

CHAPTER 9

Ecological informatics by means of neural, evolutionary and object-oriented computation

F. Recknagel & H. Cao

*School of Earth and Environmental Sciences, University of Adelaide,
Adelaide, Australia.*

1 Introduction

One distinctive feature of ecological informatics is the application of bio-inspired computation such as artificial neural networks (ANN), evolutionary algorithms (EA) and object-oriented programming for ecological data analysis, synthesis and forecasting [1]. Neural and evolutionary computation allows inducing problem solutions such as patterns, models and knowledge from complex data [2]. Object-oriented computation facilitates the synthesis of ecosystem simulation models from process-based simulation libraries as suggested by Swartzman and Kaluzny [3].

Bio-inspired computation provides a different approach for data analysis, synthesis and forecasting compared with traditional modelling techniques. Whilst traditional techniques solve some particular problem or class of problems ‘top down’ by means of ad hoc-designed statistical or algebraic algorithms, bio-inspired computation is merely based on a loosely connected family of units such as electronic neurones, chromosomes or objects, which are continually being modified to improve the computational performance. This means that bio-inspired algorithms are not rigid, but dynamically evolve ‘bottom up’ e.g. by means of principles of neuronal learning, natural selection or hierarchical inheritance.

2 Artificial neural networks

Artificial neural networks are computer programs designed for inducing problem solutions (models, knowledge) from complex data by means of principles of information processing similar to biological neurones in the human brain. A biological neurone consists of three major components: the cell body, dendrites and the axon (Fig. 1a). Connections between neurones are formed at synapses. Information is represented and transmitted by chemically generated electrical activity within the cell. Both excitatory and inhibitory inputs to the neurone enter through synaptic connections with other neurones. Input potentials are summed up within the cell body. If the total input potential is sufficient (e.g. meets a certain threshold value), then the neurone acts. Ultimately, an action potential is generated and propagated down the axon towards the synaptic junctions with other nerve cells.



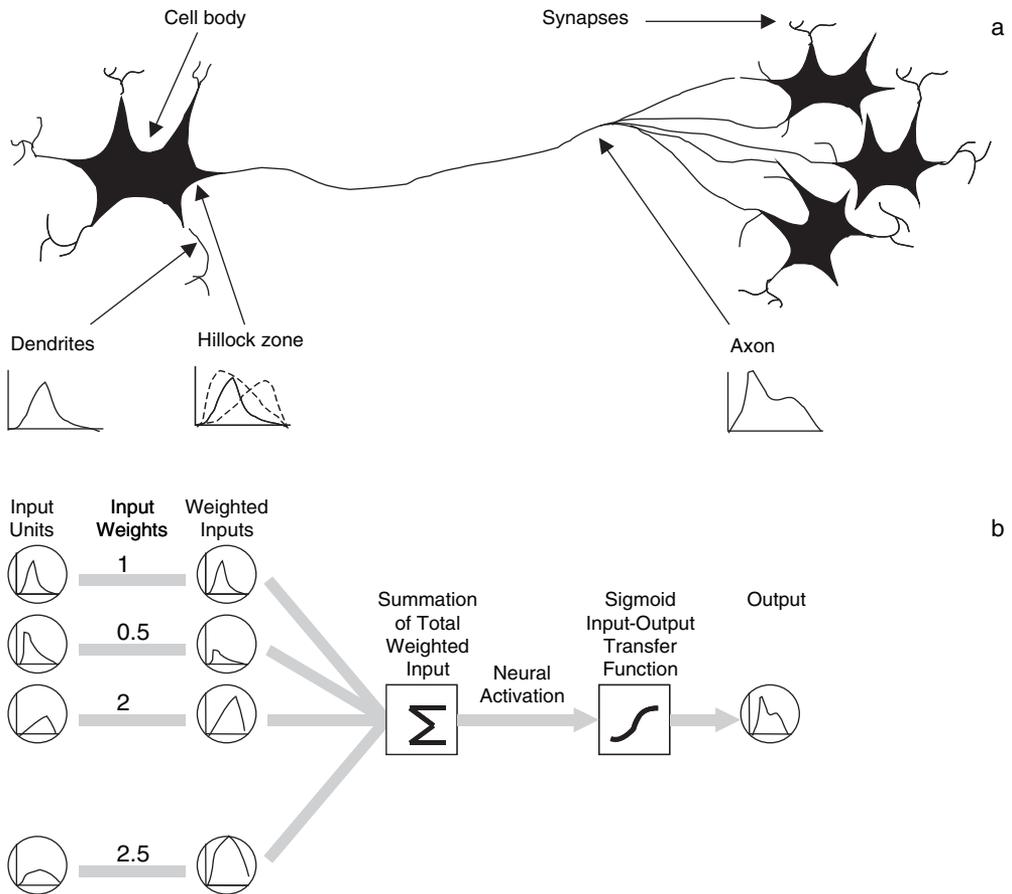


Figure 1: Conceptual structures of biological and artificial neurones.

The design of ANN (Fig. 1b) has been inspired by the structure and functioning of biological neurones. The dendrites, which act as input receptors, are represented by input units. The cell body that acts as information accumulator is represented by activation units adjusting and summing up the weights of inputs and the input–output transfer function. The axon that acts as the biological output channel is represented as the output.

Artificial neural networks gain their adaptive capability by undergoing training similar to neural learning where two basic training modes are distinguished: supervised and non-supervised training. The supervised training aims at the optimal approximation of the calculated output Y_c to the observed (desired) output Y_o . An iterative adjustment of input weights takes place to minimise the error ($Y_o - Y_c$).

After training, the generalisation of the supervised ANN is assessed by feeding it only with input values, not observed output values, and testing how close calculated outputs match observed outputs. The two most common methods for assessing generalisation are the *split-sample validation* and the *cross-validation*. The *split-sample validation* means that part of the data is reserved as a test set, which must not be used in any way during training. The test set must be representative of the problem to be modelled by the ANN. After training, the ANN is run on the test set, and

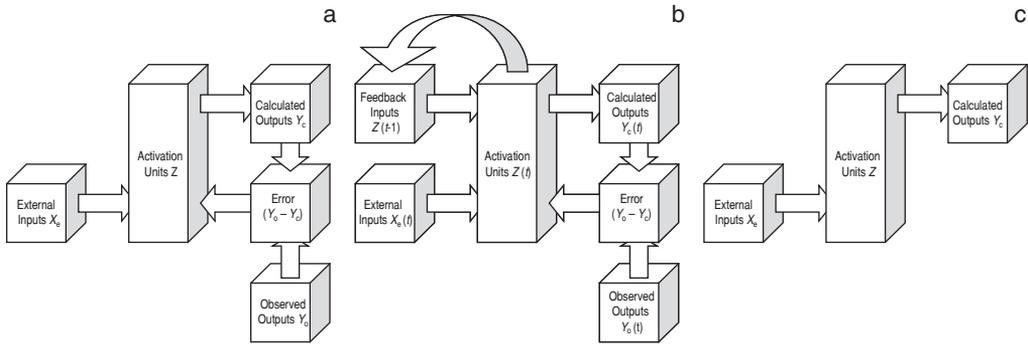


Figure 2: Basic types of artificial neural networks (ANN): (a) supervised feedforward ANN; (b) supervised feedback ANN; (c) non-supervised ANN.

the error on the test set provides an estimate of the generalisation error usually expressed by the root mean square error (RMSE) or the correlation coefficient r^2 . The disadvantage of split-sample validation is that it reduces the amount of data available for both training and validation [4]. In contrast, *cross-validation* allows you to use all the data for training. In k -fold cross-validation, the data are divided into k equal-sized subsets. The ANN is trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the generalisation error. If k equals the sample size, this is called ‘leave-one-out’ cross-validation. The disadvantage of cross-validation is that the ANN need to be retrained many times.

Depending on using external inputs only or feedback inputs as well, supervised ANN are differentiated into feedforward or feedback ANN (see Fig. 2a and b).

In contrast, non-supervised ANN process external inputs only without adjusting calculated outputs to known outputs (Fig. 2c).

2.1 Supervised feedforward ANN

The supervised feedforward ANN prove to be universal approximators of multivariate nonlinear functions [5] and are usually implemented as multi-layer perceptrons with back-propagation training. The multi-layer perceptron [6] represents input units as input layer, adjusted and accumulated input weights as hidden layer(s) and outputs as output layer. The back-propagation algorithm [7] performs the iterative adjustment of input weights (activation units) to minimise the approximation error ($Y_o - Y_c$).

Supervised feedforward ANN are widely applied in ecology either using cross-sectional data to predict discrete ecosystem states or using time-series data to predict continuous ecosystem behaviour. Successful applications by means of cross-sectional data have been demonstrated for fish communities in streams [8], macroinvertebrate communities in streams [9, 10], river salinity [11], primary productivity in estuaries [12], chlorophyll a concentrations in lakes [13], coastal vegetation [14] and bird populations [15].

Successful applications by means of time-series data have been demonstrated for marine fish and zooplankton communities [16–18], river hydrology [19], macroinvertebrate communities in streams [20, 21], freshwater phytoplankton and zooplankton communities [22–27].

A majority of the supervised feedforward ANN documented in the above-mentioned papers achieved forecasting results that were superior to conventional modelling techniques such as

multiple linear regression [8, 13]. Even though supervised ANN do not provide explicit mathematical representations of the underlying ANN model, most of the authors have conducted sensitivity analyses to identify inputs as key driving forces of the predictive ANN. An example for revealing input–output relationships by both sensitivity and scenario analysis was documented by Recknagel and Wilson [28] in their study.

2.2 Supervised feedback ANN

Supervised feedback or recurrent ANN [29] are designed to use not only external inputs for training but also activation levels of the previous training iteration, which are constantly fed back (see Fig. 2b). Their functioning can be compared with ordinary differential equations that calculate the current system state $Z(t)$ by taking into account current external inputs $X_e(t)$ and the system state $Z(t-1)$ of the time step before:

$$dZ(t)/dt = f(X_e(t), Z(t-1), P)$$

where P is a set of constant parameter.

Supervised feedback ANN prove to be very powerful for modelling time-series data where the fed back activation levels provide extra training information on the system state of the time step before.

Figure 3 shows an example of a supervised feedback ANN that has successfully been trained and tested by split-sample validation for the forecasting of the algal populations *Microcystis* and *Stephanodiscus* in the River Nakdong in South Korea [30]. The weekly measured limnological data of the river study site were interpolated to daily values. The interpolated data from 1995 to 1998 were used as training set, and the interpolated data of 1994 were used as testing set. To achieve a 4-days-ahead forecasting, a 4-day time lag was imposed between the measured inputs and the measured outputs of the training data set. The design of the feedback ANN considered the following 18 external input variables: irradiance, precipitation, discharge, evaporation, water temperature, Secchi depth, turbidity, pH, DO, nitrate, ammonia, phosphate, silica, rotifera, cladocera, copepoda (see also input sensitivities in Fig. 5), 21 hidden activation units and the two output variables: *Microcystis aeruginosa* and *Stephanodiscus hantzschii*.

After 2100 training iterations, an RMSE of 0.0017 was achieved and the generalisation of the trained ANN was tested based on testing data of 1994. Figure 4 shows the visual comparison between the observed and the 4-days-ahead predicted data for *M. aeruginosa* ($r^2 = 0.68$) and *S. hantzschii* ($r^2 = 0.73$). The results indicate a high degree of accuracy in the forecasting regarding both the timing and the magnitude of dynamics of the two algal populations, which have their distinctive seasonal patterns.

This application has demonstrated that supervised feedback ANN achieve a high generalisation degree and forecasting accuracy after training by time-lagged time series data. The typical rapid growth and blooming of the blue-green algae *Microcystis* under warm and calm conditions in mid and late summer as observed in the River Nakdong in 1994 were well reflected by the predicted data in Fig. 4a. In contrast, diatoms tend to be abundant at moderate temperatures and turbulent conditions. Both observed and predicted data for *S. hantzschii* in the River Nakdong correspond well by showing highest population densities in spring and autumn (Fig. 4b).

This case study has also convincingly demonstrated the benefits of sensitivity analyses to gain insights and test hypotheses regarding ecological relationships between input and output variables. Results in Fig. 5 compare the input sensitivities of the two different algal populations that have been interpreted in great detail by Jeong *et al.* [30]. The most obvious differences between

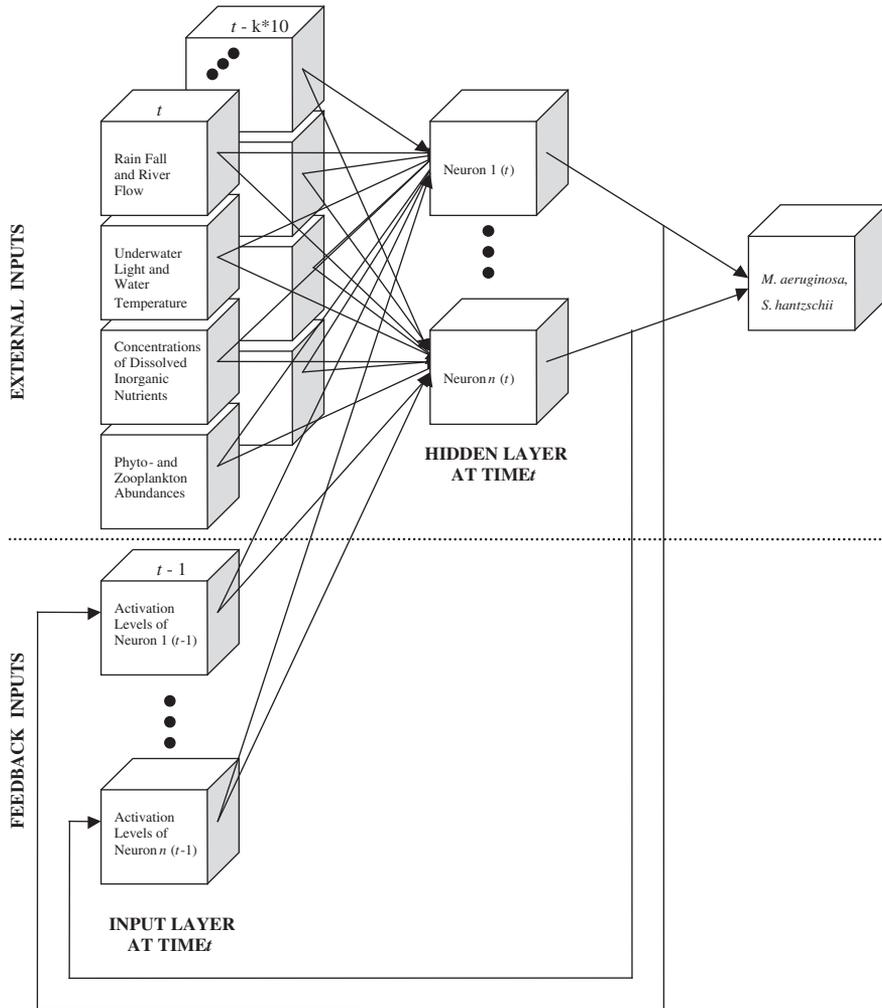


Figure 3: Supervised feedback ANN for 4-days-ahead forecasting of population densities of *Microcystis aeruginosa* and *Stephanodiscus hantzschii* in the River Nakdong (South Korea) (after Jeong *et al.* [30]).

M. aeruginosa and *S. hantzschii* can be seen in their preferred water temperature, pH and silica levels that comply with ecological theory [31, 32].

Successful applications of supervised feedback ANN have been demonstrated for time-series modelling of macroinvertebrate communities in streams [21, 33] and of phytoplankton communities in freshwater lakes and rivers [34–37].

2.3 Non-supervised ANN

Non-supervised ANN are designed to identify unknown input patterns based on similarities between inputs. The so-called self-organising maps developed by Kohonen [38] are the most

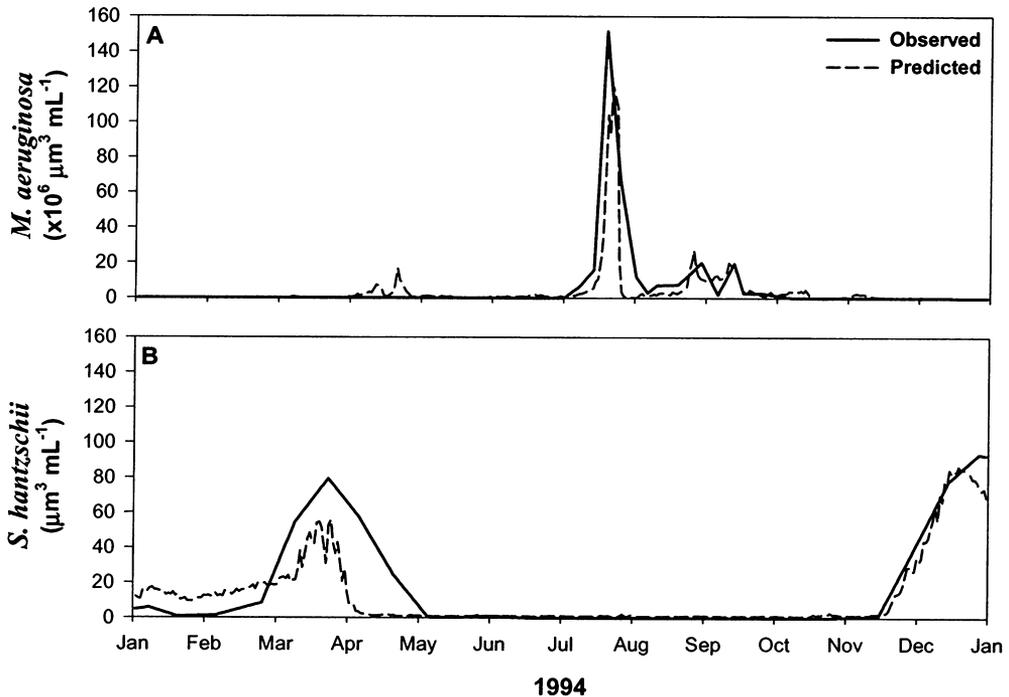


Figure 4: Four-days-ahead forecasting of population densities of *Microcystis aeruginosa* and *Stephanodiscus hantzschii* in the River Nakdong (South Korea) by means of a supervised feedback ANN (after Jeong *et al.* [30]).

popular non-supervised ANN, which can be applied to ordination, clustering and mapping of complex nonlinear data.

The principal approach of non-supervised ANN, according to Kohonen [38], is represented in a simplified manner in Fig. 6. It shows that the neurones of the non-supervised ANN learn to distinguish between similar and dissimilar features of the normalised input data, which are 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 can be expressed by Euclidean distances, which are calculated between the inputs and weights. Similarities between inputs in terms of Euclidean distances can be visualised and partitioned by the unified distance matrix (U-matrix) and the K-means map.

To illustrate opportunities of applications of non-supervised ANN to ecological time-series data, Figs 7–10 show results of a case study carried out for limnological data of Lake Kasumigaura in Japan [37]. Figure 7 represents seasonal clusters for Lake Kasumigaura as mapped by the U-matrix and K-means partitioning using the SOM Toolbox of MATLAB 5.3 [39]. The U-matrix map in Fig. 7a 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 the smallest distances belonging to a region or cluster. The black colour represents the biggest distances between neighbouring data and denotes borders between clusters. The K-means algorithm partitions the input data space into a specified number of clusters based on the U-matrix. Figure 7b represents the corresponding partitioned map for five seasons.

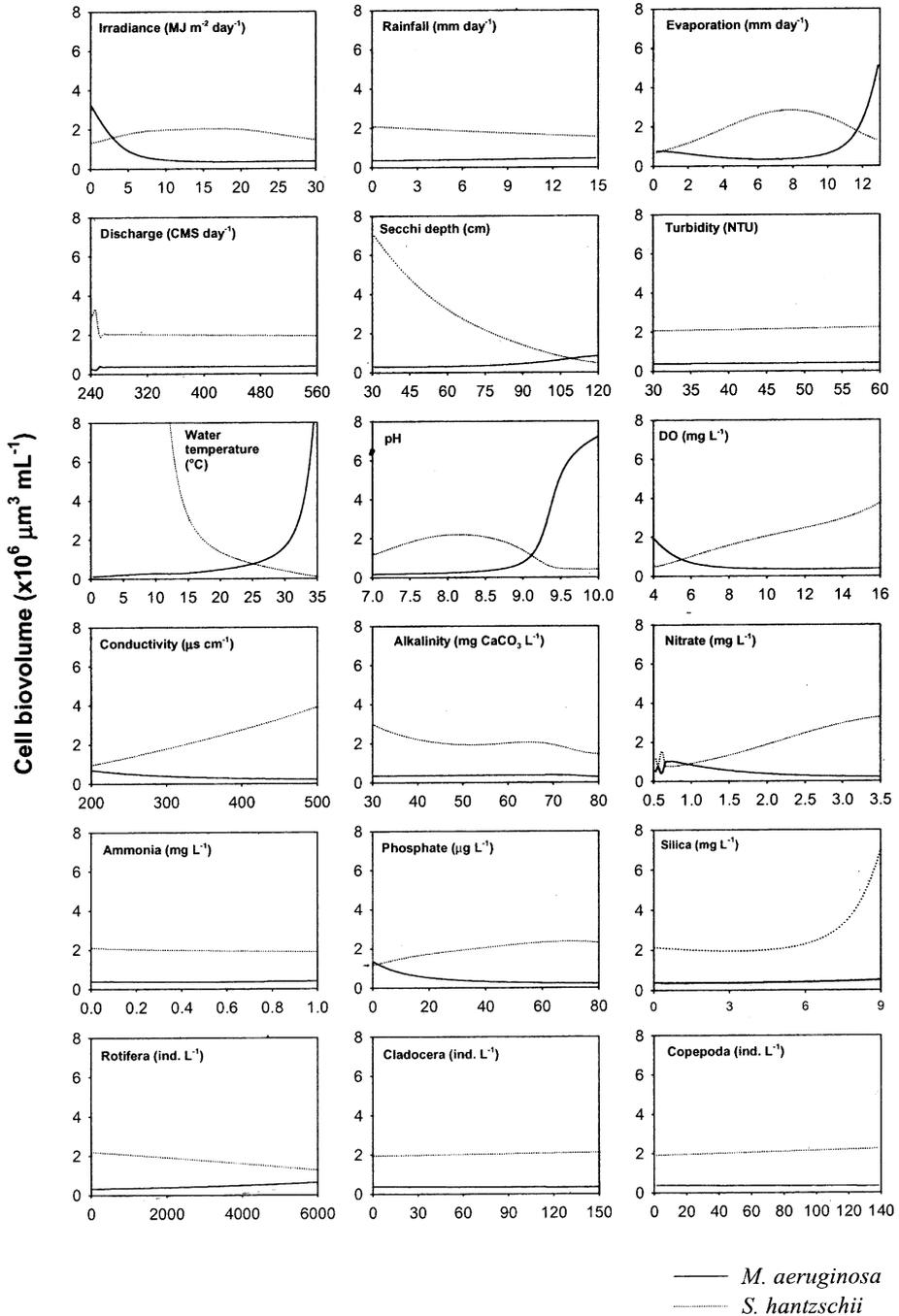


Figure 5: Input sensitivity curves for the population densities of *Microcystis aeruginosa* (dark grey) and *Stephanodiscus hantzschii* (light grey) in the River Nakdong (South Korea) by means of a supervised feedback ANN (after Jeong *et al.* [30]).

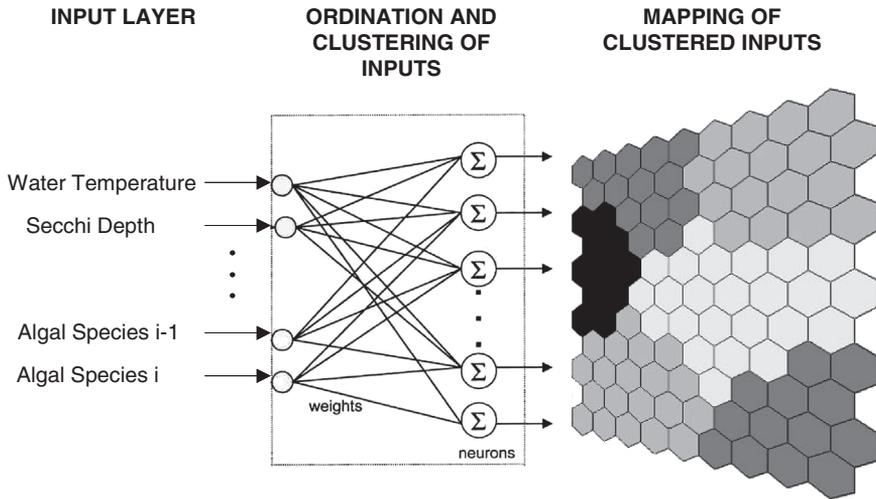


Figure 6: Conceptual diagram of the structure and functioning of non-supervised ANN.

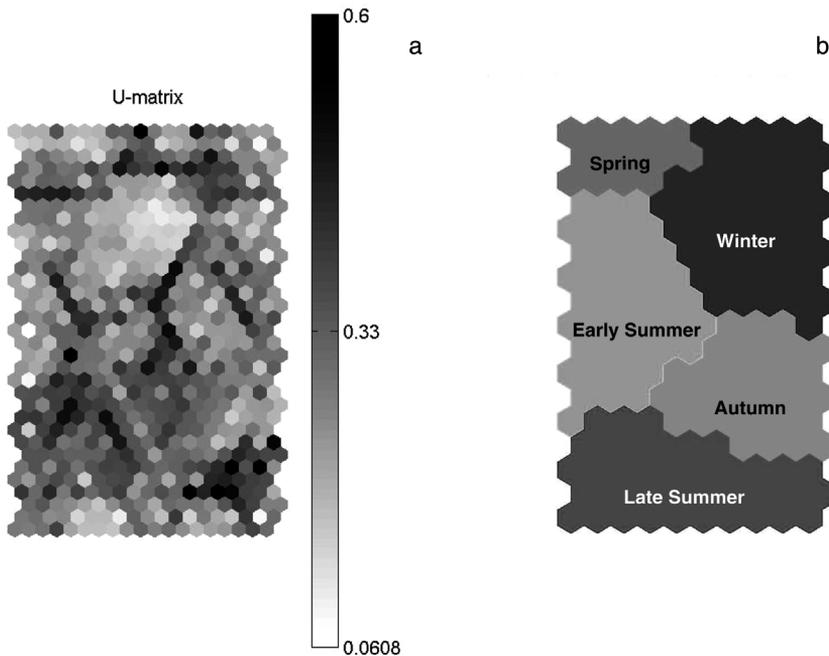


Figure 7: Ordination and clustering of seasons of Lake Kasumigaura by means of non-supervised ANN visualised as unified distance matrix map (U-matrix) (a), and as partitioned map (K-means) (b); the seasons were defined as follows: winter from 1 December, spring from 15 March, early summer from 1 June, late summer from 1 August, autumn from 1 October.

Figure 8 visualises seasonal distributions of abundances of the blue-green algae *Microcystis* and *Oscillatoria* in Lake Kasumigaura based on data for the years 1984 to 1986 (left column) and 1987 to 1989 (right column). Figure 9 represents the seasonal distributions of concentrations of $\text{NO}_3\text{-N}$ and $\text{PO}_4\text{-P}$ in Lake Kasumigaura in correspondence with the time periods differentiated in Fig. 8. Figure 8 highlights that whilst *Microcystis* declines in cell numbers by more than 50% between 1984–1986 and 1987–1989, *Oscillatoria* doubles in cell numbers. It also shows that seasonal dominance of two algal populations for the early and the late 1980s shifted for *Microcystis* from late summer to autumn, and for *Oscillatoria* from early summer to late summer. Takamura *et al.* [40] pointed out changes in $\text{NO}_3\text{-N}/\text{PO}_4\text{-P}$ ratios as possible explanations for the succession of the two blue-green algal populations during the 1980s in Lake Kasumigaura, which are indicated by the component planes in Fig. 9. From the early 1980s to the late 1980s, the $\text{NO}_3\text{-N}$ concentrations increase by 50%, whereas $\text{PO}_4\text{-P}$ concentrations dropped to 50% causing a significant change in the $\text{NO}_3\text{-N}/\text{PO}_4\text{-P}$ ratios from 8.5 to 32.

A combination of input sensitivity curves by supervised feedback ANN with component planes by non-supervised ANN proves to be an informative approach, e.g. for hypothesis testing. Whilst component planes allow mapping nonlinear relationships of output variables with pre-defined input ranges in a qualitative manner (see Fig. 10 top), input sensitivity curves reveal quantitative relationships of output variables over the whole range of input variables as learnt from training data. Both the component planes and sensitivity curves in Fig. 10 confirm theoretical assumptions that the diatom *Cyclotella* prefers low to medium water temperatures typically occurring in spring and autumn, whilst the population growth of *Microcystis* reaches the highest rates at high water temperatures in mid and late summer.

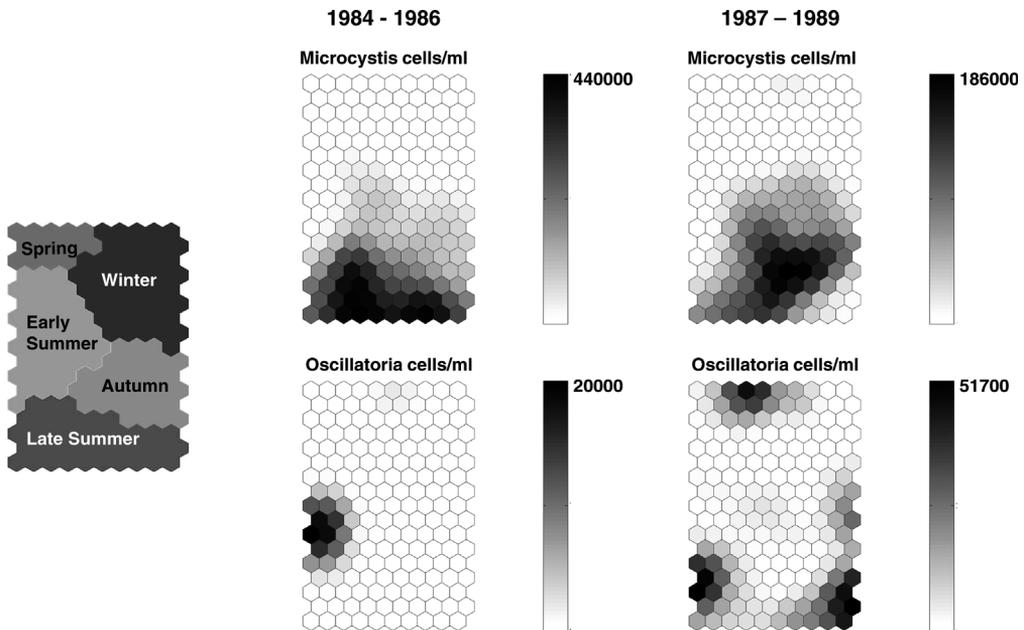


Figure 8: Component planes for seasonal abundances of *Microcystis* and *Oscillatoria* populations in Lake Kasumigaura for the years 1984–1986 (left column) and 1987–1989 (right column).

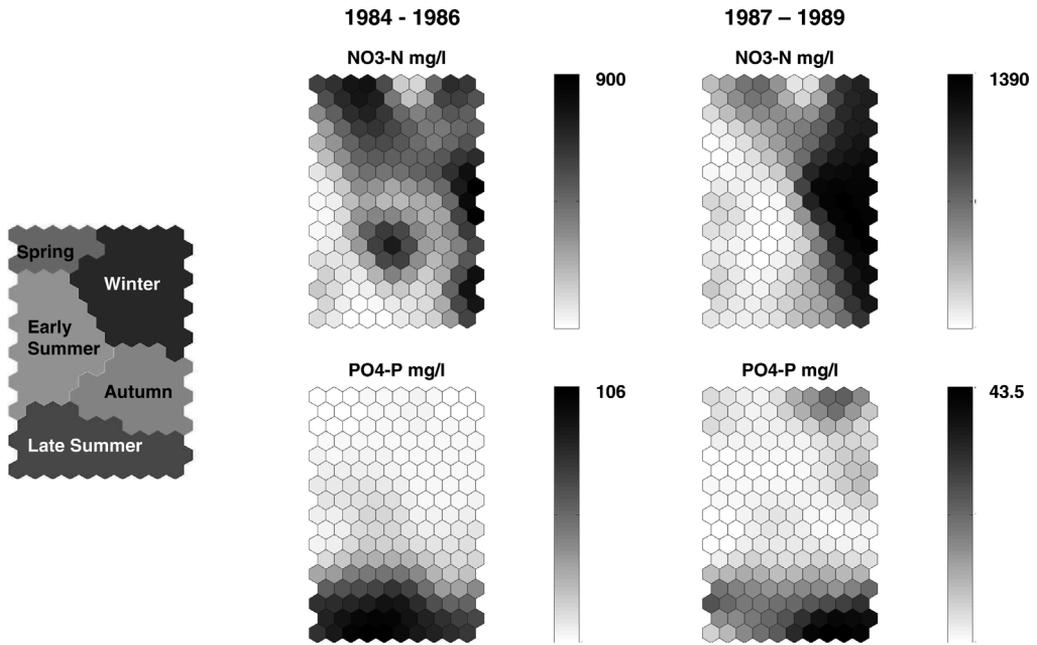


Figure 9: Component planes for seasonal concentrations of $\text{PO}_4\text{-P}$ and $\text{NO}_3\text{-N}$ in Lake Kasumigaura for the years 1984–1986 (left column) and 1987–1989 (right column).

Successful applications of non-supervised ANN have been demonstrated for cross-sectional data of macroinvertebrate communities in streams [33, 41–43] and vegetation types [14, 44] and for time-series data of plankton communities in lakes and rivers [1, 36, 37, 45].

2.4 Evolutionary algorithms

Evolutionary algorithms are adaptive methods for finding problem solutions (models, knowledge) based on principles of biological evolution by natural selection, genetic variation and ‘survival of the fittest’ (see Fig. 11). Holland [46] provided the theoretical framework for the development of genetic and EA that are being widely used for pattern recognition, forecasting, knowledge discovery, optimum control and parallel processing. Useful guidelines for history, current developments and applications of genetic and EA are provided by Goldberg [47], Mitchell [48] and Bäck *et al.* [49].

Successful implementations of EA as tools for solving complex economic and engineering problems have stimulated their application to solving ecological problems, which exhibit highest complexity. They allow inducing predictive models from ecological data sets similar to supervised ANN, but rather than lacking an explicit model representation as typical for ANN, EA are distinctively designed for assembling the explicit model represented as multivariate functions or rule sets. Therefore, EA serve as powerful tools for knowledge discovery as well.

A hybrid evolutionary algorithm (HEA) [51] has been specifically designed as a flexible tool for inducing predictive multivariate functions and rule-sets from ecological time-series data. The conceptual framework of the application of HEA to rule discovery in limnological time-series data is represented in Fig. 12. It indicates that similar to supervised ANN, the training of HEA

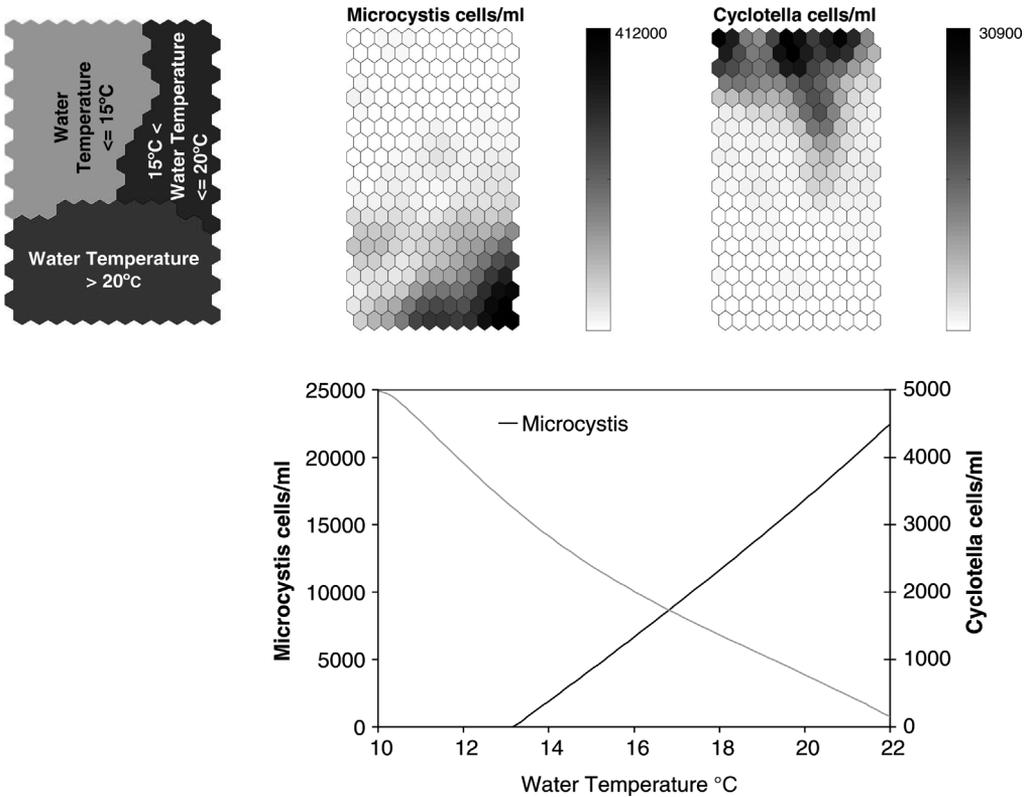


Figure 10: Component planes for water temperature preferences of *Microcystis* and *Cyclotella* populations (top) and water temperature sensitivity curves for *Microcystis* and *Cyclotella* populations (bottom) in Lake Kasumigaura for the years 1984–1993.

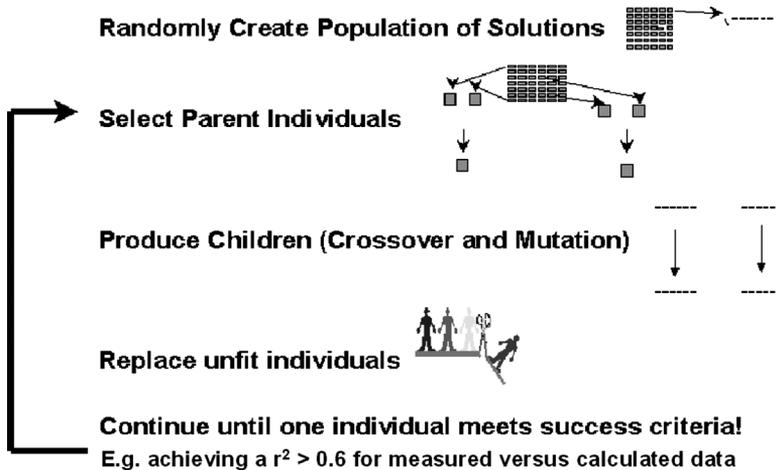


Figure 11: Conceptual diagram for the design of evolutionary algorithms (after Morrall [50]).

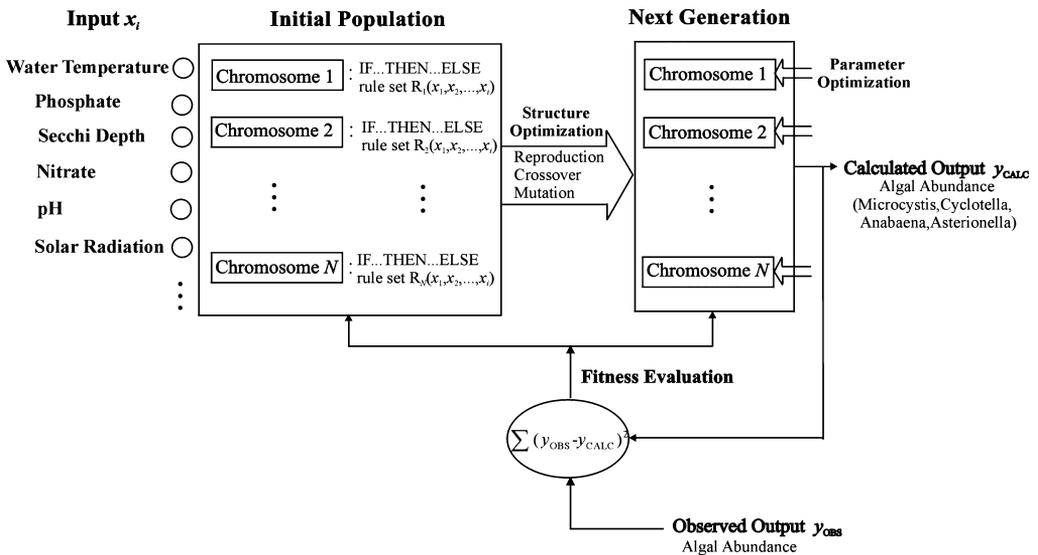


Figure 12: Conceptual framework of the application of the hybrid evolutionary algorithm HEA for rule discovery in limnological time-series data.

aims at the optimal approximation of the calculated output Y_c to the observed (desired) output Y_o . However, in contrast to supervised ANN, HEA iteratively adjusts the function or rule structure and parameter values rather than input weights to minimise the error $(Y_o - Y_c)$. A CD of the demo software version of HEA is included in this book.

The detailed algorithm for the rule discovery and parameter optimisation by HEA is shown in Fig. 13. HEA uses genetic programming (GP) to generate and optimise the structure of rule sets and a genetic algorithm (GA) [48] to optimise the parameters of a rule set. GP [52] is an extension of GA in which the genetic population consists of computer programs of varying sizes and shapes. In standard GP, computer programs can be represented as parse trees, where a branch node represents an element from a function set (arithmetic operators, logic operators, elementary functions of at least one argument), and a leaf node represents an element from a terminal set (variables, constants and functions of no arguments). These symbolic programs are subsequently evaluated by means of “fitness cases”. Fitter programs are selected for recombination to create the next generation by using genetic operators, such as crossover and mutation. This step is iterated for consecutive generations until the termination criterion of the run has been satisfied. A general genetic algorithm (GA) is used to optimise the random parameters in the rule set. More details on the design and functioning of HEA including a demo version can be found in Cao *et al.* [51].

Figures 14 and 15 illustrate the structure, input sensitivity and k -fold cross-validation of a rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass developed by HEA. The rule in Fig. 14a is the result of using 42 years of merged limnological data of the South African lakes Hartbeespoort, Roodeplaat and Rietvlei for the training of HEA. The sensitivity analysis in Fig. 14b indicates that both water temperature and Secchi depth are key driving variables for low bio-volumes of *Microcystis* of up to $14 \text{ cm}^3/\text{m}^3$ reflected by the THEN branch of the rule as well as for high biovolumes of up to $350 \text{ cm}^3/\text{m}^3$ reflected by the ELSE branch of the rule. As a result of k -fold cross-validation the parameters p_1 and p_2 have been evolved to water temperature

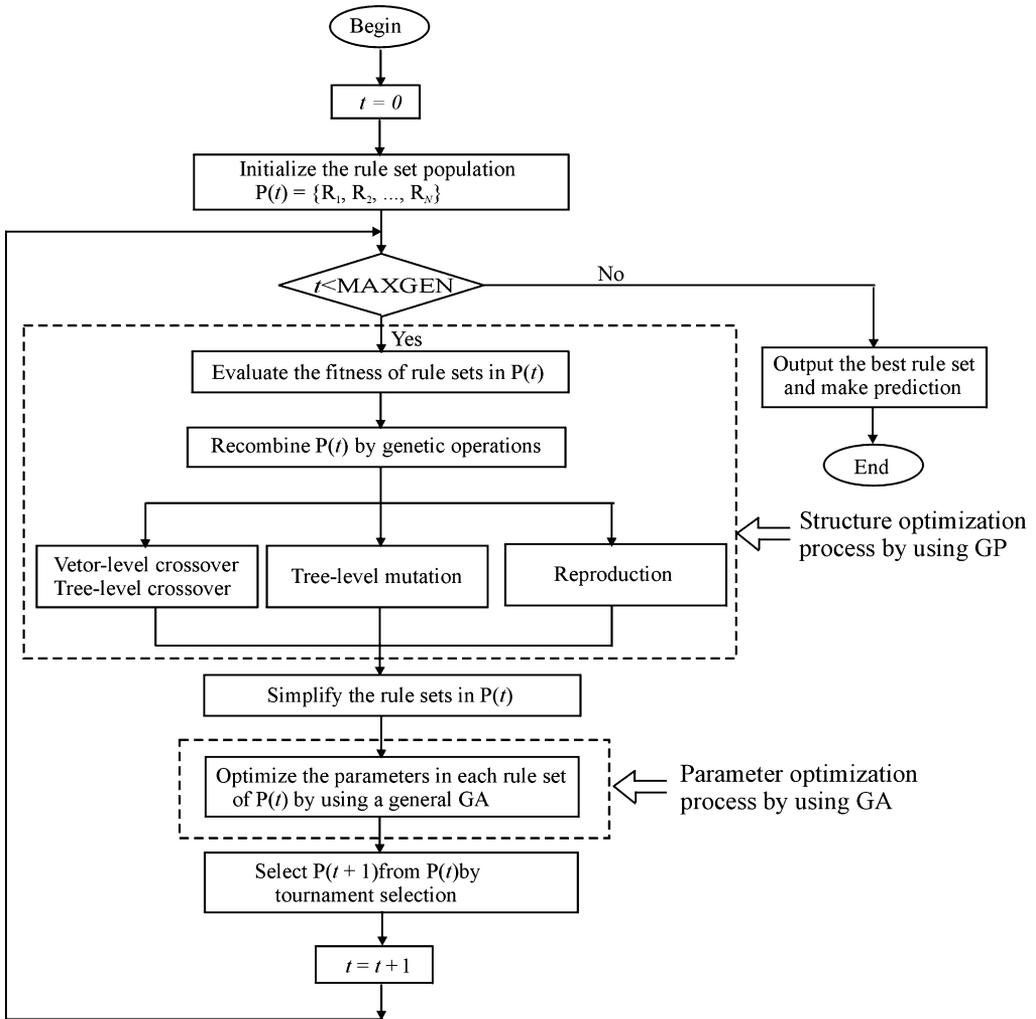


Figure 13: Flowchart of the hybrid evolutionary algorithm (HEA) for rule discovery.

functions, which provide the agent an extra mechanism for adaptation to lake specific seasonal conditions. Figure 15 demonstrates the generalisation and forecasting performance of EA by means of the rule-based agent for 7 days ahead forecasting of *Microcystis* in three South African lakes for the years 1991–2004 by achieving r^2 -values between 0.31 and 0.76.

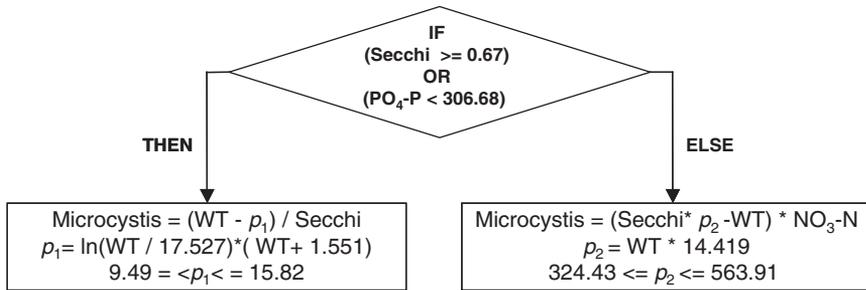
Successful applications of EA have been demonstrated for cross-sectional data of fish populations [53], macro-invertebrate communities in streams [42], for time-series data of plankton communities in lakes and rivers [54–60] [51, 61, 62] and waste water treatment [63].

2.5 Object-oriented programming

The design of object-oriented programming languages such as JAVA has largely been inspired by organising principles in ecology, such as hierarchy and taxonomy. Ecosystem simulation models



a



b

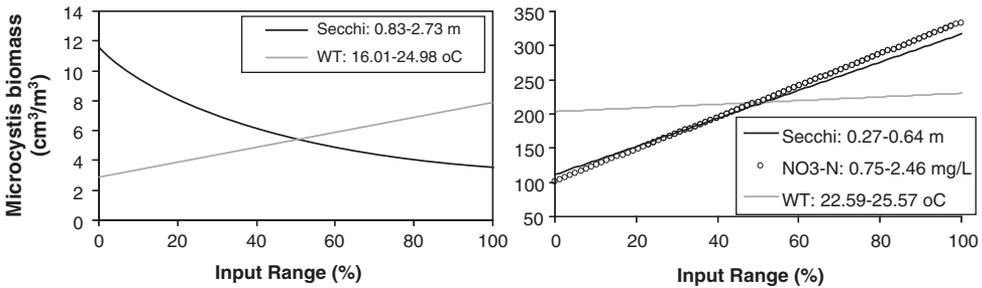


Figure 14: Structure and input sensitivity analysis of a rule-based agent for 7 days ahead forecasting of *Microcystis* biomass discovered in merged time-series data of the South African lakes Hartbeespoort, Roodeplaait and Rietvlei by HEA.

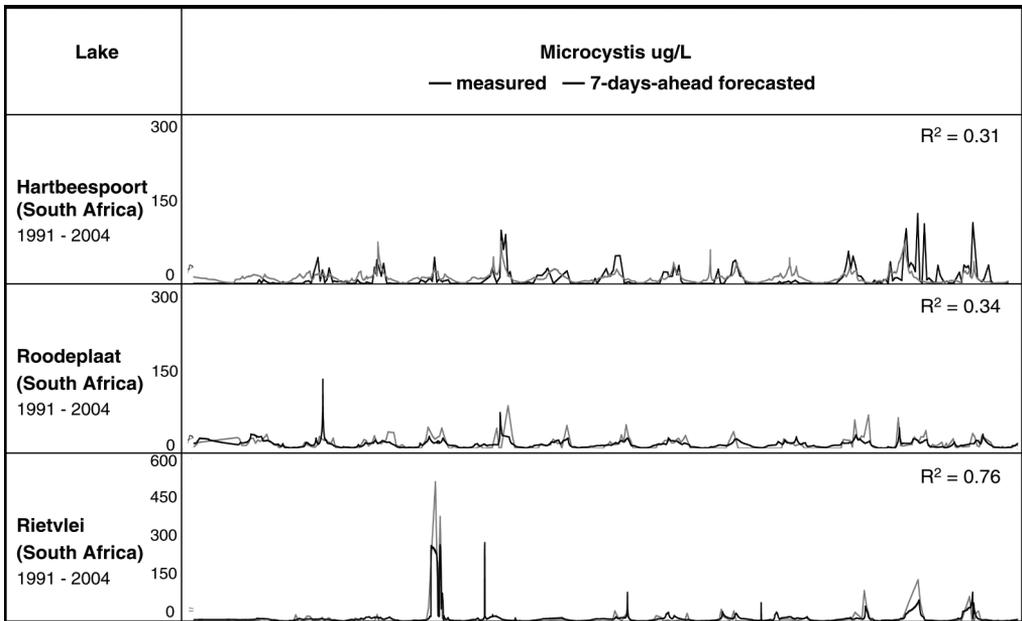


Figure 15: *k*-fold cross-validation of a rule-based agent for 7-days-ahead forecasting of *Microcystis* biomass by means of merged time-series data of the South African lakes Hartbeespoort, Roodeplaait and Rietvlei.



implemented by object-oriented programming are expected to gain improved realism and generality [64]. It allows hierarchically modularising complex ecosystem models and replacing rigid process representations by optional arrays of alternative process representations. An accordingly implemented ecosystem simulation model would also gain flexibility with regard to evolving ecological processes.

To demonstrate the advantages of object-oriented ecosystem models, the lake simulation library SALMO-OO is used as an example that has been implemented by means of JAVA.

2.5.1 Case study SALMO-OO: background

The lake model SALMO [65, 66] forms the core of the lake simulation library SALMO-OO [70] that also includes process representations for algal growth and grazing as well as zooplankton growth and mortality published previously [67–69].

Table 1 summarises the state variables and process equations provided by SALMO-OO to simulate pelagic food web dynamics and nutrient cycles. The mass balances of all state variables include influx, outflux and vertical flux during stratification. The mass balances of phosphate P and nitrate N include consumption by the three algal groups diatoms, green and blue-green algae, sediment release under anaerobic conditions as well as remineralisation resulting from algal grazing and zooplankton mortality. Specifically for nitrate, denitrification under anaerobic conditions is considered and for phosphate, sedimentation by co-precipitation, e.g. with calcite. The mass balances for the three algal groups diatoms, green and blue-green algae A[i] consider growth, grazing and sedimentation. The mass balance for herbivorous zooplankton Z considers growth, mortality by fish predation and vertical migration during stratification. The mass balance for detritus D considers grazing and sedimentation. The mass balance for dissolved oxygen O in

Table 1: State variables and processes represented by SALMO-OO.

SALMO-OO Processes	SALMO-OO State Variables					
	Phosphate (mg/m ³) d P / dt	Nitrate (g/m ³) d N / dt	Functional Algae Groups (cm ³ /m ³) d A[i] / dt	Herbivorous Zooplankton (cm ³ /m ³) d Z / dt	Detritus (mg/m ³) d D / dt	Dissolved Oxygen (mg/m ³) d O / dt
Influx	(INFL * PIN) / V	(INFL * NIN) / V	(INFL * AIN[i]) / V		(INFL * DIN) / V	(SAT * INFL) / V
Outflux	(P * OUTFL) / V	(N * OUTFL) / V	(A[i] * OUTFL) / V	(Z * OUTFL) / V	(D * OUTFL) / V	(O * OUTFL) / V
Vertical Flux between Epi- and Hypolimnion	$(P_{E,H} * VFLUX_{E,H} - P_{H,E} * VFLUX_{H,E}) / V_{H,E}$	$(N_{E,H} * VFLUX_{E,H} - N_{H,E} * VFLUX_{H,E}) / V_{H,E}$	$(A[i]_{E,H} * VFLUX_{E,H} - A[i]_{H,E} * VFLUX_{H,E}) / V_{H,E}$	$(Z_{E,H} * VFLUX_{E,H} - Z_{H,E} * VFLUX_{H,E}) / V_{H,E}$	$(D_{E,H} * VFLUX_{E,H} - D_{H,E} * VFLUX_{H,E}) / V_{H,E}$	$(P_{E,H} * VFLUX_{E,H} - P_{H,E} * VFLUX_{H,E}) / V_{H,E}$
Consumption	$\Sigma((AGRO[i]/YPA)*A[i])$	$\Sigma((AGRO[i]/YNA)*A[i])$				LO / YOA
Production						AGRO[i] / YOA * A[i]
Sediment Release	PSR	NSR				
Remineralization	$Z * ((ZMO * RATE) / YZP + ((AGRA[i] / YPA) - ZGRO / YZP) * RATP)$	$Z * ((ZMO * RATNF) / YZN + ((AGRA[i] / YNA) - ZGRO / YZN) * RATN)$				
Denitrification		NDEN				
Growth			AGRO[i] * A[i]	ZGRO * Z		
Grazing			AGRA[i] * Z		AGRA[4] * Z	
Sedimentation	KPSED * P _{T,E}		VA[i] / ZMIX * A[i]		VD / ZMIX * D	
Mortality				ZMO * Z		
Vertical Migration				ZM _{E,H} * Z _{E,H}		



Table 2: Process library for algal growth and grazing with AGRO[i]_A and AGRA[i]_A adopted from Benndorf and Recknagel [65], AGRO[i]_B and AGRA[i]_B adopted from Park *et al.* [69], AGRO[i]_C and AGRA[i]_C adopted from Hongping and Jianyi [68], and AGRO[i]_D and AGRA[i]_D adopted from Arhonditsis and Brett [67].

	SALMO-OO Algal and Zooplankton Processes			
	Benndorf and Recknagel (1982)	Park <i>et al.</i> (1974);	Hongping and Jianyi (2002)	Arhonditsis and Brett (2005)
Algal Growth	AGRO[i] _A = PHO[i] - RA[i]	AGRO[i] _B = PHO[i] - RA[i]	AGRO[i] _C = PHO[i] - RA[i]	AGRO[i] _D = PHO[i] - RA[i]
Algal Photosynthesis	$\text{PHO}[i] = \frac{((\text{PHOMAX}_i - \text{PHOMIN}_i) / (\text{TOPTA}_i * T + \text{PHOMIN}_i)) * (\text{IREDDZ} / (K_i + \text{IREDDZ})) * \text{PHONP}[i];}{\text{IREDDZ} = 1 * \exp(-\text{EPS} * \text{ZA});}$ $\text{ZA} = 0 (0.5) \text{ZMIX};$ $\text{EPS} = 0.92 - 0.46 * (\text{LTMAX} - (\text{EA}_i * \text{A}[i]));$ if $N / P \geq 0.0072$ then $\text{PHONP}[i] = \text{PHOP}[i]$ else $\text{PHONP}[i] = \text{PHON}[i];$ $\text{PHOP}[i] = P / \text{A}[i] / (K_P / \text{KAP} + P / \text{KAP} + K_P / \text{A}[i] + P / \text{A}[i]);$ $\text{PHON}[i] = N / \text{A}[i] / (K_N / \text{KAN} + N / \text{KAN} + K_N / \text{A}[i] + N / \text{A}[i]);$	$\text{PHO}[i] = \text{PHOMAX}_i * U_i * \text{PHOT}[i];$ $U_i = 3 / ((1 / \text{PHOL}[i]) + (1 / \text{PHOP}[i]) + (1 / \text{PHON}[i]));$ $\text{PHOL}[i] = (2.718 * \text{FP} / (\text{EPS} * \text{ZMIX})) * (\exp(-1 / (\text{IS} * \text{FP})) * \exp(-\text{EPS} * \text{ZMIX}) - \exp(-1 / (\text{IS} * \text{FP})))));$ $\text{EPS} = \text{KW} + (\text{EA}_i * \text{A}[i]);$ $\text{FP} = 12 - 4 * \cos(2\pi t / 360);$ $\text{PHOT}[i] = T^{1.2} \exp(T2 * (1 - T1));$ $T1 = (\text{TMAXA}_i - T) / (\text{TMAXA}_i - \text{TOPTA}_i);$ $T2 = T3 * (1 + \text{SQRT}(1 + 40 / T3)) / 400;$ $T3 = \ln(Q_{10}) * (\text{TMAXA}_i - \text{OPTA}_i);$ $\text{PHOP}[i] = P / (K_P + P);$ $\text{PHON}[i] = N / (K_N + N)$	$\text{PHO}[i] = \text{PHOMAX}_i * [1 * (1 - \exp(-\text{EPS} * \text{ZMIX})) / (\text{EPS} * \text{ZMIX})] / \text{IS} * \exp(1 - [1 * (1 - \exp(-\text{EPS} * \text{ZMIX})) / (\text{EPS} * \text{ZMIX})] / \text{IS}) * (T / \text{TOPTA}_i * \exp(1 - (T / \text{TOPTA}_i))) * (N / (K_N + N)) * (P / (K_P + P));$ $\text{EPS} = \text{KW} + (\text{EA}_i * \text{A}[i]);$ $\text{FP} = 12 - 4 * \cos(2\pi t / 360)$	$\text{PHO}[i] = \text{PHOMAX}_i * [(2.718 * \text{FP} / (\text{EPS} * \text{ZMIX})) * (\exp(-1 / (\text{IS} * \text{FP})) * \exp(-\text{EPS} * \text{ZMIX})) - \exp(-1 / (\text{IS} * \text{FP})))] * \text{PHOT}[i] * \min(N / (K_N + N), P / (K_P + P));$ $\text{EPS} = \text{KW} + (\text{EA}_i * \text{A}[i]);$ $\text{FP} = 12 - 4 * \cos(2\pi t / 360);$ if $T \geq \text{TOPTA}_i$ then $\text{PHOT}[i] = \exp(-2.3 * (((T - \text{TOPTA}_i) / (\text{TMAXA}_i - \text{TOPTA}_i))^2))$ else $\text{PHOT}[i] = \exp(-2.3 * (((T - \text{TOPTA}_i) / (\text{TMAXA}_i - \text{TMINA}_i))^2))$
Algal Respiration	$\text{RA}[i] = (\text{RATOPT}_i - \text{RATMIN}_i) / (\text{TOPTA}_i * T + \text{RATMIN}_i + 0.3 * \text{PHO}[i])$	$\text{RA}[i] = \text{RATOPT}_i * \text{PHOT}_i * \text{PHOMAX}_i$	$\text{RA}[i] = \text{RATMIN}_i * \exp(0.038 * T)$	$\text{RA}[i] = \text{RATOPT}_i * \exp(0.07 * (T - \text{TOPTA}_i))$
Algal Grazing	$\text{AGRA}[i]_A = (\text{GMAX} - \text{GMIN}) * (\exp(-R * \text{abs}(\ln(T / \text{TOPTZ}))) + \text{GMIN}) * (\text{A}[i] * \text{PFA}_i) / Z + (\text{KAG} / \text{KZ} + (\text{A}[i] * \text{PFA}_i) / \text{KZ} + \text{KAG} / \text{KZ} + (\text{A}[i] * \text{PFA}_i) / Z)$	$\text{AGRA}[i]_B = \text{GMAX} * ((\text{PFA}_i * \text{A}[i]) / (\text{KAG} + (\text{PFA}_i * \text{A}[i])) * \text{GT};$ $\text{GT} = T^{1.2} * \exp(T2 * (1 - T1));$ $T1 = (\text{TMAXZ} - T) / (\text{TMAXZ} - \text{TOPTZ});$ $T2 = 3 * (1 + \text{SQRT}(1 + 40 / T3)) / 400;$ $T3 = \ln(Q_{10}) * (\text{TMAXZ} - \text{TOPTZ})$	$\text{AGRA}[i]_C = \text{GMAX} * (((\text{PFA}_i * \text{A}[i]) - \text{FMIN}) / \text{KAG} + (\text{PFA}_i * \text{A}[i]) - \text{FMIN}) * \text{PFA}_i * (\text{A}[i] / (\text{PFA}_i * \text{A}[i]))$	$\text{AGRA}[i]_D = ((\text{GMAX} * \text{PFG} * \text{A}[i]) / (\text{KAG} + (\text{PFG} * \text{A}[i])) * \text{GT}$ $\text{PFG} = \text{PFA}_i * \text{A}[i] / (\text{PFA}_i * \text{A}[i])$ if $T \geq \text{TOPTZ}$ then $\text{GT} = \exp(-2.3 * (((T - \text{TOPTZ}) / (\text{TMAXZ} - \text{TOPTZ}))^2))$ else $\text{GT} = \exp(-2.3 * (((T - \text{TOPTZ}) / (\text{TMAXZ} - \text{TMINZ}))^2))$
Zooplankton Growth	$\text{ZGRO} = (\text{AZMAX} - (\text{AZMAX} - \text{AZMIN} / \text{GMAX} * G_{T,E,H})) * (G_{i,T,E,H} * \text{UAZ}_i - \text{RZ}_{T,E,H})$	$\text{ZGRO} = (\Sigma \text{AGRA} * \text{Fe}_i) - (\text{RZMIN} + 0.2 * ((\Sigma (\text{PF}_i * \text{A}_i)) / (\text{KAG} + \Sigma (\text{PF}_i * \text{A}_i))) * \text{GT})$	$\text{ZGRO} = (\text{AZMAX} * (\Sigma (\text{PAF}_i * \text{A}_i) - \text{FMIN}) / (\text{KAG} + \Sigma (\text{PAF}_i * \text{A}_i) - \text{FMIN})) - (\text{RZOPT} * \exp(1.066 * T) + 0.04 * \text{AGRA})$	$\text{ZGRO} = ((1 * \text{FQ}) / (20 + \text{FQ}) * \text{GT} * (\Sigma (\text{AGRA}) + \text{AGRA}_i)$ $\text{FQ} = \Sigma ((0.1 * \text{A}_i), (0.9 * \text{A}_i), (0.7 * \text{A}_i), (0.5 * \text{A}_i)) - (\text{RZOPT} * \exp(0.1 * (T - \text{TOPTZ})))$
Zooplankton Mortality	$\text{ZMO} = (\text{MOMIN} + \text{MOT} * T_{T,E,H}) * Z_{T,E,H} / (\text{KMO} + Z_{T,E,H})$		$\text{ZMO} = \text{MO}$	$\text{ZMO} = (\text{MORTZ} * Z^2) / (\text{KMO} + Z^2)$

the hypolimnion during summer considers vertical flux during stratification, production by algal photosynthesis and consumption by organic matter. Tables A1 and A2 in the appendix provide definitions of variables and parameters.

Table 2 details the alternative process equations for algal photosynthesis, respiration and grazing, as well as zooplankton growth and mortality, which are implemented as library in SALMO-OO. Whilst the equations by Benndorf and Recknagel [65] were designed and validated for oligo- to eutrophic lakes in Germany, the equations by Park *et al.* [69] were applied to some of the deep meso to eutrophic Great Lakes in the USA. The equations by Hongping and Jianyi [68] were designed and tested for the shallow eutrophic West Lake in China, whilst Arhonditsis and Brett [67] developed their model for Lake Washington in the USA.

Table 3 and Fig. 16 illustrate the way in which SALMO-OO is applied to different lake categories. Table 3 shows the three lake categories covering six lakes for which specific structures of SALMO-OO have been identified and tested. Figure 16 highlights the best performing model structure found for the warm-monomictic hypertrophic South African lakes by black coloured



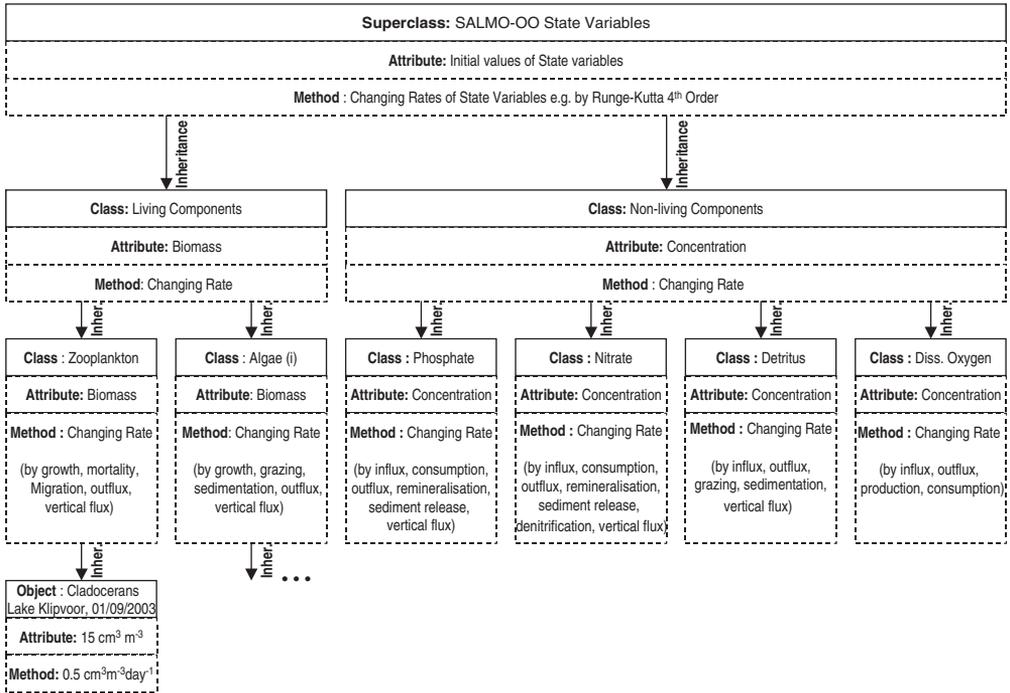


Figure 17: Conceptual diagram of the object-oriented implementation of SALMO-OO by JAVA.

hierarchically modularised into *classes* and *objects* that *inherit attributes* and *methods* from the next upper level of the hierarchy where a *class* is defined as a template for the creation (*instantiation*) of *objects* that share a common structure and behaviour. SALMO-OO was considered as the superclass at the uppermost level with attributes defined in general terms by initial values of the state variables, and methods defined by changing rates of state variables as calculated, e.g. by Runge–Kutta’s fourth-order numerical solution of ordinary differential equations [70]. The so defined and specified attributes and methods are passed on by inheritance to the classes at the next lower levels of the hierarchy up to the objects at the bottom level. The objects represent the instantaneous values of attributes and methods at each single time step relevant for the class they belong to. The instantiation of cladocerans in Lake Klipvoor on 1 September 2003 was used as an example for an object shown in Fig. 17 that materialised in the attribute of $15 \text{ cm}^3/\text{m}^3$ and the method of $0.5 \text{ cm}^3/\text{m}^3/\text{day}$.

2.5.3 Case study SALMO-OO: validation

Figures 18 and 19 document the visual validation of the simulation results for the three main algal groups, total algal biomass and phosphate of the lakes Hartbeespoort and Klipvoor. The simulation results for herbivorous zooplankton are shown as well, but there were no observed data available. Table 4 summarises the error statistics of the validation results.

Good validation results have been achieved for Klipvoor with reasonable correspondence between calculated and measured data for total algae, blue-green algae and diatoms. Whilst calculated phosphate concentrations matched well the observed data in spring and summer, they

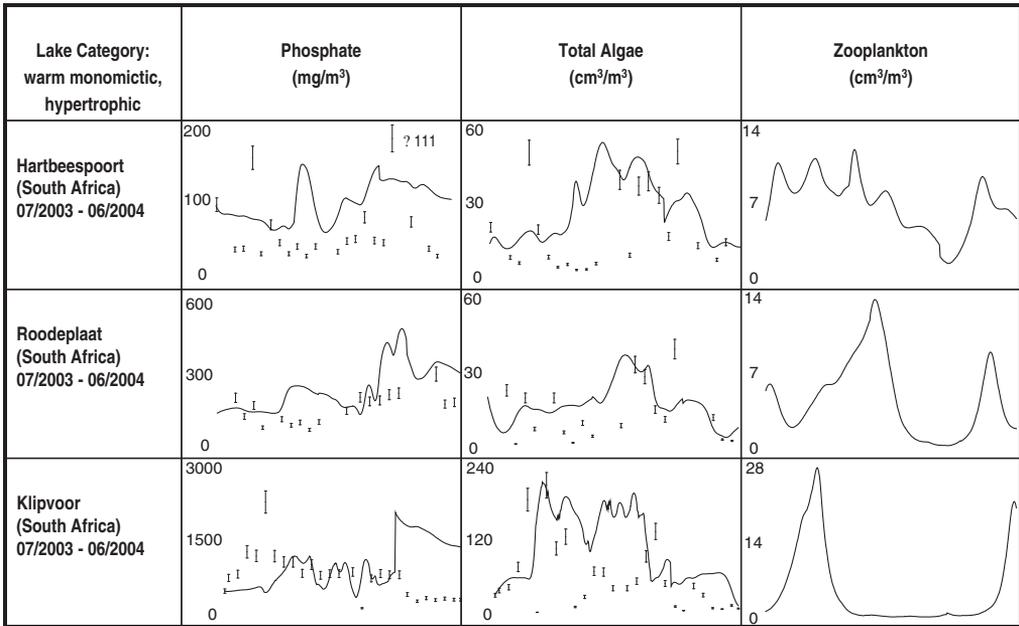


Figure 18: Validation of the best performing model structure of SALMO-OO for the warm-monomictic hypertrophic lakes Hartbeespoort, Roodeplaait and Klipvoor as specified in Table 3 and Fig. 15 based on phosphate, total algae and zooplankton (measured data from Van Ginkel *et al.*[71]).

were overestimated for autumn (see Fig. 18). The calculated biomass of green algae (Fig. 19) was heavily overestimated all year compared with the observed data. Similar validation results were produced by SALMO-OO for Roodeplaait with a reasonably good average r^2 for the phosphate and algal data. In particular, the observed seasonal dynamics of not only phosphate and total algae (Fig. 18) but also blue-green algae and diatoms (Fig. 19) were simulated in a realistic range.

Compared with the two other lakes, the results for Hartbeespoort were worse in terms of visual validation and error statistics for phosphate and total algae. The measured data show extreme ranges of oscillations between consecutive sampling points in particular in spring and late summer that hardly can be matched by SALMO-OO. Nevertheless, the simulated data fall nicely within the range of the measurements producing quite good results for blue-green algae and diatoms (Fig. 19) but do not catch up with very fast gradients as temporarily observed.

The case study SALMO-OO has demonstrated that object-oriented implementation of process-based ecosystem models allows developing flexible simulation libraries by replacing rigid process representations by optional arrays of alternative process representations. Such a library can be step-wise extended towards a comprehensive knowledge base for particular ecosystems such as lakes. The programming language JAVA proved to be highly suitable for the implementation of the simulation library SALMO-OO.

The testing of SALMO-OO for two lakes belonging to the same lake category has revealed good validation results in terms of realism and generality. Further progress with regard to improved accuracy is expected by extending the process library by additional alternative process representations and multi-objective parameter optimisation by means of EA [72].

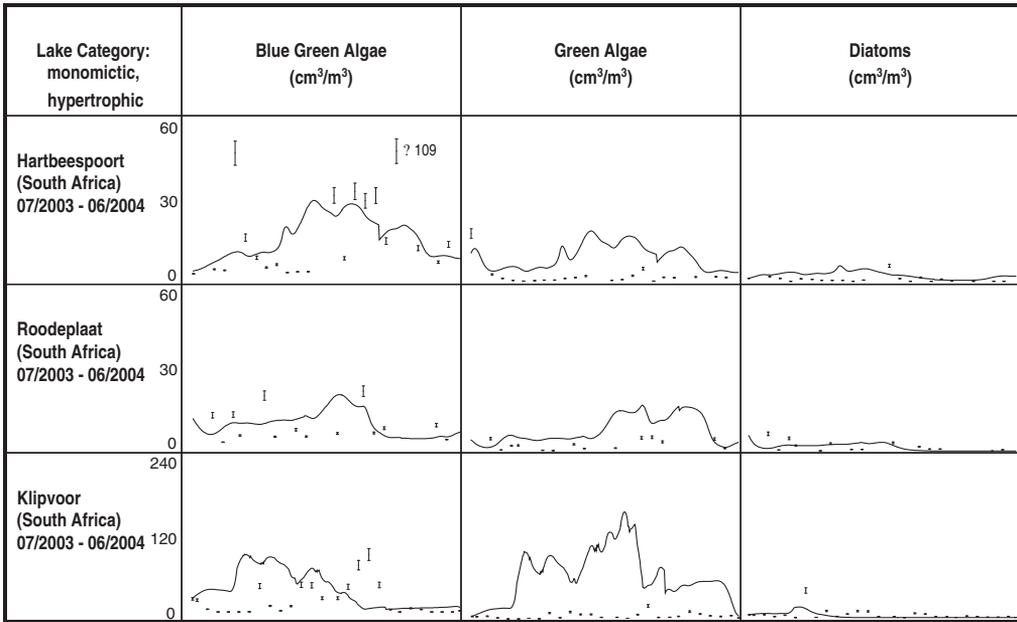


Figure 19: Validation of the best performing model structure of SALMO-OO for the warm-monomictic hypertrophic lakes Hartbeespoort, Roodeplaat and Klipvoor as specified in Table 3 and Fig. 16 based on the algal groups diatoms, blue-green and green algae (measured data from Van Ginkel *et al.*[71]).

Table 4: Error statistics for the validation of the state variables phosphate, total algae, diatoms, green and blue-green algae of the six lakes based on the correlation coefficient r^2 of the linear regression and the root mean squared error (RMSE).

		Phosphate P	Total algal biomass A	Diatom biomass A[1]	Green algal biomass A[2]	Blue-green algal biomass A[3]
Hartbeespoort	r^2	0.01	0.05	0.001	0.03	0.07
	RMSE	52.8	21.9	2.1	8.94	20.64
Roodeplaat	r^2	0.25	0.13	0.03	0.11	0.03
	RMSE	105.95	10.64	1.7	7.18	6.04
Klipvoor	r^2	0.3	0.12	0.21	0.001	0.007
	RMSE	75.8	75.8	5.6	73.04	42.4

3 Conclusions

Bio-inspired algorithms are not rigid, but dynamically evolving from data ‘bottom up’ by principles of neuronal learning, natural selection or hierarchical inheritance. Their applications in the context of ecological informatics suit well the nature of complex ecological data in achieving adaptive solutions. Further progress is imminent by hybridisation of bio-inspired algorithms.



Yao *et al.* [73] demonstrated great achievements by applications of evolving neural networks to ecological data. Parallel evolutionary computation [74] provides more realistic approaches for studying multiple food web interactions, co-evolution and multi-objective optimisation within communities. One promising approach towards structurally dynamic modelling [75] is the integration of evolutionary computation into object-oriented simulation libraries.

Appendix

Table A1: SALMO-OO variables.

Variable	Definition
A[i]	Algal biomass of the mixed lake, epi- or hypolimnion ($\text{g wet weight m}^{-3}$); i = 1: diatoms; i = 2: green algae; i = 3: blue-green algae
AGRA[i]	Algal grazing of the mixed lake, epi- or hypolimnion by herbivorous zooplankton ($\text{g wet weight m}^{-3} \text{ day}^{-1}$)
AGRO[i]	Algal growth of the mixed water body/epilimnion/hypolimnion ($\text{g wet weight m}^{-3} \text{ day}^{-1}$)
AIN[i]	Algal biomass in the inflow to the mixed lake or epilimnion (g m^{-3})
D	Detritus of the mixed lake, epi- or hypolimnion (g m^{-3})
DIN	Detritus concentration of inflow to the mixed lake, epi- or hypolimnion (g m^{-3})
EPS	Light extinction of the mixed lake, epi- or hypolimnion (m^{-1})
FP	Photo period
G	Zooplankton Ingestion rate of the mixed lake, epi- or hypolimnion ($\text{g g}^{-1} \text{ day}^{-1}$)
GT	Temperature term of the ingestion rate in the dark of the mixed lake, epi- or hypolimnion
I	Photosynthetic active solar radiation ($\text{J cm}^{-2} \text{ day}^{-1}$)
INFL	Water inflow to the mixed water body/epilimnion/hypolimnion ($\text{m}^3 \text{ day}^{-1}$)
IREZDZ	Photosynthetic active light at depth Z(J) of the mixed lake, epi- or hypolimnion ($\text{J cm}^{-2} \text{ day}^{-1}$)
KAN	Half-saturation between photosynthesis and algal biomass at nitrogen limitation (g m^{-3})
KAP	Half-saturation between photosynthesis and algal biomass at phosphorus limitation (g m^{-3})
KPSED	Coprecipitation of phosphate by calcite (day^{-1})
KZ	Half-saturation value of the inverse relationship of between zooplankton ingestion and zooplankton biomass (g m^{-3})
LO	Load of organic matter consuming oxygen in the hypolimnion ($\text{g m}^{-3} \text{ day}^{-1}$)
N	Inorganic nitrogen of the mixed lake, epi- or hypolimnion (g m^{-3})
NDEN	Denitrification in the hypolimnion ($\text{g m}^{-3} \text{ day}^{-1}$)
NIN	Nitrate concentration in the inflow to the mixed lake, epi- or hypolimnion (g m^{-3})
NSR	Nitrogen release from sediments of the mixed lake, epi- or hypolimnion ($\text{g m}^{-3} \text{ day}^{-1}$)
O	Oxygen concentration of the epi- or hypolimnion (g m^{-3})
OUTFL	Water outflow from the mixed lake, epi- or hypolimnion ($\text{m}^3 \text{ day}^{-1}$)

Table A1: SALMO-OO variables (*Continued*).

Variable	Definition
P	Phosphate concentration of the mixed lake, epi- or hypolimnion (mg m^{-3})
PIN	Phosphate concentration in the inflowing water to the Phosphate ($\text{m}^3 \text{day}^{-1}$)
PHO[i]	Photosynthesis rate of algae (day^{-1})
PHOL[i]	Light dependence of algal specific photosynthesis of the mixed lake, epi- or hypolimnion
PHON[i]	Nitrate dependence of algal specific photosynthesis of the mixed lake, epi- or hypolimnion
PHONP[i]	Relationship between algal specific photosynthesis and the primary limiting nutrient of the mixed lake, epi- or hypolimnion
PHOP[i]	Phosphate dependence of algal specific photosynthesis of the mixed lake, epi- or hypolimnion
PHOT[i]	Temperature dependence of algal specific photosynthesis of the mixed lake, epi- or hypolimnion
PSR	Phosphorus release from sediments of the mixed lake, epi- or hypolimnion ($\text{mg m}^{-3} \text{day}^{-1}$)
RA[i]	Respiration rate of algae (day^{-1})
SAT	Oxygen concentration at 100% saturation (g m^{-3})
T	Mean water temperature of the mixed lake, epi- or hypolimnion ($^{\circ}\text{C}$)
Ut[i]	Normalised function for photosynthesis limitation by light and nutrients
V	Volume of the mixed lake, epi- or hypolimnion (m^3)
VFLUX	Water exchange between epi- and hypolimnion ($\text{m}^3 \text{day}^{-1}$)
Z	Zooplankton biomass of the mixed lake, epi- or hypolimnion ($\text{g wet weight m}^{-3}$)
ZA	Actual depth (m)
ZGRO	Zooplankton growth of the mixed lake, epi- or hypolimnion ($\text{g wet weight m}^{-3} \text{day}^{-1}$)
ZMI	Zooplankton migration from hypo- to epilimnion ($\text{g wet weight m}^{-3} \text{day}^{-1}$)
ZMIX	Mean mixing depth of the mixed lake, epi- or hypolimnion (m)

Table A2: SALMO-OO parameters.

Parameter	Definition	Values (i = 1: diatoms, i = 2: green, i = 3: blue-green)
EA _i	Light extinction by algal biomass	0.1, 0.1, 0.08
FMIN	Minimum algal biomass available for grazing (g m^{-3})	0.05
GMAX	Maximum grazing rate by zooplankton (day^{-1})	1.3
GMIN	Minimum grazing rate by zooplankton (day^{-1})	0.26
IS	Saturated light intensity ($\text{J cm}^{-2} \text{day}^{-1}$)	1800
KI _i	Half-saturation constant of light for photosynthesis (g m^{-3})	29, 29, 28
KW	Extinction coefficient of water	0.2
KAG	Half-saturation constant of zooplankton for algal grazing (g m^{-3})	5



Table A2: SALMO-OO parameters (*Continued*).

Parameter	Definition	Values (i = 1: diatoms, i = 2: green, i = 3: blue-green)
KN _i	Half-saturation constant of the relationship between nitrogen and photosynthesis at minimum algal biomass (g m ⁻³)	0.0123/0.0095/0.0123
KP _i	Half-saturation constant of the relationship between phosphorus and photosynthesis at minimum algal biomass (g m ⁻³)	18/16/27
LTMAX	Maximum light transmission at 5m depth	lake specific
PFA _i	Grazing preference factor for functional algae group (-)	1.0, 1.0, 0.3
PHOMAX _i	Maximum photosynthesis under optimal conditions (day ⁻¹)	2.37/3.3/2.37
PHOMIN _i	Minimum photosynthesis at low temperature (day ⁻¹)	0.17/0.35/0
Q ₁₀	Temperature coefficient	1.9
RATMIN _i	Minimum algal respiration (day ⁻¹)	0.02, 0.02, 0
RATN	Dissolved nitrogen to total nitrogen ratio in zooplankton excrements	0.7
RATOPT _i	Optimum algal respiration (day ⁻¹)	0.06, 0.06, 0.057
RATP	Dissolved phosphorus to total phosphorus ratio in zooplankton excrements	0.7
TMAXA _i	Maximum temperature for algal growth (°C)	35, 35, 45
TMAXZ	Minimum temperature for Zooplankton grazing (°C)	35
TMINA _i	Minimum temperature for algal growth (°C)	0
TMINZ	Maximum temperature for Zooplankton grazing (°C)	0
TOPTA _i	Optimum temperature for algal growth (°C)	21, 23, 30
TOPTZ	Optimum temperature for zooplankton growth (°C)	20
VD	Sinking velocity of detritus (m/d)	0.2
VA _i	Sinking velocity of algae (m/d)	0.017/0.016/0.025
YNA _i	Nitrogen yield coefficient of algae (g wet weight g ⁻¹ N)	57, 57, 57
YOA	Oxygen equivalent of algal biomass (g wet weight g ⁻¹ O ₂)	3.75
YPA _i	Phosphorus yield coefficient of algae (g wet weight mg ⁻¹ P)	0.8, 0.41, 1

References

- [1] Recknagel, F. (ed.), *Ecological Informatics. Scope, Techniques and Applications*, 2nd edn, Springer-Verlag: Berlin, 2006.
- [2] Goonatilake, S. & Khebbal, S., *Intelligent Hybrid Systems*, Wiley: Chichester, New York, 1995.
- [3] Swartzman, G.L. & Kaluzny, S.P., *Ecological Simulation Primer*, Macmillan Publishing Company: New York, 1987.



- [4] Weiss, M. & Kulikowski, C., *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning and Expert Systems*, Morgan Kaufmann: Heidelberg, 1990.
- [5] Hornik, K., Stinchcombe, M. & White, W., Multilayer feedforward networks are universal approximators. *Neural Networks*, **2(5)**, pp. 359–366, 1989.
- [6] Minski, M.L. & Pappert, S., *Perceptrons*, MIT Press: Cambridge, 1969.
- [7] Rummelhart, D.E., Hinton, G.E. & Williams, R.J., Learning representations by back-propagation errors. *Nature*, **323**, pp. 533–536, 1986.
- [8] Lek, S., Delacosta, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S., Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling*, **90**, pp. 39–52, 1996.
- [9] Walley, W.J. & Fontana, V.N., Neural network predictors of average score per taxon and number of families at unpolluted river sites in Great Britain. *Water Research*, **32(3)**, pp. 613–622, 1998.
- [10] Huang, H., Recknagel, F., Marshall, J. & Choy, S., Predictive modelling of macroinvertebrate assemblages for stream habitat assessments in Queensland (Australia). *Ecological Modelling*, **146(1–3)**, pp. 195–206, 2001.
- [11] Huang, W. & Foo, S., Neural network modelling of salinity variation in Apalachicola River. *Water Research*, **36**, pp. 356–362, 2002.
- [12] Scardi, M., Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecology and Progress Series*, **139**, pp. 289–299, 1996.
- [13] Karul, C. & Soyupak, S., A comparison between neural network based and multiple regression models for chlorophyll-*a* estimation. *Ecological Informatics. Understanding Ecology by Biologically-Inspired Computation*, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 249–264, 2003.
- [14] Foody, G., Soft mapping of coastal vegetation from remotely sensed imagery with a feed-forward neural network. *Artificial Neuronal Networks: Application to Ecology and Evolution*, eds S. Lek & J.F. Guegan, Springer-Verlag: Berlin, pp. 45–56, 2000.
- [15] Lusk, J.J., Guthery, F.S. & DeMaso, S.J., Northern bobwhite (*Colinus virginianus*) abundance in relation to yearly weather and long-term climate patterns. *Ecological Modelling*, **146**, pp. 3–15, 2001.
- [16] Aoki, I. & Komatsu, T., Analysis and prediction of the fluctuation of sardine abundance using a neural network. *Oceanologica Acta*, **20(1)**, pp. 81–88, 1997.
- [17] Aoki, I., Komatsu T. & Hwang, K., Prediction of response of zooplankton biomass to climatic and oceanic changes. *Ecological Modelling*, **120(2–3)**, pp. 261–270, 1999.
- [18] Reick, C.H., Grünewald, A. & Page, B., Multivariate time series prediction of marine zooplankton by artificial neural networks, *Ecological Informatics. Understanding Ecology by Biologically-Inspired Computation*, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 313–327, 2003.
- [19] Poff, N.L., Tokar, S. & Johnson, P., Stream hydrological and ecological response to climate change assessed with an artificial neural network. *Limnology and Oceanography*, **41(5)**, pp. 857–863, 1996.
- [20] Schleiter, I.M., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K.D., Schmidt, H.H. & Werner, H., Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. *Ecological Modelling*, **120**, pp. 271–286, 1999.
- [21] Chon, T.S., Park, Y.S. & Cha, E.Y., Patterning of community changes in benthic macroinvertebrates collected from urbanized streams for the short time prediction by temporal

- artificial neural networks, *Artificial Neuronal Networks: Application to Ecology and Evolution*, eds S. Lek & J.F. Guegan, Springer-Verlag: Berlin, pp. 99–114, 2000.
- [22] Recknagel, F., French, M., Harkonen, P. & Yabunaka, K., Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling*, **96(1–3)**, pp. 11–28, 1997.
- [23] Yabunaka, K.-I., Hosomi, M. & Murakami, A., Novel application of a back-propagation artificial neural network model formulated to predict algal bloom. *Water Science and Technology*, **36(5)**, pp. 89–97, 1997.
- [24] Recknagel, F., ANNA – Artificial neural network model predicting species abundance and succession of blue-green algae. *Hydrobiologia*, **349**, pp. 47–57, 1997.
- [25] Recknagel, F., Fukushima, T., Hanazato, T., Takamura, N. & Wilson, H., Modelling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by artificial neural networks. *Lakes & Reservoirs*, **3**, pp. 123–133, 1998.
- [26] Maier, H., Dandy, G. & Burch, M., Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecological Modelling*, **146(1–3)**, pp. 85–96, 1998.
- [27] Wei, B., Sugiura, N. & Maekawa, T., Use of artificial neural network in the prediction of algal blooms. *Water Research*, **35(8)**, pp. 2022–2028, 2001.
- [28] Recknagel, F. & Wilson, H., Elucidation and prediction of aquatic ecosystems by artificial neural networks. *Artificial Neural Networks in Ecology and Evolution*, ed. S. Lek & J.F. Guegan. Springer-Verlag: New York, pp. 143–155, 2000.
- [29] Pineda, F.J., Generalisation of back-propagation to recurrent neural networks. *Physical Review Letters*, **59(19)**, pp. 2229–2232, 1987.
- [30] Jeong, K.-S., Recknagel, F. & Joo, G.-J., 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. *Ecological Informatics. Scope, Techniques and Applications*, 2nd edn, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 255–273, 2006.
- [31] Reynolds, C.S., *The Ecology of Freshwater Phytoplankton*, Cambridge University Press: New York, 1984.
- [32] Shapiro, J., Blue-green dominance in lakes: the role and management significance of pH and CO₂. *Internat. Revue Ges. Hydrobiol.*, **69**, pp. 765–780, 1984.
- [33] Chon, T.S., Park, Y.S., Kwak, I.-S. & Cha, E.Y., Non-linear approach to grouping, dynamics and organizational informatics of benthic macroinvertebrate communities in streams by artificial neural networks. *Ecological Informatics. Scope, Techniques and Applications*, 2nd edn, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 187–238, 2006.
- [34] Walter, M., Recknagel, F., Carpenter, C. & Bormans, M., Predicting eutrophication effects in the Burrinjuck Reservoir (Australia) by means of the deterministic model SALMO and the recurrent neural network model ANNA. *Ecological Modelling*, **146(1–3)**, pp. 97–114, 2001.
- [35] Jeong, K.-S., Joo, G.-J., Kim, H.-W., Ha, K. & Recknagel, F., Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. *Ecological Modelling*, **146(1–3)**, pp. 115–130, 2001.
- [36] Recknagel, F., Kim, B. & Welk, A., Unravelling ecosystem behaviour of lake Soyang (South Korea) in response to climate and management by means of artificial neural networks. *Journal Internat. Verein. Limnol.*, **29(3)**, pp. 1497–1502, 2006.



- [37] Recknagel, F., Kim, B., Takamura, N. & Welk, A., Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation. *Ecological Informatics*, **1(2)**, pp. 133–151, 2006.
- [38] Kohonen, T., *Self-Organization and Associative Memory*, Springer-Verlag: Berlin, 1989.
- [39] Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankangas, J., *SOM Toolbox for Matlab 5*, Helsinki University of Technology, Finland, 2000.
- [40] Takamura, N., Otsuki, A., Aizaki, M. & Nojiri, Y., Phytoplankton species shift accompanied by transition from nitrogen dependence to phosphorus dependence of primary production in Lake Kasumigaura, Japan. *Archive Hydrobiology*, **124**, pp. 129–148, 1992.
- [41] Chon, T.S., Park, Y.S., Moon, K.H. & Cha, E.Y., Patternizing communities by using an artificial neural network. *Ecological Modelling*, **90**, pp. 69–78, 1996.
- [42] Horrigan, N., Bobbin, J., Recknagel, F. & Metzling, L., Patterning, prediction and explanation of stream macroinvertebrate assemblages in Victoria (Australia) by means of artificial neural networks and genetic algorithms. *Modelling Community Structures in Freshwater Ecosystems*, eds S. Lek, M. Scardi, P.F.M. Verdonshot, J.-P. Descy & Y.-S. Park, Springer-Verlag: Berlin, pp. 252–260, 2005.
- [43] Schleiter, I.M., Obach, M., Wagner, R., Werner, H. & Schmidt, H.H., Modelling ecological interrelations in running water ecosystems with artificial neural networks. *Ecological Informatics. Scope, Techniques and Applications*, 2nd edn, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 169–286, 2006.
- [44] Giraudel, J.L. & Lek, S., Ecological applications of non-supervised artificial neural networks. *Ecological Informatics. Scope, Techniques and Applications*, 2nd edn, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 49–68, 2006.
- [45] Recknagel, F., Talib, A. & van der Molen, D., Phytoplankton community dynamics of two adjacent Dutch lakes in response to seasons and eutrophication control unravelled by non-supervised artificial neural networks. *Ecological Informatics* **1(3)**, pp. 277–286, 2006.
- [46] Holland, J.H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press: Ann Arbor, MI, 1975.
- [47] Goldberg, D.E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley: Reading, MA, 1989.
- [48] Mitchell, M., *An Introduction to Genetic Algorithms*. MIT Press: Cambridge, MA, 1996.
- [49] Bäck, T., Hammel, U. & Schwefel, H.-P., Evolutionary computation: comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, **1(1)**, pp. 5–16, 1997.
- [50] Morrall, D., Ecological applications of genetic algorithms. *Ecological Informatics*, 2nd edn, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 69–83, 2006.
- [51] Cao, H., Recknagel, F., Welk, A., Kim, B. & Takamura, N., 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. *Ecological Informatics*, 2nd edn, ed. F. Recknagel, Springer-Verlag: Berlin, pp. 330–342, 2006.
- [52] Banzhaf, W., Nordin, P., Keller, R.E. & Francone, F.D., *Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and its Applications*, Morgan Kaufmann: Heidelberg, 1997.
- [53] D'Angelo, D.J., Howard, L.M., Meyer, J.L., Gregory, S.V. & Ashkenas, L.R., Ecological uses for genetic algorithms: predicting fish distributions in complex physical habitats. *Can. J. Fish. Aquat. Sci.*, **52**, pp. 1893–1908, 1995.
- [54] Whigham, P. & Recknagel, F., An inductive approach to ecological time series modelling by evolutionary computation. *Ecological Modelling*, **146(1–3)**, pp. 275–287, 2001.



- [55] Recknagel, F., Bobbin, J., Whigham, P. & Wilson, H., Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes. *Journal of Hydroinformatics*, **4**(2), pp. 125–134, 2002.
- [56] Chan, W.S., Recknagel, F., Cao, H. & Park, H.D., Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. *Water Research*, **41**, pp. 2247–2255, 2007.
- [57] Whigham, P., Dick, G., Recknagel, F., Exploring seasonal patterns using process modelling and evolutionary computation. *Ecological Modelling*, **159**(1-2), pp. 146–152, 2006.
- [58] Muttill, N. & Lee, J.H.W., Genetic programming for analysis and real-time prediction of coastal algal blooms. *Ecological Modelling*, **189**(3-4), pp. 363–376, 2005.
- [59] Cao, H., Recknagel, F., Joo, G.-J. & Kim, D.-K., Rule set discovery for the prediction and explanation of chlorophyll *-a* dynamics in the Nakdong River (Korea) by means of a hybrid evolutionary algorithm. *Ecological Informatics*, **1**(1), pp. 43–53, 2006.
- [60] Recknagel, F., Cao, H., van Ginkel, C., van der Molen, D., Park, H. & Takamura, N., 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), pp. 191–197, 2008.
- [61] Welk, A., Recknagel, F., Cao, H., Chan, W.S. & Talib, A., Rule-based agents for forecasting algal population dynamics in freshwater lakes discovered by hybrid evolutionary algorithms. *Ecological Informatics*, **3**(1), pp. 46–54, 2008.
- [62] Recknagel, F., van Ginkel, C., Cao, H., Cetin, L. & Zhang, B., Generic limnological models on the touchstone: testing the lake simulation library SALMO-OO and the rule-based microcystis agent for warm-monomictic hypertrophic lakes in South Africa. *Ecological Modelling*, **215**(1-3), pp. 144–158, 2008.
- [63] Hong, Y.-S. & Bhamidimarri, R., Evolutionary self-organising modeling of a municipal wastewater treatment plant. *Water Research*, **37**, pp. 1199–1212, 2003.
- [64] Sequeira, R.A., Sharpe, P.J.H., Stone N.D., El-Zik K.M. & Makela, M.E., Object-oriented simulation: plant growth and discrete organ to organ interactions. *Ecological Modelling*, **58**, pp. 55–89, 1991.
- [65] Benndorf, J. & Recknagel, F., Problems of application of the ecological model SALMO to lakes and reservoirs having various trophic states. *Ecological Modelling*, **17**, pp. 129–145, 1982.
- [66] Recknagel, F., Hosomi, M., Fukushima, T. & Kong, D.-S., Short- and long- term control of external and internal phosphorus loads in lakes: a scenario analysis. *Water Research*, **29**(7), pp. 1767–1779, 1995.
- [67] Arhonditsis, G.B. & Brett, M.T., Eutrophication model for Lake Washington (USA). Part I. Model description and sensitivity analysis, *Ecological Modelling*, **187**, pp. 140–178, 2005.
- [68] Hongping, P. & Jianyi, M. Study on the algal dynamic model for West Lake, Hangzhou. *Ecological Modelling*, **148**, pp. 67–77, 2002.
- [69] Park, R.A. *et al.*, A generalised model for simulation lake ecosystems. *Simulation*, **23**(2), pp. 33–50, 1974.
- [70] Recknagel, F., Cetin, L.T. & Zhang B., Process-based simulation library SALMO-OO for lake ecosystems. Part 1: Object-oriented implementation and validation. *Ecological Informatics*, **3**(2), pp. 170–180, 2008.
- [71] Van Ginkel, C.E., Silberbauer, M.J., Du Plessis, S. & Carelsen, C.I.C., Monitoring microcystin toxin and chlorophyll in five South African impoundments. *Verh. Internat. Verein. Limnol.*, **29**, pp. 1611–1616, 2006.



- [72] Cao, H., Recknagel, F., Cetin, L. & Zhang, B., Process-based simulation library SALMO-OO for lake ecosystems. Part 2: Multi-objective parameter optimisation by evolutionary algorithms. *Ecological Informatics*, **3(2)**, pp. 181–190, 2008.
- [73] Yao, X., Liu, Y., Li, J., He, J. & Frayn, C., Current developments and future directions of bio-inspired computation and implications for ecoinformatics. *Ecological Informatics*, **1(1)**, pp. 9–22, 2006.
- [74] Capcarrere, M. Tettamanzi, A. Tomassini, M. & Sipper, M., Studying parallel evolutionary algorithms: the cellular programming case. *Parallel Problem Solving from Nature V*, eds Eiben *et al.*, Springer-Verlag: New York, pp. 573–582, 1998.
- [75] Jorgensen, S.E., Structural dynamic model. *Ecological Modelling*, **31**, pp. 1–9, 1986.

