Adaptive agents for forecasting seasonal outbreaks of blue-green algal populations in lakes categorised by circulation type and trophic state

Friedrich Recknagel, Hongqing Cao, Carin van Ginkel, Dietrich van der Molen, Hodong Park and Noriko Takamura

Introduction

Adaptive agents for specific algal populations can be powerful tools for early warning and operational control of harmful algal blooms in lakes and drinking water reservoirs. One way of developing adaptive agents for algal populations is the extraction of generic rules from ecological time series data by means of evolutionary algorithms as suggested by Recknagel (2003).

This study aims to demonstrate that merged ecological data of categorised lakes allow evolutionary algorithms to induce rule-based agents that are generic for lake categories and facilitate short-term forecasting of blue-green algal blooms. Hybrid evolutionary algorithms (HEA) are applied in a rigorous k-fold cross-validation framework to develop agents for algal populations for categorised lakes. Three case studies are presented resulting in rule-based agents for forecasting 5- or 7-days-ahead abundances of Microcystis in the shallow-polymeric and hypertrophic lakes Kasumigaura and Suwa (Japan) as well as in the warm-monomictic and hypertrophic lakes Hartbeespoort, Roodeplaat, and Rietvlei (South Africa), and abundances of Oscillatoria in the shallow-polymeric and hypertrophic lakes Veluwemeer and Wolderwijd (The Netherlands).

Both explanatory and predictive validity of the resulting agents are tested by comparisons between predicted and measured data as well as sensitivity analyses.

Key words: adaptive agents, circulation type, ecological relationships, forecasting, hybrid evolutionary algorithms, k-fold cross-validation, lake categories, Microcystis, Oscillatoria, trophic state

Materials and methods

Study sites and data

The Japanese lakes Kasumigaura and Suwa are classified as being shallow-polymeric and hypertrophic. High internal and external nutrient loadings are typical for both lakes and cause recurrent summer blooms of Microcystis (Takamura et al. 1992, Park et al. 1998, Chan et al. 2007).

The Dutch lakes Veluwemeer and Wolderwijd are shallow-polymeric and hypertrophic, and susceptible to summer blooms of Oscillatoria agardhii (Reeders et al. 1998), even though great efforts have been undertaken to control eutrophication in both lakes by external nutrient control and biomplementation since the 1980s (Jagtman et al. 1992, Meier & Hesper 1997).

The 3 reservoirs Hartbeespoort, Rietvlei, and Roodeplaat are situated in highly populated areas between Johannesburg and Pretoria and are characterised by distinctive hypertrophic and warm-monomictic conditions, causing annual cyanobacteria blooms largely dominated by Microcystis aeruginosa (Zohary & Roberts 1989, Van Ginkel et al. 2000).

Because the measurement intervals of the raw data from the 7 water bodies were highly irregular, data was interpolated to create daily values as required for modelling by hybrid evolutionary algorithms.

Data for the 7 lakes were summarised (Table 1) and classified into 3 categories (Table 2). To develop generic agents by hybrid evolutionary algorithms data of same category lakes were merged.

Methods

Hybrid evolutionary algorithm (Cao et al. 2006) was designed to discover predictive rules in ecological time-series data. It combines genetic programming (GP; Koza 1992, 1994, Banzhaf et al. 1997) to generate and optimise the structure of rules, and genetic algorithms (GA; Holland 1975, Mitchell 1996) to optimise parameters of rules. Resulting rules are subsequently evaluated by means of fitness criteria where fitter rules are selected for recombination to create the next generation by using genetic operators such as crossover and mutation. These steps are iterated for consecutive generations until the termination criterion of the run has been satisfied and the fittest rule has been determined. A detailed description of the design and
Table 1. General properties and limnological variables of the lakes Kasumigaura and Suwa (Japan); Vehuwemeer and Wolderwijd (The Netherlands); and Hartbeespoort, Rooideplaat, and Rietvlei (South Africa).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General properties</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface area km²</td>
<td>220</td>
<td>13.3</td>
<td>32.4</td>
<td>36.7</td>
<td>20</td>
<td>3.97</td>
<td>7.58</td>
</tr>
<tr>
<td>Mean depth m</td>
<td>7</td>
<td>7.2</td>
<td>5.8</td>
<td>7.8</td>
<td>9.6</td>
<td>10.6</td>
<td>5.78</td>
</tr>
<tr>
<td>Circulation type</td>
<td>Shallow polymeric</td>
<td>Shallow polymeric</td>
<td>Shallow polymeric</td>
<td>Shallow polymeric</td>
<td>Warm monomictic</td>
<td>Warm monomictic</td>
<td>Warm monomictic</td>
</tr>
<tr>
<td><strong>Limnological variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water temp. WT (°C)</td>
<td>23 / 11.7 / 32</td>
<td>21.3 / 8.5 / 28</td>
<td>10.9 / 0 / 25.1</td>
<td>Mean / Min / Max</td>
<td>11.1 / 0 / 23.9</td>
<td>20.9 / 16.5 / 5.3 21</td>
<td>16.6 / 25.4</td>
</tr>
<tr>
<td>Secchi Depth SD (m)</td>
<td>0.62 / 0.25 / 2.05</td>
<td>0.94 / 0.31 / 1.9 / 0.4 / 0.1 / 1.7</td>
<td>0.4 / 0.2 / 1.3</td>
<td>1.5 / 0.5 / 2.5</td>
<td>1.9 / 0.9 / 2.9</td>
<td>1.9 / 1 / 2.8</td>
<td></td>
</tr>
<tr>
<td>Nitrate NO₃-N (mg/l)</td>
<td>2.97 / 0.01 / 23.1</td>
<td>2.27 / 0.01 / 12.3</td>
<td>0.8 / 0.001 / 5.77</td>
<td>0.24 / 0.001 / 7.24</td>
<td>1.4 / 0.6 / 2.2</td>
<td>1.8 / 1.2 / 2.4</td>
<td>1.1 / 0.1 / 2.1</td>
</tr>
<tr>
<td>Phosphate PO₄-P (mg/l)</td>
<td>0.29 / 0.1 / 2.35</td>
<td>0.1 / 0.04 / 0.001 / 0.42</td>
<td>0.01 / 0.001 / 0.05 / 0.16 / 0.035 / 0.578 / 0.001 / 0.12 / 0.11 / 0.28 / 1.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silica</td>
<td>2.46 / 0.01 / 592</td>
<td>19.7 / 0.1 / 19.7</td>
<td>19.7 / 0.1 / 19.7</td>
<td>19.7 / 0.1 / 19.7</td>
<td>19.7 / 0.1 / 19.7</td>
<td>19.7 / 0.1 / 19.7</td>
<td></td>
</tr>
<tr>
<td>SiO₂ (mg/l)</td>
<td>940 / 130 / 860 / 70 / 1610</td>
<td>115 / 9 / 459</td>
<td>101 / 9 / 265 / 46.3 / 0.1 / 37.3 / 0.1 / 65.1 / 54.8 / 0.1 / 105.9 / 160.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

functioning of HEA, including a demo software version, is provided by Cao et al. (2006).
To determine generic rule-based agents for each lake category, HEA is imbedded in a k-fold cross-validation framework (Kohavi 1995) based on k-fold data partitioning and the consecutive use of each part of the data for both training and validation.

Results and discussion

**Microcystis agent for shallow polymeric and hypertrophic lakes**

Twenty years of daily interpolated and merged data of the limnological variables water temperature (WT), dissolved oxygen (DO), Secchi depth (SD), pH, phosphate (PO₄-P), nitrate (NO₃-N), nitrogen to phosphorus ratio (N:P), and chlorophyll-a (Chla) of the lakes Kasumigaura and Suwa (Table 1) were used to develop a rule-based *Microcystis* agent using HEA by 10-fold cross-validation. To apply the agent for 7-days-ahead forecasting, a time lag of 7 days was imposed between input and output data. The resulting rule (Fig. 1) means that when the IF condition (Chla – WT) > p1 AND (PO₄-P < 15.92) is satisfied, the THEN branch calculates the *Microcystis* cell concentrations by *Microcystis* = WT * SD. Otherwise, the ELSE branch calculates the *Microcystis* cell concentrations by *Microcystis* = Chla * p2.

The causal relationships between *Microcystis* and the input variables Chla, WT, P, and SD were analysed by their sensitivity curves (Figs. 1a and 1b). The input sensitivity for the THEN branch (Fig. 1a) indicated that both WT and SD are positively related to high abundances of *Microcystis* (up to 1 200 000 cells/mL). Both a rising WT and an increased SD provide more photosynthetic active light to the water column, promoting the growth of *Microcystis* cells. However, the input sensitivity for the ELSE branch (Fig. 1b) shows positive relationships between Chla and WT and the lower abundances of *Microcystis* (< 200 000 cells/mL).

The *Microcystis* agent for shallow polymeric and hypertrophic lakes was developed from data that reflected different time periods of the 2 lakes. The data from Lake
Table 2. Lake ecosystem categories and related rule-based agents.

<table>
<thead>
<tr>
<th>Lake Ecosystem Categories</th>
<th>Lake Examples</th>
<th>Adaptive Agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shallow Polyvictic and Hypertrophic</td>
<td>Kasumigaura(Japan)</td>
<td>Microcystis</td>
</tr>
<tr>
<td></td>
<td>Suwa(Japan)</td>
<td></td>
</tr>
<tr>
<td>Shallow Polyvictic and Hypertrophic</td>
<td>Wolderwijd(The Netherlands)</td>
<td>Oscillatoria</td>
</tr>
<tr>
<td></td>
<td>Veluwemeer(The Netherlands)</td>
<td></td>
</tr>
<tr>
<td>Warm Monomictic and Hypertrophic</td>
<td>Hartbeespoort(South Africa)</td>
<td>Microcystis</td>
</tr>
<tr>
<td></td>
<td>Roodeplaat(South Africa)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rietvlei(South Africa)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Structure of the rule-based Microcystis agent for the lakes Kasumigaura and Suwa. (a) input sensitivity of the then-branch of the rule, (b) input sensitivity of the else-branch of the rule.

Kasumigaura were collected from 1984 to 1992, whereas data from Lake Suwa were collected from 1992 to 2002. Nevertheless, the agent produced good forecasting results for Microcystis abundances, achieving an $r^2 = 0.39$ for Kasumigaura and an $r^2 = 0.48$ for Suwa. The comparison between measured and forecasted data for both lakes shows that the agent was able to capture timing and magnitudes of distinct annual patterns of Microcystis abundances reasonably well for Lake Suwa, but underestimated some of the peak events and overestimated some of the valley events of Microcystis abundances in Lake Kasumigaura (Fig. 2). A possible explanation for the variable results for Kasumigaura might be that the evolutionary algorithm optimises the rule structure and parameters of the agents largely by focusing at extreme events in the training data that were mostly contained in the data of Lake Suwa.

Oscillatoria agent for shallow polyvictic and hypertrophic lakes

Twenty-eight years of measured limnological data of the Dutch lakes Wolderwijd and Veluwemeer (Table 1) were used to develop the Oscillatoria agent for shallow polyvictic and hypertrophic lakes. As a result, the rule set (Fig. 3) was discovered. The IF condition of the rule indicates that at SD = 0.35 m, the Oscillatoria abundance is calculated by Osc = (((SIo2*(P1)+(cos(NO2-N)*(P2/SD)))+(P3/SD)) but otherwise is calculated by Osc = (cos(SD)*P4/SD). High abundances of Oscillatoria coincide with low SD, possibly caused by shading from increased algal cell numbers in the water column, and therefore shading. These effects are reflected by the sensitivity analyses showing that SD decreases with increasing abundance of Oscillatoria.

In general, from the validation of the rule-based Oscillatoria agent for lakes Veluwemeer and Wolderwijd we can conclude that the model predicts the timing of peak
events very well for both lakes (Fig. 4). However, occasionally the agent overestimates magnitudes of years with relative low peaks during post-management years 1985 and 1993 and slightly underestimates magnitudes of years with relative high peaks in the period prior to 1979 (i.e., before the implementation of lake management measures). The prediction results for Lake Veluwemeer achieved an $r^2 = 0.63$ and for Lake Wolderwijd an $r^2 = 0.62$.

**Microcystis agent for warm monomictic and hypertrophic lakes**

Forty-two years of merged and daily interpolated data of the limnological variables of the lakes Roodelaan, Hartbeespoort, and Rietvlei (Table I) were used to develop a rule-based *Microcystis* agent using HEA by 14-fold cross-validation. The IF condition of the resulting performing rule indicates 2 alternative thresholds determined by either SD or by $PO_4$-P concentration (Fig. 5). If SD is $\leq 0.67$ or $PO_4$-P is $< 306.68 \mu g/L$, then the *Microcystis* biomass is calculated by the equation $Microcystis = \ldots$
Fig. 4. Validation of the rule-based agent for 5-days-ahead forecasting of *Oscillatoria* cell concentrations in lakes Veluwemeer and Wolderwijd.

![Graph showing validation results for two lakes.]

*Oscillatoria* 100 cells/mL
- measured
- forecasted for 5-days-ahead

Lake Veluwemeer
1979 to 1993 from January to December
$R^2 = 0.63$

Lake Wolderwijd
1979 to 1993 from January to December
$R^2 = 0.62$

Fig. 5. Structure and input sensitivity of the rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass in the Roodeplaat, Hartbeespoort, and Rietvlei reservoirs.

![Rule-based agent structure and input sensitivity graphs.]

Microcystis: $(T-p)/SD$
$p_1 + \ln\left(\frac{T}{17.52}\right) + \frac{T}{2} = 1.551$

$9.49 < p_1 < 15.92$

Microcystis: $(SD * (T-T_{\text{avg}})) / NO_2-N$
$p_2 = T_{\text{avg}} + 14.419$

$32.43 < p_2 < 56.91$

(WT-pl)/SD. However if SD is $< 0.67$ or PO$_4$-P is $\geq 606.68$ µg/L, then the *Microcystis* biomass is calculated by the equation *Microcystis* = (SD*p2-WT)*NO$_2$-N. This means that the rule distinguishes between conditions for low growth of *Microcystis* reflected by relatively high SD and relatively low phosphate concentrations, and conditions for high growth of *Microcystis* reflected by relatively low SD and relatively high phosphate concentrations.

Both cases make sense because high growth of *Microcystis* coincides with large amounts of buoyant cells and floating scum at the surface, which lowers the transparency of the water measured by SD. At the same time, fast growth of *Microcystis* is favoured by high phosphate concentrations and *vice versa*. It is also interesting that WT and SD were selected as input variables of the equation (1) for low *Microcystis* biomass in the THEN branch of the rule. The sensitivity analysis for WT and SD shows that the equation of the THEN branch of the rule calculates extremely low *Microcystis* biomasses $< 14$ cm$^3$/m$^3$, a coinciding event that is positively related to low ranges of water temperature and negatively related to high ranges of Secchi depth (Fig. 5a).

By contrast, NO$_2$-N concentrations, SD, and WT were selected as input variables for predicting high *Microcystis* biomass by the ELSE branch of the rule. The sensitiv-
Fig. 6. 14-fold cross-validation of the rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass in the Roodeplaat, Hartbeespoort, and Rietvlei lakes for 1991 to 2004.

...ity analysis shows the almost linear positive relationships between increasing *Microcystis* and increasing NO$_3$-N and WT (Fig. 5b). However, the sensitivity of *Microcystis* to changes in SD is almost neutral.

A strong visual correspondence is shown between the measured and forecasted concentrations of *Microcystis* biomass for most of the years of Hartbeespoort (Fig. 6). Generally, the model predicts the right timing of peak events but sometimes overestimates magnitudes of years with relative low peaks and underestimates magnitudes of years with relative high peaks, and achieves an overall $R^2 = 0.31$ for all 14 years of data.

The testing results for Roodeplaat can be similarly interpreted as for Hartbeespoort, where its $R^2$ value of 0.34 is slightly higher. While *Microcystis* concentrations in Roodeplaat and Hartbeespoort were observed in the same range of up to 150 μg/L, the concentrations observed in Rietvlei reached up to 300 μg/L. The testing of the rule-based agent for Rietvlei resulted in the best $R^2 = 0.76$ and showed good visual correspondence for years with relative low peak events but a distinctive underestimation of the highest peak event in 1995.

**Conclusions**

In the context of this study 2 concepts have been applied for the development of adaptive agents of algal populations: (1) rule discovery by means of hybrid evolutionary algorithms (HEA) and rigorous k-fold cross-validation, and (2) rule generalisation by means of merged time-series data of lakes belonging to the same lake category.

The adaptive agents that have been discovered as an outcome of this study proved to be both explanatory and predictive. We demonstrated that the interpretation of the rules can be brought into the context of empirical and causal knowledge on population dynamics of *Microcystis* and *Oscillatoria* under specific water quality conditions. The k-fold cross-validation of the agents based on measured data of each year of similar lakes belonging to the same category revealed good forecasting accuracy resulting in $R^2$ values between 0.39 and 0.63.

The study has shown that using merged ecological data of same category lakes is a promising way to generalise rule-based agents induced by hybrid evolutionary algorithms. Future research will focus on developing agent libraries for specific algal populations and lake categories applicable for early warning and operational control of algal blooms. Special attention will be given to the use of technologically measurable water quality data as input variables in order to allow the use of adaptive agents for real-time forecasting of algal blooms.

**References**


Authors’ addresses: F. Recknagel, H. Cao, School of Earth and Environmental Sciences, The University of Adelaide, 5005 Australia. E-mail: friedrich.recknagel@adelaide.edu.au

C. van Ginkel, Department of Water Affairs and Forestry, Pretoria, 0001 South Africa. E-mail: carin.vanginkel@gmail.com

D. van der Molen, Institute of Inland Water Management and Waste Water Treatment, 8200 AA Lelystad, The Netherlands. E-mail: d.tvdmolen@riza.rws.minven.nl

H. Park, Department of Environmental Sciences, Shinshu University, Matsumoto 390–8621, Japan.

N. Takamura, National Institute for Environmental Sciences, Tsukuba 305–0053, Japan.