Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes

Friedrich Recknagel, Jason Bobbin, Peter Whigham and Hugh Wilson

ABSTRACT

The paper compares potentials and achievements of artificial neural networks and genetic algorithms in terms of forecasting and understanding of algal blooms in Lake Kasumigaura (Japan). Despite the complex and nonlinear nature of ecological data, artificial neural networks allow seven-days-ahead predictions of timing and magnitudes of algal blooms with reasonable accuracy. Genetic algorithms possess the capability to evolve, refine and hybridize numerical and linguistic models. Examples presented in the paper show that models explicitly synthesized by genetic algorithms not only perform better in seven-days-ahead predictions of algal blooms than artificial neural network models, but provide more transparency for explanation as well.

Key words | chlorophyll-*a*, *Microcystis*, short-term prediction, artificial neural network model, genetic algorithm model, rule sets, difference equations

Friedrich Recknagel (corresponding author)
Jason Bobbin
Hugh Wilson
Department of Soil and Water,
University of Adelaide,
Glen Osmond,
SA 5064,
Australia

Peter Whigham
Department of Computer Science,
University of Otago,
PO Box 56,
Dunedin,
New Zealand

INTRODUCTION

Progress in algal bloom modelling currently faces two bottlenecks: limitations in ecological knowledge for deductive modelling and limitations in data analysis for inductive modelling. While deductive modelling relies on evolving disciplinary research, inductive modelling relies on suitable techniques to cope with the complexity and nonlinearity of ecological data. Traditional deductive lake ecosystem models such as MS.CLEANER (Park *et al*. 1974), AQUAMOD (Straskraba & Gnauck 1985) or SALMO (Recknagel & Benndorf 1982) were successfully applied for the simulation of seasonal dynamics of functional algal groups and allow scenario analysis on options for eutrophication management (e.g. Recknagel *et al*. 1995). They are not yet qualified to simulate algal species dynamics in order to forecast blooms of toxic blue green algae such as *Microcystis*. Opportunities to gradually overcome these limitations are currently arising from developments in machine learning techniques.

Machine learning is a broad discipline in computer science that focuses on knowledge acquisition and processing by various automated induction techniques.

Two established approaches from this field, namely artificial neural networks and genetic algorithms, take their inspiration from aspects of biological information processing. They offer approaches for model building different to standard statistical techniques and have been demonstrated to be capable of producing robust models for different knowledge domains (e.g. Babovic & Keijzer 2000; Minns 2000).

This paper discusses novel applications of machine learning techniques for predictive modelling of the abundance of blue–green algae in a freshwater lake. Results show that artificial neural networks perform well for seven-days-ahead prediction of timing and magnitudes of algal bloom events. Once trained and validated, scenario and sensitivity analyses provide a means of explaining the model learned by the neural network regarding relationships between chemical, physical and biological parameters and abundance of specific algae (Recknagel & Wilson 2000). However, explanation through examination of the structure of the trained network has proved difficult due to the fact that neural

networks store learned models in a highly distributed manner by means of connection weights, which bear little resemblance to human understanding of rules or concepts. By contrast, genetic algorithms can derive explicit numerical or linguistic models that can easily be brought into a context with domain knowledge. Therefore genetic algorithms prove to be powerful tools for synthesizing, hybridizing and refining ecological models. The paper discusses two approaches for algal bloom modelling based on genetic algorithms: evolving constant parameters for a process-based equation for algal growth, and discovery of rule sets from data. Case studies show that these modelling approaches not only lead to reasonable seven-days-ahead predictions but, in the case of the rule-based approach, to knowledge discovery as well.

This paper compares potential and achievements of artificial neural networks and genetic algorithms in terms of forecasting and understanding of algal blooms in Lake Kasumigaura (Japan) that appear to be superior to traditional modelling techniques.

METHODS

Artificial neural networks and genetic algorithms were applied to water quality time series from Lake Kasumigaura (Japan) in order to model and predict algal dynamics. To explain and predict plankton dynamics of this shallow, hypereutrophic lake is a challenging task and is the subject of lasting research (Takamura *et al.* 1992; Recknagel 1997; Yabunaka *et al.* 1997; Recknagel *et al.* 1998).

Artificial neural network modelling of algal blooms

The prediction of the timing and magnitudes of algal blooms in freshwater lakes represents a multivariate nonlinear time-series problem. It can be modelled by feedforward artificial neural networks based on the back-propagation algorithm (Rumelhart *et al.* 1986), where input layers contain nodes for the limiting factors of algal growth such as surface and underwater irradiance,

nutrient concentrations, zooplankton abundance, and output layers contain the cell numbers of algal species highly abundant in the studied lake.

Figure 1 represents the architecture of the artificial neural network model ANNA (Recknagel 1997) that was designed to predict abundance and succession of algal species. It has been trained and validated for a variety of freshwater lakes in Europe, Asia and Australia. The example shown in Figure 1 is based on data and conditions of Lake Kasumigaura (Japan). For the performance of ANNA it proved to be crucial to train neural network ensembles considering time lags by multi-vector input layers (Recknagel *et al.* 1998).

Recently, applications of recurrent artificial neural networks for time-series modelling of algal dynamics in lakes (Walter *et al.* 2001) and rivers (Jeong *et al.* 2001; Jeong *et al.* 2002) proved to perform even better. They have been designed to mimic the deterministic modelling paradigm where the system state at time t is calculated by means of the system state at time $(t-1)$ (Pineda 1987). Assuming that the weights of neurons of the hidden layer represent the 'hidden' state of the system the copied weights of time $(t-1)$ are considered as feedback inputs for the determination of weights of the neuron at time t.

Genetic algorithm modelling of algal blooms

The development of genetic algorithms has been inspired by processes of natural evolution and was first clearly described by Holland (1975). He explored algorithms, operating on strings of bits that he called chromosomes. To apply genetic algorithms to solve a problem, a representation for potential solutions is encoded on chromosomes (the representation genotype) and an evaluation or objective function is defined in order to measure the performance of the chromosomes against the defined problem. Depending on the nature of the application, chromosomes can be strings of bits (110110001), lists of real values (0.3, 0.1, 0.2, 0.9, 0.7), permutations of elements (A2, A5, A1, A3, A6, A4) or some other representation. It must be noted that no one representation is suitable for all problems.
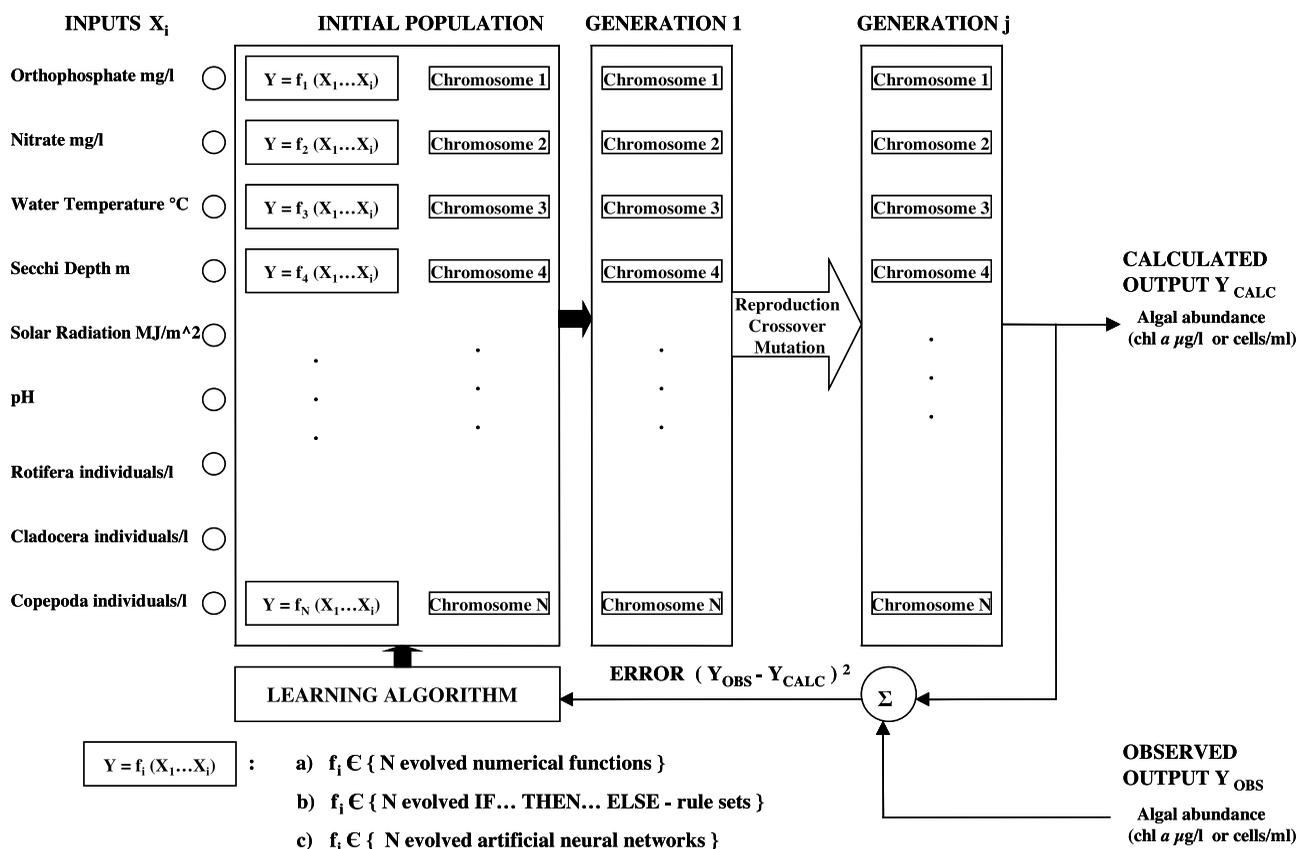
Figure 1 │ Architecture of the artificial neural network model ANNA for the prediction of algal blooms in Lake Kasumigaura.

Genetic algorithms are based on a two-step process of heritable variation and selection. Variation occurs in the diversity of behaviours exhibited by individuals in the population, where this behaviour is derived from the makeup of the genotype representation. A measure of the fitness of each individual (the phenotype) gives a measure of the performance against the set problem. Selection removes those individuals from the population that do not have an appropriate fitness, leaving behind those individuals that can pass on their genotype to subsequent generations. During this process of heredity, variation in the genotype can occur, either through swapping portions of the genotype between a set of parents (crossover) to construct new individuals, or through random mutation of portions of the genotype. Variation in the reproductive process is the source of change at the genetic level, which translates into new

innovation at the phenotypic level. Selection serves as a filtering mechanism to ensure that individuals of low fitness are removed along the way. The systems described here use a proportional-based fitness selection method, where the probability of selection is proportional to the fitness of the individual compared to the fitness of the entire population. Therefore, if one individual is twice as fit as another individual, it will have twice the probability of being selected for transmission into the next generation. The fitness measure is problem-specific: in this paper the fitness is based on the accuracy of the predicted solutions based on a set of training data using a root mean square error (RMSE). The closer an individual can predict the training sequence the lower the RMSE is, and therefore the higher the fitness.

The crossover operation used in this paper is based on a single-point crossover. Basically, pairs of chromosomes

are recombined by swapping parts of them from a randomly selected point in order to create two new chromosomes. There are numerous other techniques for mixing representations of two or more parents to form new individuals: however, there is no one preferred operator for all problems. In general, genetic algorithms are robust search methods and will give adequate performance using a standard crossover and mutation method for creating variation. Mutation occurs with a very low probability and makes a small change in the genotype of an individual. For example, when a binary representation is used for mutation, each bit in the string has a small probability of being changed from a zero to one or vice versa. Mutation ensures that new genetic material is continually introduced to the population and therefore helps to ensure that the population does not converge to a local minima. An excellent introduction to genetic algorithms may be found in Goldberg (1989), and a history of evolutionary computation, of which genetic algorithms are one example, may be found in Fogel (1998).

Multivariate time-series modelling of the abundance of algae by genetic algorithms considers the same inputs and outputs as the artificial neural network approach. In contrast to neural networks, genetic algorithms are able to consider a combination of certain arithmetic operators, rules or neural network designs as chromosomes in order to evolve a numerical, linguistic or neural network model. The chromosomes repeatedly undergo recombination by reproduction, crossover or mutation in order to minimize the output error: the sum of the differences between the actual output vector and the desired output vector.

Figure 2 represents the basic structure of the genetic algorithm model GAMA for predictions of algal blooms based on data of Lake Kasumigaura.

RESULTS AND DISCUSSION

The neural network model ANNA (Figure 1) and the genetic algorithm model GAMA (Figure 2) have been trained and validated with water quality time series recorded from surface samples at a central site of Lake Kasumigaura from 1984 to 1993 (Takamura *et al.* 1992; Hanazato and Aizaki 1991). In both cases, (1) total chlorophyll-*a* and cell counts of the blue–green algae *Microcystis* have been modelled as outputs, and (2) two years of data, for 1986 and 1993, were excluded from training, to be used as independent data for predictions and model validation.

In Figure 3 results of seven-days-ahead predictions of chlorophyll-*a* and *Microcystis* for 1986 and 1993 by application of ANNA and GAMA are represented. The diagrams in Figures 3(a) and (b) show the prediction results of ANNA in comparison with measured data. ANNA was built considering the inputs as represented in Figure 1, one hidden layer with ten nodes and the sigmoid transfer function. Training was conducted using the scaled conjugate gradient method. One hundred training runs were performed using random initial starting weights and random training subsamples generated by a bootstrap method to account for sources of variance arising from initial conditions and sampling variability. The optimum error for stopping training, when training and test errors started to diverge, was determined experimentally in order to prevent overfitting. The median RMSE over 100 runs for chlorophyll-*a* was 31.6. Figures 3(a) and (b) display the maximum output value from the distribution of test results at each time step, thus simulating a worst case scenario model for bloom prediction.

Figure 3(a) shows that the model ANNA predicts trends of chlorophyll-*a* dynamics rather than specific peaks of the validation years 1986 and 1993. It fails to simulate the spring peak and underestimates the late summer peak in 1986. It predicts the spring and summer peaks in 1993 with a delay of several days and lower magnitude. The application of ANNA proved to be more successful in order to predict the abundance of *Microcystis* in 1986 and 1993. Figure 3(b) shows that the timing of fast algal growth towards the large summer peak in 1986 was predicted very well, even though the peak was reached a few days in advance. The model ANNA has predicted a moderate late summer bloom of *Microcystis* in 1993 that had not been observed.

The learned time-series model of ANNA is stored in a highly distributed manner by means of connection weights. To gain an explicit explanation of prediction
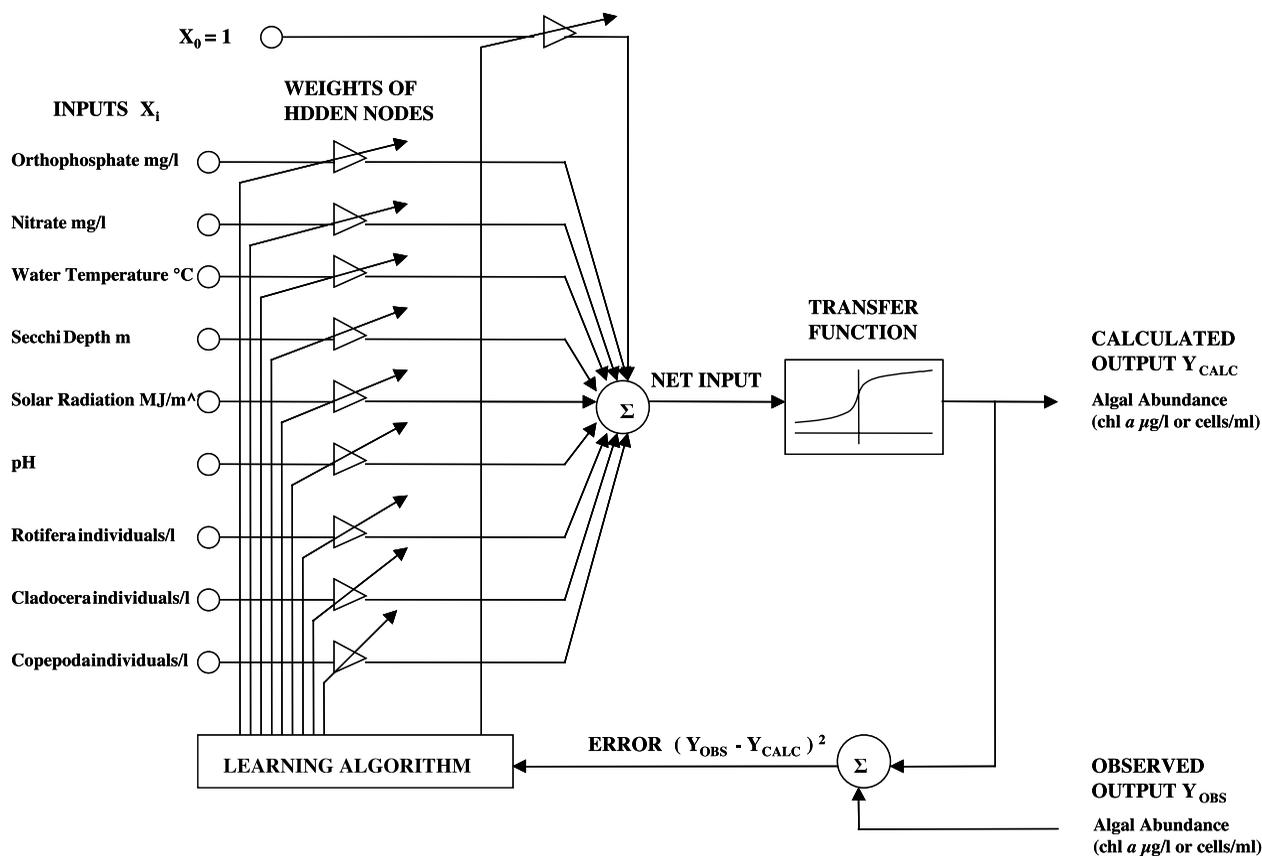
Figure 2 | Conceptual diagram for the application of genetic (evolutionary) algorithms for modeling algal abundance (GAMA) by: (a) evolving numerical functions, (b) evolving linguistic rules and (c) evolving artificial neural network designs.

results by examining the structure and weights of the trained model proves to be difficult. However, Recknagel & Wilson (2000) have shown that scenario and sensitivity analysis provide useful information about relationships between ecological driving variables, seasonality and algal abundance simulated by ANNA.

The model GAMA was applied to evolve empirical and process-based equations for the prediction of chlorophyll-*a* (Whigham & Recknagel 1999, 2000). *Inter alia* the following equation, based on normalized data, was discovered empirically and validated in Whigham & Recknagel (1999):
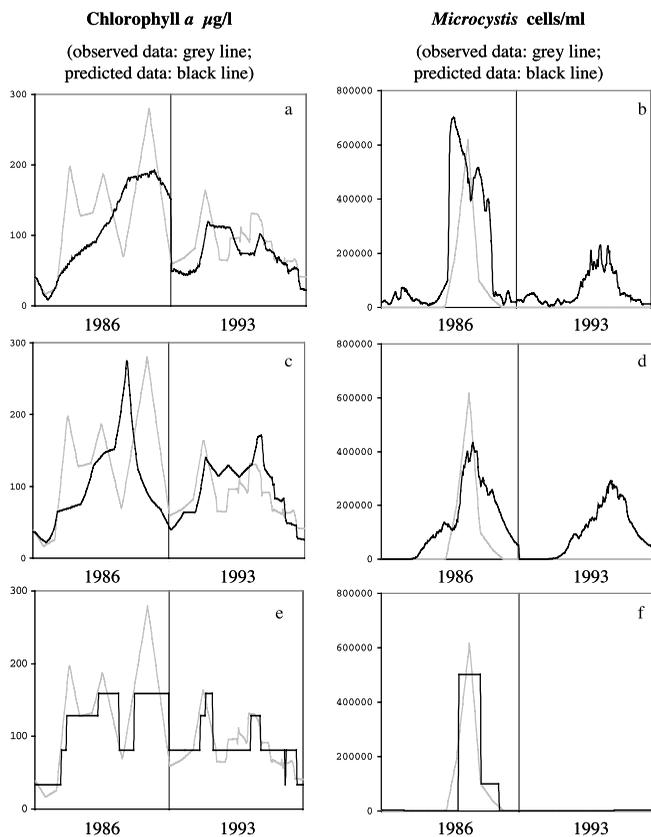
$$\text{Chl-}a = \cosh(T)/(6S + 2N + 2.251)$$

where $T =$ water temperature, $S =$ Secci depth and $N = NO_3\text{-N}$ nitrate.

This example showed that one compact multiple non-linear equation extracted from time series by GAMA allowed the approximate prediction of the seasonal dynamics of algal biomass for two unseen years. The RMSE for the prediction was 37.08, which was comparable in accuracy to other methods on the same dataset (Whigham & Recknagel 1999). However, as with statistical regression models, it did not provide any explanation about the nature of the underlying processes responsible for algal growth. Therefore GAMA was applied to evolve the following process-based finite-difference equation adopted from the dynamic lake model SALMO (Recknagel & Benndorf 1982):

$$A(t + 1) = A(t) + A(t)$$
$$*(PHOT - RESP) - A(t)$$
$$*(COP + CLAD)*0.0001 - A(t)*(e/5)$$

Chlorophyll *a* μg/l

(observed data: grey line; predicted data: black line)

***Microcystis* cells/ml**

(observed data: grey line; predicted data: black line)

Figure 3 | Validation of seven-days-ahead predictions for chlorophyll-*a* and *Microcystis* in 1986 and 1993 in Lake Kasumigaura. (a), (b): Applications of the model ANNA; (c), (d): applications of the model GAMASalmo; (e), (f): applications of the model GAMARules. (Please note that *Microcystis* was only observed at extremely low concentrations in 1993.)

$$PHOT = [a/b*T]*(0.025*L/(c + 0.025*L))$$
$$*(P/A(t)/(d/X + P/X + d/A(t) + P/A(t)))]$$

$$X = 5.76*A(t) \times 0.41$$

$$RESP = [(0.057/b*T) + 0.3*PHOT]$$

where A(t) = chlorophyll (μg/l) or *Microcystis* (cells/ml) at time t, PHOT = photosynthesis, RESP = respiration, L = photosynthetic active light, P = PO_4-P phosphate, X = auxiliary term, COP = biomass of crustacea copepoda (individuals/l), CLAD = biomass of crustacea cladocera (individuals/l), and a–e = constant parameters. The model GAMASalmo was applied in two stages: (1) to optimize the related parameters a–e within their range of estimation errors, and (2) to evolve new algebraic terms within the suggested equations. Results of stage (1) are documented in the framework of this paper. For the constant parameter optimization (constants a–e), the genotype used a floating-point number representation for the constants. Additionally, the constants were constrained to remain within ± 20% of the original physically based measures. A population of 1000 individuals, each representing the constants a–e, were evolved, based on the training data for 100 generations using crossover (90%) and mutation (1%).

Even though GAMASalmo predicts the magnitude of the chlorophyll-*a* summer peak in 1986 very well it predicts the timing of the peak three weeks too early (Figure 3(c)). The seven-days-ahead prediction (RMSE = 51.7) has missed out the spring peak and roughly indicated the early summer peak in 1986. For 1993, the spring peak was simulated well in magnitude and timing whilst the summer peak was somewhat over-estimated (Figure 3(c)). As chlorophyll-*a* is only a lump-like indicator for algal abundance, GAMASalmo was applied to predict dynamics of *Microcystis*, a toxic blue–green algae that is frequently high in abundance in Lake Kasumigaura. The results for 1986 in Figure 3(d) show that GAMASalmo tends to predict the right timing of the summer peak but underestimates the magnitude. GAMASalmo predicted a moderate summer bloom of *Microcystis* for 1993 that had not been observed (Figure 3(d)). The wrong prognosis for 1993 by GAMASalmo might be attributed to changed underwater light that has not been considered yet in the algal growth equation. Results of scenario and sensitivity analyses by Recknagel & Wilson (2000) support this assumption.

In an attempt to discover linguistic rules for the prediction of algal abundance the evolutionary algorithm GAMARules was applied to the database of Lake Kasumigaura according to Figure 2(b) where populations of 500 individuals and 150 generations were used (Bobbin & Recknagel 2001a). The evolutionary algorithm used is based on a method known as evolutionary programming, which is broadly similar to the genetic algorithms method described earlier. The principal differences are in the details of the mutation operators, selection scheme and the self-adaptive nature of the approach. The algorithm can be applied to a discrete rule-based representation,

which can be understood. It also faces a regression problem where the rule sets are required to predict a real-valued dependent variable. The algorithm automatically optimizes the real threshold values used in the rule premises and consequence parts, along with the discrete structure of the rule set. The interquartile range of the RMSE for 20 independent runs of the predictive rules discovered for chlorophyll-*a* was 36.9–40.7 with a median of 37.6, which is as good as those results achieved by other methods.

The following is an example of an evolved rule set for the prediction of chlorophyll-*a* concentrations:

IF $S >= 110$ cm

THEN Chl-*a* = 34.38 µg/l

ELSE IF $N >= 577$ mg/l AND IF $P <= 34$ mg/l

THEN Chl-*a* = 96.69 µg/l

ELSE IF $P <= 34$ mg/l

THEN Chl-*a* = 54.85 µg/l

where $S =$ Secchi depth, $P =$ PO$_4$-P phosphate and $N =$ NO$_3$-N nitrate.

Results in Figure 3(e) indicate that the rule set is able to accurately predict the timing of the major chlorophyll peaks in 1986 and 1993. Even though it fails to meet the magnitudes of the three peaks in 1986, it simulates realistically the chlorophyll dynamics in 1993. In a next step GAMARules was applied to evolve rule sets for the prediction of *Microcystis* dynamics in Lake Kasumigaura. The following rule set was discovered in the database 1984 to 1985 and 1987 to 1992, and validated by means of data for 1986 and 1993 (see also Bobbin & Recknagel 2001b):

IF $P >= 81.7$ µg/l AND $P <= 126$ µg/l

THEN *Microcystis* = 500,000 cells

ELSE

IF pH $>= 9.72$ THEN *Microcystis* = 500,000 cells

EXCEPT IF $T <= 19.5°C$ AND $T >= 5.67°C$

THEN *Microcystis* = 3,000 cells

ELSE

IF $N/P >= 47.2$ AND $N/P <= 55.2$ THEN *Microcystis* = 0

ELSE

IF $S >= 95.5$ cm THEN *Microcystis* = 3,000 cells

EXCEPT IF $T >= 5.67°C$ AND $T <= 15.7°C$

THEN *Microcystis* = 0

OR IF $N >= 1110$ µg/l THEN *Microcystis* = 0

ELSE

IF $P >= 15.6$ µg/l AND $P <= 116$ µg/l

THEN *Microcystis* = 100,000 cells

EXCEPT IF $T >= 26.7°C$ THEN *Microcystis* = 500,000 cells

EXCEPT IF $S <= 160$ cm then *Microcystis* = 100,000 cells

ELSE

IF $N >= 757$ µg/l AND $N <= 1690$ µg/l THEN *Microcystis* = 0 cells

EXCEPT IF $T >= 15.7°C$ AND $T <= 26.7°C$ THEN *Microcystis* = 3,000 cells

EXCEPT IF $P <= 15.6$ µg/l THEN *Microcystis* = 0 cells

ELSE

IF $T >= 15.7$ AND $T <= 26.7°C$ then *Microcystis* = 100,000 cells

EXCEPT IF $S <= 160$ cm AND $S >= 74.4$ cm THEN *Microcystis* = 0 cells

ELSE

IF $T >= 5.67°C$ AND $T <= 15.7°C$ THEN *Microcystis* = 3,000 cells

ELSE

Microcystis = 100,000 cells.

The prediction results for *Microcystis* in 1986 and 1993 based on these rule sets are shown in Figure 3(f). It shows evidence that GAMARules is the only model in this case study that correctly predicts both timing and magnitude of the *Microcystis* peak in 1986 and the non-occurrence of *Microcystis* in 1993.

The rule sets that have been discovered from eight years of limnological time series of Lake Kasumigaura by evolutionary algorithms provide explicit information on the physical and chemical conditions responsible for high concentrations of chlorophyll-*a* and abundance of *Microcystis*. The proper interpretation and validation of these rules will contribute to a better understanding of the ecology of phytoplankton in this lake.

CONCLUSIONS

The case study on seven-days-ahead predictions of chlorophyll-*a* and *Microcystis* in Lake Kasumigaura by three types of machine learning time-series models has shown that:

1. artificial neural network models can be powerful short-term predictors of the timing of algal bloom events but are difficult to generalize and do not provide explicit explanation of underlying processes;

2. genetic algorithms are able to evolve predictive equations, either randomly synthesized or in the framework of existing process equations. The model GAMASalmo was based on deterministic algal growth equations and has resulted in reasonable forecasting of chlorophyll-*a* dynamics but only in rough estimations of *Microcystis* abundance;

3. evolutionary algorithms are able to discover predictive rules in ecological data sets. The model

GAMARules has demonstrated that it can achieve high accuracy in predicting timing and magnitudes of chlorophyll-*a* peaks as well as of *Microcystis* peaks. It proves to be a powerful tool for the discovery and testing of causal knowledge;

4. genetic and evolutionary algorithms are flexible tools for the synthesizing and hybridizing of numerical, as well as linguistic, ecological models.

Overall the results of the present study have proven that multivariate time-series modelling by machine learning techniques has the capacity to forecast the timing and magnitudes of highly nonlinear and sudden ecological events, such as algal blooms, to a reasonable extent where traditional deterministic modelling techniques currently fail. Their further improvement and implementation into early warning systems for harmful algal blooms is the subject of ongoing research.

REFERENCES

Babovic, V. & Keijzer, M. 2000 Genetic programming as a model induction engine. *J. Hydroinformatics* **2**, 35–60.

Bobbin, J. & Recknagel, F. 2001a Inducing explanatory rules for the prediction of algal blooms by genetic algorithms. *Environ. Int.* **27**, 237–242.

Bobbin, J. & Recknagel, F. 2001b Knowledge discovery for prediction and explanation of blue-green algal dynamics in lakes by evolutionary algorithms. *Ecol. Model.* **146**(1–3), 253–262.

Fogel, D. 1998 *Evolutionary Computation: The Fossil Record*. IEEE Press, Piscataway, NJ.

Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, New York.

Hanazato, T. & Aizaki, M. 1991 Changes in species composition of cladoceran community in Lake Kasumigaura during 1986–1989: occurrence of *Daphnia galeata* and its effects on algal biomass. *Japan. J. Limnol.* **52**, 45–55.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. Addison-Wesley, New York.

Jeong, K.-S., Joo, G.-J., Kim, H.-W., Ha, K. & Recknagel, F. 2001 Predicting phytoplankton dynamics in the flow-regulated Nakdong River (South Korea) by means of feedforward and recurrent neural networks. *Ecol. Modell.* **146**(1–3), 115–129.

Jeong, K.-S., Recknagel, F. & Joo, G.-J. 2002. Prediction and elucidation of population dynamics of the blue-green algae *Microcystis aeruginosa* and the diatom *Stephanodiscus hantzschii* in the Nakdong River-Reservoir system (South Korea) by a recurrent artificial neural network. In: *Ecological Informatics: Scope, Concepts, Case Studies* (ed. F. Recknagel) Springer Verlag, Berlin, Heidelberg, New York (in press).

Minns, A. W. 2000. Subsymbolic methods for data mining in hydraulic engineering. *J. Hydroinformatics* 2, 3-13.

Park, R. A., O'Neill, R. V., Bloomfield, J. A., Shuggart, H. H., Booth, R. S., Goldstein, R. A., Mankin, J. B., Koonce, J. F., Scavia, D., Adamas, M. S., Clesceri, L. S., Colon, E. M., Dettman, E. H., Hoopes, J. A., Huff, D. D., Katz, S., Kitchell, J. F., Kohberger, R. C., La Crow, R. J., McNaught, D. C., Peterson, J. L., Titut, J. E., Weiler, P. R., Wilkinson, J. W. and Zahorcak, C. S. 1974. A generalized model for simulating lake ecosystems. *Simulation* 33–50.

Pineda, F. 1987 Generalization of backpropagation to recurrent neural networks. *Phys. Rev. Lett.* **19**, 2229–2232.

Recknagel, F. 1997 ANNA–artificial neural network model for predicting species abundance and succession of blue-green algae. *Hydrobiologia* 349, 47–57.

Recknagel, F. & Benndorf, J. 1982 Validation of the ecological simulation model SALMO. *Int. Rev. Ges. Hydrobiol.* **67**(1), 113–125.

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

Recknagel, F., Fukushima, T., Hanazato, T., Takamura, N. & Wilson, H. 1998 Modelling and prediction of phyto- and

zooplankton dynamics in Lake Kasumigaura by artificial neural networks. *Lakes Reservoirs Res. Mngt.* **3**, 123–133.

Recknagel, F. & Wilson, H. 2000 Elucidation and prediction of aquatic ecosystems by artificial neural networks. In *Artificial Neural Networks in Ecology and Evolution* (ed. S. Lek & J. F. Guegan), pp. 143–155. Springer-Verlag, New York.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986 Learning representations by back-propagation errors. *Nature* **323**, 533–536.

Straskraba, M. & Gnauck, A. 1985 *Freshwater Ecosystems: Modelling and Simulation*. Elsevier, Amsterdam.

Takamura, N., Otsuki, A., Aizaki, M. & Nojiri, Y. 1992 Phytoplankton species shift accompanied by transition from nitrogen dependence to phosphorus dependence of primary production in Lake Kasumigaura. *Arch. Hydrobiol.* **124**, 129–148.

Walter, M., Recknagel, F., Carpenter, C. & Bormans, M. 2001 Predicting eutrophication effects in the Burrinjuck Reservoir (Australia) by means of the deterministic model SALMO and the recurrent neural network model ANNA. *Ecol. Modell.* **146**(1–3), 97–113.

Whigham, P. & Recknagel, F. 1999 Predictive modelling of plankton dynamics in freshwater lakes using genetic programming. *Proceedings. of the International Congress MODSIM'99, Hamilton, New Zealand*, vol. 3, pp. 691–696.

Whigham, P. & Recknagel, F. 2000 Evolving difference equations to predict chlorophyll-*a* for a freshwater lake. *Proceedings of the 2000. Proceedings of the Congress on Evolutionary Computation, July 2000, La Jolla, USA*.

Yabunaka, K.-I., Hosomi, M. & Murakami, A. 1997 Novel application of a back-propagation artificial neural network model formulated to predict algal blooms. *Wat. Sci. Technol.* **36**(5), 89–97.