Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation

Friedrich Recknagela,⁎, Hongqing Caoa, Bomchul Kimb,1, Noriko Takamurac,2, Amber Welka

aUniversity of Adelaide, School of Earth and Environmental Sciences, Adelaide, 5005, Australia
bKangwon University, Department of Environmental Sciences, Chunchon 200-701, South Korea
cNational Institute for Environmental Studies, Tsukuba 305-0053, Japan

ARTICLE INFO

Article history:
Received 25 November 2005
Received in revised form
31 January 2006
Accepted 15 February 2006

Keywords:
Recurrent supervised artificial neural networks
Non-supervised artificial neural networks
Hybrid evolutionary algorithms
Lake Kasumigaura
Lake Soyang
Cyanobacteria
Diatoms
Time series modelling
Ordination
Clustering
Forecasting

ABSTRACT

Precious ecological information extracted from limnological long-term time series advances the theory on functioning and evolution of freshwater ecosystems. This paper presents results of applications of artificial neural networks (ANN) and evolutionary algorithms (EA) for ordination, clustering, forecasting and rule discovery of complex limnological time-series data of two distinctively different lakes. Ten years of data of the shallow and hypertrophic Lake Kasumigaura (Japan) are utilized in comparison with 13 years of data of the deep and mesotrophic Lake Soyang (Korea). Results demonstrate the potential that: (1) recurrent supervised ANN and EA facilitate 1-week-ahead forecasting of outbreaks of harmful algae or water quality changes, (2) EA discover explanatory rule sets for timing and abundance of harmful outbreaks algal populations, and (3) non-supervised ANN provide clusters to unravel ecological relationships regarding seasons, water quality ranges and long-term environmental changes.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Continuing in situ measurements of limnological, hydrological and climate variables of lakes and rivers retain most momentous information about complex ecological relationships between plankton populations, physical and chemical water properties, as well as climate and environmental changes over time. The extraction of this precious ecological information from long-term time series data advances the theory on functioning and evolution of freshwater ecosystems. Recent
that were limited in coping with the multiple non-linear nature of data. By contrast data ordination and clustering by non-supervised ANN (Kohonen, 1989, 1995) proves to be applicable to highly complex and non-linear data including limnological time-series (e.g. Chon et al., 1996; Recknagel et al., 2006).

The results show that both recurrent supervised ANN and hybrid EA allow 7-days-ahead forecasts of seasonal succession and abundances of blue-green algae and diatoms in quite distinctive lakes with reasonable accuracy. The sensitivity analysis of the predictive rule sets for four algal species discovered by hybrid EA provided interesting insights into algal specific relationships with physical and chemical water quality properties. The sensitivity curves from supervised ANN complemented well the ordination and clustering of the two algal populations regarding their temperature, nitrogen, phosphorus preferences as well as pH tolerances by means of non-supervised ANN. These results revealed from data corresponded well with related hypotheses postulated by Reynolds (1984).

The pattern analysis of periods with distinctively different water quality conditions of the two lakes by non-supervised ANN has discovered behaviours of phyto- and zooplankton likely in response to according management efforts.

2. Materials and methods

2.1. Study sites and data

Lake Kasumigaura is situated in the southeastern part of Japan and receives flow from 56 rivers and streams. Its catchment area of 2135 km² consists of paddy areas but is largely urbanised and industrialised. The lake was turned from a temperate climate with monsoonal rain in mid-summer. Using real lake data the present research aimed at: (1) forecasting outbreaks and seasonal succession of blue green algae and diatom populations in both lakes by means of recurrent supervised ANN and hybrid EA; (2) determining relationships between blue green algae and diatom populations and physical and chemical lake properties by means of rule sets discovered by hybrid EA and sensitivity analysis carried out by recurrent supervised ANN; (3) analysing complex interactions between algal populations, seasons and inter-annual water quality changes by means of non-supervised ANN. Results from the sensitivity analysis by supervised ANN were brought into a context with data ordination and clustering by non-supervised ANN in order to test hypotheses on complex interactions of algal populations with environmental conditions as largely postulated by Reynolds (1984). The terms supervised or non-supervised respectively mean in this context that the learning algorithm of an artificial neural network is either guided by known output patterns or is not guided by known outputs but learns the patterns from features of the inputs.

Ecological time-series data of lakes have previously been ordered and clustered by conventional multivariate statistics (e.g. Varis et al., 1989; Varis, 1991; Van Tongeren et al., 1992)...

### Table 1 – General characteristics of Lake Kasumigaura and Lake Soyang

<table>
<thead>
<tr>
<th></th>
<th>Lake Kasumigaura</th>
<th>Lake Soyang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface area, km²</td>
<td>219.9</td>
<td>45</td>
</tr>
<tr>
<td>Maximum volume, km³</td>
<td>662</td>
<td>2900</td>
</tr>
<tr>
<td>Maximum depth, m</td>
<td>7</td>
<td>110</td>
</tr>
<tr>
<td>Mean depth, m</td>
<td>3.9</td>
<td>42</td>
</tr>
<tr>
<td>Water residence time, years</td>
<td>0.55</td>
<td>0.7</td>
</tr>
<tr>
<td>Catchment area, km²</td>
<td>1597</td>
<td>2675</td>
</tr>
<tr>
<td>Circulation type</td>
<td>Non-stratified</td>
<td>Warm</td>
</tr>
</tbody>
</table>

### Table 2 – Limnological properties reflected by the databases of Lake Kasumigaura and Lake Soyang

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Min/Max</td>
<td>Mean/Min/Max</td>
<td></td>
</tr>
<tr>
<td>PO₄, µg/l</td>
<td>14.44/1/235</td>
<td>3.4/0.15/20</td>
</tr>
<tr>
<td>NO₃, mg/l</td>
<td>0.52/0.001/2.38</td>
<td>1.02/0.3/2.2</td>
</tr>
<tr>
<td>Si, mg/l</td>
<td>3.66/0.015/12.49</td>
<td></td>
</tr>
<tr>
<td>Chl-a, µg/l</td>
<td>73.05/0.69/279.5</td>
<td>3.7/0.4/45.1</td>
</tr>
<tr>
<td>DO, mg/l</td>
<td>10.23/4.88/18.21</td>
<td>9.36/5.3/13.1</td>
</tr>
<tr>
<td>Turbidity NTU</td>
<td>1.36/0.5/10.35</td>
<td></td>
</tr>
<tr>
<td>Secchi depth, m</td>
<td>0.87/0.28/3.8</td>
<td>4.12/0.7/10</td>
</tr>
<tr>
<td>pH</td>
<td>8.75/7.12/10.13</td>
<td>7.36/2.9/9.1</td>
</tr>
<tr>
<td>Water temperature, °C</td>
<td>16.48/2.1/32</td>
<td>15.2/4.6/29.5</td>
</tr>
<tr>
<td>Phytoplankton cells/ml</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anabaena</td>
<td>6008/1/112112</td>
<td>465/1/17000</td>
</tr>
<tr>
<td>Oscillatoria</td>
<td>20160/1/502302</td>
<td></td>
</tr>
<tr>
<td>Microcystis</td>
<td>38563/1/644117</td>
<td>162/1/9130</td>
</tr>
<tr>
<td>Cyclotella</td>
<td>5160/1/75420</td>
<td></td>
</tr>
<tr>
<td>Asterionella</td>
<td></td>
<td>670/1/20600</td>
</tr>
<tr>
<td>Zooplankton ind./l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cladocera</td>
<td>170/1/2446</td>
<td>4/1/56</td>
</tr>
<tr>
<td>Copepoda</td>
<td>156/1/640</td>
<td>7/1/80</td>
</tr>
<tr>
<td>Rotifera</td>
<td>229/1/2542</td>
<td>32/1/47</td>
</tr>
</tbody>
</table>
brackish into a freshwater lake 5 years after a floodgate to the Pacific Ocean was implemented in 1963.

Lake Soyang is situated in the northeastern part of South Korea and fed by the Soyang River contributing 90% of the inflowing water. Nutrient loadings to Lake Soyang are predominantly caused by non-point sources from paddy and forest areas, and temporarily by in-lake fish farming using net cages. Table 1 summarises characteristics of the two lakes. Table 2 provides details of the limnological databases of the two lakes.

Data of Lake Kasumigaura were collected with a column sampler of 2 m at the centre of the Takahamairi Bay. Data of Lake Soyang were collected in meter steps at the central station and averaged over the upper 10 m for the present study. As the measurement intervals of the raw data from both lakes were highly irregular and sampling dates different for physical, chemical and biological data the data was interpolated to create consistent daily values as required for the development of ANN models.

### 2.2. Methods

The ecological time-series of Lake Kasumigaura and Lake Soyang underwent pattern analysis and predictive modelling by the integrated application of three well-established bio-inspired computational techniques: recurrent supervised ANN, non-supervised ANN and hybrid EA (see Fig. 1). The integrated approach in Fig. 1 allowed: (1) to gain better understanding of ecological relationships in lakes by interpreting results from the pattern analysis by non-supervised ANN and the sensitivity analysis by recurrent supervised ANN.

---

**Fig. 1 - Integrated application of recurrent supervised ANN, non-supervised ANN and hybrid EA for pattern analysis and predictive modeling of ecological time-series data.**
in a close context, and (2) to determine best performing forecasting models for outbreaks of algal populations from alternative results by supervised ANN and hybrid EA, and their representation by non-linear rule sets.

2.2.1. Super- and non-supervised artificial neural networks
A recurrent supervised ANN (Fig. 2) was applied to predict 7-days-ahead seasonal succession of *Microcystis* and *Cyclotella* for Lake Kasumigaura, and of *Anabaena* and *Asterionella* for Lake Soyang. These pairs of algal data were selected from each lake according to most dominant species of the functional groups cyanobacteria and diatoms. Recurrent supervised ANN were introduced by Pineda (1987) mimicking the principles of deterministic modelling by ordinary differential equations. They consider both, current external inputs as well as feedback inputs of copied neuron weights at time $t-1$ in order to determine current weights of neurons at time t. They prove to be very efficient for time series modelling (e.g. Walter et al., 2001; Jeong et al., 2006).

For Lake Kasumigaura the ANN were trained with daily input values for PO$_4$, NO$_3$, Si, Secchi depth, Chl-$\alpha$, pH and water temperature and daily output values for *Microcystis* and *Cyclotella* of the years 1984 to 1985 and 1987 to 1993. The prediction results for 1986 were validated with independent daily input and output values not used for ANN training, and assessed by the mean square errors (MSE). A comprehensive sensitivity analysis was conducted by means of the recurrent supervised ANN to discover relationships between the input variables and the *Microcystis* and *Cyclotella* populations.

For Lake Soyang the ANN were trained with daily input values for PO$_4$, NO$_3$, Secchi depth, turbidity, pH and water temperature and daily output values for *Anabaena* and *Asterionella* of the years 1990 to 1996 and 1998 to 2000. The prediction results for 1997 were validated with independent daily input and output values not used for ANN training, and assessed by the $r^2$ values of the linear equation without intercept. A comprehensive sensitivity analysis was conducted by means of the recurrent supervised ANN to discover relationships between the input variables and the *Anabaena* and *Asterionella* populations.

A non-supervised ANN as introduced by Kohonen (1989) was applied according to Kohonen (1995) to ordinate, cluster and map water quality and phytoplankton data of both lakes with respect to seasons and ranges of nutrients, pH and water temperature conditions (see Fig. 3). In the context of the
present study it was applied to ordination, clustering and mapping of water quality and phytoplankton data with respect to seasons and eutrophication management according to Kohonen (1995). The principal approach is represented in a simplified manner in Fig. 3. It shows that the neurons of the non-supervised ANN learn to distinguish between similar and dissimilar features of the normalised input data, which can be mapped as clustered inputs. The term non-supervised in this context means that the learning algorithm is not guided by known output patterns but learns the patterns from features of the inputs. Those features are expressed by Euclidian distances, which are calculated between the inputs and can be visualised as unified distance matrix (U-matrix) and as partitioned map (K-means). Fig. 4 shows the seasonal clusters for Lake Kasumigaura as mapped according to Table 3 by the U-matrix and K-means partitioning using the SOM Toolbox of MATLAB 6.5.1 (Vesanto, 1999). The U-matrix map (Fig. 4a) visualises the relative distances between neighbouring data of the input data space as shades of grey. The light areas in the U-matrix visualise neighbouring data with smallest distances belonging to a region or cluster. The black colours represent the biggest distances between neighbouring data and denote borders between clusters. The K-means algorithm partitions the input data space into a specified number of clusters based on the U-matrix. Fig. 4b represents the corresponding partitioned map for the 5 seasons defined in Table 3.

![Fig. 3 - Structure of the non-supervised ANN.](image)

![Fig. 4 - Ordination and clustering of seasons of Lake Kasumigaura as defined in Table 3 by means of non-supervised ANN and visualised as unified distance matrix map (U-matrix) (a), and as partitioned map (K-means) (b).](image)
Table 3 summarises the criteria for ordination and clustering of time-series data of Lake Kasumigaura and Lake Soyang by means of the non-supervised ANN.

<table>
<thead>
<tr>
<th>Classification criteria</th>
<th>Lake Kasumigaura</th>
<th>Lake Soyang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasons:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>15th March to 30th May</td>
<td>15th March to 30th May</td>
</tr>
<tr>
<td>Early</td>
<td>1st June to 30th July</td>
<td>1st June to 30th July</td>
</tr>
<tr>
<td>Summer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Summer</td>
<td>1st August to 30th</td>
<td>1st August to 30th</td>
</tr>
<tr>
<td>Autumn</td>
<td>September to 30th</td>
<td>September to 30th</td>
</tr>
<tr>
<td>Winter</td>
<td>1st December to 14th March</td>
<td>1st December to 14th March</td>
</tr>
<tr>
<td>Water quality ranges:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO4-P</td>
<td>&lt;5; 5 and &lt;25; &gt;25</td>
<td>&lt;5; 5 and &lt;25; &gt;25</td>
</tr>
<tr>
<td>NO3-N</td>
<td>&lt;0.5; 0.5 and &lt;1; &gt;=1</td>
<td>&lt;1; 1 and &lt;1.5; &gt;=1.5</td>
</tr>
<tr>
<td>pH</td>
<td>&lt;7.5; 7.5 and &lt;8.5; &gt;=8.5</td>
<td>&lt;7; 7 and &lt;8.5; &gt;=8.5</td>
</tr>
<tr>
<td>Secchi depth</td>
<td>&lt;0.75; 0.75 and &lt;1.5; &gt;=1.5</td>
<td>&lt;0.75; 0.75 and &lt;1.5; &gt;=1.5</td>
</tr>
<tr>
<td>Water temperature</td>
<td>&lt;15; 15 and &lt;20; &gt;=20</td>
<td>&lt;15; 15 and &lt;20; &gt;=20</td>
</tr>
</tbody>
</table>

Fig. 5 – Conceptual diagram of the hybrid evolutionary algorithm (HEA) for the discovery of predictive rule sets in water quality time-series.

3. Results

3.1. Forecasting and explanation of seasonal algal abundances and succession

The recurrent supervised ANN (Fig. 2) were specifically designed and trained for the two lakes in order to forecast abundances of representative blue-green algae and diatom attributes as self-organization, self-learning, intrinsic parallelism, generality, and have been successfully applied to pattern recognition, optimum control and parallel processing (Goldberg, 1989; Bäck et al., 1997).
populations for 7-days-ahead. Fig. 6 shows forecasting results for Microcystis (Fig. 6, top, left) and Cyclotella (Fig. 6, bottom, left) of the testing year 1986 of Lake Kasumigaura. Whilst the predicted timings and magnitudes of the summer peak of Microcystis corresponded well with the measured data ($r^2=0.93$; $r^2=0.97$), the ANN predicted the spring peak of Cyclotella actually with a realistic magnitude but 3 weeks too early ($r^2=0.26$). The ANN also forecasted slight summer and autumn peaks of Cyclotella even though only a late autumn peak was observed in 1986. The EA predicted a realistic timing of both spring and autumn peaks of Cyclotella but underestimated the magnitude of the spring peak by 35% ($r^2=0.59$).

The forecasting results of the summer peak of Anabaena (Fig. 6, top, right) for Lake Soyang were reasonable regarding its timing but overestimated by the ANN ($r^2=0.49$) and underestimated by the EA ($r^2=0.84$) within the observation error. However the predicted timings of two spring peaks of Asterionella (Fig. 6, bottom, right) for Lake Soyang were several weeks too early and partially overestimated by the ANN ($r^2=0.49$) but matched well the observed timing with a partial overestimation by the EA ($r^2=0.68$).

From the above results it can be concluded that the forecasting of the diatom populations Cyclotella and Asterionella is more challenging compared to the blue-green algal populations Microcystis and Anabaena. This finding is also reflected by the complexity of rule sets discovered by HEA (see Table 4) that is higher for diatoms than for blue-green algae. While rule set 1 for Microcystis and rule set 3 for Anabaena in Table 4 are relatively simple consisting of one single rule and their IF conditions are associated with only one input variable respectively, the rule set 2 for Cyclotella and rule set 4 for Asterionella have a much more complicated structure consisting of 4 and 3 IF condition branches respectively. Fig. 7 visualises the activations of rule set 1 for Microcystis (Fig. 7a) and rule set 3 for Anabaena (Fig. 7b) respectively. Fig. 6 visualises the activations of rule set 1 for Microcystis (Fig. 7a) and rule set 3 for Anabaena (Fig. 7b). While highest abundances of Microcystis in Lake Kasumigaura are triggered by the threshold value $P_0\geq 52.11$ $\text{µg/l}$, highest abundances of Anabaena in Lake Soyang are triggered by the water temperature $W_T\geq 20.66$.

3.2. Relationships between algal abundances and water quality conditions

Sensitivity analyses by means of the recurrent supervised ANN (Fig. 2) as well as ordination and clustering by means of non-supervised ANN (Fig. 3) were carried out based on data from 1984 to 1993 of Lake Kasumigaura, and data from 1988 to 2000 of Lake Soyang in order to study complex relationships between blue-green algae and diatom populations, and water quality conditions (see Table 3).
Table 4 - The best performing algal specific rule sets discovered by HEA in terms of the minimal testing error obtained in 100 runs with 7-days-lagged input data

<table>
<thead>
<tr>
<th>Algal population</th>
<th>Best rule sets</th>
<th>Training error</th>
<th>Testing error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microcystis</td>
<td>IF (PO4&lt;52.11) THEN Microcystis=WT*(WT−14.05)+PO4<em>WT ELSE Microcystis=Chl−a</em>(pH/15.32+WT−152.37)</td>
<td>512.95</td>
<td>392.47</td>
</tr>
<tr>
<td>Cyclotella</td>
<td>IF (PO4≤35.71 AND Si≥exp (PO4)) THEN Cyclotella=3.67 ELSE IF (WT≤16.36) THEN Cyclotella=335.95 ELSE IF (Po4≥85.04) THEN Cyclotella=exp(pH/pH+exp (pH/3.96) ELSE IF (Si*exp(WT)≥74.60) THEN Cyclotella=exp(pH/pH) ELSE Cyclotella=exp(pH/pH+exp (pH/3.96)</td>
<td>923.17</td>
<td>863.07</td>
</tr>
<tr>
<td>Anabaena</td>
<td>IF (WT&gt;20.66) THEN Anabaena=(Turb<em>70.73−274.1)<em>exp(11.93</em>ln(Chl)/pH) ELSE Anabaena=37.27</em>Chl−a*Chl/pH</td>
<td>193.14</td>
<td>73.57</td>
</tr>
<tr>
<td>Asterionella</td>
<td>IF (exp(S.D.)≥55.95) OR (WT&gt;17.30) THEN Asterionella=18.34 ELSE IF (WT&gt;12.88) THEN (PO4≥4.7) THEN Asterionella=120.56 ELSE IF (S.D.<em>exp(S.D.)&lt;37.65) THEN Asterionella=Turb</em>exp (pH/pH)<em>Chla ELSE Asterionella=Turb</em>Chl−a*59.83</td>
<td>115.47</td>
<td>67.71</td>
</tr>
</tbody>
</table>

Fig. 8 illustrates relationships between water temperature and Microcystis and Cyclotella in Lake Kasumigaura, and Fig. 9 between water temperature and Anabaena, Microcystis and Asterionella in Lake Soyang. These figures clearly show that in both lakes blue-green algae Microcystis and Anabaena reach their highest abundances at temperatures higher than 20 °C but diatoms Cyclotella and Asterionella at temperatures below 16 °C. Fig. 10 represents relationships between pH and Microcystis and Cyclotella in Lake Kasumigaura, and Fig. 11 relationships between pH and Anabaena and Asterionella in Lake Soyang. Even though the pH ranges (Table 2) indicate distinctive alkaline conditions in Lake Kasumigaura but neutral to alkaline conditions in Lake Soyang, the results in Figs. 10 and 11 demonstrate the coincidence of high abundances of blue green algae and high pH values of 8 to 9 in Lake Kasumigaura and 7 to 8.5 in Lake Soyang. By contrast the diatom Cyclotella in Lake Kasumigaura peaks at pH of 9 to 9.3 whilst the diatom Asterionella in Lake Soyang seems to occur at neutral to slightly acidic conditions. Figs. 12 and 13 reflect preferences of diatoms to higher concentrations of NO3–N in both lakes but tolerance of blue-green algae to relatively low concentrations of NO3–N. The relationships of algal groups to PO4–P concentrations show the opposite trend compared to NO3–N for both lakes. The diatoms Cyclotella in Lake Kasumigaura (Fig. 14) and Asterionella in Lake Soyang (Fig. 15) reach highest abundances at relative low PO4–P concentrations. However the blue-green algae Microcystis in Lake Kasumigaura (Fig. 14) and Anabaena in Lake Soyang (Fig. 15) show a distinct preference for highest PO4–P concentrations.

3.3. Relationships between algal abundances, seasons and long-term water quality changes

Ordination and clustering by means of non-supervised ANN (Fig. 3) were carried out based on data from 1984 to 1986 and from 1987 to 1989 because of relative NO3–N deficiency, 1987 to 1989 because of relative NO3–P sufficiency but PO4–P deficiency. The period from 1992 to 1993 of Lake Soyang was selected to reflect conditions of relative PO4–P sufficiency but NO3–N deficiency, and 1993 to 1998 to 1999 of Lake Soyang in order to study complex relationships between functional algal groups, seasons (see Table 3) and water quality changes. The period from 1984 to 1986 of Lake Kasumigaura was selected to reflect conditions of relative PO4–P sufficiency but NO3–N deficiency, and 1992 to 1993 of Lake Soyang was selected to reflect conditions of relative PO4–P sufficiency but NO3–N deficiency as a result of intensive fish farming, and 1998 to 1999 because of relative NO3–N sufficiency but PO4–P deficiency as a result of terminated fish farming.

Fig. 16 shows the seasonal abundance clusters of the green algae Scenedesmus and the diatom Cyclotella for two periods with distinctive nutrient conditions in Lake Kasumigaura. The results suggest that both algae experience a shift of their predominant occurrence from winter in the period 1984 to 1986 to spring and early summer in the period 1987 to 1989 with two times higher abundances in the second period. Fig. 17 shows patterns of two blue-green algae where the highest abundance of Microcystis shifts from late summer to autumn and declines by 50% between periods 1 and 2. A different trend can be observed in Fig. 17 for Oscillatoria that shifts its dominance from spring in period 1 to late summer in period 2 by tripling its abundance.

Fig. 18 shows that abundances of both cladocera and copepoda for the two periods of Lake Kasumigaura are highest in late summer and autumn with slightly increased numbers in period 2. The seasonal patterns of nutrient concentrations in Lake Kasumigaura in Fig. 19 show a 30% increase of NO3–N from period 1 to 2 but at the same time a 30% decrease of PO4–P. While NO3–N peaks in winter in period 1 it peaks in spring and early summer in period 2. By contrast PO4–P peaks in early and late summer in period 1 but peaks in winter in period 2.

Fig. 20 shows seasonal abundance patterns of Anabaena and Asterionella during and after intensive fish farming in Lake Soyang. It indicates that the abundance of Anabaena peaked in early and late summer during fish farming but peaked only in late summer and decreased by 50% after fish farming. By contrast Asterionella became most abundant in autumn during fish farming but peaked with in spring after fish farming at tenfold higher abundance.

The seasonal patterns of NO3–N and PO4–P were visualised in Fig. 20 for the two different periods of management of lake Soyang. The results demonstrate that NO3–N concentrations were highest in winter and spring during
fish farming but shifted to spring and early summer with 30% increased maxima. However, highest PO$_4$–P concentrations occurred in winter and early summer during fish farming but only in winter at 30% lower magnitude after fish farming.

4. Discussion

4.1. Forecasting and explanation of seasonal algal abundances and succession

The results in Fig. 6 have demonstrated that recurrent supervised ANN and evolutionary algorithms EA have the capacity for forecasting outbreaks and seasonal succession of different algal populations in distinctive lakes for 1 week in advance. The blue-green algae *Microcystis* which were observed to peak at 650,000 cells/ml in mid-August 1986 in the shallow hypertrophic Lake Kasumigaura, were forecasted by the EA to peak at 600,000 cells/ml with the right timing, but were underestimated by the ANN with a magnitude of 550,000 cells/ml for late-August. The blue-green algae *Anabaena* which were observed to reach their highest abundance of 5000 cells/ml in late-August 1997 in the deep mesotrophic Lake Soyang were forecasted by the EA to peak at 3500 cells/ml and by the ANN to peak at 6000 cells/ml both in early-September. The diatom *Cyclotella* that were observed with a maximum of 55,000 cells/ml in mid-April 1986 in Lake Kasumigaura were forecasted by the EA to peak at 30,000 cells/ml with the right

Fig. 7 – Visualisation of the rule set activation of (a) *Microcystis* for Lake Kasumigaura in 1986, and (b) *Anabaena* for Lake Soyang in 1997 using the rule sets shown in Table 4.

Fig. 8 – Clustering of *Microcystis* and *Cyclotella* abundances regarding temperature classes in Lake Kasumigaura using non-supervised ANN (top) and sensitivity curves of *Microcystis* and *Cyclotella* abundances over the temperature range of Lake Kasumigaura using supervised ANN (bottom).
timing, but were forecasted by the ANN with 60,000 cells/ml in mid March. Finally Asterionella that were observed to peak at 4000 cells/ml in mid May 1997 in Lake Soyang were forecasted by the EA to peak at the right time with 6000 cells/ml and were forecasted by the ANN but with 6000 cells/ml in late April. The slightly better forecasting performance of EA compared with ANN compares well with findings of Recknagel et al. (2002).

Overall the results are encouraging towards the development of early warning systems for blue-green algal blooms based on real time forecasting by EA and supervised ANN. In view of the extreme non-linear and rapid population dynamics of freshwater algae both modelling techniques have demonstrated their capability to forecast the right seasonality of outbreaks of blue green algae and diatoms in these distinctively different freshwater lakes. Forecasting results for algal abundances with the above-demonstrated accuracy may prove to be reasonable for early warning systems, which utilise real-time monitoring data. Such a system has been successfully tested for forecasting chlorophyll-a of a coastal bay (Muttil and Lee, 2006) by means of on-line electronically measurable variables such as water temperature, dissolved oxygen, and solar radiation.

4.2. Relationships between algal abundances and water quality conditions

The results in Figs. 8–15 have shown that ordination and clustering by non-supervised ANN and sensitivity analyses by supervised ANN can be integrated to a powerful tool for analysing complex ecological relationships in data. It has revealed from data that blue-green algae and diatoms have distinctive relationships with water temperature, pH, NO₃-N- and PO₄-P-concentrations despite differences in the trophic state and morphometry of a lake. For both lakes clusters of the automatically mapped blue-green algae abundances corresponded well with the sensitivity curves indicating fastest growth at water temperatures higher than 20 °C. By contrast clusters of the automatically mapped diatom abundances corresponded well with the sensitivity curves indicating fastest growth at water temperatures below 15 °C namely between 9 and 12 °C. These results comply with the temperature preferences postulated for specific algal assemblages (e.g. Reynolds, 1984; Shapiro, 1990), and discovered for Microcystis and Stephanodiscus in River Nakdong by Jeong et al. (2006). With regards to pH Microcystis in Lake Kasumigaura was mapped in the range between 8 and 9, and mapped together with Anabaena in Lake Soyang in the range between 7 and 8.5. As the pH of freshwater is determined by its CO₂ budget (Stumm and Morgan, 1970) alkaline conditions are likely for a hypertrophic lake such as Kasumigaura as the availability of dissolved CO₂ can be seasonally limited when the primary productivity is highest (Schindler, 1971). This may explain the slight upward shift of the pH range at which blue green algae predominately occur in Lake Kasumigaura compared to Lake Soyang. However Figs. 10 and 11 clearly show that high abundances of blue-green algae in both lakes coincide with distinct alkaline conditions. Relationships between diatoms and pH revealed by Figs. 10 and 11 show that high abundances...
Fig. 10 – Clustering of *Microcystis* and *Cyclotella* abundances regarding pH classes in Lake Kasumigaura using non-supervised ANN (top) and sensitivity curves of *Microcystis* and *Cyclotella* abundances over the pH range of Lake Kasumigaura using supervised ANN (bottom).

Fig. 11 – Clustering of *Anabaena* and *Asterionella* abundances regarding pH classes in Lake Soyang using non-supervised ANN (top) and sensitivity curves of *Anabaena* and *Asterionella* abundances over the pH range of Lake Soyang using supervised ANN (bottom).
Fig. 12 – Clustering of Microcystis and Cyclotella abundances regarding NO₃⁻N classes in Lake Kasumigaura using non-supervised ANN (top) and sensitivity curves of Microcystis and Cyclotella abundances over the NO₃⁻N range of Lake Kasumigaura using supervised ANN (bottom).

Fig. 13 – Clustering of Anabaena and Asterionella abundances regarding NO₃⁻N classes in Lake Soyang using non-supervised ANN (top) and sensitivity curves of Anabaena and Asterionella abundances over the NO₃⁻N range of Lake Soyang using supervised ANN (bottom).
Fig. 14 – Clustering of *Microcystis* and *Cyclotella* abundances regarding PO₄-P classes in Lake Kasumigaura using non-supervised ANN (top) and sensitivity curves of *Microcystis* and *Cyclotella* abundances over the PO₄-P range of Lake Kasumigaura using supervised ANN (bottom).

Fig. 15 – Clustering of *Anabaena* and *Asterionella* abundances regarding PO₄-P classes in Lake Soyang using non-supervised ANN (top) and sensitivity curves of *Anabaena* and *Asterionella* abundances over the PO₄-P range of Lake Soyang using supervised ANN (bottom).
Fig. 16 – Seasonal abundance clusters of the *Scenedesmus* and *Cyclotella* for two periods 1984 to 1986 and 1987 to 1989 with distinctive nutrient conditions in Lake Kasumigaura.

Fig. 17 – Seasonal abundance clusters of the *Microcystis* and *Oscillatoria* for two periods 1984 to 1986 and 1987 to 1989 with distinctive nutrient conditions in Lake Kasumigaura.
Fig. 18 – Seasonal abundance clusters of cladocera and copepoda for two periods 1984 to 1986 and 1987 to 1989 with distinctive nutrient conditions in Lake Kasumigaura.

Fig. 19 – Seasonal clusters of NO$_3$-N and PO$_4$-P concentrations for two periods 1984 to 1986 and 1987 to 1989 with distinctive nutrient conditions in Lake Kasumigaura.
of Cyclotella in Lake Kasumigaura coincide with pH values greater than 9, whilst Asterionella in Lake Soyang seems to be most abundant at neutral pH. Talling (1976) concluded from a series of experiments that both diatoms and blue-green algae appear to be photosynthesis tolerant regarding high pH and low CO₂ concentrations. Shapiro (1984) postulated that blue-green algae are physiologically adapted to cope well with low CO₂ concentrations and out-compete eukaryotic algae at high pH. Reynolds (1984) confirmed these findings for Microcystis aeruginosa and Anabaena flos-aquae. By virtue of the present results and the literature findings it can be concluded that Microcystis and Anabaena behave very much like K-selected species as they are specialised for distinct environmental conditions such as higher water temperature and pH whereby Cyclotella and Asterionella tolerate a much broader range of conditions typical for r-selected species. As postulated by Reynolds (1984) the differentiation between K- and r-selected algal species relies very much on their capability to cope with sinking and grazing losses. Both Microcystis and Anabaena are considered K-selected as they minimise sinking losses by regulating their buoyancy and avoid grazing losses by forming large cell colonies or filaments as well as contain toxic substances (Reynolds, 1984). By contrast Cyclotella and Asterionella are considered r-selected by having relatively high sinking losses because of their dense silica cell walls and being largely exposed to grazing (Reynolds, 1984).

The elucidation of relationships between algal populations and nutrient conditions in lakes with different trophic states was another aim of the current study. With regards to NO₃-N concentrations clusters in Figs. 12 and 13 gave evidence that Microcystis and Anabaena reached highest abundances in both lakes when NO₃-N was lowest in concentration. On the other hand highest abundances of Cyclotella in Lake Kasumigaura and Asterionella in Lake Soyang were clustered at medium concentrations of NO₃-N. These findings were supported by corresponding sensitivity curves showing for both lakes that blue green algae peaked at low NO₃-N concentrations e.g. 0.4 mg/l in Lake Kasumigaura and 0.9 mg/l in Lake Soyang but diatoms peaked at higher NO₃-N concentrations e.g. 0.8 mg/l in Lake Kasumigaura and 1.3 mg/l in Lake Soyang. There are two possible explanations for these results: (1) blue-green algae diminish NO₃-N concentrations significantly by high NO₃-N uptake during maximum growth and diatoms are competitively excluded because of both non-favouring NO₃-N levels and water temperatures at that times, and (2) some blue-green algae out-compete diatoms at times of lowest NO₃-N concentrations by assimilating dissolved atmospheric nitrogen N₂ through heterocysts (Fay et al., 1968) as being clearly demonstrated for Anabaena but not yet for Microcystis (Reynolds, 1984). The hypothesis (1) may explain best the results in Fig. 12 for Lake Kasumigaura whilst hypothesis (2) is more likely to reflect conditions in Lake Soyang where summer stratification further accelerates sinking losses of diatoms. Opposite trends were revealed in the two non-P-limited lakes for relationships between blue-green algae, diatoms and PO₄-P in Figs.
Both Microcystis and Anabaena showed affinity to highest PO_4-P concentrations which read in Lake Kasumigaura 25 to 40 μg/l and in Lake Soyang 4 to 5 μg/l but Cyclotella and Asterionella became most abundant at low PO_4-P concentrations of 2.5 to 5 μg/l. The results for Microcystis and Anabaena correspond well with their intolerances of low PO_4-P concentrations postulated by Reynolds (1984). It was found by Mackereth (1953) that Asterionella cells have a phosphorus storage capacity equivalent to 24 times the absolute cell minimum that may explain Asterionella’s and Cyclotella’s tolerance to low PO_4-P concentrations as observed in this study. Although most findings of this research correspond well with current knowledge on algal specific relationships with pH and nutrient conditions it must be pointed out that these relationships are distinctively bi-directional as e.g. algal metabolism in turn changes nutrient and CO_2 budgets as well. Therefore patterns of relationships between physical, chemical and biological properties of ecosystems automatically mapped from historical data by non-supervised ANN may reflect not necessarily the cause but the result of such relationships.

4.3. Relationships between algal abundances, seasons and long-term water quality changes

The results in Figs. 16–21 have demonstrated that ordination and clustering by non-supervised ANN can reveal seasonal and long-term patterns for ecological relationships in lakes. The data for Lake Kasumigaura were seasonally ordained and clustered for both the period 1 from 1984 to 1986 and the period 2 from 1987 to 1989 between which a significant increase of the TN (total nitrogen)/TP (total phosphorus) ratio from 10 to approximately 20 had been observed (Takamura et al., 1992). The resulting patterns revealed that Cyclotella peaked during winter in period 1 and with two times higher magnitude during spring in period 2. The shift of highest abundance of Cyclotella from winter to spring and period 1 to 2 seems to be determined by higher NO_3-N concentrations in period 2 and its known preference of NO_3-N sufficiency discussed above for findings in Fig. 12. The green algae Scenedesmus appeared to be most dominant in spring and early summer in both periods with a 30% increased abundance in period 2. Microcystis clearly dominated in late summer in period 1 but shifted its dominance to autumn in period 2 with a 60% decrease in abundance. By contrast Oscillatoria shifted its dominance from early summer in period 1 to late summer in period 2 by more than doubling its maximum abundance. Results in Fig. 17 show competitive seasonal exclusion between Microcystis and Oscillatoria for both periods with a distinct PO_4-P limitation of Microcystis in autumn of period 2 (see Fig. 19) caused by high growth and PO_4-P consumption of Oscillatoria in late summer. While cladocera tended to have highest abundances in late summer of both periods, copepoda peaked at the transition from late summer to autumn. As in both periods high cladocera abundance coincided with abundant Microcystis there seems to be an indication for feeding of decaying Microcystis cells by cladocera as observed in Lake Kasumigaura by Hanazato and Yasuno (1987) and

Fig. 21 – Seasonal clusters of NO_3-N and PO_4-P concentrations for two periods 1992 to 1993 and 1998 to 1999 with distinctive management conditions in Lake Soyang.
Hanazato (1991). Results in Fig. 18 show also seasonal exclusion by predation of cladocera by copepoda for both periods.

The data for Lake Soyang were seasonally differentiated and clustered for both the period 1 from 1992 to 1993 with intensive fish farming and the period 2 from 1998 to 1999 with no fish farming. These periods were chosen as stopping fish farming in period 2 eased eutrophication of the lake by approximately 30% lower PO4–P concentration (Kim et al., 2000). The seasonal clusters in Fig. 20 indicate a distinct response behaviour of blue-green algae and diatoms to the changed management between period 1 and 2. The abundance of Anabaena peaks in period 2 only in late summer at a 50% lower maximum but occurs in early summer at insignificant level only. One possible reason for this seasonal shift and reduced abundance of Anabaena can be found in Fig. 21 indicating that high PO4–P concentrations in summer were typical for period 1 providing PO4–P sufficiency as required by blue-green algae but occurred in winter only at 30% lower concentrations in period 2. On the other hand there is a shift in the dominance of Asterionella from autumn in period 1 to spring in period 2 with a 10 times increased abundance (Fig. 20, bottom). Distinct spring blooms of Asterionella in period 2 may have been triggered by a combination of the slightly increased NO3–N concentrations in spring as a result of no fish farming, and its known PO4–P storage capacity (Mackereath, 1953) charged during the windrily PO4–P enrichment of the lake.

5. Conclusions

The current study has demonstrated that complex limnological time-series data can beneficially be processed by novel computational techniques in order to provide: (1) 1-week-ahead forecasting of outbreaks of harmful algae or water quality changes by recurrent supervised ANN and EA, (2) explanatory rule sets for timing and abundance of harmful outbreaks algal populations by EA, and (3) clusters to unravel ecological relationships regarding seasons, water quality ranges and long-term environmental changes by non-supervised ANN.

The performance of EA with regards to forecasting accuracy and explanation of timing and magnitude of algal population outbreaks was superior compared to recurrent supervised ANN.

The combination of quantitative input sensitivity analyses by recurrent supervised ANN with qualitative differentiation and clustering by non-supervised ANN proved to be highly indicative for testing hypotheses on algal specific preferences for water quality and environmental conditions.

Overall the research has demonstrated that both ANN and EA provide a useful framework for comparative studies between lakes with largely different conditions that will further facilitate “basic research on complex interactions (that) will lead to explanations for the variability and unpredictability that presently hamper lake management efforts...” Carpenter (1988). Future work will focus on the determination of predictive and explanatory rule sets for specific algal populations by EA being generic for a certain category of lakes such as shallow hypereutrophic or stratified mesotrophic in order to support agent-based forecasting of emergent algal populations in lakes of the same category (Recknagel, 2003).

Acknowledgements

The authors are grateful to the Lake Kasumigaura Survey Group of the National Institute for Environmental Studies, Tsukuba, Japan, and the Lake Soyang Survey Group of the Kangwon National University, Chunchon, South Korea for providing data. The authors thank Colin S. Reynolds for his valuable comments on the first draft of the manuscript.

The research was funded by the Australian Research Council (DP0345279).

REFERENCES

Kim, B., Choi, K., Kim, C., Lee, Y.H., 2000. Effects of the summer monsoon on the distribution and loading of organic carbon in a...