2. *Forecasting models for* C. raciborskii *in the Lake Wivenhoe based on online data from 2007 to 2010*

After demonstrating the suitability of electronically measurable water quality variables as predictor variables for 7-day-ahead forecasting in Fig. 5, online water quality monitoring data from September 2007 to December 2010 from Lake Wivenhoe was used for modelling and forecasting the timing and magnitude of *Cylindrospermopsis raciborskii* population dynamics using HEA. Fig. 10 shows results by including online measured Chl_*a* (left column) and excluding Chl_*a* (middle column) as input variable. In both cases the correlation is good with r^2 values greater than 0.65. Similarly good correlation has been achieved between measured and forecasted timing of maximum growth rate of cell concentrations, but better estimation of peak cell concentrations were obtained by excluding Chl_*a* as an input variable. Interestingly WT and EC thresholds for triggering excessive and diminished concentrations of *C. raciborskii* found by both models matches the thresholds for models of Wivenhoe based on historical data (left columns of Figs. 4 and 5). To learn more about lag times in the response of the *C. raciborskii* population to changing water quality conditions, a systematic screening experiment has been conducted testing all combinations of time lags between 0 and 10 days for the 5 online monitored input variables WQ, DO, pH, TURB and EC. The best forecasting results have been achieved by considering a 7 days delay between *C. raciborskii* and EC, and a 1 day delay between *C. raciborskii* and DO. However no significant forecasting improvements have been observed by time lags imposed on the remaining 3 input variables. Fig. 10 (right column) shows a slightly improved forecasting result being achieved by the lagged inputs of EC and DO in comparison with un-lagged inputs (Fig. 10, left and middle column). The threshold values of WT and EC for separating high and low abundances of *C. raciborskii* selected by the model including time-lags (Fig. 10, right column) match the previously discovered thresholds by models for the Lake Wivenhoe.

4. Discussion

This study has demonstrated the ability of evolutionary computation to perform inductive reasoning and forecasting through exploration of the rich information content of complex water quality monitoring data from aquatic ecosystems such as drinking water reservoirs. In particular it has shown the strength of evolutionary computation in: (1) predictive modelling of fast

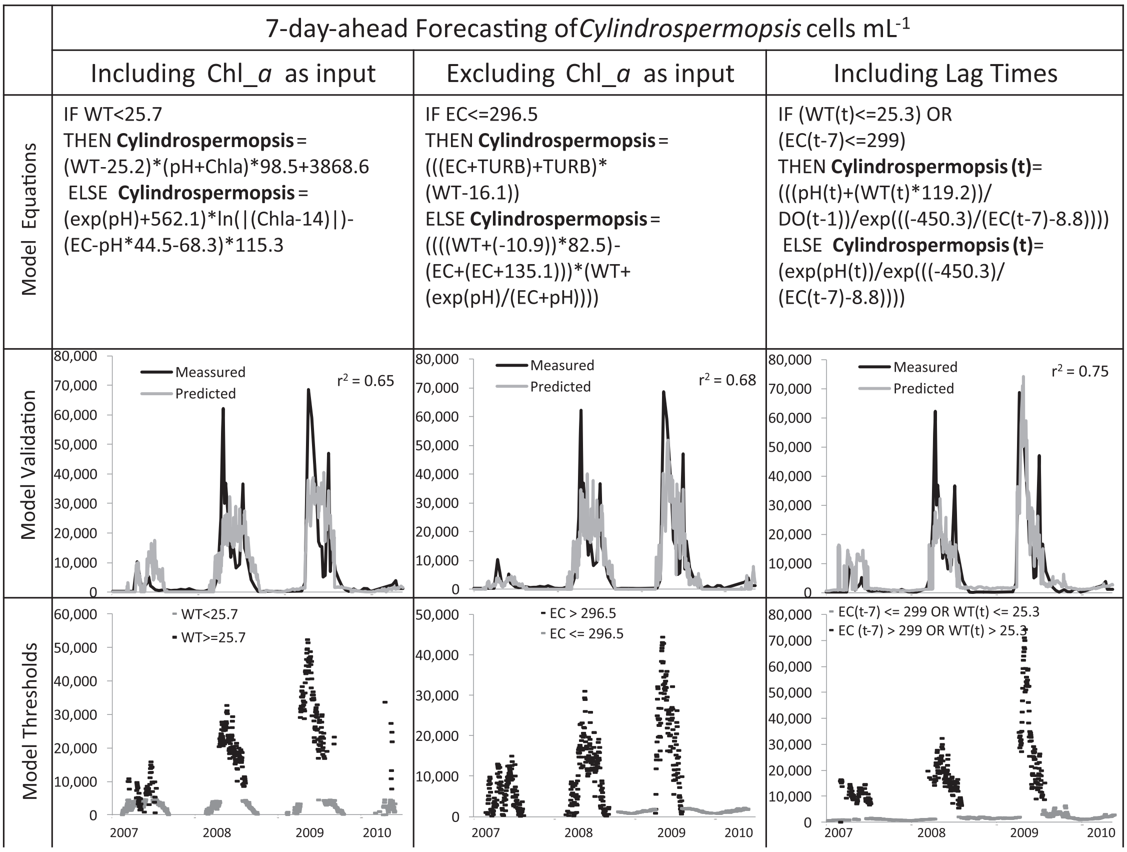


Fig. 10. 7-day-ahead forecasting of *C. raciborskii* in the Lake Wivenhoe utilising online water quality data including chlorophyll_a (left column) and excluding chlorophyll_a (middle column) and including time-lagged inputs (right column) documented by the model equations (upper row), model validation (middle row) and thresholds (bottom row).

processes such as population growth of *Cylindrospermopsis raciborskii* in response to multivariate environmental driving variables, and (2) discovering knowledge about synergies between environmental driving variables and population dynamics of *C. raciborskii.* A particular novel outcome of this research is the revealing of ecological thresholds that indicate alternative states of the three drinking water reservoirs: 'mass development or restrained development of *C. raciborskii*'. Both, ecological thresholds and interrelationships between water quality variables and population dynamics of *C. raciborskii* contribute to improved understanding and informed management of this species.

The research has shown that electrical conductivity, water temperature and turbidity are indicative predictors for development of *C. raciborskii* blooms at each of the three study sites. These findings have further been refined by assessing the relationship between *C. raciborskii* cell concentrations and key water quality variables extracted and generalised from the 10 best performing models by HEA. These results suggest that the optimum water temperature range for population growth of *C. raciborskii* is $25.6 \leq WT < 32.7$ °C which compares well with findings of Briand et al., 2002 who suggested an optimum temperature range of $25 \leq WT < 35$ °C based on laboratory experiments. The overarching effect of electrical conductivity on population growth of *C. raciborskii* seems to be inhibitory within the range of 140–206 $\mu S\,cm^{-1}$ and above 280 $\mu S\,cm^{-1}$ with the highest cell concentrations occurring when EC is between 206 and 280 $\mu S\,cm^{-1}$. The range of EC observed in the 11 data sets reflects impacts of sustained drought and large inflow events typical of the subtropical climate in south east Queensland.

Despite the finding that population growth of *C. raciborskii* decreases with increasing EC levels, results show that peak cell concentrations remain high. Therefore it can be concluded that it tolerates a broad range of salinity as previously suggested by Briand et al. (2002) and Moisander et al. (2002). Sensitivity results for Lakes Wivenhoe and Samsonvale suggest self-shading effects of *C. raciborskii* reflected by linear relationships between growing cell concentrations and increasing turbidity. Nevertheless the fact that highest cell concentrations occur at highest turbidity levels corresponds with findings by Shafik et al. (2001) that *C. raciborskii* can grow well in low underwater light conditions within the surface mixed layer.

Positive relationships have been discovered for the three lakes between growing cell concentrations of *C. raciborskii* and increasingly alkaline pH. This relationship may refer to the rapid CO_2 consumption during cyanobacteria bloom events (Shapiro, 1990) but also to the potential of cyanobacteria to use bicarbonate as their carbon source (Kaplan et al., 1991). As expected for mesotrophic lakes the nutrient pool of total phosphorus is exponentially diminished by *C. raciborskii* reflected by highest cell concentrations at lowest TP concentrations. However the relationship with total nitrogen appears less dynamic and may suggest that *C. raciborskii* does not solely rely on the internal TN pools of the lakes but utilises atmospheric N sources by nitrogen fixation (Bouvy et al., 2000; Moisander et al., 2008).

Linear positive relationships have been discovered between SiO_2 and total phosphorus and *C. raciborskii* for Wivenhoe and Somerset. These relationships may point at competitive exclusion between diatoms and *C. raciborskii* suggesting that in absence of diatoms the SiO_2 consumption is relatively low and therefore high SiO_2 concentration may correspond with high cell concentrations of *C. raciborskii*. However the linear inhibitory relationship discovered for Samsonvale suggests possible coinciding mass developments of diatoms and *C. raciborskii.*

5. Conclusions

Overall this study has demonstrated the capacity of the hybrid evolutionary algorithm HEA to extract and synthesise interrelated information from complex long-term monitoring data of aquatic ecosystems. HEA supports both, developing highly accurate models for short-term forecasting of population dynamics of fast growing microorganisms such as cyanobacteria and knowledge discovery on highly complex synergies between changing environmental and climate conditions and microorganisms.

More specifically this research has resulted in following outcomes:

1. Rule-based models developed by HEA for the three sub-tropical reservoirs provided accurate 7-day-ahead forecasts of population dynamics of *Cylindrospermopsis raciborskii* by using (i) all available water quality data as inputs, or (ii) by only using electronically measurable water quality data.
2. Rule-based models developed by HEA from on-line measured water quality data from Lake Wivenhoe performed highly accurate 7-day-ahead forecasts of population dynamics of *C. raciborskii*. These models can now be implemented and tested for real-time forecasting and early warning of cyanobacteria blooms.
3. Threshold conditions of EC and WT that trigger high cell concentrations of *C. raciborskii* as revealed by the models compared well between the three lakes and with findings in the literature.
4. Interrelationships between physical and chemical water quality variables and *C. raciborskii* revealed by sensitivity analyses compared well between the three lakes and with findings in the literature.

Acknowledgements

This research has been funded by the Australian Research Council (ARC-L0990453) in cooperation with the Queensland Bulk Water Supply Authority (Seqwater) and the South Australian Water Corporation. The authors thank anonymous reviewers for valuable comments that have significantly improved the manuscript. [SS]

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.

Branco, C.W.C., Senna, P.A.C., 1994. Factors influencing the development of *Cylindrospermopsis raciborskii* and *Microcystis aeruginosa* in the Paranoa Reservoir, Brasilia, Brazil. Algol. Stud. 75, 85–96.

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.

Calandrino, E.S., Perl, H.W., 2011. Determining the potential for the proliferation of the harmful cyanobacterium *Cylindrospermopsis raciborskii* in Currituck Sound, North Carolina. Harmful Algae 11, 1–9.

Caldwell, W.S., 2001. Hydrologic and Salinity Characteristics of Currituck Sound and Selected Tributaries in North Carolina and Virginia, 1998–99. U.S. Geological Survey, Water-resources Investigations Report 01-4097.

Cao, H., Recknagel, F., Welk, A., Kim, B., Takamura, N., 2006. Hybrid evolutionary algorithm for rule set discovery in time-series data to forecast and explain algal population dynamics in two lakes different in morphometry and eutrophication. In: Recknagel, F. (Ed.), Ecological Informatics. Scope, Techniques and Applications. second ed. Springer–Verlag Berlin, Heidelberg, New York, pp. 330–342.

Cao, H., Recknagel, F., Orr, P.T., 2013. Enhanced functionality of the redesigned hybrid evolutionary algorithm HEA demonstrated by predictive modelling of algal growth in the Wivenhoe Reservoir, Queensland (Australia). Ecol. Model. 252, 32–43.

Chan, W.S., Recknagel, F., Cao, H., Park, H.-D., 2007. Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. Water Res. 41, 2247–2255.

Chapman, A.D., Schelske, C.L., 1997. Recent appearance of *Cylindrospermopsis*(cyanobacteria) in five hypereutrophic Florida lakes. J. Phycol. 33, 191–195.

DeSouza, R.C.R., Carvalho, M.C., Truzzi, A.C., 1998. *Cylindrospermopsis raciborskii* (Wolosz.) Seenaya and Subba Raju (Cyanophyceae) dominance and a contribution to the knowledge of Rio Pequeno Arm, Billings Reservoir, Brazil. Environ. Toxicol. Water Qual. 13, 73–81.

Dokulil, M.T., Mayer, J., 1996. Population dynamics and photosynthetic rates of a *Cylindrospermopsis-Limnothrix* association in a highly eutrophic urban lake, Alte Donau, Vienna, Austria. Archiv. Fuer Hydrobiol. suppl. 117, 179–195.

Fastner, J., Ruecker, J., Stueken, A., Preusel, K., Nixdorf, B., Chorus, I., Koehler, A., Wiedner, C., 2007. Occurrence of the cyanobacterial toxin cylindrospermopsin in Germany. Environ. Toxicol. 22, 26–32.

Hamilton, P.B., Ley, L.M., Dean, S., Pick, F.R., 2005. The occurrence of the cyanobacterium *Cylindrospermopsis raciborskii* in Constance Lake: an exotic cyanoprokaryote new to Canada. Phycologia 44 (1) 17–25.

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.

Isvanovics, V., Shafik, H.M., Presing, M., Juhos, S., 2000. Growth and phosphate uptake kinetics of the cyanobacterium, *Cylindrospermopsis raciborskii* (Cyanophyceae) in throughflow cultures. Freshw. Biol. 43, 257–275.

Kaplan, A., Schwarz, R., Lieman-Hurwitz, J., Reinhold, L., 1991. Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria. Plant Physiol. 97, 851–855.

Kim, D.-K., Cao, H., Jeong, K.-S., Recknagel, F., Joo, G.-J., 2007. Predictive function and rules for population dynamics of *microcystis aeruginosa* in the regulated Nakdong River (South Korea), discovered by evolutionary algorithms. Ecol. Model. 203, 147–156.

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

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.

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.

Padisak, J., 1997. *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. Archiv. Fuer Hydrobiol. Suppl. 107, 563–593.

Recknagel, F., French, M., Harkonen, P., Yabunaka, K., 1997. Artificial neural network approach for modelling and prediction of algal blooms. Ecol. Model. 96, 1–3, 11–28.

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. Internat. Verein. Limnol. 30 (2) 191–197.

Recknagel, F., Ostrovsky, I., Cao, H., Zohary, T., Zhang, X., 2013. Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake Kinneret, Israel elucidated by evolutionary computation and wavelets. Ecol. Model. 225, 70–86.

Shafik, H.M., Herodek, S., Presing, M., Voros, L., 2001. Factors affecting growth and cell composition of cyanoprokaryote *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya and Subba Raju. Archiv. fuer Hydrobiol. Algol. Stud. 140, 75–93.

Shapiro, J., 1990. Current beliefs regarding dominance by blue–greens: the case for the importance of CO_2 and pH. Verh. Int. Verein. Limnol. 24, 38–54.

Sprober, P., Shafik, H.M., Presing, M., Kovacs, A.W., Herodek, S., 2003. Nitrogen uptake and fixation in the cyanobacterium *Cylindrospermopsis raciborskii* under different nitrogen conditions. Hydrobiologia 506–509, 169–174.

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

Wood, S.A., Stirling, D.J., 2003. First identification of the cylindrospermopsin producing cyanobacterium *Cylindrospermopsis raciborskii* in New Zealand.N.Z. J. Mar. Freshw. Res. 37, 821–828.